David M. Frohlich

Fast Design, Slow Innovation

Audiophotography Ten Years On

EXTRAS ONLINE

✿ Springer

Fast Design, Slow Innovation

David M. Frohlich

Fast Design, Slow Innovation

Audiophotography Ten Years On

Springer

David M. Frohlich
Faculty of Arts and Social Sciences
University of Surrey
Guildford, UK

For Betty

Preface

The idea for this book started several years ago from an invitation by Helen Desmond at Springer to write a second edition of the 2004 *Audiophotography* book. I was busy at the time, but taken with the idea of delaying until 2014, by which time 10 years would have elapsed. This seemed to be a good period over which it would be good to look back at what happened since 2004. As it turned out, the current book was written over a longer period than anticipated – between January 2014 and August 2015. This was due to a combination of work pressures in 2014 and an unforeseen sight problem which required a period of recovery that put the writing schedule back. Readers should be aware therefore that 'Ten years on' should be interpreted flexibly as meaning 10 or 11 years on, depending on when exactly the new chapters of the book were written, and bearing in mind that the original book was completed early in 2004 and took 18 months to write.

These flexibilities in the interpretation of time pervade more than the title of the book. In starting the research, I quickly realised that I would need to make decisions about when and how the original *Audiophotography* vision was published, and that various related publications to the book itself date back to 1999. Unpublished work was also important in beginning to influence colleagues within HP, who later went on to carry out related studies before and after 2004. Indeed, the earliest research in which the audiophoto concept emerged was a series of focus groups in 1996. This puts the notion of audiophotography as it occurred to me, around 19 years ago. Similar anomalies abound in the subsequent work I reviewed, because of lags in the time taken to publish any study. Sometimes sequential studies appear in publication in the same year, and in principle could appear in reverse order if earlier studies are temporarily withheld from publication or delayed due to review and publishing processes. So the timing of studies in the literature is only an approximate indication of when they were done. Products are even more difficult to date, since documentation of releases is hard to find and details of obsolete products tend to hang around on the web. Nevertheless, the approximate timing of research and development activities is indicated by whatever time stamps one can find, and still serves to show the tremendous fertility and volatility of ideas released into different scientific and commercial forums.

The attribution of ideas to authors is also an approximate science, as is the causal links between ideas. Academic papers are usually co-authored, and citation of previous work does not necessarily imply direct development of an idea. So the tracing of developments in audiophotography had to be done somewhat delicately, and through awareness of what substantive elements were being 'advanced'. Furthermore, boundaries between research groups started to break down when authorship straddled institutions, or staff later moved from one group to another. And groups said to be developing an element of audiophotography may not have thought of their work in quite that way. Despite these challenges of interpretation, it is clear from the review that a surprisingly small number of groups around the world have continued to explore the dynamics of sound and image recording since 2004, and have extended the claims and application of my original work. I am happy to have been an ongoing part of this community over the years, and look forward to continued collaboration with colleagues and friends in the area, both old and new.

Guildford, UK   David M. Frohlich
August 30, 2015

Acknowledgements

In contrast to work on the original *Audiophotography* book, this has been a relatively solitary journey. Nevertheless I have a number of people to thank for their input along the way, starting with previous colleagues at HP Labs. In particular I would like to thank Guy Adams, Andy Hunter, Stephen Pollard and Fraser Dickens for discussions relating to Chap. 7, and Wayne Davies for gaining permission to use various photographs of HP products and prototypes. HP also returned the copyright for the original book to me, allowing Part I of this book to be reprinted here. In addition, I would like to thank the following people for comments on earlier drafts of individual chapters, including Elise van den Hoven, Risto Sarvas and Alan Brown. Alan was especially generous in lending me his business books and discussing their meaning over a beer or two. I am also grateful to Matthew Sansom at Surrey for his help in organising a related exhibition of audiophotography content and technology, and recommending an accessible way of archiving the audiophotographs in Part I of the book as a Vimeo album. This was also a vehicle for discussing the boundaries between art and domestic audiophotography, and its relation to video.

Concurrent work on the Interactive Newsprint, Light Tags and Storytelling for Development projects, as well as discussions within the EU COST FP1104 network on 'New possibilities for print and packaging', also contributed to my understanding of Audioprint Products in Chap. 9. I thank all colleagues involved in these collaborations and at Digital World more generally for their support and insights on augmented paper and digital storytelling. Special thanks go to Jocelyn Spence, Connie Golsteijn, Janko Calic, Victoria Alexander, Kristina Langhein, Peter Lancaster and Radu Sporea, whose support and ideas have been especially important to me over the period of writing the book. I am also enormously grateful to my commissioning editor, Helen Desmond at Springer, whose patience and support at difficult moments have been invaluable. Her colleague James Robinson has also helped with editing and permissions in a way which allowed me to concentrate on the content of the book first and foremost. I am also grateful to my son Joel and daughter Lily for their help with creating the cover image for the book.

Contents

Chapter 1
Introduction

Knowledge-based innovation has the longest lead time of all innovations. There is, first, a long time span between the emergence of new knowledge and its becoming applicable to technology. And then there is another long period before the new technology turns into products, processes or services in the marketplace Drucker (1985, p. 98)

1.1 Slow Innovation

Corporate research labs like to work on technology that is 3–5 years away from introduction in the market. Any shorter, and it is not enough of a step change from what is out there now. Any longer, and the conditions for its deployment may have changed too much for it to be useful. Companies are also impatient. Who wants to wait more than 5 years for a new product to get to market, especially in the high tech industry?

This book is about a new photographic technology started in a corporate research lab. It was introduced to the market 15 years ago and was 4 years in the making before that. In that time, the conditions for its use have changed out of recognition from its early beginnings, which were pre-cameraphone, pre-internet and pre-social media. And the use of this technology is not yet a mass-market practice in consumer photography, although there are signs of its emergence as a new social media form. Some would conclude the technology is a failure. And who wants to write about failure?

But what if innovation is not like this: hit or miss? What if innovation is more like a nexus of ideas, *all* of which have a ripple effect on other ideas? What if *nothing* is a failure, but rather a lesson pointing to a better idea? The traditional view of innovation is very technology-centric, and based on the notion of 'technology transfer' into a commercial product or service. But many business and market factors are involved in commercialization, and few new product releases are an instant success. In this context, it becomes difficult to tell whether and when new technology ideas have become innovations, and at what point they have succeeded or failed. In fact short-term failure appears to be the default mode by which technology and product ideas get refined and improved, and the route by which all technologies evolve. Most popular stories of innovation focus on a minority of cases where adoption was

fast and unequivocal (e.g. Kirkpatrick 2011). But in fact there is usually a 'long nose' of about 20 years to the refinement and adoption of new technologies (Buxton 2008). Interestingly for the current argument, Buxton argues that "*Any technology that is going to have an impact in the next 5 years is already 15 years old, and still likely below the radar*" (op cit).

This being the case, perhaps we would learn more from the vast majority of cases where technology failed to take off as expected and was iteratively refined and re-designed. This is the premise and topic of this book, using *audiophotography* as its' case study.

1.2 Audiophotography

Ten years ago I outlined a vision for a new branch of photography incorporating sound (Frohlich 2004). I proposed that audiophotography could be a new medium and practice, lying somewhere between photography and videography. Still images with sounds offered all kinds of opportunities for capturing memories, expressing emotions, and telling stories in new ways, yet were not properly supported by the camera, display and print technology of the day. The possibility of playing back sounds from printed photographs (audioprints) further differentiated audiophotos from video clips, and seemed to open up a new design space for augmented paper products such as audio-enabled photo frames and albums, cards, documents and actual talking books. Put in a historical context, this revived a belief by Edison that families would want to make phonographic as well as photographic family albums, and seemed to me to be on a par with the addition of sound to moving images. If consumers preferred 'talkies' over silent movies, why would they NOT prefer audiophotos over silent photos, given the choice?

The evidence I presented was based on a series of small-scale user studies with families from the UK and US, who were given the chance to make ambient, musical, talking and conversational photographs, on print and on screen. Families discovered different properties and values for each type of sound when combined with domestic photographs, and their preferences for paper or screen-based playback varied with the social context of use. In answer to my own question above, families wanted sound capture as an optional feature of a camera or display device, and something that could be mixed or muted at playback time. This became the recommended manifesto for supporting audiophotography at the end of the book, with sound as a creative new dimension of photography available at capture time or later in the history of the photograph; to enhance its meaning and impact in different ways. The context and scope of the work was limited to domestic rather than art or professional photography, and the era was marked by the advent of the digital camera (and dictaphone), and the early rise of the internet. The company supporting the work was Hewlett Packard (HP) whose business still includes printing, imaging and computing.

1.3   Ten Years On

It is now 2014/2015 as I write this extended second edition of the audiophotography book and vision. When I wrote the first version, I fully expected that we would all be capturing and sharing sounds with our photographs by now, as routinely as we capture and share photos alone. Yet we are not doing this. So why not?

Different theories of innovation have different answers to that question. Management theories might suggest that HP's market strategy was flawed. Did HP target innovators and early adopters for the use of audio with pictures, or were they working with an established majority of more conservative digital camera users that were just not interested in innovating themselves (Moore 2014)? If HP had correctly identified the value proposition for the correct market, did they really implement with sufficient focus and market leadership to make audio the next big thing in photography (Drucker 1985)? Then there is the question of standards. In the absence of an accepted file format for audiophotos, could enough HP customers share audiophoto material with non-customers to make it worthwhile to capture (Gawer and Cusumano 2002)?

Sociological theories of innovation might suggest that there was not sufficient fit to existing photographic practices (e.g. Flichy 2007). In the absence of a solution for audioprints, how would families integrate sound into their traditional activities of printing audiophotographs for physical distribution and placing them in frames or albums? If they had moved to online distribution and display, how would they share audiophotos by email or on photo websites? Typically such practices change slowly over long periods of time, with feedback from the market influencing, or 'socially constructing', new versions of the product (Biijker et al. 2012). So had customers been given sufficient time to experiment with different audiophoto products and their associated information and communication technologies, to domesticate them into their everyday lives (e.g. Haddon 2006; Shove et al 2007)?

Media theorists might make a similar point to this last one. Every new medium requires a period of assimilation and experimentation with the content as well as the technology (e.g. McLuhan 2001; Murray 2011; Balsamo 2011). Was it even clear to existing customers, let alone new customers, what they were supposed to do with sound and what emerging genres of audiophotography they were supposed to follow. Related precedents were available in the wider culture, such as photographic portraits, music videos, podcasts, business presentations, talking greeting cards and so on. But how were early adopters expected to adapt these or invent their own without a reference to a community of practitioners?

Of course there is a much a simpler explanation than all of these. Perhaps I was wrong about the value of audiophotos and the attraction to the consumer. Perhaps I 'configured' my users incorrectly (Woolgar 1990). My studies were small-scale after all. However, that goes for most of the studies in my professional fields of design research and interaction design, and small-scale studies have always served me well in the past. So like the innovation theorists, I prefer to blame someone or something else!

But theoretical assignment of blame doesn't get us very far in reflecting on the past, present and future of audiophotography. Rather than asking why something never happened in a fixed amount of time, it is better to ask 'what actually happened next?' This is a more productive question, born of inductive scientific curiosity. It is also closer in spirit to the view of innovation as a nexus of evolving ideas. With the metaphor of a ripple effect in mind, it becomes immediately obvious that you could look at the effect of the audiophotography vision on the immediate colleagues and company through which it was originated. Or you could look further afield to its effect on the wider world. Within these private and public domains, you could examine the effect on research and development activities separately. These interacting dimensions are illustrated in the ripple diagram of Fig. 1.1.

The internal and external circles of Fig. 1.1 reflect an important distinction in the literature between closed and open innovation. This hinges on the extent to which innovation activities are carried out 'in-house' or not. In recent years there has been growing recognition that open innovation can be cheaper and more effective for companies if managed correctly, drawing not only on the world's best researchers and engineers across the globe but also on the ingenuity of a company's most innovative customers (e.g. Chesbrough 2003; von Hippel 2005). It applies both formally and informally to HP, which formally promotes a strong in-house R&D capability, balanced with a programme of global university and partner engagement (Savage 2011). Further forms of open innovation work informally as in house staff are influenced by scientific literature, conferences and workshops, and allowed to publish their own work, subject to publication approval.

In the audiophotography case, HP allowed a steady stream of publications of the work to emerge from 1999 onwards and encouraged the writing of the 2004 book to promote the approach (Frohlich and Tallyn 1999; Frohlich et al. 2000, 2004;
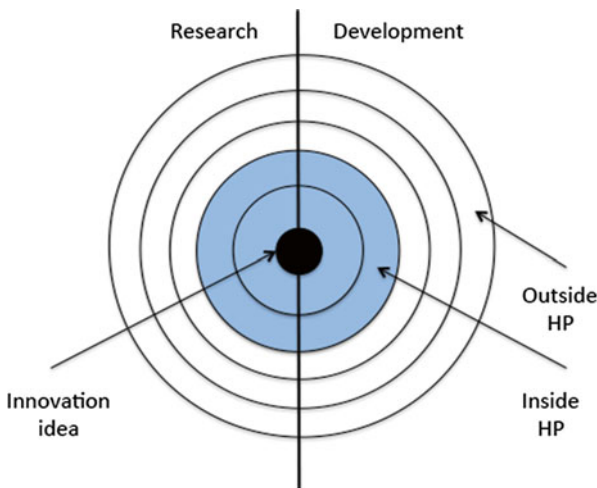
Fig. 1.1   The innovation ripple effect

Frohlich 2004). Further openness happened through collaboration across institutions and staff turnover. For example, work on an *audiophoto desk* for playback of sound from printed photographs was a collaboration with photographer Tony Clancy at the University of Central Lancashire and engineers John Robinson and Enrico Costanza at the University of York. Furthermore, Ella Tallyn and myself left HP by 2005 for other university contexts at Kings College London and Surrey, leaving Guy Adams as the only original co-author at HP.

The impact on research and development can be considered separately because they are essentially different activities with different objectives and outputs. The aim of research is to generate new knowledge with either internal reports or external scientific publications as outputs. The aim of development is to refine the design and manufacture of products which are themselves the output. Core research ideas and findings may have a life of their own, independent of product development, while products may evolve without necessarily using or generating related research. Of course company research labs are usually set up to inform and generate new product development opportunities, but the translation of research into product is not a guaranteed process – as this book demonstrates. In fact, there is a research-practice gap affecting this translation, especially in the new technology field (Norman 2010). Early prototoypes with small-scale feedback are a long way from finished products. However, they may inspire other prototypes, which eventually find their way into the market via other companies. Therefore, looking at both halves of the audiophotography ripple effect in Fig. 1.1, inside and outside HP, should reveal the full extent of its impact and evolution.

1.4   How to Read This Book

The rest of this book is in two parts. Part I introduces the Audiophotography vision and technology. It contains five of the original chapters republished from the 2004 book *Audiophotophotography: Bringing photos to life with sounds* (Frohlich 2004). The chapters are unabridged, apart from the correction of figure and table numbers corresponding to the new chapter numbers of this book, and the modification of other cross-references. This means that 'Related work' sections have been left in tact, and are current only to 2004. More recent work is described in Part II. The chapters in Part I step through the properties of ambient, talking, musical and conversational photographs, before concluding with a discussion of paper versus screen preferences. Audiophotos are quoted throughout, and can be played from the following website instead of from an accompanying CD: https://vimeo.com/album/3548546. The chapters discuss the role of different kind of sounds in six photographic activities identified in a 'diamond framework' showing the interplay of photographs, audience, subject and photographer. This framework is reproduced in Fig. 1.2 below to introduce new readers to these activities and their relationship to each other.

Part II of the book then takes up the story of what happened next in HP and beyond. The influence of the audiophotography work on research and development within HP is summarized together in Chap. 7. However its' influence on external
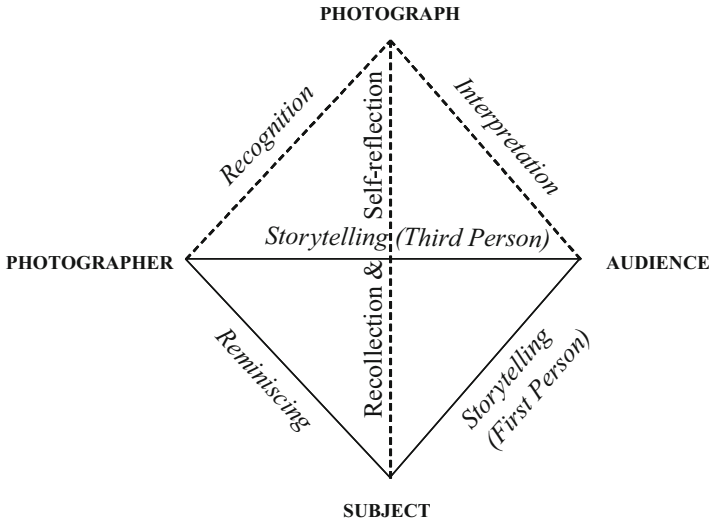
Fig. 1.2 The diamond framework (Fig. 3.5 in Frohlich (2004))

research and development is summarized separately in Chaps. 8 and 9. Related work in the broader research literature (Chap. 8) is traced mainly by following its forward citation in academic articles and books. The same technique cannot be used for tracing related products in Chap. 9, so a separate review is done of *current* audiophoto products and services that embody the original ideas across different platforms and technologies. This glosses the history of these products and underscores Buxton's call for better ways of archiving the evolution of digital products (in Melanson 2011). In fact, while academic citation guarantees that new authors were aware of earlier work in further research, the lack of commercial product citation in new product development does not allow design influences to be publically tracked in the same way. Hence I do not strictly review the influence of audiophotography on related product development, but rather look for its *functionality* in a snapshot of existing products. While limited, this has the practical benefit of exposing how it is possible to do audiophotography today in a way that was not possible in 2004.

A final discussion chapter takes stock of the lessons of this case study for audio-photography and innovation. I will argue that design and innovation naturally work over very different (fast and slow) time frames, and that different factors are involved in stabilizing their relationship in research and development contexts. Design is defined in ordinary language to mean the art of conceiving a plan of something before it is made, while innovation is defined as making changes in something established by introducing new methods, ideas or products (Oxford English Dictionary). Design is naturally faster than innovation although there are interesting calls for slow design and fast innovation that will be considered in the discussion. Hence the book will be of interest to students, scholars and developers of digital media content and services, and to research and business managers concerned with

promoting innovation in all its forms. This is essentially a behind-the-scenes book about how a new digital medium is born in the research labs of corporations and universities, and the harsh reality of the consumer market.

References

Balsamo A (2011) Designing culture: the technological imagination at work. Duke University Press, Durham

Bijker WE, Hughes T, Pinch T, Douglas DD (2012) The social construction of technological systems: new directions in the sociology and history of technology. MIT Press, Cambridge

Buxton W (2008) The long nose of innovation. Business Week, 2 Jan 2008. http://www.business-week.com/stories/2008-01-02/the-long-nose-of-innovationbusinessweek-business-news-stock-market-and-financial-advice

Chesborough HW (2003) Open innovation: the new imperative for creating and profiting from technology. Harvard Business School Press, Boston

Drucker P (1985) Innovation and entrepreneurship. 2004 reprint by Elsevier/Butterworth-Heinemann, Amsterdam/Boston

Flichy P (2007) Understanding technological innovation: a socio-technical approach. Edward Elgar, Cheltenham

Frohlich DM (2004) Audiophotography: bringing photos to life with sounds. Kluwer Academic, Dordrecht

Frohlich DM, Tallyn E (1999) Audiophotography: practice and prospects, CHI'99 extended abstracts. ACM Press, New York

Frohlich DM, Adams G, Tallyn E (2000) Augmenting photographs with audio. Pers Technol 4:205–208

Frohlich DM, Clancy T, Robinson J, Costanzo E (2004) The audiophoto desk. In: Proceedings of 2AD: 2nd international conference on appliance design, p 139

Gawer A, Cusumano MA (2002) Platform leadership: how Intel, Microsoft and Cisco drive industry innovation. Harvard Business School Press, Boston

Haddon L (2006) The contribution of domestication research to in-home computing and media use. Inf Soc 22(4):195–203

Kirkpatrick D (2011) The Facebook effect: the inside story of the company that is connecting the world. Simon & Schuster, New York

McLuhan M (2001) Understanding media. Routledge, London

Melanson D (2011) Microsoft's Bill Buxton exhibits gadget collection 35 years in the making. Engadget,UK,9May2011.http://www.engadget.com/2011/05/09/microsofts-bill-buxton-exhibits-gadget-collection-35-years-in-t/

Moore G (2014) Crossing the chasm: marketing and selling disruptive products to mainstream customers, 3rd edn. Harper Business, New York

Murray JH (2011) Inventing the medium: principles of interaction design as a cultural practice. MIT Press, Cambridge

Norman DA (2010) The research practice gap: the need for translational developers. Interact Mag 17(4):9–12, ACM Press, New York

Savage N (2011) HP's open innovation strategy: leveraging academic labs. MIT Technology Review, 8 Feb 2011. http://www.technologyreview.com/news/422636/hps-open-innovation-strategy-leveraging-academic-labs/

Shove E, Watson M, Hand M, Ingram J (2007) The design of everyday life. Berg, New York

von Hippel E (2005) Democratizing innovation. MIT Press, Cambridge

Woolgar S (1990) Configuring the user: the case of usability trials. Sociol Rev 38(S1):58–99

Part I
Audiophotography Defined

Chapter 2
Ambient Photographs

It definitely made you more aware of your sense of hearing and more aware of sounds… You had to think more about capturing the sound and taking a photo to go with it, rather than taking a photo and then the sound (Debbie, audiocamera trial).

Sounds are an integral part of our physical environment and psychological experience. In the context of recording, ambient sounds are defined as those which happen spontaneously without being consciously created *for* the recording. These can be both natural and man-made. Natural sounds include the sounds of the elements such as water, wind and rain, and of animals such as birds, insects and mammals. Man-made sounds include the noises made by people themselves, such as footsteps, speech and music, and the sound of technology such as traffic, machinery and television. The balance of natural and man-made sounds changes as we move across geographic and cultural boundaries, between rural and urban areas, and into and out of buildings. Ambient sounds in general may be growing louder and more artificial over time, with corresponding losses in natural sounds (Schafer 1977). In fact, each part of the world can be considered to have a soundscape as dynamic and fragile as its landscape and wildlife, and equally in need of protection and management (Schafer 1976).

This chapter examines for the first time what kind of ambient sounds people would like to save with their photographs. It is based on an exercise in which families were given the ability to record sounds on a conventional camera. I report what kind of sounds were recorded and how they were associated with the images taken around the same time. I also examine the comments of participants as they explain why they took particular audiophotos, what they think of them, and how they would use them according to the diamond framework described in Chap. 1. These data give an initial insight into the role of ambient sounds in all the activities shown in the framework. They also provide an opportunity to test early speculations that ambient sounds may enhance memory for a photograph, and that ambient photographs may even be preferable to home video clips under some circumstances. But to put this study onto context, I begin with a brief review of related work.

2.1   Related Work

In general, there is very little scientific work on the sentimental value of sound, and none which connects this to photographs. For example, social scientists have studied the physical and psychological properties of sound to understand how we hear and identify different sounds. However, they have failed to explore the personal meanings of sounds. At another level, members of the world soundscape movement, such as Schafer referenced above, have begun to classify and map geographic soundscapes, and campaign for an improvement in the quality of the sonic environment. While this work calls attention to certain general preferences for natural over man-made sounds, it seems to stop short of mapping personal preferences within a soundscape (Truax 1978).

In contrast, a number of artistic projects with sound have begun to do this, and reveal a rich set of meanings and preferences for particular kinds of sounds. In one such project, the sound artist and recordist Peter Cusack collected people's favourite sounds of London. His method was to invite proposals from members of the public and then go out and capture examples of these sounds for himself. Some of the resulting recordings were broadcast on a local radio-art station in 1998 and released on CD in 2001 (Cusack 2001). Favourite sounds of the city turn out to be highly ideosyncratic, and just as likely to include man-made sounds as natural sounds. The CD collection contains a number of natural sounds you might expect, such as a dawn chorus of blackbirds, a fountain, and the rise and fall of barges moored on the Thames. However, it also contains the less predictable sounds of underground trains, the bell of a London bus, Big Ben chiming and a street market packing up. Some sounds are extremely delicate and specific, such as the sound of a bicycle rolling over loose concrete slabs on a particular part of a canal towpath. Others are quite generic such as the sound of the turnstiles on entry to a football match or the sound of a coffee making machine. All these sounds are likely to have particular associations for the people proposing them, although these are not explored directly within the project. In a related piece, Cusack draws on his own memories of the Lea Valley in East London to create a collection of representative sounds (Cusack 2000). This time he included spoken stories of the area told by locals, alongside familiar sounds of the place. These were recorded around the canal running through the valley, and include bird and frog calls, dogs barking, footsteps under railway arches, dripping water, gentle motor boat sounds and conversation outside a waterside pub. The sounds for both these projects are loosely associated with a series of images of London and the Lea Valley which are included in the audio CD booklets.

Another noteworthy project is the *Lost and Found Sound* programme on National Public Radio in the United States. This is produced by the Kitchen Sisters (Davia Nelson & Nikki Sylva) and Jay Allison, and features stories about sound recordings or sound-related memories phoned in by radio listeners. For example, a recent story concerned the Green Street Mortuary Band and its tradition of giving past residents of San Fransisco's Chinatown a musical send off. While these stories might be triggered by a single radio caller, some stories are based on a host of submissions on a particular theme. A dramatic example of this became the *Sonic Memorial Project*, in

which listeners were invited to send in their sound memories of the World Trade Centre just after its destruction on September 11th 2001. This invitation resulted in a flood of stories and recordings relating to different personal memories of the buildings over time. Some callers described the sounds of the day itself, commenting on the silence which followed the collapse of the towers. Others noted the sounds of the buildings in use, such as the beat of the revolving entry doors, the creak of the walls in a high wind, or the sound of latino music played by the South American janitors in the evenings. A final group of callers described their memories of the original site of the centre where music would play from loudspeakers outside the electronics shops of 'Radio Row'.

These examples show that ambient sounds clearly trigger memories and stories of a location or event in much the same way that photographs do. The question that has not yet been answered by these projects is how such sounds might be captured alongside photographs, and integrated with the practice of domestic photography in all its forms. It is to this question that I now turn.

2.2   The Audiocamera Trial

An audiocamera trial was organised to allow people to capture sounds with photographs. Ambient sounds could be captured around the time of taking each photograph, or other sounds could be consciously added later by recording into the audiocamera device. Four PC-owning families with children were given the opportunity to capture audiophotos with this camera on their summer holidays, and to review them at home before commenting on a range of new product options in what we call a 'feedback group'. In addition to discussing audiophoto usage, the families were specifically asked about the value of sound capture and its comparison with home video. The study was carried out in collaboration with Ella Tallyn.

Middle-income families with children living at home were chosen for the study because they happened to be the target market for early digital cameras. In our study, they comprised two families of five and two families of four. The parents and all but four of the children used the unit and took part in the ensuing discussions. This meant that we had eight adults and six children as full participants. The participating children ranged in age from 11 to 16. Each family owned several point-and-shoot cameras and a multimedia home PC.

Families were given an audiocamera to use on their 1996 summer holiday. This unit comprised a Minolta compact camera combined with a slimline Lanier Dictaphone with an omnidirectional external microphone (see Fig. 2.1). Each unit allowed users to record photographs and sounds in any combination, but did not allow them to associate the two at capture time. Individual sound recordings could be marked with a beep on the tape. Before the trial families were visited at home and given a written explanation of the trial process at this stage, together with instructions and guidance on how to operate the unit. After the trial families sent their photos to us for developing. We then revisited each family at home for a preview and editing session in which we

Fig. 2.1 The audiocamera
unit used in the trial

- discussed their experiences of using the capture units,
- observed and discussed their initial review of the printed photos alone and then in combination with their raw sound clips, and
- obtained their instructions regarding the arrangement and association of their best audiophotos for a holiday album.

Afterwards we edited their sound clips as instructed and supplied the family with a cassette tape copy for using at leisure with their prints. At the same time we used extracts from their best materials to make up audiophoto album demonstrations for the feedback group.

Two separate feedback groups were held for two families at a time in HP Labs Bristol. We structured the discussion to obtain their general impressions of the audiophoto medium, before asking them to summarise their private use of the edited tape and prints. We then showed the families a number of ways of reviewing their audiophoto albums on screen and paper. These are explained further in Chap. 6, which examines the reactions and comments of families in this part of the study.

The rest of this chapter looks mainly at the findings of the first part of the trial in which families recorded ambient audiophotos, reviewed and commented on them, and selected favourites for an album. It begins by examining the kinds of audiophotos that were captured initially, before considering what families said about them.

2.3 Audiophotos Captured

A total of 261 photos and 158 sound clips were captured by families in the trial. Their structure and content reveals a lot about the attractiveness of ambient sounds and audiophotos. In particular, this set of audiophotos reflects the kinds of ambient sounds and images that participants thought they wanted to keep, before they had a

chance to see the results later on. Although the set includes audiophotos which are later judged to be poor or ineffective (as with any set of ordinary photos) it does not include sounds and images in which the photographers had no interest. The set as a whole therefore reflects the kind of image and sound associations that people may want to capture on an audiocamera.

2.3.1 Types

The different types of sound recorded are shown in Fig. 2.2. This pie chart was generated by classifying each sound clip as containing one or more of the types of sound shown in the key. Hence, these are not mutually exclusive types of sound clips, but rather properties or components of all the clips recorded.

The chart shows that the three most common properties of these recordings were *ambient sounds* of the environment in which a photo was taken, *ambient conversation* taking place between participants in the scene, and *voiceover* spoken to the camera sometime after the photograph was taken. The chart also distinguishes *ambient music* from more natural environmental sounds, and three other types of deliberate speech recordings. These were *stage directions* spoken by the photographer to subjects in the scene (such as 'say cheese'), *interviews* between the photographer and the subjects, and *live commentary* designed to describe an event as it unfolds. These last three forms were less common but nonetheless important to the overall effect of the sound when they occurred. A mix of different kinds of sound can take place within clips as well as between them. This is illustrated in Fig. 2.3.

Hence, Fig. 2.3 shows a pair of audiophotos of the same people, containing ambient music and ambient conversation. However, in Fig. 2.3a the ambient music of the street band is more prominent than the background conversation. In Fig. 2.3b the reverse is the case, since the hubbub of conversation in the restaurant is louder and precedes the
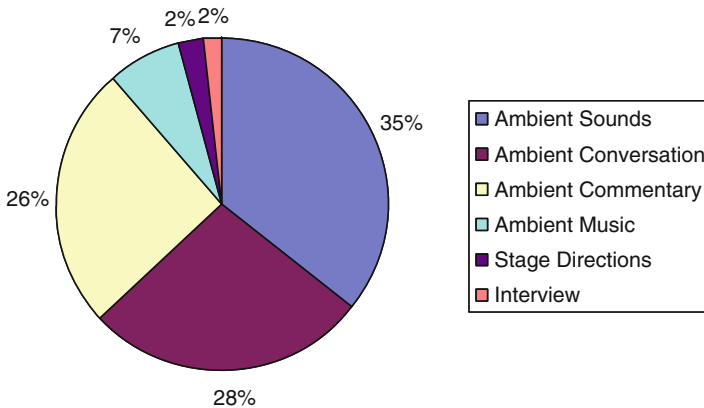
Fig. 2.2 Types of sounds recorded on the audiocamera

Fig. 2.3 Different mixes of ambient music and ambient conversation in audiophotos containing both types of sound

faint introduction of background music half-way through the clip. Note also that the composition of the images in these photos also affects the salience of the music. Only the street café photo shows the source of the music in the image – behind the main subject. This also illustrates how the sound and image work together to create a combined audiophoto effect, since the music of Fig. 2.3 serves to call attention to the band in the background of the image, changing the salience of its visual features.

2.3.2 Durations

Figure 2.4 shows a graph of the duration of sound clips recorded in the trial. Each bar in the graph represents the number of clips lying within each 5 s interval along the bottom. Hence, the first bar on the left shows that there were 22 clips between 5 and 10 s long. Although the most popular duration was 10 s, the average duration overall was 24 s, with about 95 % of clips lasting under a minute.

The shortest clips were voiceovers which served only to attach a voice label to the photo. Examples include 2 s clips to say 'Sunny Zakopane' or 'A bendy bus', and were equivalent to the kind of label that might be handwritten on the back of a photo or on an album page. More extensive voiceovers varied in length from between 5 s and about 30 s depending on the subject matter.

Pure ambient sound clips were about 10 s long. These included the sound of a horse and carriage (8 s), a tram (8 s), rain on a window (6 s), the rattle of a toy car (8 s), and the sound of a stream (13 s). Pure ambient conversation clips tended to be longer than this to allow for the closure of someone's turn or the completion of a phrase. For example, there were numerous conversations recorded around a table ranging from 18 to 38 s. In this respect, the clip represented in Fig. 2.3b was a typical duration of 29 s.

The majority of clips were between 15 and 40 s and contained unfolding combinations of sounds. Those containing music and talk were not cut-off in mid-sentence or bar and tended to reach a natural closure. This may have extended their duration compared to a clip of ambient sounds that could be stopped at any point. The

combination of different types of sounds often unfolded over time, making up a sound event. For example, Fig. 2.5 shows two people jumping into a swimming pool. The sound clip lasts for 29 s and moves through a sequence of sound types. These start with ambient conversation and a count-down to the jump, the ambient sound of their splash into the water, more ambient conversation in the form of exclamations, and a comment from the photographer which serves to label the photo.
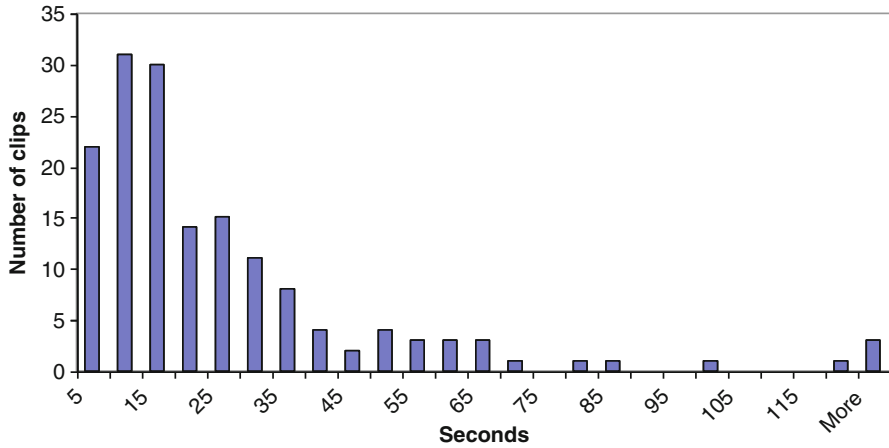
Fig. 2.4 The duration of sound clips in the audiophoto set
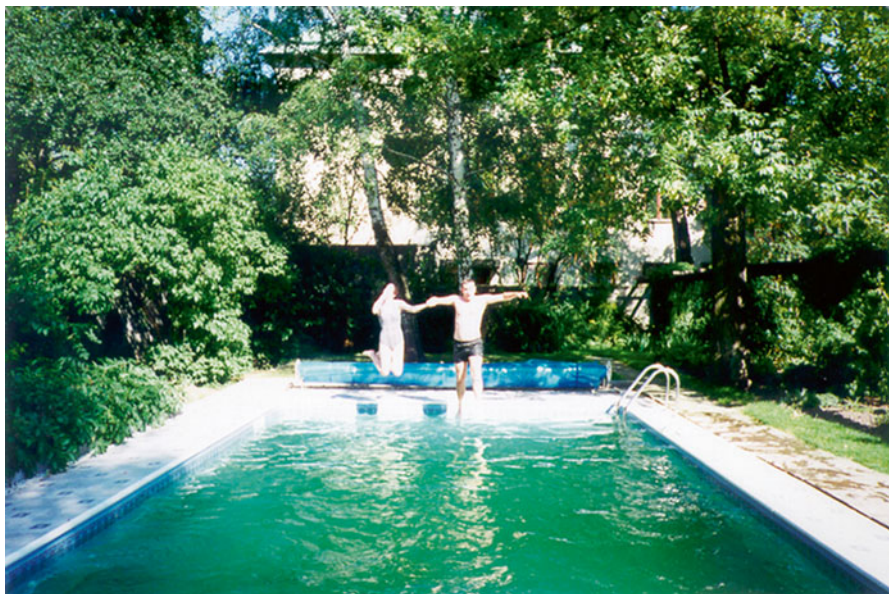
Fig. 2.5 A sequence of sound types in a sound event lasting 29 s

The longest sound clips were music recordings made at a concert (4 min 57 s), live commentary on an open-topped bus tour (3 min 19 s) and a small number of sound-only recordings of conversations and comments, lasting 1 or 2 min. Although these were long in comparison to other clips recorded on the audiocamera, they are relatively short for recordings made on a dictaphone. This reveals the constraining influence of the camera on sound recording in the trial, and the fact that sound clips were almost always designed to go with corresponding images. It is to the nature of this correspondence that we now turn.

2.3.3 Associations

The capture of sound and image was completely unconstrained on the audiocamera unit, and our instructions encouraged participants in the trial to think about recording sounds and images out of sequence and adding sounds later. However, in practice, most sounds were recorded around the time of image capture, and in strict chronological order.

Recording seemed to begin from a variety of positions around each shutter depression on the camera, but could not be said to precede or follow it by any predictable duration. This meant that a lot of the sound and image associations were obvious even to us as spectators of the audiophotos, since the two media essentially indexed the same event and proceeded in sequence across the recorded film and tapes. Any ambiguities were cleared up later in conversation with the families.

Figure 2.6 shows that about two thirds of sound clips were associated with a single image in the corpus. This association could usually be recognised by the sound of a single shutter click occurring within a sound clip, and by the connection between the content of images and sounds. Of the remaining one third of clips, most contained one or more sound clips for an *image sequence*. Finally, there were a small number of occasions in which participants attached several sounds to the same photograph.

Simple sound to image associations are easy to visualise and comprise audiophotos like that shown in Fig. 2.5. Other kinds are more difficult to visualise. Composite audiophotos with one sound for many images included action shots with a running commentary like a sequence of four surfing photos with a running commentary, a sequence of sailing photos with ambient sounds, and three photos of an obstacle race with 90 s of instructions and cheering sounds. A related form of *audiophoto narrative* was also generated when people took 2 or 3 photos of a more static scene that was indexed by a single piece of ambient sound or voiceover. The Restaurant picture of Fig. 2.3b was one of two photos taken at the time of that ambient sound. Other examples include the recording of a single voiceover for multiple photos of a beach.

Sometimes multiple sound clips were recorded for a single image. Figure 2.7 shows an example of this type of association. Here a comment was recorded first about a castle that was visited and photographed as shown. A further two audio clips followed, containing an interview between the photographer and other members of the family shouted between the bottom and top of the castle tower, and their joint conclusions when re-united at the bottom. These were rare because it was more usual for people to
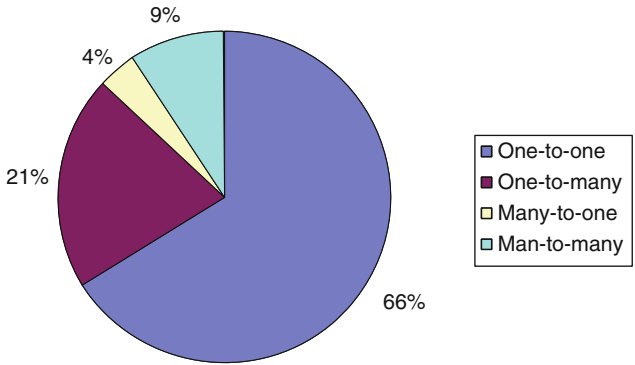
Fig. 2.6 Audio:image association in the ambient photo set

Fig. 2.7 An audiophoto comprising three separate sound clips for the image shown

record multiple phases of an event live in a single recording, as in Fig. 2.5. Note also that the distinction between the sound clips in Fig. 2.7 disappears at playback, because the clips have been joined together to play through as a single piece.

Many-to-many sound and image associations are a variation of the other composite forms. An example is shown in Fig. 2.8, which contains two ambient sound recordings for three street market photos. The photographer's intention here was usually to create a mood in multiple sound clips for a set of photos that might be viewed together like a collage, rather than to illustrate a sequence of photos to be viewed one at a time. This has been indicated by printing the images as a collage in

Fig. 2.8 An audiophoto collage of two sound clips for three images

the Figure. Technically these might be decomposed into simpler forms, but we preserve them here to indicate the behaviour of capturing an entire scene with looser collections of sounds and images.

2.3.4   Silent Photos

Because of the complexity of associations between sound and images in the collection, the 158 sound clips corresponded to many more than 158 of the 261 photos taken. Nevertheless, a residual set of 43 silent photos were left over when these associations had been taken into account. There was no particular theme to the content of these photos that might explain why they were left silent. However, a combination of technical features and comments by families explained why sound was not recorded on certain photos.

The majority of silent photos occurred amongst sets for which there was already sufficient ambient sound coverage. A good example of this was in one family's

record of a trip to London. They took a variety of tourist photos with conversation, ambient sounds and some commentary from an open-topped bus, but six out of nine photos were technically silent. Another major reason for silent photos was sound equipment failure. This accounted for 11 out of 15 silent photos at the end of one family's film, after the microphone batteries had run out. This resulted in consequent disappointment since the family believed they were still capturing sounds with these images. A batch of beach photos at the end of another family's photo set were also silent, due to salt water corrosion of the battery contacts in the dictaphone. Other reasons for not taking sound included the lack of any strong or distinctive sounds worth recording in certain environments, for example in the home, and the extra effort required to think about sound and control the dictaphone in combination with the camera. These factors will be discussed further below.

2.3.5   Solitary Sound Clips

In addition to photos without sounds there were also a number of sounds without photos recorded on the audiocamera. These might be called *audiographs*. Ten sound clips were recorded in this way during the main part of the trial, and a further six clips were captured by one family who asked to borrow the unit again afterwards.

   As with silent photos, solitary sound clips were sometimes recorded at the end of sequences when the other medium had run out. For example, the teenage children of one family recorded a mock argument and a summary of the day at the end of one tape, after the film in their camera had finished. Other solitary clips were taken to record sounds that were considered attractive in their own right, such as street music and singing, or sounds that were difficult to capture with corresponding images, such as laughter. In one case, there was also a second take of an interview that was later rejected by the participants.

   All these examples begin to indicate the value of sounds in extending what can be captured on a camera, and the way in which families quickly fell into the habit of recording sounds for this purpose. More direct evidence for the value of sound comes from how families reacted to their own audiophoto material, in terms of their preferences and comments.

2.4   Reactions to the Audiocamera and Audiophotos

2.4.1   Initial Reactions

Many of the participants initial reactions were to the audiocamera unit itself and the experience of recording sounds with photos. The unit was generally too complicated to operate and slowed down the process of taking photos. This was annoying when participants were trying to capture something quickly, since they had to remember

to switch on both the camera and the dictaphone, and decide when to start and stop the sound recording in relation to taking the photograph. As one mother put it:

In some ways it was distracting and you tended to rush the photo (Viv)

Most participants reported getting better at this with practice. Sometimes they would decide not to record sound with an image, while at other times they would take a picture to index a sound recorded earlier. With extended use, people appeared to treat the sound less as an adjunct to picture taking and more as a central or even primary medium to be captured with an image. This appeared to increase people's awareness and atunement to sounds in the environment, as indicated in the opening quote of this chapter. Such attunement appeared to increase the importance of sound during playback as well as during the capture period:

For me the sound initially augmented the visuals. But then as you run through it and you've got the sound going, the visual begins to augment the sound. You start with one, and become driven by one of the two media that are going on. (Paul)

This awareness of sounds appeared to make adult participants self-conscious about the sounds and comments they were making in front of the audiocamera. Some participants mentioned not liking the sound of their own voice, or feeling stupid recording comments into 'thin air'. All this was less of a problem for children who either ignored the audiocamera unit entirely or enjoyed playing up to it. Self-consciousness appeared to diminish for adults over time as they got used to the presence of the audiocamera, and even began to play up to it occasionally for better effects. For example, the very next audiophoto following the restaurant scene in Fig. 2.3b was of two adult members of the family singing silly lyrics to the 'The Yellow Rose of Texas' after the meal.

Another discovery made by participants using the audiocamera was that it raised privacy issues about the appropriateness of recording sound in certain situations. While it was fine to record sounds within a group of friends or family who knew they were being recorded, it was more of a problem with strangers. Such people would not expect a camera to record their conversation and might be offended to learn that it had been used to do so.

This problem is similar to that encountered with camcorders, except that camcorders are commonly known to record sound as well as moving images. Again, participants appeared to get used to this constraint by avoiding sound recording in delicate social situations, or other locations where it seemed inappropriate. For example, one family had visited the site of Austwitz concentration camp in Poland, and felt strongly that they didn't want to record any ambient sound there.

Despite some of the reservations expressed about the process of recording audiophotos, the families were intrigued to see the results for the first time. Their first listening was accompanied by a good deal of laughter and conversational banter, and was followed by a flood of comments to us about what they liked and disliked about the material. There was more consensus about the kinds of sounds they disliked most, with positive preferences for sounds and audiophotos being more idiosyncratic.

The sound of the camera shutter going off within a sound clip was unanimously judged to be annoying. Within a gentle ambient sound clip it destroyed the mood of the moment. Families asked for this sound to be edited out in the process of constructing an album of their best audiophotos. Similar intrusions sometimes occurred in the form of unexpected ambient sounds like traffic going past, or someone shouting in the background. For example, one clip of a distant steam train puffing past was spoiled by the sound of a relative yelling at a dog in the foreground. Other disliked sounds included wind noise on the microphone that tended to drown out other sounds, nondescript sounds whose source could not be recognised, and recordings that were not true to life. All these have their parallels in ordinary photography where images can turn out to be masked by overexposure, to depict subjects too distant to recognise, or to contain colours that are not true to their source.

One surprise discovery was the relative unattractiveness of voiceovers to the people who had recorded them. Families who had recorded short voice labels or comments on their photos tended not to like these themselves at playback. They usually remembered the context of each photo, which after all was taken less than a month before, and felt that the comments were redundant.

In contrast, participants liked ambient sounds that were distinctive, uninterrupted and true to the actual sound as it was heard and remembered. If this sound also matched the content of an associated photo then it was usually perceived to work well as an audiophoto. A good example of this is shown in Fig. 2.9, which shows a view of a horse and cart ride. The sound is a clear recording of the cart rattling down the road with the clip-clop of the horses hoofbeats in the foreground. While ambient commentary was also recorded on about a quarter of audiophotos, it was really these

Fig. 2.9 Distinctive sounds appear to make the best audiophotos

more distinctive types of natural sounds that delighted the families most, on listening to their audiophotos for the first time. Further evidence of this effect was provided through the selection of 'favourite' audiophotos for albums we made up for each of the four families after their initial review of material.

2.4.2   Favourites

In selecting about ten audiophotos for a favourites album, families explicitly revealed their preferences for sound and image combinations in the set. In the course of making this selection, individual family members also discussed the reasons for their preferences, and sometimes volunteered their overall favourite audiophoto in the set.

In general, there was a preference for ambient sounds over other types. There was also a tendency to prefer more complex sound and image combinations in the favourite set. This suggests that looser or more complex associations may be considered more interesting or evocative than simple one sound one image associations. Nevertheless simple audiophotos made up the majority of favourite as well as non-favourite audiophotos.

In discussing their favourite audiophotos, families confirmed these overall preferences. They listed the following types of ambient sounds among their favourite recordings:

- Children's voices
- Crowds
- Rain
- Running water
- Chinking cups
- Street music
- Foreign language being spoken

These sounds turned up in various individual 'all-time' favourite audiophotos. Three of these are discussed here in order to illustrate their composition and explain why people found them so attractive. The first all-time favourite is shown in Fig. 2.10 below.

Figure 2.10 was taken and selected by the mother of one family (Viv), and shows her husband and dog in the sea. A pure ambient soundtrack has captured the lapping of the water as Viv walks through the shallows. Although this is a simple one-to-one association of sound and image which contains little additional information in the sound channel, it was favoured because of the way the sound added to the mood of the photo and reminded Viv of the peace of that particular summer evening on a Devon beach. As spectators on this scene we can imagine something of the external atmosphere involved, but to Viv herself it appears to be her own internal feelings of the moment that are triggered by the sound and image. The fact that it was these very sounds she heard, from that particular perspective, is probably important to the

Fig. 2.10   All-time favourite audiophoto of Viv, depicting her husband and dog in the sea one summer evening, while she walked through the shallows

Fig. 2.11   All-time favourite audiophoto of parents Mervyn & Sue, depicting their son Philip making a mistake in his commentary on their mountain climbing photo

effectiveness of the audiophoto in triggering this recollection, but it is the recollection itself that makes the audiophoto so special.

Another all-time favourite audiophoto is shown in Fig. 2.11. This was singled out by the parents of Philip, shown on the far left of the photo, who recorded the associated voiceover describing where the photo was taken. The rest of the family found Philip's mistake at the end of the voiceover very funny, because he had rehearsed what he was going to say several times before making this recording. Philip's clear displeasure at the end of the clip added to this effect. Although this mixture of embarrassment and humour make the audiophoto appealing in its own right, its main value for the parents lies in the fact that it reveals something of Philip's personality that he might not want to reveal about himself, namely that he likes to get things exactly right. In this respect, the audiophoto operates as an enhanced self-portrait.

A final all-time favourite audiophoto is shown in Fig. 2.12. This was chosen by Debbie, another mother in the trial. The figure shows a collection of photos of Debbie's friend and daughter, taken at her friend's house on a weekend visit. Her friend was learning to play the piano at the time, and the background sounds were recorded while she practiced over the weekend, as shown in one of the photos. These

Fig. 2.12 All-time favourite audiophoto of Debbie, depicting her friend learning to play the piano at home, on a weekend visit with her family

sounds end with Debbie wishing her daughter goodnight. This extract is an audio-photo collage of sounds and images. The images could be made to reveal themselves one-at-a-time at playback, or be viewed together as shown in the figure.

While the individual photos were taken at different times, Debbie chose to combine them with the music and conversation to create an impression of her friend's house. She used to live with her friend who often filled the house with the sound of her cello playing, and the sound of the piano reminded her of that. The inclusion of her daughter situates the photo in more contemporary times, in which Debbie has a new life with her own family. The selection of what sounds to record and how to combine them with images, opened up new creative possibilities for the families. This piece is an example of how some participants began to exploit these to good effect. Again the overall aim was to create something that operated as an effective reminder of thoughts and feelings of the moment, or revealed something of the character of the people in the photographs.

2.4.3   The Value of Ambient Sound

The above discussions of favourite audiophotos begin to reveal some of the properties of ambient sound that participants valued most in the trial. For example, the ability of sound to add mood and humour to an image, and to foster increased creativity in photo taking, appeared to be key values of the medium. This was confirmed and extended in direct discussions of the value and attraction of audio.

The most common comment about the addition of ambient sound to photographs was that it *brings the photographs to life*. The notion of enlivening a still image through sound appeared to have a number of facets. Ambient sounds were seen to add a general mood or atmosphere to the scene, as in the market square shown in Fig. 2.8, the café and restaurant of Fig. 2.3, or the seascape of Fig. 2.10. In some cases, a sequence of events unfolds within these clips, lending a further impression of the life going on within them. This is particularly evident in Fig. 2.5 where the depicted action of jumping into the swimming pool is introduced and completed by the sound clip. The sound of voices in a clip also served to bring the human subjects of a photo to life, and to reveal something of their thoughts and feelings of the moment through the recorded conversations. By these means, the photos themselves were seen to inherit a kind of human agency, even without the inclusion of deliberately staged or annotated commentary. As one person put it:

The photo speaks for itself (Viv)

In bringing a photo to life, sound also appeared to supply *additional stimulation to trigger people's memory* of the original events. This was shown by the fact that the families recalled and discussed the events in greater detail when looking at the photos with the sounds than without them. In fact the very act of listening to the associated sounds made the families look longer at each photo, and attend to the extra details of the event that were contained in the sounds. For example, sounds

might be made by objects outside the frame of the photo, such as by a waterfall behind the photographer in one beach scene. In other cases they contained repetitive or typical sounds of the location. For many of the participants these sounds seemed to trigger an existing sound-memory for events that might be just as salient as their visual-memory yet untouched by photographs alone. As one participant put it:

I have an audio memory in my brain. (Debbie)

Some audiophotos were even taken for the audio memory alone. For example, one family reported taking the audiophoto in Fig. 2.13 only for the sound of the marching band. Ordinarily, they wouldn't have taken a picture of a marching band at all. In this case the photo was used to index the sound, which acted as the primary cue for the family's memory of the scene.

Another major value of sound was *the element of humour* it often introduced. There was a noticeable increase in the amount of laughter experienced looking at the photos with sound rather than without them. Sometimes this was due to the strangeness of the sounds encountered in certain locations. For example, the music of Fig. 2.13 was found by the same family to be playing down a salt mine visited later in the holiday. At other times the humour was strongly linked to the embarrassment of being caught saying something revealing, as in the private conversation between one of the teenage girls in the trial and her friend about which members of a pop band they liked the most. This is shown in Fig. 2.14. The girls hated the audiophoto while the rest of the family loved it. This made for an interesting discussion at playback time.

Fig. 2.13 An audiophoto taken only because of the sound

Such embarrassment also extended to the commentary and staging of audiophotos which went wrong in funny ways. The all-time favourite photo shown in Fig. 2.11 is of this kind, where the son makes a mistake in his commentary on the mountain. Another example of a humorous 'mistake' was created by one father who tried hard to get his dog to bark for an audiophoto. The resulting picture of his dog is accompanied by increasingly desperate calls for her to bark, ending with some heavy panting as he holds the microphone to her snout. Hence, the general feeling was that:

> *Its much more fun with the sound. (Thomas)*
> *Its something that would prompt you to play around a bit and have more fun (Paul)*

Two of the families did begin to play around with sound recording to exploit the humourous potential of sound deliberately. For example, a surfing sequence was done as a spoof sporting commentary by one father for his son. This turned out to be especially funny to everyone except the son, because of the number of times it took him to catch the wave. The favourite photo shown in Fig. 2.12 of a friend playing the piano was later sent to the friend as a joke. It contained her own playing, which she did not know was being recorded, preceded by a fast classical piano solo taken from Rachmaninov's Piano Concerto Number 3. Finally, the same father and son team as in the surfing commentary, created the audiophoto shown in Fig. 2.15. This shows the father accidentally hitting his son several times, while putting up a windbreak on the beach. The ambiguity of the sound and image combination is exploited here to tell a false story-of-the-image. All these examples not only rein-

Fig. 2.14 An audiophoto recorded by two teenage girls

Fig. 2.15 A joke audiophoto made by a father and son at the beach

force the humorous value of sound, they also show how sound encouraged people to take more creative photographs and audio-photo combinations.

Perhaps the most revealing finding about the overall value of sound was the disappointment that was expressed at any sound loss. Whatever the value, it was missed when taken away. This happened on about three occasions when the microphone batteries ran out within the trial period, and when the audiocamera unit was removed at the end of the trial. One family had got so used to the habit of recording natural sounds with their photographs they asked to borrow the unit again after the trial.

As a final indication of the value of ambient sound, we asked families directly if they would like to use an audiocamera again if they could. Twelve out of the 14 trial participants said yes. Most participants felt that that sound capture would be a good option on a camera, to use whenever there was some interesting sound going on in the background. An associated perception was that this might increase the overall use of the camera, because there would then be two kinds of events, visual and auditory, to trigger recording.

2.5   Practice

It is clear from the trial that users of the audiocamera began to listen more atten-
tively to sounds in their environment, and to watch for audio as well as photo
opportunities. A number of difficulties in realising these opportunities were
revealed by comments on the process of capturing, editing and reviewing audio-
photos, and by observing review behaviours. These are drawn together in this
section to reveal the potential of the medium for transforming current practices,
and to identify usage problems which need to be addressed by the enabling
technology.

2.5.1   Capture

The prototype unit was generally considered to be too difficult to use, despite hav-
ing very few controls and settings. The requirement to switch on the dictaphone
separately to the camera meant that some people missed action shots or composed
their photos badly. The consensus view was that an audiocamera should be as auto-
matic as possible to preserve the point-and-shoot feel of a compact camera:

There was too much to think about for a snap (Debbie)
   There's no time to fiddle around (Sue)

A number of people commented on the difficulty of anticipating the quality of
the sound being recorded at capture time, based on a series of disappointments on
first hearing the results. People seem to make the same mistake with sound capture
as they do with image capture, in assuming that whatever they can see or hear will
be picked up faithfully by the recording equipment.

Participants also expressed a variety of opinions about the overall quality of
sound playback. In general, the sound quality achieved on the prototype unit was
adequate for most people, although our impression is that high fidelity stereo sound
would accentuate the feeling of being taken back into the recorded scenes by the
ambient sound clips. People tended to complain about sound quality when the
recording diverged too far from the remembered sound of the source; such as when
a boat engine was said to sound like a building site. We also found more tolerance
for poor quality sound clips by members of the capture group who share the same
memory of the recorded event. Presumably their memory helps with the interpreta-
tion of sounds which might be unintelligible to outsiders. From these effects it
appears that the expectation for sound quality is the same as that for image quality.
A relatively high baseline quality is expected by most users, but some users prefer
higher quality recordings.

2.5.2 *Editing*

The fact that sound clips were recorded selectively around the time that photos were taken meant that the families did not need to cut out extensive portions of recorded sound or to significantly alter the order and association of sounds and photos. Most of the families' requirements for editing were expressed to us in terms of exceptions to a "keep everything" rule. For example, we were asked to take out the bits of sound where "nothing happened", such as when Thomas recorded his sister waking up, or when one couple waited for a cow to moo. Other bits of sound which were singled out for deletion were either interruptions to the clip, such as the sound of the camera shutter, or sources of embarrassment. However, as discussed above, it was often the embarrassments and mistakes in the sound tracks which added humour to the photographs and we found unaffected members of the family arguing to keep these in.

Arguments within the family about what sound or images should be kept or thrown away were common, and reflected a deeper issue about editing rights. Individual members of the family appeared to have different rights to editing depending on whether they personally took the audiophoto. Family members also had different opinions about the audiophotos depending on who they were going to be shown to outside the family. The situation was analogous to deciding which conventional photos should go into the family album. Nobody wanted to look foolish within the formal family record, and everybody wanted to show their own version of an event to personal friends.

Another important factor for editing was that participants began to record longer sound clips with practice. This was particularly noticeable with one family who used the unit for two cycles, shooting a 36 exposure film each time. On the first cycle they recorded about 21 min of unedited sound, whereas on the second cycle they recorded about 36 min. Both sets of sounds were edited down to about 17 min. This implies that families may be willing to do more sound capture and editing to achieve better effects over time.

2.5.3 *Review*

Most of the comments people made about the value of ambient sound were made in the context of thinking about its impact for them personally. To this extent, we can say that there was a positive perception of ambient photographs enhancing *private review* by photographers and subjects in the trial (see again Fig. 1.2). Although we have not formally collected the feedback of a private audience on this material, readers might reflect on their own experience of these audiophotos in that role. Speaking for myself as a researcher reviewing the materials for the first time, the sounds added a good deal of context to each photo that helped me not only to interpret the photos more easily, but also to picture the mood, emotions or interpretations of the participants more effectively. I felt I was getting more of an insight into the

character of the subjects of the photos with the sounds than without them. Spoken commentary was particularly effective in this respect, even though it was much less effective for the participants themselves.

Other comments made about the value of sharing audiophotos begin to extend these findings to the practice of *social review*. For example, several people felt that distant family and friends would appreciate a spoken voice message with a photo, either as an *audiophoto postcard* from the capture location, or as an *audiophoto reprint* upon return home. It was thought that a spoken voiceover on a photo would be perfect for this purpose:

> *Sound (commentary) would be good if you aren't there to explain (Anna)*
> *Commentary would be good for sending to people who aren't there (Debbie)*

There was some ambiguity about people's perception of how they would like to share ambient audiophotos directly with others. At one extreme, some participants spoke about wanting to make up a full-blown slide show of images and sounds to show to family and friends as a presentation. At the other extreme, several people could foresee themselves showing the photos to others, such as a secretary at work, without the sounds, so as not to overcomplicate the process of sharing.

In fact this prediction came true within the methodology of the trial, since we left separate cassette tapes of sound clips with each family after the second home visits and before the feedback groups. At the feedback groups, some people reported showing their photos to others without playing the sounds. In between these extremes, the majority of people expressed a desire to selectively control or curtail the amount of sound that would play with each photo, so that they could adjust it for the audience at hand. As one participant explained:

> *A choice of sound would be good. It depends on the subject matter and the audience (Christine)*

The reasons for this ambiguity and sensitivity to the audience become clearer if we observe what happened when families began to share the audiophoto material within and between themselves. This happened on the second home visit when we gave families a chance to review their audiophotos together, and in the feedback groups when they were asked to show their audiophotos to another family. The first occasion represented a reminiscing situation in which photographers and subjects of the audiophotos interacted with each other. The second occasion represented a storytelling situation in which photographers and subjects interacted with an external audience.

By observing audio playback from a cassette tape in the home visits, we found that families automatically modified the way they handled and discussed the photos in order to accommodate the audio. They switched spontaneously from passing individual photos along the family group (without audio playback), to a process of one person turning over photos in the pack whilst others looked over their shoulder. This led to a more intimate group experience, with much laughter and conversation occurring over the rolling soundtrack.

However, this arrangement removed individual choice of how long to look at each photo. In fact, the photo holders tried hard to turn over the photos in pace with the sounds, to preserve the correspondence between image and audio. This some-

times caused the families to dwell on photos with long sound clips for longer than they wanted to, and to speed through photos without sounds too quickly. As one participant said:

> *There is no time to linger over the photos when you are listening to the sounds at the tape pace (Philip)*

The same effect was observed in the feedback sessions later when families tried to show each other their favourite audiophotos from a pack of selected prints. Each print with sound had an audio track number written on the back, which could be played from a minidisc player connected to nearby speakers. Prints were turned over by a single member of the authoring family rather than passed around the group. Furthermore, photo holders tended to turn over the pack systematically and let the audio clips play through, rather than using the minidisc controls to pause the audio or randomly access audio clips out of sequence.

Whereas in the home visits this behaviour had led to a good deal of laughter and enjoyment, in the feedback session it led to frustration with the audio getting in the way of conversation about the photos. Ambient sounds appeared to interfere with storytelling:

> *Its very frustrating 'cos I want to tell Viv all about where we were and what we were doing but I can't do that with the water noise on (Sue)*

Participants pointed out how different this felt to sharing audioprints within the family, when listening to the ambient sounds was more fun. Here, ambient sounds appeared to enhance reminiscing:

> *If you do this with somebody else its quite different. 'Cos the first one we heard as a family was hilarious really wasn't it? We had Lucy with us, and she was covering her face, and the girls were so embarrassed at the sound of their voices, and it was actually real fun. But that wouldn't come over in an (outside) group… (Viv)*

These findings show that introducing sound data into the existing activity of photo sharing dramatically affects its organisation and enjoyment. Ambient sounds seemed to enhance the experience of *reminiscing* with members of the original capture group, but potentially interfered with the experience of *storytelling* to others who weren't there at the time.

However, the extent to which sounds enhanced or interfered with photo sharing behaviours appeared to depend on the type of sound and its method of control. We noticed more frustration in families who recorded more voiceover with their photos. Ambient sounds and music appeared to be easier to talk over than forms of recorded speech. We also found that families were able to control the timing and pace of sound playback better on a PC album, where photo had to be clicked on before they began to play. The minidisc controls, although simple, were too much to think about in conversation where attention was focussed on the images and participants rather than the audio technology (see Chap. 6 for further details).

2.5.4   *Comparisons Between an Audiocamera and a Camcorder*

To explore the issue of whether many of these values are provided better by video recording, we asked participants at the end of the trial whether they would prefer a camcorder or an audiocamera at the same price. Eleven out of fourteen people preferred the audiocamera.

Since none of the families in the trial were camcorder owners, one might at first glance think that this result was due to the fact that they didn't understand the benefits of camcorders. However, all of the four families had experience of capturing and watching video or cine film, and were well aware of the costs as well as the benefits of camcorder use. Indeed they had deliberately chosen not to buy camcorders because of this. This suggests that their preference for the audiocamera reflected an informed judgement about its true benefits over and above a conventional camcorder. In discussing this preference, participants identified a number of benefits of audiophotos.

Several people in the trial said they just liked the aesthetic qualities of photographs better than those of video for capturing the mood of an event or moment. For them, enhancing the richness of a photograph with sound was inherently more attractive than switching to what they see as a completely different (video) medium.

Another factor in the preference for an audiocamera was that it allowed the user to participate more fully in the scene being recorded. Some participants contrasted this with the lack of participation when using a camcorder:

> *We don't go on holiday to create a film. We are much more instant people. We prefer to live life rather than record it (Geoff)*
>
> *One of the reasons we haven't got a video camera is that it is too intrusive and you end up not participating the event that you are in (Mervyn)*

People reported using the audiocapture unit selectively, like a point-and-shoot camera. This enabled them to get back into the scene quickly after use. Furthermore sound recording could be left on with the user still participating in the scene itself, as shown in a number of audiophotos where the user's voice was in the sound clip. This is in stark contrast to video recording where the user is usually distanced from the scene for the entire duration of (longer) recording periods.

One family member commented that the audiophoto medium is easier to master than video for beginners. This is substantiated by any casual look at the families favourite audiophotos which are remarkably good for a first attempt at using the new medium.

> *You need less skill and aptitude as a beginner to use this (audiocamera) than you would with a video. Video gives you more flexibility but aesthetically the results of the last audiophoto set were better than a first go at video. (Paul)*

The final appearance of video is critically dependent on the editing of raw video material, which is something camcorder users seldom get around to doing. In contrast we found that the editing requirements for audiophotos were very simple because of the way sound clips were recorded selectively around the time of image capture.

Some trial participants recognised that this might be an advantage of audiophotos over home video, as long as sound editing and handling didn't turn into a big job:

> *You had to be more thoughtful about what to select with the audiocamera. It's a one take opportunity.... With a camcorder you can take a chance and keep it running more. (Debbie)*
> *Its different from a camcorder. You are looking for the moment (Geoff)*
> *Editing is the downfall of video for personal use. Would this be the same? I couldn't be bothered to edit each film we produce. Half the time we don't get round to putting photos in albums (Paul & Viv)*

A final reason for the selection of an audiocamera in preference to a camcorder relates to the way the captured material is reviewed. All the trial participants without exception wanted to review audiophotos in the same interactive manner that they review ordinary photos – by discussing some audiophotos in more detail than others depending on the audience. While this was not always possible within the review scenarios experienced in the trial, a number of people seemed to recognise that this would be an advantage for audiophotos over watching home video presentations which are notoriously boring to outsiders.

2.6   Discussion

Although modest in scale, the audiocamera trial has revealed a good deal about the properties of ambient sound for memory and communication, and its role in domestic photography. In this section I draw together these lessons with reference to the diamond framework.

In general, we found that the families taking part in the trial adapted quickly to the use of a camera with sound, taking a large number of sound clips with their photographs in all kinds of contexts and forms. This device appeared to unleash a widespread interest in sound, as revealed in the art and radio projects mentioned at the beginning of the chapter. There was even evidence that this interest began to overtake picture taking. Hence the act of capturing sounds began as an adjunct to capturing pictures but soon became an equivalent or even primary task in the process of recording memories.

This transformation occurred as users became better attuned to the sonic properties of their environment, and the aesthetic properties of the combined medium. In reviewing this medium, families were delighted to discover that sounds brought many of the pictures to life, gave them mood and atmosphere, and provided additional information by which to interpret and remember the moment. This prompted them to think about how to improve the method of capture on the prototype unit, and the editing of sounds and associations afterwards, so as to consolidate and expand these values in more creative ways. This positive experience resulted in the majority of participants wanting to record audiophotos again, and led some people to regret the absence of the audiocamera when it was withdrawn.

Not all of these effects were due to the recording of pure ambient sounds taking place exactly as the photographs were taken. In fact a major finding of the trial was that ambient sounds are made up of multiple components lying before, during or after the images to which they relate. In fact, the ambient audiophotos contained music, conversation and commentary as well as environmental sounds, and comprised sound and image *sequences* that overlapped in different ways.

To the extent that these other sound types were recorded around the moment(s) of image capture, they remain part of the ambience of a scene or event and can still be distinguished from music or voiceover added later, and from the conversation about an image that occurs when it is shown to others. Nevertheless, the fact that these components occur or can be recorded as part of the ambient sound track adds to the complexity of its role in photography and to the expressiveness of the audiophoto medium itself. To understand this role and expressiveness better, it is useful to relate the study findings to the personal and social contexts for interacting with photographs, as shown in Fig. 1.2.

2.6.1 The Role of Ambient Sound in Personal Image Review

Ambient sound was found to be beneficial for the personal *recall* and *recollection* of a scene or event by those involved in it. This confirms the expectation of people in our pilot study. Ambient sound has also been found to be useful for the *interpretation* of images by those not involved.

In the first case, participants told us directly that the recording and playback of ambient sounds with images improved their memory of the original event. In particular, natural sounds, ambient music and background conversation all served to capture more of the atmosphere of the time, and appeared to help people recover or remember their feelings of being there again. This was most clearly illustrated in the seascape audiophoto shown in Fig. 2.10, where the simple sound of walking through shallow water helped to recreate the mood of an entire summer evening on holiday. Ambient sound undoubtedly provides more information about the scene that people may use to access their memory of it, but it is the *emotional* aspect of the sound that they appear to recognise and value most. This partly explains why voiceover was seen as less valuable than natural sound by the participants themselves. Although voiceover adds information about where and when the photograph was taken, it does little to mentally transport the participant back there, compared to ambient sounds of the moment.

Thus, ambient sound appeared to enhance memory by helping participants to emotionally re-live the event, rather than by furnishing details of the event to recollect cognitively. This effect is similar to that of *telepresence* in remote communication, where some systems and media are more effective in creating an illusion of being in a remote location than others (Lombard and Ditton 1997). Through photography, people can be transported to a previous time of their lives, and experience the

illusion of being back at a particular location and moment. I will call this illusion *retropresence* to distinguish it from telepresence. Through audiophotography, the illusion of retropresence can be heightened by the use of ambient sounds, which literally surround you with the same auditory stimulation you experienced at the time. The same effect will be created by the ambient sound track of a video clip, although its literal depiction of movement through the moment may be less conducive to an overall sense of retropresence at the moment of review. This is an issue for future research.

In the second case, the interpretation of an unfamiliar scene was assisted by the playback of sound or comments providing additional contextual information. This could be seen in the reactions of families to each other's audiophotos, and in our own reactions to the trial materials as they arrived for editing and analysis. The images were easier to read with ambient sounds providing extra clues about the identity of people and objects in the scene, and comments providing a personal description of their significance. By definition, these audiophotos revealed more of the story-of-the-photo (Berger 1982), either through the unfolding of sounds stretching beyond the moment of image capture, or directly through a voice comment. This made them highly suitable for sending to others who weren't there at the time, as observed by several participants in the trial. The same effects also enriched the depiction of people in the images. The inclusion of the voice of a subject was found to be particularly revealing of their character, as in Fig. 2.11. This is not surprising with hindsight, since the essence of someone's personality is likely to be conveyed at least as effectively by what they say or how they say it, as by what they look like.

2.6.2 The Role of Ambient Sound in the Social Interaction

There were mixed effects of ambient sounds on the social interaction around images. As with personal review, these effects depended on *who* was involved.

In reminiscing conversations amongst themselves, families were observed to laugh more and generally enjoy the experience more with sound than without it. The element of humour was a key value of ambient sound noted by participants themselves, but it manifested itself most in social interaction where families could share the jokes and tease one another about how they sounded or what they had said.

In contrast, storytelling conversations between families were less fun. Indeed, people complained that they had difficulty talking over the background sounds and, in some cases, wanted to switch them off. People were trying to provide a live narrative commentary on their audiophotos, outside the sounds and commentary recorded with the images. These appeared to compete with each other, at least in the case of recorded speech. Hence ambient sound appeared to restrict rather than enhance the process of sharing photos with those outside the capture group, because it drew attention away from what the author and audience wanted to say about the

image. The fact that the same sounds can be beneficial in one social setting and detrimental in another, is an important finding for audiophotography, to which we will return in later chapters.

In short, we have found that families take readily to the practice of recording ambient sounds with photographs, and produce audiophoto effects they prefer to conventional home video. This kind of audio adds considerably to what Schaverian (1991) calls the *life in the image*, by flooding the senses with some of the most distinctive sounds heard at the time of image capture. While music, commentary and conversation were all present in ambient sound to contribute to these effects, their contribution was usually secondary to that of environmental sound which was the most common type overall. In the next three chapters we consider each of these other kinds of sound in turn, as something added to the life of an image *after* it is captured. This will deepen our understanding of other kinds of audiophotos, before going on to consider the effect of screen and paper formats on the playback experience.

References

Berger J (1982) Appearances. In: Berger J, Mohr J (eds) Another way of telling. Butler & Tanner, London

Cusack P (2000) The horse was alive the cow was dead: sounds stories and people from the Lea Valley, East London. Audio CD. London Musicians' Collective, London

Cusack P (2001) Your favourite London sounds. Audio CD. London Musicians' Collective, London

Lombard M, Ditton T (1997) At the heart of it all: the concept of telepresence. J Comput-Mediat Commun (on-line serial) 3(2)

Schafer MR (ed) (1976) European sound diary. No. 3. The music of the environment series. World soundscape project. ARC, Vancouver

Schafer MR (1977) Tuning of the world. Random House/University of Penn Press, New York

Schaverien J (1991) The revealing image: analytical art psychotherapy in theory and practice. Jessica Kingsley Publishers, London

Truax B (ed) (1978) Handbook for acoustic ecology. No. 5 The music of the environment series, World soundscape project. ARC, Vancouver

Chapter 3
Musical Photographs

For me the voice captured the memory, the music captured the mood (John, audio annotation trial).

In a survey of 29 funeral homes in the south west of England it was found that people are choosing more popular music to be played at funeral services (The West Briton 2002). Traditional hymns are being replaced by classic songs such as Elton john's *Candle in the wind* and Simon and Garfunkel's *Bridge over troubled waters*, or by contemporary hits such as *Angels* by Robbie Williams. At the top of the funeral charts calculated in the survey were pieces from Hollywood movies about bereavement and loss. Number one was Bette Midler's *Wind beneath my wings* from the film *Beaches* while number two was Celine Dion's theme from *Titanic*, *My heart will go on*. Outside the top ten were occasional choices for more light-hearted music such as *I'm the king of the swingers* from *The Jungle Book*, *Party Atmosphere* by Russ Abbot, and *Always look on the bright side of life* from the Monty Python film *The life of Brian*. The point of this story is that ordinary people are rather good at selecting music based on its personal meaning to them or to their close family, and using that music to create particular emotional and symbolic effects. Those familiar with any of the examples above will recognize symbolic connections with enduring love and the celebration of life in adversity, and these meanings will be amplified when there is some personal association between the music and the person who has died. Indeed the funeral directors interviewed about the survey results, stated that people are turning to popular music to *personalise* the funeral service and make it relevant to their loved one who may not be familiar with church hymns and choruses.

This chapter explores whether people can do the same sort of thing with photographs. Perhaps music can be chosen to add personal meaning to a photograph and influence the kind of effect it has on oneself or others. For example, music associated with the time or place the photograph was taken might serve to enhance the memory of the event. These possibilities are examined here by encouraging consumers to annotate a small set of their own photographs with music from their own personal collections. However, before reporting the details of this study, we pause briefly to consider what is already known about this activity from previous studies or design practice.

3.1 Related Work

Very little academic research on music has considered its sentimental value and use. Research questions have centred more on how we perceive and remember pure musical phrases (psychoacoustics), and on how musical tastes and styles differ across cultures and time (musicology). For the best insights into the personal value of music, we need to turn again to popular culture.

An experiment in personal music selection has been carried out through an ongoing BBC radio show called Desert Island Discs. The show celebrates its 73rd anniversary this year (2015) and must surely be the longest music-psychology study in history. The format for the half hour show is to invite a celebrity guest to choose eight records they would take with them to a desert island. The guests are interviewed about their lives in the context of the music selection, and always give a reason for each piece of music before it is played (in shortened form). Only four main interviewers have hosted the show over the years: Roy Plomley (1942–1985), Michael Parkinson (1986–1988), Sue Lawley (1988–2006) and Kirsty Young (2006- present day 2015). The very success of this format says something about the appeal of hearing about other people's choice of music as well as hearing the music itself. It also reveals the suitability of music as a vehicle for talking about the past.

In commentaries on the show, Roy Plomley and Sue Lawley discuss patterns in the choice of music across shows (Lawley 1990; Plomley 1975, 1982). The top 10 or 20 music choices appear to change over time, particularly as new popular music replaces old. Classical music selections are more stable and include well-known pieces such as Beethoven's *Ode to Joy* (Symphony No. 9, last movement) and *Pastoral* Symphony (No. 6), Mozart's *Clarinet Concerto* (second movement, Adagio), Dvorak's *New World Symphony* (largo), Elgar's *Pomp & Circumstance* (March No. 1), Brahm's *Violin Concerto* (last movement) and Rachmaninov's *Piano Concerto No. 2* (last movement). However, the reasons for choosing the very same piece of music can be quite different across people. Sometimes the music is chosen for its connection with the past and ability to conjure up a happy memory. At other times, the music is chosen primarily for entertainment or because of a preference for a particular composer or performer. In all cases the choices are *personal* and idiosyncratic. They therefore reveal something about the personality of the guest that might otherwise have been difficult to elicit by direct interview alone.

These points can be illustrated by some music choices made by Brian May of the rock group Queen, speaking on the show during the week I was originally writing this chapter (broadcast on BBC Radio 4, Sunday 15th September 2002). His choices fell into three categories. Some tracks were chosen because of their aesthetic qualities and ability to lift his mood. These included:

- *Since you've been gone*, Rainbow – "This always gets me and lifts me"
- *Highway to Hell*, AC/DC – "This is pure rock. It clears you out and gets you back to basics".
- *We will rock you*, Queen – "It brings people together and makes them feel strong and uplifted".

It turned out that Brian had been subject to depression in the past, and wanted to use music to avoid this on his desert island.

Other tracks were chosen because of their association with important people in his life. These included:

- *Maybe baby*, Buddy Holly – "An early inspiration"
- *To know him is to love him*, Anita Dobson & Brian May – "My darling wife. I couldn't be without her".

Discussion of these two choices revealed a secret love of pop music in a quiet and conservative home setting, and the break up of a first marriage.

Finally, three tracks were chosen because of their nostalgic value. Each one reminded Brian of a different time in his life:

- *Saturn-Bringer of old age*, Planets Suite by Holst – "I wrote a monolog when I was a kid and used to perform to this"
- *Back on my feet again*, The Babies – "My driver used to play this as a pick-me-up on the car radio on the way back to the hotel after an all-night drinking session".
- *Tracks of my tears*, Smokey Robinson and the Miracles – "This took me all through my college days after a break-up with my girlfriend"

The meaning of each piece is indicated by its associated quote, and reveals something about Brian's past attitudes and behaviour. The Planets Suite movement relates to his love of astronomy, which he developed as a child and later studied at university. The monolog he wrote was about the night sky in winter. The Babys track refers to a period of excessive partying on tour with the band, while the Smokey Robinson track relates to his time at university. He played *Tracks of my tears* when getting over an important relationship with a girlfriend, and pretending to be OK when he was not. The lyrics of the chorus articulate this feeling exactly, and expressed for him something he couldn't put into words himself. This also prompted him to talk about his reason for wanting to write songs, as a way of connecting to people 'Heart to heart'.

Chorus:

> Take a good look at my face
> You'll see my smile looks out of place
> If you look close its easy to trace
> The tracks of my tears

These examples suggest that music has a variety of meanings and associations in one's life, and can be used privately to trigger memories and alter moods. Furthermore, the *discussion* of these meanings and associations seems to be revealing of personality. All this is relevant to the photographic context in which images serve a similar function. In this context it is interesting to reflect on what the associated images might be for Brian May's music selection. While the people selections could be linked to standard photographs of the performers, such as you might find on a CD cover, the nostalgic selections could not. It would be more meaningful to

combine the *Saturn* music with a photograph of Brian as a child, or of the night sky, than with a picture of an orchestra performing the Planets Suite. Similarly a photograph of his driver or car would be best with *Back on my feet again*, while a photo of his long lost girlfriend or his smiling face at that time might be appropriate to *Tracks of my tears*.

Unfortunately, the format of this radio show does not support such speculation, and there is little scientific work on the connection between music and images. In a recent conference on *Musical Imagery*, the ability of people to hear and sing tunes in their heads was discussed and studied, as was the ability to read music or other visual symbols and hear inner sounds (Godoy and Jorgensen 2001). However, the ability to hear actual sounds and mentally picture a scene was not discussed in any detail, nor was the ability to look at an image and imagine the sounds or music that go with it (although see Mountain 2001 for an exception).

That people can do both these things is indicated by some of the findings in Chap. 2 of this book and by a book on music in modern art (Maur 1999). Hence, Maur (1999) shows that a number of famous painters and musicians were heavily influenced by each others work, taking inspiration from one medium and translating it into the other. For example, Kandinsky was inspired by Schoenberg's use of dissonant sounds to develop a more abstract form of cubist painting than had been done before. Furthermore, in Chap. 2 of this book, we saw that ambient sounds could either trigger a visual memory of an event or be recalled from a photograph of the event. The history of the cinema also attests to the ability of musicians to set music to silent films. In fact many cinema pianists and organists of the time could improvise music which watching the film live. This is exactly the kind of thing that is done today over a longer timescale by film music composers. As this is such a relevant activity to the selection of music for still images, we will consider the film composers' skill as a final backdrop to our study.

Composers in general talk a lot about finding inspiration outside music for their compositions (Mountain 2001). They draw on analogies with human emotions, historical characters and stories, animals and landscapes. For film music composers, these analogies are given visual form and pace through film footage. Note also that this footage is generated in relation to a screenplay and accompanied by ambient sounds and dialogue telling the story of the film. The process of producing such material takes time, and can be more or less integrated with the creation of the music. Hence we find that a perennial concern of film music composers is which comes first; the image or the music? This issue is discussed in a series of interviews with famous film composers, talking about their work (Russell and Young 2000). While it is more usual for a composer to be shown rough-cut footage as a stimulus for the music, some directors are open to discussion with the composer about the intention of a scene before it is shot or edited. For example, Elmer Bernstein was able to create a 'temp' score for Martin Scorsese, before editing ***The age of innocence***:

> *Scorsese is one of those directors who will talk through the sequences and the way he would like the music. He said that listening to music is what made him want to become a director. For him the image and the music are inseparable. He finds it very difficult to edit a film cold,*

so he will tend to bring in the music as he's editing the film. In the case of The age of inno-
cence, I wrote some themes for him which he liked and I then suggested we make a temp
score based on these themes, so he was always working with what became the final score
(*Bernsein quoted in* Russell and Young 2000, p. 43).

Nevertheless, one of the key skills developed by film composers is to look a sequence of images and imagine the kind of music that goes with it. For some film composers this can be an immediate experience:

Whenever I see a film, a good film, I feel the orchestration right away (*Maurice Jarre*
quoted in Russell and Young 2000, p. 55).

The intention of the music is usually to define and accentuate the *mood* of the scene or characters, and to assist the *narrative* flow of the film in telling the story. Individual characters may have separate themes associated with them, which communicate their own emotions to the audience. At other times, the music may cut against the (primary) emotions of the characters in the film, to communicate another (secondary) emotion to the audience – as in the warning of an unforeseen danger. These techniques have been shown to be effective both in influencing the emotional experience of the audience and their understanding and recollection of the film (e.g. Boltz 2001; Thayer and Levinson 1983). Finally, the music is only one of the sound elements in a modern film and must be combined with ambient sound, dialogue and voiceover. The control and balancing of these various elements is the job of the sound designer who must understand each contribution and fit them together in time to the visual events (Sonnenschein 2001).

Whether or not these kind of skills can be applied by ordinary people to domestic photographs remains to be seen in the rest of this chapter and book. However, we now know from this review that such skills can be acquired, and might be combined with already well-developed abilities to select music for personal and social effect. So let us put this to the test.

3.2   The Audio Annotation Trial

Using a methodology similar to the audiocamera trial in Chap. 2, we gave a small number of families the opportunity to annotate their photographs with music and voiceover recordings. The study was carried out with Adam Drazin and Tony Clancy, and was designed to provide insights on both musical and talking photographs. This chapter reports the findings on music annotation while Chap. 4 reports the findings on voice annotation.

The main difference between this study and the audio-camera trial is that participants were not given a technology prototype for an unsupervised trial period. Current methods of annotating photographs with digital music or sound recordings were considered too complex to use by untrained consumers or too simple to accommodate their requirements. Indeed, one aim of the study was to generate user requirements for such a system. Instead, we visited families at home to supervise

their *choice* of music and photo combinations that we later digitized and assembled into audiophoto albums. We also took in audio equipment to record voiceover narrations for other photo content that we again digitised and assembled later. This exercise comprised a pseudo-trial period, after which there was a return visit by families to the Labs. At this feedback session, families reviewed and discussed electronic versions of the albums they had designed, in a variety of forms.

3.2.1   Participants

This approach was followed with eight households in the Bristol area. The households were all PC-owning families in which one member had a special interest in photography or music. This interest was measured through attendance at a photography, album making or music club.

 These criteria were used to ensure some level of technological sophistication, together with a propensity to benefit from either voice or music annotation of photos. For example, people involved in photography and album making might be willing to consider using voice annotation with photographs. Alternatively, those with an interest in music might consider playing music from photographs as well as from vinyl records or CDs. We also hoped to assess the general appeal of these activities, by involving other members of the household, and looking at the reactions of photo enthusiasts to music and music enthusiasts to voiceover.

3.2.2   Home Visits

Each of the eight households generated one musical and one talking photo album in the home visit, apart from one couple who generated a pair each on separate visits. There was usually one main contact and interviewee from each household, with partners or other family members present at about four of the visits. The first part of the home visit involved a general discussion about photo and music use in the home. This was designed to explore photo album making behaviours and the role of music in remembering the past. The concept of audio annotation of photos was introduced in the second part of the home visit for comment. Participants were then asked to select a set of photos to annotate with music, and a set to annotate with voice. We left this choice open as far as possible, to allow people to direct the activity towards photo content they felt might benefit from audio annotation. For practical purposes we asked people to limit the size of each set to about 10 photographs.

 Once the photographs for music annotation were selected, participants began to work through the set with reference to their own music collection available in the house. Sometimes a piece of music came to mind immediately as they looked at a particular photo or group of photos. In these cases, they would search directly for

that piece to hand over to us later. At other times, participants would be less sure of the particular piece they wanted, but usually had in mind the mood and sentiment it should express. In these cases, they would browse parts of their collection looking for inspiration or ask us for advice. When answering these questions we encouraged them to make their own selection. Each choice was documented on a simple two-column form on which we wrote photo names in the first column and music track names in the second. The discussion of music choices was also recorded to capture the participants' reasons for particular selections. A different procedure was used for voice annotation, and is described in Chap. 4.

3.2.3   Album Creation

At the end of each home visit we took away the music and voice album specification forms, and borrowed all the music and photo materials relating to them. Music samples and voiceover materials were digitized by recording them into a sound editing application on a PC. Photographs were scanned into an image editing application. The composite sound and image files were then assembled into audiophoto albums on the PC, using Macromedia Director. This is a professional multimedia authoring system that had sufficient flexibility to support the variety of sound and image combinations chosen by participants in the trial. It is worth noting here that simpler, photo album software of the day provided either no support for audiophotos, or only poor support in the form of one sound per album or one sound per image.

This procedure raised a number of music copyright issues. Music copying within the European Union is allowed for personal use, and could be supported in musical photo editing software. However, the distribution of musical photographs to other people begins to infringe copyright law by denying revenue to artists from recipients. This would inhibit the development of photoware for musical photographs. The study procedure itself was legal in the sense of facilitating the composition of personal musical photographs for live sharing and discussion. However, it left participants with a dilemma about how to share their new music-photo albums after the study, and me with the job of requesting music licences for all the music in this chapter. This is an important issue to which I will return later.

The composition process resulted in the creation of 9 musical and 9 talking photo albums containing about 10 photographs each. In addition to these personal albums, we made up a demonstration set comprised of one musical and one talking audiophoto from each album. This set was used to create screen and paper-based demonstrations, showing a variety of non-PC formats for the same material. These included presentations on a TV, a palmtop PC, a tablet PC, and in the form of an audio enabled print, card and physical photo album. These are described in further detail in Chap. 6, which deals with the findings of this comparison.

3.2.4   Feedback Groups

Three separate feedback groups were then held for two or three families at a time in HP Labs Bristol. Each group session started with pairs of participants reviewing their own album material together. Pairs sat up to a flat panel display with active stereo speakers on a café table, and clicked through their PC-based albums using only a mouse. A laptop computer driving this experience was placed out of sight beneath the table. The interface to each album was kept simple and uniform, as shown in Fig. 3.1.

Most album pages showed one photo at a time at nearly full screen size, above a set of multimedia control buttons displayed at the bottom of the screen. Forward and back buttons had to be clicked to advance or reverse the display of pages, and audio tracks played automatically on presentation of their corresponding page. Sound could be muted using the mute button, but not paused. Where one audio track was specified to play over several photographs, it simply played continuously while any of those photographs were displayed. In a few cases where participants had specifically requested it, multiple photographs were displayed on the same page to one or more audio tracks. After reviewing each other's albums, pairs were asked to select their favourite musical and talking photographs from each album. They wrote down their selections with corresponding reasons on a form.

The initial sharing period was followed by a group discussion of the uses and values of music and voiceover with photographs. This was an opportunity for

Fig. 3.1 Screenshot of a PC-based musical photo album

participants to reflect on the experience of creating and reviewing their personal albums, and comment on the problems and benefits involved. Groups were then shown a demonstration set of audiophotos in a series of alternative screen and paper formats. At the end of this demonstration each participant expressed their preferences for musical and talking photo formats on a form, before discussing their answers.

Findings on the *format* discussions for music and voiceover are described in Chap. 6, along with a similar comparison made for ambient audiophotos in the audio-camera trial. Findings on the *content* selections and discussions are described separately for music and voiceover in this chapter and the next. The analysis in these chapters follows that used in Chap. 2 for ambient sound, where technical features of the audiophoto corpus are described ahead of the users' perspectives on the material. In fact, the analysis section headings are kept the same to enable easier comparison between chapters on each of these three major audiophoto types. Conversational audiophotos are treated somewhat differently in Chap. 5 where conversation is explored as another method of eliciting voiceover narration.

3.3 Musical Photos Captured

A total of 48 music clips were combined with 94 photographs to make up nine music-photo albums. The range of musical and photographic material used was extensive, and people approached the task of selection in a number of different ways. For example, some participants chose a recent collection of photos and set each one to a separate piece of contemporary music. Others chose an older photo collection and set it to a single piece of nostalgic music. Some people even selected ten of their favourite photographs and ten of their favourite pieces of music and tried to match them with each other. This made for a very interesting set of audiophotos, albums and discussions, as participants began to explore the possibilities of the new medium and find its value.

3.3.1 Types

The different types of music selected are shown in Fig. 3.2. This chart was generated by classifying each music clip as belonging to one or other of a standard set of a music categories used by music shops. It shows that classical music was the most commonly selected type of music for this group, but that a wide range of other types were also used.

Three typical example combinations of music and photos are shown in Figs. 3.3, 3.4, and 3.5. The photographic components consist of a landscape shot, a group photo and an individual head-and shoulders portrait. These are combined respectively with a piece of classical music, a modern dance track and an old-
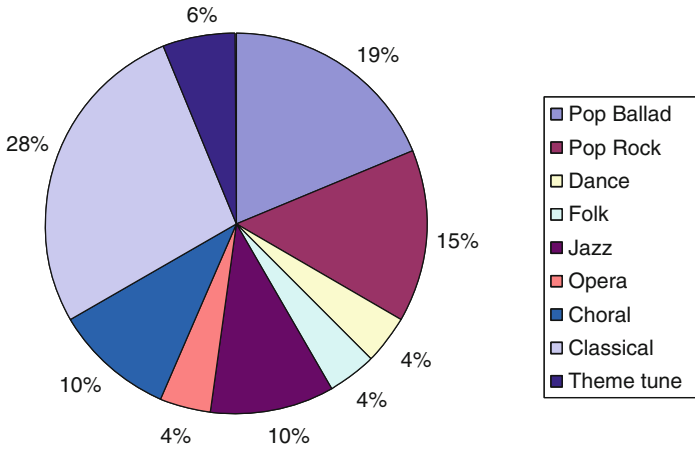
Fig. 3.2 Types of music chosen with the photographs

Fig. 3.3 Ocean view to Dvorak's 'New World Symphony'

time pop ballad. Hence the audiophoto effects of each are very different. Figure 3.3 conveys the mood of a rolling sea that occasionally crashes on the rocks in the foreground – as the movement of Dvorak's New World Symphony reaches a series of crescendos. Figure 3.4 communicates the energy and fun of a night out with the lads in some holiday location – through the pounding of a

Fig. 3.4 Group of friends to 'Hey Boy Hey Girl' by the Chemical Brothers

Fig. 3.5 Wife's face to 'Portrait of my love' by Matt Munro

dance beat behind the series of happy faces. Figure 3.5 shows the smiling face of an older lady serenaded by a romantic ballad about beauty and love. We will return to each of these examples later, but for now it is sufficient to note the variety of content used in the study and the powerful effect of the music in modifying and accentuating the perception of each photograph.

3.3.2   Durations

The duration of audio clips used for music was understandably greater than the duration of audio clips used for ambient sound. The average duration of a music clip was about 3 min (182 s) compared to 24 s for an ambient clip. However, the spread of durations was similar as shown in Fig. 3.6. Ninety five percent of the clips varied from about 1–5 min. This variation is partly due to inherent differences in the length of published music tracks, especially those including short dance or jazz numbers and longer classical movements. In addition, this distribution was skewed towards the lower durations, by people who expressed a preference for specific parts of a track to be combined with their photos.

The request for parts of a track to be associated with a photograph was especially important in the case of long classical pieces or songs with extensive lyrics. This is because the mood and meaning of these kinds of tracks can change dramatically within the track. For example, the symphony represented in Fig. 3.3 starts off quietly and builds to a loud finale in 7 distinct cycles over 6 min. By including all these cycles in this audiophoto the music echoes the repetitive movement of the sea depicted in the photograph. Selecting the quieter or louder part of each cycle would have emphasised the calmer or more dramatic aspect of the sea respectively.

A further consideration is that each piece of music may not be played back in its entirety. This can be seen from the section on music-photo review below, which shows that participants often skipped onto the next audiophoto after just a few seconds of conversation about it. This means that the most important *work* of each track, in triggering a memory or conveying a mood, is done in the opening sequence.
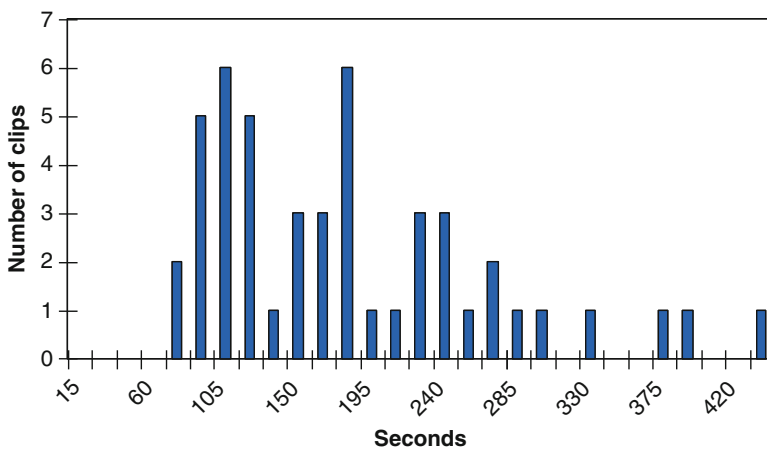
Fig. 3.6 The duration of musical sound clips

3.3.3   Associations

None of the photos selected in the study were combined with more than one music clip. This contrasts with the ambient sound photos in Chap. 2 where a small subset were combined with several sound clips (see again Fig. 2.6). Instead, the photos in this study were combined with a single music clip, as shown in Fig. 3.7. In two thirds of cases there was a one-to-one association of sound to image. In the other third of cases, there was a one-to-many association, meaning that a single piece of music was combined with two or more photographs.

The ocean scene and group photo shown in Figs. 3.3 and 3.4 were both associated with a single music track. However, the portrait shown in Fig. 3.5 goes with an additional full length photo of the same woman. Both were associated with the same Matt Munro song. Usually a pair of photos like this depict the same scene or subject from different angles, in landscape and portrait mode, or at consecutive points in time. Other examples of *musical pairs* of photos in the study were a pair of animals taken on the same trip, a pair of buildings found near to each other, and different parts of the same waterfall.

Some associations of sound and images involved more than two photos. These tended to be sequences of images relating to the same episode in time or even to an entire photo album. For example, Fig. 3.8 shows four photos taken on a trip to Japan, combined with a single song by a Welsh male-voice choir.

The fact that music was often used to link pairs, groups or albums of photographs together in this way is reminiscent of the use of film music to achieve continuity between shots or scenes in a movie. Conversely the combination of music tracks with individual photographs serves to set them apart from each other, even when they belong to the same set or episode. We turn now to the participants' own reasoning about these choices, and their reactions to the resulting material. This will show how conscious they were of using particular techniques, and what value they perceived them to have.
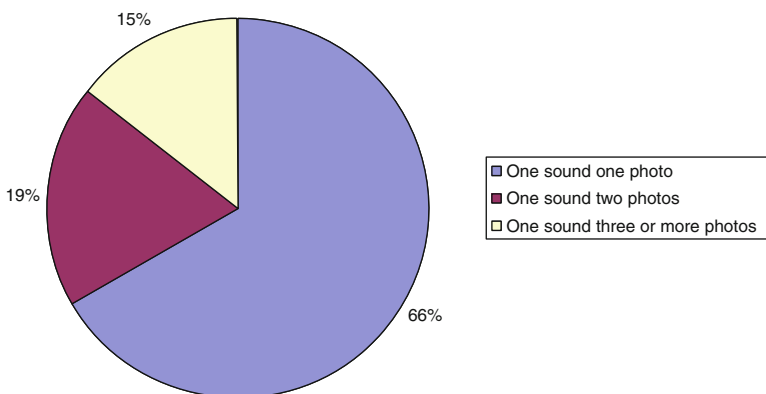
Fig. 3.7 Audio:image associations in the musical photo set

Fig. 3.8 Four images of Japan set to a male-voice choir

3.4   Reactions

3.4.1   Initial Reactions

Most people were extremely skeptical of the value of combining music with photo-graphs at the beginning of the study. As the quotes below indicate, they tended not to associate the two media in their minds and struggled to imagine what the combi-nation would feel like. Consequently, the music annotation exercise was performed under sufferance, primarily to please us.

> *I wouldn't naturally put them together. I'm not convinced* (*Rob*)
> *I don't know. I've been trying to think about it after we talked the last time. I'm looking forward to seeing it because at the moment I can't* (*Traci*)
> *I can't picture it in terms of what we are doing at the moment* (*Byron*)
> *I don't tie up photographs with music specifically. I don't like concerts on television* (*Gordon*)

Only two participants spoke favourably about the possible benefits of combining music and photos. One of the older men in the study drew a comparison with film

music, while the youngest man of the study immediately saw it as another way of organising and playing his music by its date of release:

If one sees a film with the soundtrack off, *however emotional it is*, *one can watch it fairly dispassionately. But you put the music on with that film and watch it and however stoic you may be*, *you'll find tears coming into the corner of your eyes or get other emotions. So certainly music can have a terrific effect in that respect and if you could find the right music to go with the photos I could well imagine that it would be more nostalgic*, *much more evocative* (*John*).

It sounds good yes, *so I'd choose a song for each photo would I? If you've got a photo during the time you were into that sort of music*, *like at school or something*, *that might work* (*Nick*).

After reviewing their own music photo albums on the PC for the first time, we asked participants what they thought of them. Most people were pleasantly surprised and a few were positively shocked at how much they liked them. The majority of comments were positive, and mentioned the power of music in capturing a mood or atmosphere for the photos. This can be seen in the opening quote for this chapter, and in the comments below:

I liked the music ones (*Traci*)

I thought they worked surprisingly well (*Byron*)

I thought it was really good actually. I think putting pictures and music together does actually work (*Nick*)

I think music does give the atmosphere (*Gordon*)

The most dramatic conversion was experienced by Rachel. Rachel was quite outspoken at the beginning of the trial about the lack of appeal of combining music with photos and the difficulty of associating them in her mind:

It doesn't appeal to me personally really. I don't necessarily find the photos that I take are evocative of a certain musical experience. Although some pictures do evoke some kind of music or sounds I wouldn't necessarily want to sit here and have it played at me. Pictures are good enough in their own right(*Rachel*)

However, for her album, Rachel chose two family photographs and a series of 4 landscapes. The landscapes were set to classical or choral music that appeared to match the images perfectly. One of these musical landscapes is shown in Fig. 3.9. The combined effect of listening to these photographs was to change her mind completely:

I was impressed. I came thinking I wasn't going to be. The whole concept was quite moving really (*Rachel*)

As experimenters, we also experienced something of the same conversion effect. Despite anticipating the potential value of music, rather like John quoted above, and being familiar with ambient sounds and commentary on photographs, nothing prepared us for the range of emotions triggered by some of the families' material. We found John's portrait of his wife Betty shown in Fig. 3.6 to be particularly touching, and were amazed at how well Rachel and others were able to match the mood of an image with appropriate music.

3.4.2 Favourites

After reviewing each other's music-photo material in pairs, participants were asked to choose a favourite from each set. They were also asked to write down the reasons for their selections. In this section we concentrate on the selection of personal favourites from people's own music-photos, and contrast this briefly at the end of the section with their selections from other people's material.

Preferences were based purely on aspects of the content, rather than on the type of music, the duration of the clip or the type of sound-to-image association. Two distinct qualities were mentioned as reasons for preferring particular music photos over others. The first was the goodness of fit between the image and the music. Wherever the music appeared to capture and amplify the mood of the photograph it was liked. For example, Rachel liked all her music-photos but singled out that in Fig. 3.9 as her favourite. Her reason was that:

> *The music and images work together so well. The Rutter music perfectly reflects the mood of the photos* (*Rachel*)

A similar reason was given by Byron who chose the Japanese photos and Welsh music shown in Fig. 3.8 as his favourite. He singled out the first image of a wooden walkway as working particularly well with the music:

> *Because the music seems to fit the tranquil mood of the garden, despite the Japanese/Welsh mix* (*Byron*)

Fig. 3.9 Sunset over the Isle of Skye to 'A Gaelic Blessing' by John Rutter

The second quality that appeared to distinguish favourite music photos was the personal significance of the image or music. The more significant the music or photo, the more it was liked as a combination. For example, the group photo shown in Fig. 3.4 was Nick's favourite because:

Its one of my favourite photos anyway and I thought the music went very well with it (*Nick*)

In another example, one participant chose a favourite music photo because of the significance of the music. It was a Christian rock song called *Surrender* and was combined with multiple photographs from a holiday in Italy. The lyrics of the song were important to him because they described his recommitment to God following a recovery from cancer.

In contrasting the choice of favourites from within one's own or another person's music-photo set, we found little consistency of selection. In other words, the same music-photos tended not to be chosen by both members of a pair of participants. Instead, the choice of favourites was highly idiosyncratic. This is understandable in terms of the personal significance of the material, since everyone had his or her own reasons for judging it to be significant to them. For example, Nick's girlfriend Kerry chose a different favourite from Nick's music-photo album than the favourite Nick chose for himself. Nick chose the shot of his friends shown in Fig. 3.4 while Keri chose the shot of Nick shown in Fig. 3.10 below. She considered the latter to be a good audio-portrait, much like the favourite ambient-photo shown in Fig. 2.11, and felt it captured Nick's personality and love of music well:

Fig. 3.10   Nick on his record decks, set to a DJ re-mix

It's very real, even though this was taken over three years ago, you will still find him doing this now- everyday! (*Kerry*)

However, we expected more consistency in the judgement of fit between a photo and a music track. As we discussed this with people later in the feedback groups we found that this too was highly subjective. It depends on the mood they read into the image, which is coloured by their experience of being there in the case of one's own photographs. This might be quite different to what is read into another person's photograph from the image alone. For example, the connection between the music and image in Fig. 3.11 is somewhat obscur, but was chosen as the authors' favourite music-photo. To Harry who took the photo, the music went with the mood of East Berlin when he visited it before the fall of the Berlin wall. The multi-lingual cabaret song is both aggressive and inviting, much like the sign and the gateway in the photograph, although these allusions might be difficult to appreciate for someone who never experienced the city at this time.

3.4.3 The Value of Music

After selecting their favourite music-photos, participants were asked what they liked about adding music to photographs and what they would use it for. Taken together, these responses suggested four key values of music perceived by families in the study.

Fig. 3.11 The Brandenberg Gate in Berlin, set to the film song 'Cabaret'

First, adding music was seen as a way of re-creating and reinforcing the mood of the scene, as experienced at the time of capture. The music for the ocean view and the group of friends in Figs. 3.3 and 3.4 appears to have been selected for this effect, as well as the choral music for the images of Japan and Skye in Figs. 3.8 and 3.9. It provides a similar value to that of ambient sound in adding atmosphere and bringing the photo to life, although in a more controlled and personal way. This is because music can be selected to be more representative of the *feelings* of the moment rather than of the literal *atmosphere* of the moment, as given by ambient sound. Hence there was a sense for many people that they were often selecting the music for themselves, and that the choice of music was personal and not necessarily to be shared or understood by others:

> *Well I chose my music for something I liked, something I enjoyed listening to, but also something that would be sort of upbeat and positive and in general reflect the mood of the photos or the place of the photos* (*Vicky*)
>
> *The music would be evocative of your own feelings at the time and it would be your personal choice. When I go home and show this to my husband he wouldn't necessarily fit those particular pieces of music with this picture* (*Rachel*)
>
> *To me its just a personal souvenir that's been enhanced by the music, and as such it wouldn't be something that was a showpiece to be put in a museum or an art gallery. It's something which is personal and helps to improve it; to make it real* (*John*)

As with the use of ambient sound reported in Chap. 2, there was some attempt to subvert the atmospheric value of music, to create effects that deliberately conflicted with the mood of the photograph. This led to a second value of music in creating a new mood for the photo, designed more for amusement or entertainment value than for capturing a past mood. Harry's Brandenberg Gate audiophoto shown in Fig. 3.11 has something of this value, although the contrasting effect of sound and image may also be true to Harry's perception of Berlin at the time. In another example, a pair of mopeds in Italy were set to a piece of opera by Verdi. This was done by Gordon as a deliberate joke, to undermine the seriousness of Verdi's music with a modern image, taken near to the place in which the music was first performed. A more straightforward use of music for entertainment, was also discussed by Vicky as a way of making the photo album as interesting as possible for others:

> *Music sort of adds to the dull part. If you are not really into photo albums you can just sit back and go with the music* (*Vicky*)

A third value of music was to evoke memories by association with nostalgic music of the time the photo was taken. Music associated with a particular holiday or person was said to have this kind of value, and would remind people of a whole collection of events around a time period. A good example of this is shown in Fig. 3.12.

Figure 3.12 is a picture of Traci's daughter taken on a trip to Bristol zoo, but set to a song from her favourite video of the time ('Veggie Tales'). This combination not only makes for a good audio-portrait of the daughter, it also reminds the mother of a period of home life which was punctuated by this music several times a week. Music listened to repeatedly at home, on car journeys and on the radio was also mentioned as having this nostalgic value.

Fig. 3.12 A daughter set to the song 'Thankfulness' on a children's video

Finally, a fourth value of music was to communicate some hidden message through the music itself. This use of music was surprisingly common in the music-photo collection, and second only to the selection of music to match a mood of the moment. In some ways it is the complete opposite of this latter value, since here the music is designed for someone else rather than for oneself. However, the effects may be just as subtle and poetic.

This can be seen from a series of what can only be described as 'love photos' comprising a picture of the loved one together with a song with romantic lyrics. Figure 3.5 is a good example of this, created by a husband for his wife. A more contemporary example is shown in Fig. 3.13 and was created by Nick for his girlfriend. It is a picture of his girlfriend set to the track Wonderwall by Oasis and contains lyrics that might have been difficult to speak to her directly. Other love photos were created by the wives in the study for their husbands, including a picture of Rachel and her husband set to Alison Moyet's *Is this love*?, and one of Jane and Byron set to *Wind beneath my wings* by Bette Midler. We also found that parents sometimes used love songs with images of their children, as with a pair of photos of Jane and Byron's daughter set to *Your song* by Elton John.

In all these cases, the lyrics of the songs took on extra significance when paired with the photographs, and appeared to speak about the people in the photos. Outside this class of romantic music-photos, we found the same use of music to communicate more subtle meanings. For example, Dvorak's *New World Symphony* was chosen with the ocean image in Fig. 3.3 not only for its mood, but also for its connection to the New World:

So I think that goes with that because it is the New World. It's California. Its all the things that go with the States (Jane)

Fig. 3.13 A love photo, comprising a picture of a girlfriend set to 'Wonderwall' by Oasis

3.5 Practice

Although participants in the study did not have chance to make their own music-photo albums, the acts of specifying content, reviewing it together and discussing it later gave rise to a number of comments about the album creation process. In this section we examine these comments in the context of what people told us about their existing music and photo practices, to understand how music and photographs might be used together in the future.

3.5.1 Capture

The most obvious way to create music-photos is to annotate existing photographs with music. Indeed this was the method promoted in our study. However, it was mentioned above that a number of participants approached the task of combining music and photographs somewhat differently. In particular, some of the music enthusiasts in the trial pulled out favourite pieces of music to combine with favourite photos. This shows that it is possible to approach the combination of music and photographs in two ways, either by adding music to photographs or by adding photographs to music. Both views are represented in the following two quotes. These indicate the spread of opinion on whether photographs or music should be the

primary medium. Eventually, these views could lead people to capture photographs or music differently, in order to improve the match with the complementary medium

> *As Byron and I have both said, if we'd had more time we'd have picked something more appropriate. But its definitely the photograph itself which is the centre point. It's not the talking or the music – those are purely supplementary* (*Jane*)
>
> *This is what I was getting at before. Sometimes you've got to approach it from one side, sometimes the other. If you've got a definite visual image in your mind then you may decide to fit music to it. If you've got a definite musical image, say the 20's or something like that, then you might find a photograph of the Charleston going on, to fit with the music. It depends which point you are making* (*Gordon*)

The majority view was that photographs were the primary medium, as expressed in Jane's quote above. In this context, three scenarios for adding music emerged from the discussions. Most participants in the trial captured many more photographs than they ever got around to putting in albums. The best of these were quickly reprinted and distributed to close family and friends, or slowly archived in a chronological album. In addition, the photo enthusiasts in the trial, tended to draw on their best photos to make themed albums or scrapbooks based around specific trips or events. Some of these were given away as personal gifts. Each of these photographic creations might benefit from music, which could itself be added at the time of creation. Hence, music could be attached to a reprint at the point of sending it to a friend. Alternatively, music could be selected for attachment to an album or scrapbook at the point of its physical assembly. The challenge here would be to support music selection and capture in a way that does not add too much time to the album-making activity, and in a format that can be integrated with the selected photographs. Another approach would be to turn the album-making activity into a social event, so that people could enjoy spending time together selecting and playing music for their photographs. This idea was suggested by Rob in the study, who was aware of the way his wife hosted album-making events:

> *I know one of the things which works well in what Traci does, is that most of the people that want to make albums just don't have the time. So once a month they all get together and glue together. So it's sort of a social thing. You've got ten people round the table with all the tools. I don't know how possible that would be to have ten people working on this at the same time* (*Rob*)

Finally, people said very little about the practicalities of music selection and digitisation in the trial. Selection was *felt* to be difficult, although people managed rather well to locate suitable music tracks from within their own music collections. This activity sometimes involved a search for a particular track within an identified genre or album, and might have benefited from computer assistance. Digitisation was not mentioned because the trial methodology did not require participants to digitise their own music. This is currently being made easier for consumers by digital media players which can copy music from an audio CD for personal use. Digitisation from records or cassette tapes is not so easy and involves connecting a record deck or tape player to a computer for copying. Such difficulties may decrease in the future if personal music collections become centralised in a digital file-store, and connected to commercial music services on the World Wide Web.

3.5.2 Editing

Participants seemed pleased with their music-photo albums for the most part, and expressed few desires to change them. The main comments made about editing related to cropping the music clip to focus on the most important part of a track for the desired effect. Many of the classical pieces were considered to be far too long on playback, and even some of the more contemporary songs had long introductions that could have been cut out. For example, Nick expressed a desire to cut straight to the chorus of one track while Jane noted that they never reached the opening lyrics of another.

We also found that the one-at-a-time presentation of photographs used in most of the albums was annoying to the photo album enthusiasts who were used to arranging multiple prints on a page. In fact Traci and Vicky both requested a small number of multi-photo screen pages for their own audiophoto albums. They might have increased this number if they had been able to manipulate the layout themselves and view the effects.

3.5.3 Review

At the beginning of each Feedback group participants were given the chance to view their music-photo and talking-photo albums together. This was usually done between pairs of people who knew each other, although Rachel and Rob were exceptions to this since they had never met before the study. We also had Jane, Byron, Nick and Kerri share two sets of albums together in a foursome. For any pair or group, the photographs were usually of events not shared by the other parties. However, Jane and Byron and John and Betty created joint albums of photos whose memories they both shared. This meant that we were able to record a rich set of interactions around the material, representing all the possible sharing combinations shown in the diamond framework in Fig. 1.2.

By examining these interactions and participants comments about them afterwards we quickly discovered that the review of photographs with music was very different to the review of photographs with voiceover. The main difference was in the ability to talk over music but not voiceover:

We didn't talk over the voices, we just talked over the music (*Jane*)

Indeed, not only was it possible for participants to talk over the music-photos, the music seemed to positively encourage this. As in the lively reminiscing conversations observed with ambient sound photos, we found that the review of music-photos was animated and noisy. The musical atmosphere appeared to break the ice between those who didn't know each other so well, and generally improve the ambience of the occasion. This was true for reminiscing conversations *and* storytelling ones. In contrast, voice-photo review sessions were very quiet and subdued, as all parties paused to listen through to the recorded comments (see Chap. 4 for further details).

The style of music-photo review was as follows. We noticed that the owners of each album tended to take control of the mouse and decide when to turn each page. In reminiscing conversations both parties made occasional comments or exchanged questions and answers while listening to the music and looking at the photos. These conversations contained short periods of silence between comments, as participants absorbed the album content. In storytelling conversations, the album owners took the initiative in telling the story of certain photos. These stories were prompted and steered by the audience, as in ordinary photo talk (see Chap. 5). One minor difference was that the music itself was sometimes noteworthy, and was allowed to run to the end of a bar or phrase before being curtailed.

However, for the most part, the kinds of stories that were told were similar to the comments recorded as voiceovers for the talking-photo albums. This was illustrated most dramatically by Jane and Byron, who happened to use the same photographs for both types of albums. They reviewed their music-photos first with Nick and Kerri, telling the same stories as were heard later on their talking-photo album. The similarity was so striking that Jane often skipped through the talking-photo album without letting the comments play through because she had just said the same thing live to the background music.

An example of a live and recorded story on the same photograph is shown in Table 3.1 below. The photograph itself is shown in Fig. 3.14, together with its accompanying music clip. This type of conversation around music-photos meant that the music tended to be listened to for the duration of the talk on each photograph, rather than for the duration of the music track itself. From a rough calculation on each music session, the average length of time spent listening to each track was about 40 s rather than 3 min. This fits with the average duration of recorded voice commentary, which was about 30 s (see Chap. 4). This commentary was inflated here by about 10 s worth of prompting and reaction from the live audience.

Table 3.1 A live and recorded story told for the photograph shown in Fig. 3.14 (Jane)

Live story to music-photo	Recorded story in talking-photo
{**Verse**: *If I was a sculptor but then again no*} **Jane**: *This was before we moved house to (Henley) cottage, and we were clearing out the roof. And we found, what was* **Byron**: *My grandfather's top hat and silver topped cane.* **Jane**: *That was when leg warmers were in* **Kerri**: *Yes I was going to say they look warm.* **Jane**: *Look like (ballet) shoes don't they?* {**End of chorus**: *My gift is my song and this one's for you*} (32 s)	**Jane**: *This is one of my favourites and this is at our previous house and just before we were due to move. We went up into the loft and found all sorts of bits and pieces, including Byron's grandfather's top hat and silver topped cane. No sooner as we got them down and they were used for dressing up purposes as these things always are. Rhiana was four and a half at this stage and immediately dressed up alongside her fashionable gear, i.e. leg warmers, and proceeded to do a song and dance routine in our conservatory* (40 s)

Fig. 3.14 A daughter set to 'Your Song' by Elton John

3.5.4 Comparisons with Home Video

On the whole, participants in the audio annotation trial did not connect the music or talking photos they created with home video. This was in contrast to the audio camera trial in which it was natural to compare the camera to a video recorder. Here, participants viewed what they were doing as an extension of photography and a way of enhancing photograph albums. The most common comment about audio annotation itself, was that both music and voice annotation should be supported on the same set of photographs. This issue will be taken up again in the next chapter:

To make it even better it would have been ideal to have it in together, music and voice (*Vicky*)

I thought as a medium it worked, but not to separate voice from music (*Gordon*)

When people were prompted to compare music-photos and talking-photos with conventional home video they tended to express a preference for the photographic medium. This is revealed in the following quotes which echo the sentiments of the comparisons made in Chap. 2. The effects of combining music and voices with photographs were seen as more professional than those obtained by capturing ambient sounds on moving images.

The (audiophotos) looked better. I suppose the only videos I've looked at in this setting are home videos and they're chronic aren't they? This is much more professionally put together, much more effective (*Rachel*)

Its like someone saying 'Now for a little video set'. It's great for five minutes, wonderful if you are a grandparent, maybe and hour – but after that, no. But if you have actually snapshot it and put it to music, you've got two interest points. You're actually telling people about it, which is why journaling is so important for photo albums, and music sort of adds to the dull part (Vicky)

3.6   Discussion

This investigation began with a sense that ordinary people might be rather good at selecting music for photographs, and that this would entail engaging with the kinds of concerns held by film composers in designing music for cinema. It was also assumed, from the Desert Island Discs radio show, that music would have a variety of personal meanings for people, and might stimulate private nostalgic feelings or more intimate social interaction.

In fact, many of these starting assumptions were borne out in the results of our study. We found that despite an initial scepticism about the music selection task and their own abilities to perform it, participants were surprised at the quality and power of the music-photos they produced. Perhaps the most impressive aspect of this finding was that the music was selected 'blind' so to speak, without playback or sampling with the photographs, and only reviewed sometime later when the two media streams had been assembled according to specification. The results are likely to improve further with the ability to review music-photo combinations on the fly, and with the benefit of experience.

As in the case of selecting funeral music, participants in the trial were sensitive to the aesthetic effects of the music, and to its symbolism for the audience in mind. The selection and discussion of favourite music-photos revealed that people were concerned to match the mood of the photographs and music aesthetically, and to choose music of personal significance to them or others. This significance factor was sometimes determined simply by how much they liked the music, but more often than not there was a deeper and more symbolic reason for selection, such as its association with past listening contexts, or with its title and lyrics. This meant that the connections between the music and the person selecting it were idiosyncratic and as potentially revealing of personality as those observed on Desert Island Discs. However, in contrast to what happens on this show, participants in the trial rarely discussed their reasons for music selection when reviewing music-photos with others. These connections remained implicit in the material, and it was left to the audience to work out the meaning of the music against a backdrop of conversation about the image. This gave the music-photos a certain poetic and cinematic feel to them, particularly when viewed in the absence of the author.

The cinematic analogy is appropriate not only to the music-photo material, but also to the mindset of the participants. They tended to look at their sets of photographs as sequences of images, and used the music either to tie short sequences together or to set individual photographs apart. As with professional film composers,

participants were able to choose the music to reinforce the primary emotions of characters in the scene, or of themselves as photographers. They were also able to trigger new secondary emotions in the audience for the photographs, as in the creation of joke music-photos. Some participants even voiced film music composition concerns about which medium should come first when combining music and images, and began to ask for the incorporation of other kinds of sounds, such as voiceover, in the overall sound design.

3.6.1   The Role of Music in Personal Image Review

Regarding the prediction that music may enhance a personal memory for a photograph, we found this to be true in two ways. In some of the music-photos, participants used nostalgic music, listened to around the time the photographs were taken, to remind them of that period. This appeared to work best for people when there was a history of repeated listening, as with theme tunes of regular TV programmes or popular music played on the radio. In these cases, an individual's memory for the scene depicted in the photograph appeared to be enhanced by association with music of the time.

Another form of memory enhancement was through the use of music to match the mood of the scene or characters in the scene, including the photographer. This was the most common use and value of music in the study, and served to 'colour' a photograph with the emotion of those who were present at the time. This is an altogether more intimate function than the attachment of nostalgic music or the capture of ambient sound of the moment, since it goes beyond the recording of sensory stimulation to reflect the inner feelings of the moment. To the extent that it is does this successfully, we can speculate that mood music will heighten the sense of retropresence to the past scene for those involved, and amplify the memories that follow.

Further benefits for the personal review of images were provided by music of other kinds. These benefits went beyond the enhancement of memory, into the realms of entertainment and communication. Pure entertainment value was delivered by music that *didn't* match the mood of the moment but created some other mood in the audience for the photograph. This was observed in a number of joke-photos. Communication values were realised by music containing hidden messages in the title or lyrics. These were most commonly seen in love-photos designed by men or women for their partners or children, as illustrated in Figs. 3.5 and 3.13. Such messages were usually designed to travel between the photographer and the subject of the photo, although with practice, we can imagine others being designed for a more general audience.

Indeed, the interpretation of photographs by those not present at the time was also improved by music. At one level, the entertainment and information values of the photographs were simply increased by the additional medium – making them more interesting. But at a deeper level, participants appeared to recognise from their own behaviour, that there was some method to the selection of music for photographs,

and that this method often involved re-creating the mood of the moment. To this extent, even complete strangers in the study were able to review each other's music-photos with noticeably more interest and potentially more *empathy* than usual.

Given the benefits of music for photo interpretation, it is unfortunate that this behaviour would be inhibited today by music copyright law. Creators of musical photographs comprised of published music tracks could not send copies of such photographs to others without breaking the law. This increases the urgency for some form of micro-payment system for copies of full music tracks, or for a change in the law to accommodate different forms of music and multimedia consumption. One possibility here would be to waive charges for *partial* music tracks used with photographs, since participants in the trial either selected fragments of a track or only listened to full tracks for a fragment of their duration.

3.6.2   The Role of Music in Social Interaction

The possibility that music might enhance social interaction around photographs was also realised in the trial. It appeared to liven up both reminiscing *and* storytelling conversations, which occurred over the top of the music itself.

Part of this effect may have been due to the influence of music in any social setting as a method of raising the atmosphere and as a catalyst for conversation. As proof of this influence, imagine a party without music. However, as we have seen from the previous section, the personal impact of specific music on the photographs at hand went beyond the playing of party music on a home hi-fi. This experience was more akin to the viewing of film music and images, with all the associated emotions.

The particular effect of music *annotations* on photo sharing may therefore be to trigger precisely those emotions that are appropriate to the story of each photograph. This may serve to stimulate story-telling or simply to elaborate it in a non-verbal way. The use of music in this supporting role to conversation, was exactly what we observed in the miss-match between the duration of each music track and the length of time it was listened to. Listening time was tied directly to the duration of conversation about corresponding photographs, and usually fell well short of the duration of each track.

References

Boltz M (2001) Musical soundtracks as a schematic influence on the cognitive processing of filmed events. Music Percept 18(4):427–454

Godoy RI, Jorgensen H (eds) (2001) Musical imagery. Swets & Zeitlinger, B.V. Lisse, The Netherlands

Lawley S (1990) Sue Lawley's desert Island discussions. Hodder & Stoughton General Division, London

Maur K (1999) The sound of painting: music in modern art. Prestel, London

Mountain R (2001) Composers and imagery: myths and realities. In: Godoy RI, Jorgensen H (eds) Musical imagery. Swets & Zeitlinger, B.V. Lisse, The Netherlands, pp 271–288

Plomley R (1975) Desert island discs. William Kimber & Co Ltd., London

Plomley R (1982) Plomley's pick of Desert Island Discs. Weidenfeld and Nicolson, London

Russell M, Young J (2000) Film music screencraft. RotoVision SA, Crans-Pres-Celigny

Sonnenschein D (2001) Sound design: the expressive power of music, voice and sound effects in the cinema. Michael Wiese Productions, Studio City

Thayer J, Levinson R (1983) Effects of music on psychophysiological responses to a stressful film. Psychomusicology 3:44–54

The West Briton (2002) Hymns are fading from charts for funerals. Thursday, 8 Aug 2002, p 15

Chapter 4
Talking Photographs

Spread before you your photo or photos, memorabilia and any notes. Ask yourself this critical question: 'If I am to give voice to this material, what would I have it say?' Think in terms of what information is important to capture for sharing with the next generation. Focus on what you value, and let your insights help you determine where to put your emphasis (Campbell Slan 1999, p. 77).

While most of us find it difficult to enough to sort our photographs into conventional albums, a small but active community of people in the United States pursue the hobby of scrapbooking. This is referred to in the quote above, and involves the use of photos and other memorabilia to tell life stories in scrapbooks. Scrapbookers use special paper, glue and plastic covers to preserve the photographs in their books, they personalise the presentation of materials by cropping and decorating the pages in various ways, and handwrite labels and stories in the margins. In this way the scrapbook is designed to stand alone as a document for future generations in a family to review as a piece of family history. An example scrapbook page taken from Joanna Campbell Slan's book, cited above, is shown in Fig. 4.1.

A related hobby to scrapbooking is journaling. This involves writing down one's life experiences as a way of expressing thoughts and feelings and documenting one's life story. Hence the journal itself is usually more personal than a scrapbook, and designed primarily for authors themselves to read later in life. More important than the product, is the process of producing it, which is said to be therapeutic and life enhancing (e.g. Weldon 2001).

In this chapter we consider whether some of the personal and social values of writing in journals and scrapbooks might be captured in talking photographs. That is, if it were made possible to record voice labels or commentaries over photos, would people use this capability to capture their life stories for themselves, or record accounts of their family history for others? And if they did that, would others want to listen to the results?

An additional consideration of the chapter is the *ease* with which voice annotation might be done, compared to writing. We note that scrapbooking and journaling are specialist hobbies and not part of the common practice of domestic
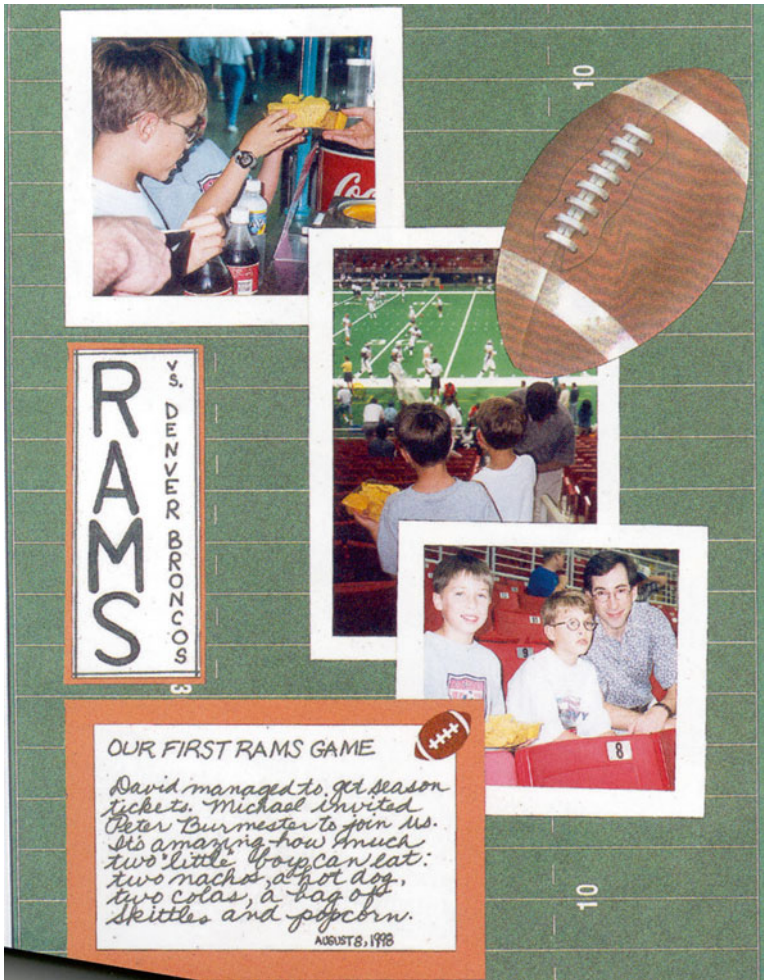
Fig. 4.1 A page of a photo scrapbook containing handwritten notes (Fig. 7.10 in Campbell Slan 1999)

photography. In contrast, talking about one's photos to others is an integral part of this practice. Furthermore, storytelling and reminiscing in conversation appear to confer all the same practical and therapeutic benefits as personal history writing (c.f. Haight and Webster 1995; Harvey 1996). If verbal storytelling skills could be harnessed more easily around photographs, either by voice annotation when photographs are first reviewed, or by recording subsequent photo-sharing conversations, then the practice and benefit of personal storytelling might be brought into the realm of domestic photography. This chapter examines the prospect for voice annotation while Chap. 5 examines photo-conversation recording.

4.1   Related Work

Previous attempts to record voices with images fall into two categories. First, *voiceover narration* is often used in fiction films and documentaries to complement the visual storytelling. Second, the use of voiceover narration with photographs is also finding its way into emerging forms of *digital storytelling*, in which digital technologies are used to record personal stories.

Voiceover narration in films can be done from the point of view of a character in the film, or of an outside commentator on the story scene (c.f. Kozloff 1988). An example of 'first person' narration by a character is the voice of Philip Marlowe in the famous detective films. His voice was used to set the scene and communicate his own thoughts and feelings about the unfolding case, as in the following quote:

> *The joint looked like trouble, but that didn't bother me. Nothing bothered me. The two twen-ties felt nice and snug against my appendix.*

Other examples include the reading of a letter shown on screen in the voice of its writer, or to preface a flashback or flash-forward in the story.

An example of third person narration by an outside onlooker is often to be found at the beginning and end of a film to introduce the story and draw conclusions from it. In the following quote, the invisible narrator introduces the main character of the film and sets up the rest of what follows as 'his story':

> *His name was Jeremiah Johnson and they say he wanted to be a mountain man. Story goes that he was a man of proper wit and adventurous spirit suited to the mountains. Nobody knows whereabouts he came from but it don't seem to matter much… This here's his story.*

Other third person narrators are used in news programs and documentaries to flesh out the details of events and explain their meaning. Because such narrators appear to know more than the characters in the events, can travel to different events in space and time, and are never seen on camera, they may be referred to as the *voice-of-God*. In some documentaries the voiceover is combined with camera movement over a still image. For example, in programmes on the American Civil War, group photos are often used as a backdrop from which to describe historic characters or events.

All these techniques add an extra dimension to visual storytelling that plays on the time course of the story, and the relationship between the characters, the story-teller and the audience. All voiceovers create a kind of secret complicity and inti-macy between the storyteller and the audience, and the origin of the voice defines whether they lie inside or outside the story-world.

A few attempts have been made to apply these techniques to digital photographs. In a recent study of an electronic photo viewer, Balabnovic et al. (2000) asked peo-ple to organise their own photographs into sequences over which they could narrate a story. The system allowed users to do this in a number of ways. These included selecting an existing chronological set of photos as a basis for voice annotation (photo-driven), or assembling a new set for the story at hand (story-driven). It also allowed users to narrate after the photos had been selected or to begin speaking first and select photos within the ongoing narration. The authors found that everyone

chose to select photographs before narrating over them, at least one photo in advance. However, as they selected the photographs, users often moved between a photo-driven and a story-driven style of storytelling. For example, they might start recording comments on an existing set of photos, but then be prompted to bring in other photos to elaborate a story. A further finding was that users did not use the voice recording facilities when showing their photographs to others face to face, nor did they record voice comments for their own use later. They preferred to explain the photographs live in conversation, and only resorted to recording a voiceover when asked to select photographs to send to remote family or friends.

Finally, a story-first approach to photographs and other memorabilia is used in an ongoing movement to encourage 'digital storytelling' (Lambert and Mullen 1998). This is really a form of scrapbooking using digital and screen-based technology rather than paper and pens. The movement arose from the work of the late Dana Atchley who was a solo performer and storyteller based in San Francisco. In the early 1990s he developed a one-man show called *Next Exit* in which he used digital media technology to project photographs and home movie clips as a basis for telling personal anecdotes. This approach was subsequently embraced and adapted by others in the area, to encourage more private forms of storytelling with digitised photographs, video clips and voice narration. A particular adaptation was to capture aspects of the verbal story within a digital scrapbook or presentation. To this end Joe Lambert and Nina Mullen set up the San Francisco Digital Media Centre in 1994, and continue to run workshops teaching business and community groups how to do this. A recent book of their experiences has been published, although it lacks detail on exactly how people choose to use voiceovers in their digital scrapbooks (Lambert 2002). For example, we cannot tell what voice, content or structures are used in voiceovers recorded at the workshops, or what relationship they have to other visual media. The same details are also missing from the analysis of user behaviour in the study by Balabnovic et al. (2000).

These issues are now taken up with reference to the audio annotation trial introduced in the last chapter. Having already described the findings for music annotation, I focus here on how participants chose to combine voiceover narration with photographic content.

4.2 The Audio Annotation Trial

Nine people were presented with the ability to record voiceover content on their photographs in the context of the audio annotation trial described in Chap. 3. Essentially, voiceover recordings were made during a home visit, on a small selection of photographs. These were later assembled into PC-based 'talking photo' albums, alongside additional musical photo albums and demonstrations, in preparation for a review session in HP Labs Bristol. Details of the overall study methods and music album procedures are given in Chap. 3. However, it is necessary here to clarify the procedure for voice recording during the home visit.

Participants were instructed to select a set of about 10 photographs that might benefit from voice annotation. They then worked through the photos in order, speaking comments they wanted to associate with particular images or groups of images. These comments were recorded on minidisc as part of the ongoing record of the interview, and were later extracted for digitisation and association with the corresponding images. We also kept a log of these associations on a two-column form, detailing individual photos by name and the topic of any connected comments. Some participants planned the outline of what they were going to say in advance of recording, but most simply began speaking and recorded their comments in a single 'take'. All participants were given the chance to listen to their voiceovers, but most listened to only a small fragment and declined to change anything. The fact that we remained present in the room throughout the recording session may have led some participants to bias their comments towards our own interests and status as strangers, although we stressed the importance of choosing a target audience of their own for the recordings.

Analysis of the talking photo set, together with participants' reactions to it, is described below. As before, these data are treated separately and in order, according to the section headings used to report the findings on ambient and musical photos.

4.3   Talking Photos Captured

Nine talking photo albums were created in the trial. These were made up of 105 voiceovers on 112 photos. The majority of associations between voiceovers and photos were therefore one to one, as discussed below. However, as with the musical photographs, there was great variety in the audio material used and in the intentions behind the creation of each album. Some albums were used to tell the story of a single family trip. Others were used to relate interesting aspects of favourite photos, or to comment on the importance of various people in a photo set. Because the methodology involved sharing these creations with others, participants got to see at least one other person's material in full, and the demonstration albums showing a cross-section of audiophotos from everyone. They were also able to discuss their views on these materials in a group. This method was effective in circulating different opinions about the value of voiceover, and generating some consensus by the end of the process. In this section I try to represent the variety before working towards the consensus in later sections.

4.3.1   Types

Figure 4.2 shows the kind of things discussed in the 105 voiceovers recorded in the trial. This was based on a content analysis of what people mentioned in each voiceover. We found that most of the basic questions that can be asked of an event or scene were covered in the voiceovers. These questions include Who, What, Where, When and Why. These topic categories were not mutually exclusive. This
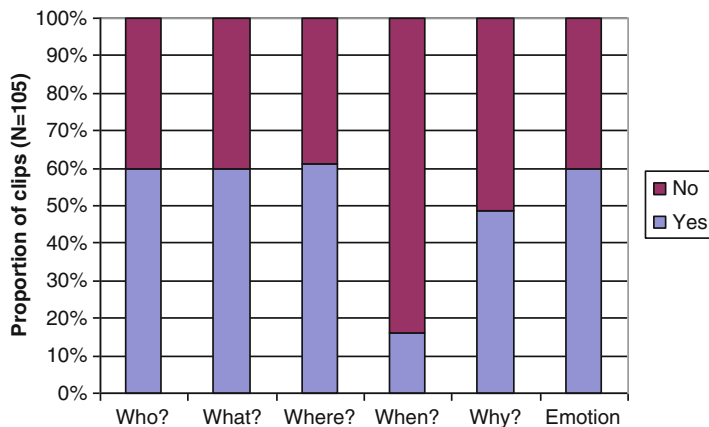
Fig. 4.2 Content analysis of voiceovers recorded over photographs

meant that several topics could be mentioned in the same commentary, and explains why the percentages don't add up to 100. Each percentage should be read on its own as an indication of the proportion of voiceover clip mentioning a specific topic. For example, in 60 % of voiceovers, participants described *who* was in the photograph. The only question we couldn't map onto the recordings was 'How?'. The method of doing something in the photo was seldom discussed, or was incorporated into a more general description of 'What' was happening. Instead, we found that an additional question was addressed concerning the author's *emotional* relationship to the photo. Hence in 60 % of voiceovers, people gave a value judgement about how they or others felt at the time, or now feel about the photograph itself. This is an interesting discovery in view of the importance of emotion in music annotation. It suggests that here again, voiceover might be used to attach emotion to a photograph as well as to describe factual details of the depicted scene.

Four talking photographs are shown in Figs. 4.3, 4.4, 4.5, and 4.6. These illustrate the ways in which various questions about the photographs are addressed in the voiceovers. Figure 4.3 is the talking version of Fig. 3.14 in Chap. 3. In fact the voiceover itself is transcribed in column two of Table 3.1. It is Jane's story of her daughter's hat and cane discovery whilst clearing out the loft before a house move. As such, it involves a reference to every question that can be asked about the photograph, including the author's emotional relationship to it. This makes it one of the most complete and complex voiceovers in the collection.

The following question topics are addressed by the voiceover for Fig. 4.3:

- The subject of the photo is Rhiana when she was four and a half (*Who*?),
- The activity she was engaged in was dressing up with Byron's grandfather's hat and cane (*What*?)
- The photo was taken in their previous house (*Where*?)
- The photo was taken just before they moved from the house (*When*?)

Fig. 4.3 A story
addressing all the questions
of the photograph

Fig. 4.4 Details of a group activity in the first person

Fig. 4.5 Flashback to an activity that used to occur in this location

Fig. 4.6 An externally facing voiceover about the politics of Albania

- Rhiana was dressing up was because they had discovered the props whilst clearing out the loft (*Why*?)
- The photograph of Rhiana is one of Jane's favourites (*Emotion*)

Note that the questions are not addressed in order, but rather worked into a chronological story form beginning with Jane's evaluation of the photograph. Details of the time and location are not literal, but relative to other times and locations, and the fact that Rhiana is Jane's daughter is not given in this particular voiceover because it was introduced in a previous one. A final complexity is that the story is told in two voices. The discovery of the top hat and cane is told in the first person as a joint activity, while Rhiana's dressing up behaviour is reported in the third person. In general, the 1st person voice is used in 57 % of voiceovers. This figure rises to 77 % if the speaker is a subject of the photograph.

Figure 4.4 is another example of a voiceover addressing multiple questions about the photograph. However, this time, it is not so clear who is depicted or when exactly the photograph was taken. Also, the voiceover is not structured as a story. Instead, it takes the form of a series of observations about the most striking feature of the photo; which is the fact that everyone is wearing the same 'tank top' sweater. Thus, the main activity is said to be "*Us looking very smart with our tank tops on*", during a holiday in Malia. The broader context for wearing them and taking the picture emerges later in the voiceover. They were bought cheaply as a gift by one of the group and worn for at least three nights out on the town, of which this was "*a particularly good one*". The point of the voiceover is not really to tell the story of the tank tops being purchased and worn repeatedly, but to use the joint wearing of the tank tops as an example of how *together* they were as a group and how much fun they had. In fact, the fun elements are underscored by no less than four separate references to positive emotions in the voiceover, while togetherness is emphasised by narration in the first person.

Figure 4.5 gives another indication of the complexity and richness of voiceovers recorded in the study. As in the previous two examples, the voiceover addresses some of the standard questions that can be asked about the photograph. The difference here is that those questions relate to two distinct time periods in which the scene was visited; the current time at which the photo was taken and previous times "*when my children were small*". In other words, the voiceover incorporates a *flashback*. In the first part of the narration John talks about the difficulty of capturing the particular view of the bridge and village shown in the photograph. Through this account he is able to say where the photograph was taken and something about why. In the second part of the narration John goes further back in time, to talk about a previous habit of coming to the same place with his children. In this flashback he uses a past participle in the first person with phrases like "*we used to take bits of bread and throw them in the stream*". With this technique John describes a new set of actors and activities for the scene that are not shown in the photograph, but are triggered by it. This example shows that photographs can be used to talk about memories associated

with any aspect of their content, and may not necessarily relate to the time the photograph was taken.

The practice of speaking about things outside the photograph at hand is also illustrated in Fig. 4.6. This is our final example of voiceover types, and relates to a picture of gun turrets in an Albanian landscape. It is one of a series of photographs of a trip to this country by Rob in the trial. Although some of Rob's other voiceovers are more personal and descriptive of people and events in the pictures, this one is not. Instead it talks about the gun posts in the image and the fact that they point inward towards the population rather than outward towards the sea. Where, what and why questions about the image are covered in the voiceover, but in a way that describes the politics of the country rather than Rob's personal memories of the scene. Insofar as they can be counted, these kinds of externally-facing voiceovers accounted for about 18 % of those recorded in the study. These included other stories about the character of a country or its citizens, and references to the personalities of people in the photographs.

4.3.2 Durations

The length of voiceover clips varied from a few seconds to a few minutes as shown in Fig. 4.7. However, the average duration was 29 s, with 95 % of all clips lying between eight and 50 s. These figures are similar to those found for ambient sound clips in the audio-camera trial, but considerably shorter than the 3-min music clips also observed in the audio annotation study.

The talking photographs already shown in Figs. 4.3, 4.4, 4.5, and 4.6 are a fairly typical length. The voiceovers in Figs. 4.4 and 4.6 are around 30 s long, while
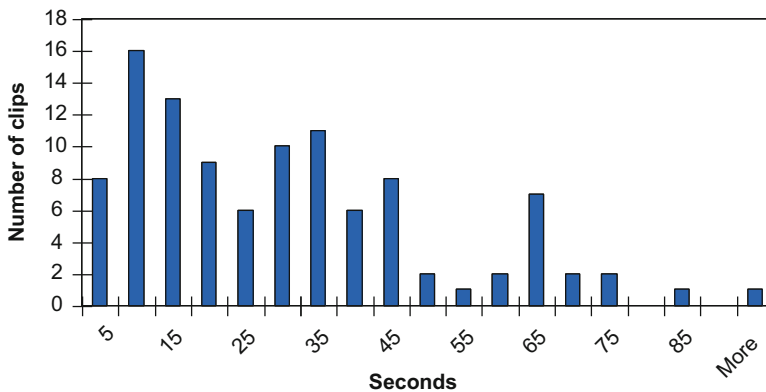
Fig. 4.7 The duration of voiceover sound clips

Fig. 4.8 A 66 s voiceover
with parts from two
different speakers

those in Figs. 4.3 and 4.5 are somewhat longer (41 and 60 s respectively). The
1-min narration in Fig. 4.5 is due to the fact that two stories are told about the
photograph, one for the time it was taken and one in flashback. These parts could
almost be separated into two separate voiceovers of about 30 s each. Other long
voiceovers have this multiple story form, although sometimes it is from the contri-
butions of two different speakers. For example, the talking photo shown in Fig. 4.8
has a 66 s voiceover with contributions from both Byron (46 s) and Jane (20 s)
about their mutual friend shown in the image.

At the other end of the spectrum, some voiceovers were under 10 s long. These
tended to be simple voice labels for the photographs, such as you might write on the
back of a print. An example of one of these is shown in Fig. 4.9. This is a picture of
Buckingham Palace in London with a 7 s voiceover saying so. Voice labels of this
kind typically addressed only one of the questions shown in Fig. 4.2 – usually either
Where or Who.

Another reason for very short voiceover recordings was that they were embed-
ded in a narration relating to a series of photographs. Individual portions of this
narration still related to particular photographs, but were short. They could also
cut in at the beginning or cut off at the end. This issue concerns the relationship
between the voiceovers and the photographs, and will now be explored in the
next section.

Fig. 4.9 A short seven-second voiceover serving to label the photograph

4.3.3 Associations

The majority of sound and image associations for the talking photo set were one-to-one. Hence participants recorded a single voiceover for a single photograph in 94 % of cases. For this reason I have not included a graph of associations, as in previous chapters. Many of the voiceover recordings were self-contained stories or reflections on the photographs, as in most of the above examples. However, sometimes these stories were designed to fit together across photographs, even though their individual parts still bore a one-to-one association with constituent photographs. An example of this is given in Fig. 4.10, which is the first part of an album describing a trip to Disneyland. After a scene-setting voiceover on the first photograph in the set, Vicky recorded a series of shorter voiceovers describing individual Disney characters encountered on the visit. Note that the recorded speech on each photograph is understandable in its own right, but works better when viewed in the context of the series. In rare cases, we found that the speech on individual photographs could be truncated at the beginning or end of each clip and therefore become less understandable on its own. This is actually true of Fig. 4.9, which is a voice label embedded in a longer account of a trip to London. If you listen carefully to the opening of this clip you can hear the talking begin in the middle of a sentence. We decided not to classify these voiceovers as one sound to many photographs unless authors expressed an intention to have the speech run over more than one photograph at a time.

Where participants did express a desire for a voiceover to run over more than one photograph, it was always in the context of requesting more than one photograph per screen page. The album enthusiasts Vicky and Traci did this more than others in the trial, since they were used to arranging multiple prints on a single photo album page. An example of one of Traci's screen-based versions of this with voiceover is shown in Fig. 4.11. The voiceover relates to two photographs taken inside Hamley's toy store in London. It therefore falls into the 6 % of one sound to many photo associations in the talking photo set.

Fig. 4.10 The first two talking photos of a series from a trip to Disneyland

Fig. 4.11 A talking photograph with a voiceover relating to two images simultaneously

4.4 Reactions

4.4.1 Initial Reactions

The idea of recording voice commentary on photographs was less alien to participants than the notion of attaching music. Nevertheless, they expressed some skepticism about its value, and a certain amount of dismay at having to record their own voices. Most people said that they didn't like the sound of their own voice, but would be happy to listen to the voices of others:

This prejudice continued beyond the recording process at home, and coloured the participants' initial reactions to the talking photograph albums in the Feedback Groups. In general, they preferred listening to other people's talking photos than to their own. This also made them also more inclined to favour their own music photos over their own talking photos. Both effects are apparent in the following quote by Nick:

> *I preferred the music photos. It might have been because it was my own voice though. I think Jane's sounded better than mine* (Nick)

However, the physical sound of one's own voice was only one reason for this preference for other people's voiceovers. In fact, several people said they got used to hearing their own voices after a while. Another reason was that the content of other people's talking photographs was simply more interesting. By definition, participants already knew the content of their own talking photographs, and were naturally more interested in what other people had to say about *their* photographs. This was revealed in a number of comments questioning the point of recording voiceovers for yourself, and praising the voiceovers of others:

> *You know in your head the memories that you have of that particular holiday, so you don't need the voiceover* (Rachel)
>
> *I felt the photographs with the commentary were much more interesting than the music because it enabled me to see what Nick was thinking and feeling at the time* (Jane)

Alongside the preference for other people's talking photos, was a realisation that one's own talking photographs would be more interesting to other people, and especially to those who weren't present at the time the photograph was taken. This was stated directly by Gordon in the trial, when asked if it was worth the effort of recording voiceovers on his photographs.

> *It depends on whether the person I'm showing it to was with me at the time I took the photo, because if they were there, there is no point to all the explanation. I would do this to show to people who weren't there at the time* (Gordon)

For others, the same realisation was expressed in terms of a shift in the intended audience for their talking photographs. When asked at the home visit who they were recording the voiceovers for, participants were evenly split in saying they were for themselves or for their close family and friends. However, when asked at the

Fig. 4.12 Nick and Kerry's favourite talking photo of Kerry

Feedback Group who they thought would benefit most from the albums, they tended to mention only geographically distant family and friends:

> *I would send it. We've got an uncle that lives in the States so if we'd been able to send him something like this to look at the kids, a self-explanatory thing, that would be good (*Rachel*)*
> *I would use it more to send to people than to sit someone down (*Traci*)*

4.4.2 Favourites

After seeing the talking photo material for the first time, participants voted on their favourite combinations from their own and each others' albums. They also wrote down the reason for these choices. There appeared to be three main reasons for selecting favourites from any album. These included the personal significance of the photograph, the interest of the voiceover, and the mention of emotion.

As with the selection of favourite music photos, the personal significance of the material was important to the selection of talking photos. For example, some participants chose a favourite talking photo from their own set because they simply liked the photograph and its associated memories. This could be seen in the selection of favourites by three of the mothers in the trial who selected talking photos containing images of their children. Jane chose Fig. 4.3 showing her daughter with the hat and cane, Traci chose Fig. 4.11 showing her daughter and nanny together at

Hamleys, and Vicky chose an audiophoto of her children meeting Mini Mouse at Disneyworld. A similar choice was made by John, who chose a talking photo about his grandchildren. Figure 4.12 shows Nick's favourite talking photo of his girlfriend Kerry who was also present at the feedback group. Kerry also chose this audiophoto as her favourite, perhaps as much for the personal significance of what Nick said about her in the voiceover, as for the image itself. Note that this is the talking version of the same photograph that is set to music in Fig. 3.13.

The interest of the voiceover was also used as a criterion for selecting a favourite talking photo. This was a slightly different criterion to the fit between sound and image, which was important to the selection of favourite music photos. It seemed to have more to do with the form and content of the voiceover itself. For example, Jane selected Byron's talking photo of a Japanese temple, simply because "Byron's description is so clear and interesting". This voiceover is contained within Fig. 4.13, and is the counterpart to a musical version shown in Fig. 3.8. Byron himself chose a favourite talking photo of a friend's daughter from his own album "Because it

Fig. 4.13   A favourite talking photo comprising a voiceover on a Japanese temple

describes the picture so well". In a final example, Rachel chose Rob's talking photo of the Albanian gun posts shown in Fig. 4.6 as a favourite for the following reason:

*It's a gorgeous photo in its own right, but the message behind it is so telling (*Rachel*)*

The mention of value or *emotion* in the voiceover also seemed to increase its attraction and the likelihood of selecting it as part of a favourite talking photo. This could be seen statistically in content of voiceovers selected as favourites from participants' own albums. Seventy eight percent contained references to the emotion of the image or scene. This compares with an incidence of 60 % in the rest of the corpus. However, additional evidence came from the *reasons* for selection, which often referred to the fact that the voiceovers revealed something about how the author was feeling or thinking at the time. This was expressed by Rachel as a reason for selecting her own favourite talking photo, shown in Fig. 4.14. The photograph was significant to Rachel because it contained the image of a cross. The voiceover allowed her to explain this significance and therefore to reveal something of her own faith and personality.

Fig. 4.14 A self-revealing talking photo

4.4.3 The Value of Voiceover

From discussion of the talking photo material in the Feedback Groups it became apparent that participants attached different values to voiceover, depending on whether it was theirs or not. If they were the author, they valued the ability to personalise the photographs or add a message to someone else. If they were the audience, they valued the way in which the voiceovers added interest and interpretation to the photographs, and revealed the personality of the sender.

In the first case, we have already seen how a voiceover was valued by an author for the way that it allowed them to *personalise* a photograph. This was introduced above, in connection with Fig. 4.14. Rachel's favourite talking photo shown in the figure was actually a self-portrait in which she appears in both the photograph and the voiceover. In her case, she liked the voiceover because it allowed her to speak about her faith.

Equally personal voiceovers were recorded on images that did *not* contain the speaker (see for example Figs. 4.3, 4.4, and 4.5). These can also be seen as self-portraits of a sort, in which the speaker can enter the life of the photograph through the voiceover. Clearly they are most revealing when the author speaks about their own feelings, thoughts or beliefs. However they are also revealing even when such things are not mentioned, simply because of the way the voiceover is spoken. This can be seen in Rob's Albanian gun post voiceover in Fig. 4.6, whose content is impersonal and doesn't mention emotion. His voice itself is still full of character, and serves to personalise a photograph that is utterly devoid of people and life. In this way, voiceovers came to act as *voice signatures* for photographs, through which people could reveal more or less of themselves to others, depending on what they 'wrote' into the recording.

Writing is actually a good analogy for how people used and valued voiceovers as a communication medium. Again for authors, voiceovers were seen as another way of *adding messages* to photographs, akin to writing a letter to send with reprints, or journaling in the margins of a photograph album page. In both these situations, the writing explains the reason for sending a photograph to a particular audience. Voiceovers did a similar job for their authors. They also appeared to give authors a chance to voice thoughts and feelings that are hard to articulate face-to-face. For example, Traci told us that she designed her favourite talking photo in Fig. 4.11 as part of a series for Rebecca her nanny who is shown in the picture. On an earlier photo she stated that Rebecca came over from the States to help out during a period when Rob was ill, and "*was very much an encouragement to us*". This choice of words, and their location within a series of happy photographs taken during a sad time, underscores the value that Traci placed on Rebecca's help, and constitute a kind of multimedia thank you card. Although Traci had tried to tell her this directly, she felt that the talking photographs would be more effective in communicating her feelings.

This kind of subtle communication is similar to that observed for some of the music-photos that communicated a hidden message in their lyrics. The most direct

use of music for this purpose was in love photos of the kind shown in Fig. 5.5. The same phenomenon was observed again in talking photos such as Nick's message to Kerry in Fig. 4.12 and John's talking version of Fig. 3.5 to his wife, shown in Fig. 4.15. This shows the same image of Betty's face, accompanied this time with some of John's thoughts about her. For people who may find it difficult to talk about their feelings for their partners directly, voiceovers like this may provide a more natural alternative to letters, email or text messages.

In the context of receiving photographs voiceovers were seen to *add interest* to photographs that might have been more boring on their own. This was mentioned directly by John who was thinking about being handed a large pile of someone else's holiday photos to look through. In this scenario he seemed to judge the talking photos to be more interesting than a live explanation by the author, although the majority of participants felt that their greatest benefit was realised when the author was absent. There may also be an overall enhancement of interest by virtue of selecting certain photographs for annotation, thereby filtering out the least interesting ones. Finding out more of the story of the photo and the memory of the author appeared to be the key sources of interest added by voiceover:

> *This to me is part of the boredom when people get their holiday photos out and they bring out a big pile of envelopes with hundreds of photos in. Whereas with this, as we have all been saying, it seems that the voice makes it more interesting* (John)
>
> *In Vicky's the music was excellent, but it needed the voice to tell me what was going on* (Traci)

Discovering the thoughts and feelings of the author in relation to the photographs, was the final value of talking photographs to those receiving them. This is the flip side of the value of expressing such feelings by senders. There were many

Fig. 4.15 A spoken love photo from a man to his wife

indications that participants liked this revelatory aspect of the voiceovers, not least of which was the use of this criterion in selecting favourite talking photos (see above). But perhaps the most succinct comment on this value came from Betty who noticed that the playback of a voiceover on a photograph shifts the receivers' attention from the photograph to the speaker. To this extent a talking photograph raises questions not only about the scene and what is happening within it, but also about the person recording the voiceover on the scene and their relationship to it:

With the voice you immediately think of the person that's speaking (Betty)

4.5 Practice

The ease or otherwise of recording and using talking photographs is discussed in this section. Some initial judgements are made from observing the process of voice recording during the home visits, and the process of reviewing the results during the feedback groups. Participants also made comments about their likely use of the medium after the review sessions.

4.5.1 Capture

On the whole, participants took easily to the role of recording voiceover on their photographs. This was despite their professed dislike of hearing their own voice on playback. Recording may have been made easier by our presence as strangers in the house, since we inadvertently became a live audience for their commentaries. However, participants were encouraged to think of another, more realistic, audience for their comments, and seemed to have little trouble in visualising someone and thinking of something to tell them.

The real problem was in selecting the right someone, and the appropriate set of photographs to annotate. I have already noted above that there was an initial tendency for participants to record voiceovers for themselves or their close family, such as children and partners. Once they had seen the results of these attempts, they reported wanting to shift their audience to distant family and friends. This can be seen in terms of the three album contexts for annotation mentioned in the discussion of music capture in Chap. 3. Attempts to annotate chronological photo albums with voiceover were judged less successful than attempts to annotate sets of reprints or themed albums to give away. Future voice annotation would therefore be directed towards the latter two contexts, as illustrated in the following quote:

Its probably just like a gift to someone. That's what I would use it for (Traci)

In the context of designing talking photos to give away, participants made two kinds of suggestions for enhancing capture. Either the voice comments could be

made more spontaneous by capturing them along with the photographs on camera, or they could be made more professional by scripting voiceovers recorded later. The first suggestion might be more appropriate to reprints, while the second would enhance themed albums or scrapbooks:

> *If I was taking a role of film and I could talk and say something about the photographs*, *then I might want to do that* (Byron)
> *I'd record it at night when the kids were in bed. I would probably sit down and write down what I wanted*, *then record* (Vicky).

There was also a feeling that a sequence of photos worked better for voiceover and needed to be selected carefully before recording. This was encapsulated in a discussion between Traci and Vicky in the study. Vicky used a series of photographs of her children with different characters at Disneyworld for her talking photo album, while Traci used more of a mixed set relating to different days out. Both women judged Vicky's album to be more effective because it told a single story over the photo set.

> Traci: *I mean it sounded like I was just saying it off the top of my head*, *whereas Vicky's sounded like a story*
> Vicky: *I think my photos probably helped compared to yours*, *because yours would have been hard to carry a theme through or create a dialog that was sort of a continuous story*

4.5.2 Editing

Apart from the observations made above about capturing talking photos differently, participants said very little about editing them. Most voiceover recordings were done in one take, and were judged good enough to use in their original form. This finding on the acceptability of voiceovers should not be surprising, since it is the very same attitude that participants had to their photographs. While some photos were judged better than others, few were judged bad enough to require modification.

What was requested was the ability to combine voiceover with music on the same photographs. This was mentioned briefly in the context of reporting the music findings, and demands further attention here. This combination can be seen as quite a natural thing to request by people who were asked to review the value of music and voice annotation in the same study. Some of these people even experimented with putting both types of sound on the same set of photographs and could directly compare the effects. These included Byron and Jane, John and Betty, Gordon, and Nick. Aesthetically the mix is more compelling than the individual sounds, since voice and music serve to personalise a photograph in different ways and communicate different kinds of information about the associated mood and memory.

The main problem with combining music and voice has to do with the relative loudness of each and the difficulty of listening to both tracks together, especially when there are lyrics in the music track. Participants acknowledged this interference

problem in responding to an extra music-and-voiceover album we had made up from Jane and Byron's albums. This highlights the need for careful sound balancing in mixed tracks, and perhaps for manual controls at playback that could be used to change this balance interactively.

4.5.3 Review

The presence of voiceover on a photograph tended to inhibit conversation around it, when participants tried to share their albums with each other. This was in stark contrast with the review of musical photographs which seemed to stimulate conversation. In a typical session, the author of the talking photo album would control the turning of each screen-page, in time to the ending of each full voiceover clip. While the voiceover played, participants would listen quietly, waiting until the end to make further comments or ask questions. This pattern was broken only in rare cases where the same photograph had been seen before as part of a music album, and the author had already told the recorded story live (see again Table 3.1).

This kind of review behaviour meant that participants tended to look quietly at each talking photograph for about 29 s – the average duration of voiceovers themselves. However, because of the wide range of clips around this time, some pairs of participants spent far longer staring at a single image in silence. The record was achieved by John and Betty who listened patiently to John's verbal description of their association with a waterfall for 103 s. This is a long time to sit without talking in any social setting, and is made all the more noticeable by the fact that their focus of attention was on a single static image.

To put this into perspective, the average pause length in conversation is usually 1 s (Jefferson 1988) while the average duration of a *moving* image shot on television is about 5 s. So the tendency to suspend conversation for so long around a photograph is very unusual indeed, and must be due to some powerful social force at work. One possibility is that participants were treating the voiceover *as* conversation. They would therefore tend to wait for its contribution to finish before speaking themselves. This is in marked contrast to something like watching the television, where speaking over television voices appears to be more acceptable.

4.6 Discussion

In short, the opportunity to record voiceover on photographs was welcomed by participants in the trial, despite some misgivings about the sound of their own voices. Everyone had a lot to say about the photographs they selected and were keen to experience what a talking photograph would feel like. However there were various problems with voiceovers which limited their usefulness to listeners.

For example, the content of voiceovers did not always tell the story-of-the-photo in Berger's terms (Berger 1982a). Stories were something that participants strived for in the activity of selecting and commenting on their photographs. But they were not always evident in what they actually said. Typically, participants would answer two or three classes of questions about each photograph, such as who was in it, where the photo was taken and what was happening. Furthermore, when they did describe a temporal sequence of events, this was just as likely to refer to a period covered by multiple photographs or not really covered at all (for example with a flashback), as it was to refer to the local period of time in which the photo was situated. These accounts could also be told in the first or third person. In short, any voiceover pointed to the *context* for a photograph, but it could do this at any level, from more than one perspective, and in more or less detail depending on the intended audience. This kind of selective description was identified by Rachel in the trial, at the end of her voiceover on the stone cross of Fig. 4.14. She concluded by saying:

So that was a potted history, a snapshot of our time in Cornwall (Rachel)

The notion that a voiceover is itself a snapshot of all the things that *could* be said about a photograph is reminiscent of work by Sacks (1970) on the news-worthiness of stories in conversation. By analysing the presentation of news in conversation, he found that what is news to one audience is common ground to another. Speakers are aware of this and act to select, mark and phrase information interactively, in order to maximise its news-worthiness.

Participants in our study appeared to do something similar with their voiceovers, giving a *version* of what could be said which they believed would be news to the intended audience. This arrangement worked well as long as the intended audience was someone else who could be clearly visualised at the time the voiceover was made. However, if the intended audience could not be visualised, or was the author him/herself, the voiceover was less newsworthy and effective.

In the worst cases, it was impossible for authors to record any version of events that was not already known to them. This made them disinclined to use voice annotations for archiving photographs for future personal use. This is despite the fact that they might benefit from such annotations long term, when their own memory for events fails. One way out of this paradox might be to archive voice annotations recorded for others, within a personal photo collection. These might be used to retrieve photographs from the collection and to remind authors of forgotten details with the passage of time. A similar technique has been developed for collecting text annotations from email messages with photo attachments (Lieberman and Liu 2002).

4.6.1   Role of Voiceover in Personal Image Review

Thus, although voiceovers were recorded for almost all the uses of photographs shown in the diamond framework of Chap. 1, participants settled on one use above all as having the highest value. This was the ability to add personal and contextual details to a photograph in order to aid its *interpretation* by a remote audience. This finding explains why the ambient commentary in the audiocamera trial was felt to be less valuable than ambient environmental sounds. Participants in that study tended to use such commentary as a voice label for themselves rather than as a voice message for others, and were consequently disappointed in the results. In both trials, voice labels were not found to enhance the memory of a photograph for the people involved. This pattern of findings is similar to that observed by Balabanovic et al. (2000) using different methods. In their study, subjects only recorded voiceover on photographs they imagined sending to others.

The reason for these effects becomes clear when we consider the specific reported values of voiceover in our study. A talking photograph was perceived by its receivers to be more interesting than a silent photo and to be more revealing of the thoughts and feelings of its author. These were the end-user benefits. In addition, the authors and senders of talking photographs valued the way they allowed them to 'frame' a photo with contextual information and to personalise it with something of themselves. Such personalisation could be done directly, by including a direct emotional reference to how the author felt about the image or activity depicted. It was also done implicitly by attaching the author's voice itself as a signature or self-portrait of sorts. These dual benefits of voiceover to senders as well as receivers imply a value to the processes of recall and recognition of an image, as well as to its private interpretation (see again Fig. 1.2). However, this value has more to do with the representation of the self to oneself and others, than with the remembering of the event itself. Voiceovers then, appear to influence the communication and interpretation of images, and this can be just as much of a benefit to senders as to receivers.

4.6.2   Role of Voiceover in Social Interaction

Despite the benefits of voiceovers in the interpretation of photographs, they were found to be less than useful in live social interactions. Their presence was found to inhibit conversation around photographs. This led to the bizarre situation of two or more participants staring at a photograph for long periods of time while listening in silence to recorded voice comments. Because speakers are so sensitive to the shifting interests and responses of recipients in conversation, live commentary on a photograph is likely to be altogether more newsworthy and interesting than recorded commentary.

This may also go some way to explaining the frustration with ambient sound in the audiocamera trial of Chap. 2. Many of the ambient sound clips included voice

commentary and recorded conversation. Participants found it difficult to talk over these, especially in storytelling conversations. Any ambient commentary arising in these situations would therefore be more difficult to talk over because of the competition between live and recorded explanations.

References

Balabanovic M, Chu LL, Wolff GJ (2000) Storytelling with digital photographs, Proceedings of CHI. ACM Press, New York, pp 564–571

Berger J (1982) Stories. In: Berger J, Mohr J (eds) Another way of telling. Butler & Tanner, London

Campbell Slan J (1999) Scrapbook storytelling: save family stories and memories with photos, journaling and your own creativity. EFG Inc., F & W Publications

Haight BK, Webster JD (eds) (1995) The art and science of reminiscing: theory, research, methods and applications. Taylor & Francis, London

Harvey JH (1996) Embracing their memory: loss and the social psychology of storytelling. Simon & Schuster Co., Needham Heights

Jefferson G (1988) Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In: Roger D, Bull P (eds) Conversation: an interdisciplinary perspective. Cambridge University Press, Multilingual Matters, pp 166–196, Chapter 8

Kozloff S (1988) Invisible storytellers: voice-over narration in American fiction film. University of California Press, Berkeley

Lambert J (2002) Digital storytelling: capturing lives, creating community. Digital Diner Press, Berkeley

Lambert J, Mullen N (1998) Digital storytelling: the creative application of digital technology to the ancient art of storytelling. Institute for the future report

Lieberman H, Liu H (2002) Adaptive linking between text and photos using common sense reasoning. In: Proceedings of the 2nd international conference on adaptive hypermedia and adaptive web systems, Malaga, Spain, May 2002

Sacks H (1970) On doing being ordinary. In: Atkinson JM, Heritage J (eds) Structures of social action: studies in conversation analysis. Cambridge University Press, Cambridge, pp 413–429

Weldon M (2001) Writing to save your life: how to honor your story through journaling. Hazelden, Minnesota

Chapter 5
Conversational Photographs

> *Had I showed for example all my wedding photos – I'd have*
> *had a summary of the day from my perspective which in years*
> *to come would be a nice memory for me or an interesting story*
> *for our daughter (Brid, Storytelling photograph study)*

Imagine for a moment you are a photograph. You would have a pretty interesting social life for an object. You will be conceived instantaneously in an environment of ambient sound and activity. You will be born into a family with at least one parent, either at hospital (the photo processing lab) or at home (the home photo printer). Your family will fuss over you at birth, passing you around and talking about your resemblance to people and things in the past. Then you will get proudly shown off to family and friends who will comment on your charm and good looks. Eventually your novelty will wear off, and you will be put to bed for a while, possibly in your own space (album), until such time as one of your parents decides to wheel you out with your brothers and sisters to see how you are getting on, or to show you off again at social gatherings. If you are really lucky, you will get to stay out forever (frame) to attract the comments of visitors. Granted this is something of a shallow life, which revolves around appearances and gives you very little scope for initiative. But hey, you don't have to go to school or work, and you still get to meet a lot of interesting people and make a variety of friends for life.

This scenario illustrates for domestic photography what Schaverian (1991) means by the *life of an image* in art therapy. I have represented this life as lines of social interaction around a photograph in the diamond framework of Chap. 1 (Fig. 1.2). These lines show that there is a sonic environment around printed photographs which is every bit as rich and varied as that which exists around the original experience that the photograph represents. Just as we can choose to preserve the original ambient sound context of this experience with the photograph, so we can choose to preserve the later sound context – not once, but many times during the life of the photo. Although the later sound contexts may include ambient sounds, they are likely to contain a greater proportion of conversation, this time about the photograph or experience itself. I call this kind of conversation *photo-talk*, to distinguish it from ambient conversation around the time of image capture. Just as we considered the properties and value of ambient

photographs in Chap. 2, we now consider the properties and value of *conversational photograph*s, by which I mean photographs combined with recordings of photo-talk.

Because photo-talk occurs spontaneously around photographs and contains conversation about the images, it serves as a free source of voiceovers. Given the benefits of voiceovers described in the previous chapter, at least in aiding the interpretation of photographs by a remote audience, it would be useful to capture them automatically. This would overcome the time and effort required to consciously record voiceovers for each new set of photographs. In the example above, perhaps the conversation between family members about a new photograph, could be recorded and transmitted with the photograph itself to remote family and friends. Alternatively, perhaps a collection of conversations around a single photograph could be recorded and kept with the photograph for future personal reference by the author or subject. These possibilities are explored here and contrasted with previous findings on the properties of voiceover on photographs.

The approach to studying conversational photographs is somewhat different to that used to study other kinds of audiophotos. The use of a single trial methodology is possible, but would affect the very behavior we are trying to study. It would involve asking people to share photographs with each other, and then to select parts of the recorded conversation to keep with the photographs for posterity. The ensuing conversation might be very different from that which would have taken place spontaneously without intervention. A more suitable approach would be to record spontaneous episodes of photo-talk, and then to ask participants to identify those parts of the conversation they would like to keep with the photographs involved. This is a considerable challenge, since an experimenter would either have to be on hand to do this sporadically over long periods of time, or delegate the tasks of recording and selecting to participants themselves.

In fact, a variation of both approaches is reported here. A study of naturalistic photo sharing was used to examine the character of photo-talk itself, and its potential to serve as an effective form of voiceover for the photographs involved. The findings also establish a basic understanding of ordinary printed photo sharing against which to evaluate comments on print-based and screen-based methods of sharing audiophotos in the next chapter. Participants in this study were asked to fill in photo-diaries tracking the occurrence of photo sharing conversations over a 3-month period, and to record as many conversations as possible on audiotape. A subsequent laboratory study of photo sharing was used to assess the value of recording particular kinds of conversational photographs identified in the first study. Participants here reviewed a video of their photo sharing session and ticked off, from a contact sheet, those photographs for which they would have like to have saved the corresponding conversation. To set the scene for both studies, I briefly review existing work in this area, in order to build on what is already known about these sorts of conversations.

5.1 Related Work

While almost nothing is known about photo-talk itself, a related form of talk without photographs has been extensively studied. I will refer to this form as *conversational storytelling*. This involves telling stories of one's own or other people's lives in order to make points in the course of a conversation. This is one kind of talk amongst many that have been studied within the fields of sociolinguistics and conversation analysis (Goffman 1981; Drew 1994). The relevance of conversational storytelling to photo-talk will become clear as we apply what is known about it to a common sense view of storytelling with photographs. I present this view as a 'straw man' theory about photo-talk that is de-bunked by the work on conversational storytelling, and by work reported later in this chapter.

Thus from any experience, such as a holiday abroad, people may take photographs to remind them of the things they did. Any individual photograph can trigger a memory of what actually happened and this can be used in conversation to illustrate a story describing the course of events. Because the photograph always indexes the same event and memory, it should lead to roughly same story being re-told time and again in different conversations. Furthermore, because the order of photographs in a set usually corresponds to the order in which they were taken in time, the stories they trigger will tend to be narratives describing the unfolding of activities over time. I will call this account the *slide-show model of photo-talk*, because it suggests a single person standing up before an audience and taking people through a series of photographs and memories in a fixed presentation. Although the audience might call out questions of clarification from time to time, or gasp in amazement at twists in the storyline, the presentation would run its course until the last photo had been explained. In fact this model is implicitly assumed in current digital photo sharing applications. They usually distinguish between an ad hoc photo browsing mode for personal use, and a more linear slide-show mode for presentation to an audience. Stories are assumed to be cued by the presentation, and to unfold with the photographs.

Conversational storytelling turns out to be an altogether more difficult and complex achievement. In the first place, the storyteller must find a way of introducing the story in the course of an ongoing conversation and phrase it in such a way as to reach a conclusion. Failure to do so would result in the storyteller being perceived as mad, rude or self-indulgent, and the point of the narrative would be lost (Polanyi 1985). Consequently there is usually 'entrance' and 'exit' talk in the course of telling a story in conversation that marks it out from other activities such as remarks, questions, requests and so on. This marking is particularly important if the story is to take more than one turn, since it forestalls the listener from interrupting the story in its telling (Sacks et al. 1974). Sacks (1968a, 1974) describes the talk which does the initial marking as a *story preface*, and shows how it sets up the listener to listen for a certain point in the story's completion. For example, in the following story preface, a woman (A) gets ready to tell her friend (B) about a car wreck she saw on the freeway (see Fig. 5.1). However, the point of the story is that it was not reported in the local paper. The relevance of the paper is introduced at the outset of the story. Consequently, B

```
A: Say did you see anything in the paper last night or hear
   anything on the local radio, Ruth Henderson and I drove
   down to Ventura yesterday
B: Mm hm
A: And on the way home we saw the:: most gosh-awful wreck
B: Oh::::
```

Fig. 5.1 A story preface (Sacks 1968b). A *colon* (:) indicates drawl on the last syllable before it

does not respond to A's initial question about reading the papers in the second turn. She waits until the story of what happened on the drive to Ventura is completed with "It wasn't in the paper last night, I looked" some 19 turns later (not shown). In each of these examples the figure labels are used to introduce a number of conventions for transcribing talk, developed within conversation analysis. Only those which are helpful to the analysis are used. For further details of these and other transcript conventions see the opening sections of Atkinson and Heritage (1984).

While the introduction of stories relating to photographs might be prefaced by the presentation of a particular photograph, it is unlikely that this act alone would perform all the subtle work of a conventional story preface. It is much more likely that any stories about photographs will be embedded in a stream of non-story talk, and require introduction and conclusion in the usual way.

Another finding about the conclusion of stories further complicates this picture. In conversation, people often tell *second stories* in response to a first. A second story not only demonstrates that they have been listening and understand the point of the first story (Sacks 1968b), it also serves as a form of reciprocal self-disclosure (Duck and Pittman 1994; Sprecher 1987). For example, in response to the story prefaced in Extract 1 above, B offers the second story shown partially in Fig. 5.2. It describes her witnessing a different disaster and also failing to find it reported in the papers. Although these stories are not full of intimate life details, their exchange serves to maintain a friendship between these two women that includes sharing their observations and questions on life (c.f. Duck 1994). In the context of a friendship, a one way sharing would be strange. The telling of a series of (first) stories from photographs without allowing the audience to respond would also be strange, especially if the audience is well known to the storyteller. Photo-talk then, might be more or less littered with second stories by audience members. These would have nothing to do with the photographer's memory of events but everything to do with the audience's memory of similar events and observations.

So far we have seen that stories about photographs, as about anything else, might be embedded in a stream of talk that includes non-story talk, introductions to the stories, conclusions or punch lines, and second stories told by the audience. A final set of findings from the literature shows that individual stories themselves are subject to variation within and between tellings. This means that the same story may never be told in the same way twice. This has to do not only with the influence of the audience on the story, but also with the influence of the storytelling on the author's memory for the story.

```
 B: You know, I looked and looked in the paper- I think I told
 you f- for that uh f-fall over at the Bowl that night. And I
 never saw a thing about it, and I   looked in the next couple
 of evenings.
  A: Mm hm
  B: Never saw a th- mention of it.
```

Fig. 5.2 A second story preface (Sacks 1968b). A *dash* (–) indicates a word cut off in mid-flow

From any event it is possible for people to generate a wide variety of accounts or descriptions. This can be seen from witness statements to the police following a crime, or from the verbal accounts of crimes in courtroom interactions (Atkinson and Drew 1979). In fact, the system of cross-examination in court can elicit quite different accounts of the same event from the same witness, depending on the choice of words and line of questioning used (Drew 1985). In general, people try to design their versions of events to be relevant and understandable to the person they are talking to, and that person usually cooperates with them to understand what they are saying (Grice 1975). These two factors of *recipient design* and *audience participation* work together to shape the account or story that is told for any event.

Stories are further affected by the presence and participation of others who share the memory of the story subject. In fact, where two or more people share the memory, the story is often told collaboratively with input from each of them. A brief example of *collaborative storytelling* is shown in Fig. 5.3. This shows part of a joint account of the film ET, being generated by a group of undergraduates who have just watched it (Edwards and Middleton 1986). Instances of this kind of talk are challenging for theories of human memory which overlook the role of other people in the process of remembering. They are cited as evidence for new theories of *collective remembering* which suggest that memory itself can be reconstructed and shaped through storytelling with others (Edwards et al. 1992; Edwards 1997). Eventually, people may come to believe more strongly in the stories they tell of an event than in the memory of the event itself (e.g. Tversky and Marsh 2000). They may also learn to recognise and participate in the re-telling of familiar family stories as a way of fostering group rapport and passing on family values. Norrick (1997) has called this *re-membering*, because it involves the strengthening of family or group membership through the joint re-telling shared stories.

Only two published studies have attempted to look at these kinds of behaviours in the context of storytelling with photographs. One of these was mentioned in Chap. 6 in connection with recording voiceover on digital photos (Balabanovic et al. 2000). This study also included a 10–15 min period of face-to-face photo sharing using a handheld photo viewer. Participants were found to move between a 'photo-driven' and a 'story-driven' method of sharing photographs. Either the photographs themselves appeared to trigger the conversation or the conversation suggested a story that might be illustrated with new photographs. This begins to place the style of talk somewhere *between* a slide-show and ordinary conversation.

```
Karen: Elliott is sitting at his school desk and they are doing
       experiments with frogs are they
Diane: and he lets all the frogs out
       {General hubbub of agreement}
Tina:  sets all the frogs out yeah
Lesley: and what's that little girl that he fancies
John:  it's when he's watching E.T.'s watching television and John
       Wayne kisses the heroine in the film
Diane: oh so Elliott kisses [her
John                        [and then Elliott kisses the little
       girl
```

Fig. 5.3 Collaborative storytelling about the film 'E.T.' (Edwards and Middleton 1986). *Square brackets* ([) on consecutive lines are used to indicate overlapping speech

The focus on photo selection in this study obscured analysis of the conversation itself, and lacked consideration of story prefacing, second stories and audience participation. The use of pairs of participants also excluded the possibility of looking at collaborative storytelling, in which two or more people tell a shared story to others. An earlier study of printed photo sharing between parents and young children, aged between 4 and 6, discovered some of these phenomena operating in photo-talk (Edwards and Middleton 1988). For example, parents acted to bring the children into a conversation about the past events represented in the photographs, increasing audience participation and encouraging collaborative storytelling. The main shortcoming of the study was the fact that the setting was not characteristic of ordinary photo sharing situations with a variety of adults and children of more equal status in the conversation.

In the next section, I describe a new study of photo-talk recorded in situations with just this kind of diversity. The aim of the study is to see whether the findings of studies on conversational storytelling apply to photo-sharing conversations, or whether the latter are special in some sense, because of their dependence on photographs. This will reveal the extent to which the slide-show model applies, in spite of known phenomena that cause speakers to mark and change their stories for the audience at hand, and listeners to collaborate, steer and respond to stories in various ways. The implications of the findings will then be used to target particular kinds of photo-talk that might be recorded as free voiceover for photographs. A follow-up study is then reported which elicits reaction to the use of this kind of talk in conversational photographs.

5.2   The Photo-Talk Study

Eleven families in Northern California were interviewed at home about their photo organisation and sharing practices (Frohlich et al. 2002). Each family owned a personal computer, and at least two of the following three digital photography products: a scanner, a digital camera, and a photo-quality printer. In addition, they used at least 10 rolls of conventional film a year. These criteria ensured that the

participants were involved in both analogue and digital photography and could answer questions on both subjects in the interviews.

At the end of the interviews, families were given photo diaries and audio tape recorders with which to record subsequent photo-sharing conversations. The diary comprised a simple table in which each row represented a single photo-sharing episode, and each column entry represented a piece of information about it. The information was written in by hand, and included details such as when and where the episode took place, the name of the photo set, whether the photos were in loose or album form, the names of the other participants, and whether they were present at the time the photos were taken. Families were asked to fill in one entry in the diary each time they got their photos out to share with someone, or each time someone else presented their photos to them. They were also asked to tape record these conversations if possible, with the permission of the other people involved. This procedure was followed by each family for a 3-month period over the Christmas of 1998.

Some basic properties of photo-talk are evident from the resulting diary data. This will be reviewed first before examining the details of the recorded conversations themselves. This approach is similar to that used in Frohlich et al. (1997), in which various quantitative aspects of a set of conversations are used to motivate the qualitative analysis of individual conversations.

5.3   Background Features of Photo-Sharing Conversations

A total of 127 photo-sharing conversations were recorded in full in the photo diaries. This amounts to about one conversation a week for each family. The majority of conversations involved collections of printed photographs, despite the fact that families were using digital photography products. This can be seen in Fig. 5.4, which shows the proportion of different types of photographs shared. Most sessions involved either loose prints or prints in albums, with only seven recorded episodes of screen-based photo sharing. This suggests that participants may have been printing out digital photos to share them face-to-face, in preference to sharing them on the PC or the camera itself.

The form in which photographs are shared is likely to influence the kind of conversation that takes place around them, since this will affect how many people can see individual images at the same time and who can take control of them at any point. Other contextual factors that may influence the conversation include the location in which the photos are shared and the number and type of participants. On these dimensions we found that 82 % of conversations took place in participants' homes; usually in a kitchen, dining area or living room. The average number of participants present was four. These people tended to be family (40 %) or friends (50 %) rather than acquaintances (10 %).

A final feature of these conversations was recorded for its relevance to the diamond framework in Fig. 1.2 and turned out to be a critical factor in the ensuing talk. Participants were asked to indicate in the diary, whether or not each partner in a photo-sharing conversation shared the memories represented in the photographs.

This effectively determined whether the conversations involved reminiscing or storytelling. The findings are summarized in Fig. 5.5. We found that the majority of conversations involved reminiscing between people who had all been present at the time the photographs were taken, while a sizable minority involved storytelling with people who had not been present at that time. An additional discovery was that a third set of conversations involved a mixture of reminiscing and storytelling because they included more than one person of each kind. Typically a couple sharing the memories of the photographs would show them to another couple.

While this picture of photo-sharing may not be representative of the full spectrum of domestic photographic practices, it does resonate with what has been reported by other methods (e.g. Chalfen 1987). In the next section we go beyond both interview and diary self-reports of photo-sharing, to examine the content and dynamics of photo-talk from 80 conversations recorded on tape.
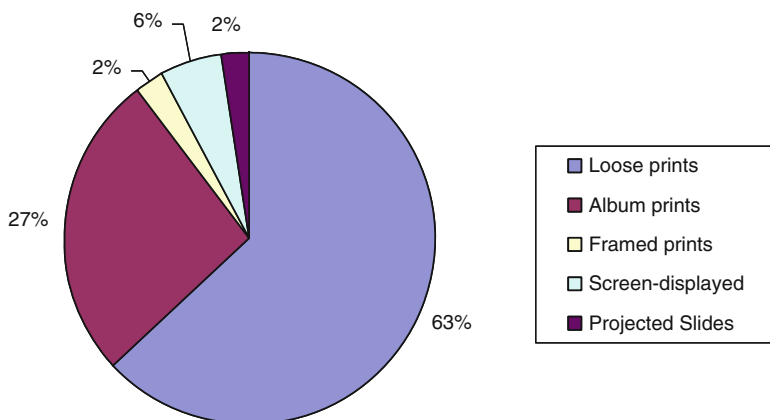

Fig. 5.4 Proportion of conversations using photographs of different forms
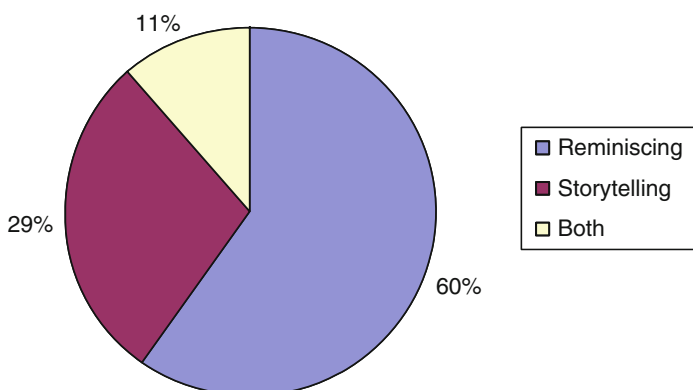
Fig. 5.5 The proportion of reminiscing and storytelling conversations with photographs

5.4   The Content and Dynamics of Photo-Talk

The character of the recorded conversations was dramatically affected by whether or not the participants shared the memory of the events depicted in the photographs. Essentially, stories of the photographs emerged only in the storytelling conversations, or in the joint storytelling-and-reminiscing conversations, between people who did not all share the memories represented by the photographs. Together these conversations made up 40 % of the total number of conversations recorded in the study (see again Fig. 5.5). In the remaining 60 % of reminiscing conversations, participants tended to make a series of short remarks and observations about the photographs. These were often directed towards finding the memory together, only to leave it unelaborated once it had been discovered.

An example of this kind of reminiscing talk is shown in Fig. 5.6. This conversation took place within a family of five who were looking over a new set of prints developed from an old roll of film. It was Christmas 1998 when the conversation was recorded. The main topic of the conversation was deciding when the photos were taken. Each member of the family chips in their own contribution to this topic, often in the form of questions, until the family as a whole arrives at an estimate of Christmas 1996. Many of these contributions are difficult to hear because they are made in overlap with an existing speaker (shown in square brackets). Although the question of '*When*?' the photographs were taken is addressed by this talk, it is done so in a way that is both protracted and unclear. Furthermore, the participants did not go back to address other questions about individual photographs shown in the conversation: making it altogether unsuitable as a form of voiceover on the photographs themselves.

One reason why a range of questions are not addressed in reminiscing conversations is that the participants already know the answers to most of them. In fact, telling the story of a photograph to someone who was with you when you took it would be bizarre, unless they had forgotten the event altogether. Even in these cases, there is evidence that people are reticent to explain a photograph directly. This can be seen in Fig. 5.7, which shows a fragment of conversation between two sisters, Jenny and Clare. They are looking at photographs of Jenny's daughter's birthday party at which they were both present. Jenny has forgotten **why** she is shown making a phone call in one of the photographs. Instead of telling her, Clare feeds her clues about the event that cause Jenny to remember that she was phoning her daughter's new pager. In this way, the two sisters arrive at the answer together. This kind of collaborative question-answering is typical of reminiscing conversations, but tends to stop short of collaborative story-telling itself. The instruction to "Remember" is used here as elsewhere to encourage others to find and rehearse their own version of events from memory rather than to listen to someone else's version.

In the absence of storytelling about the events depicted in photographs, reminiscing talk tends to focus on individual questions of context, on the superficial appearances of people and things in the photographs, or on the technical features of the photographs themselves. Question-answering has already been illustrated in Extracts 5.6 and 5.7. An example of commenting on appearances is shown in

```
Mar: No :   you know what? is that last, year's, presents?
     though?
Emm: Yeah those are last [ y e a r : s : . ]
Joh:                      [What's that big b]ox
Nie: Yeah well we must have had that roll in there
     fe[r a long ti[me]?
Emm:   [No that's la[s ]t °years.°]
Mar:               [ JO::HN? ]don'[t  bi]te my shoul:der:]
Joh:                              [sorry.] I didn't bite yu-]
Mar: Yes you di::[d
Nie:             [But loo-]
Emm:             [That's  la]st  years  because  these  are  in
     yuhknow thuh fireplace:
Joh: But we had um
Emm: They're normally li[ke
Mar:                     [Uh granma n' granpa's stockings are
     there
Nui: Was that the y[ear we had two? christma]sses?
Emm:               [ Was that  last year ]
Mar: That was two years ago
Nie: It was [christmas ninety si]x then
Emm:        [ Two  years  ago ]
Mar: Yep cu[z (granma prop]er) [poppa didn't come] this
Joh:      [ Y E S c u z-]      [ c u z-  m o m ]
Mar: la[st chr]istmas=
Joh:   [ m o m ]
Joh: Thuh calenders
Mar: Uh [ha]
Nie:    [Ye]ah (0.4) the do[:g wasn't even full grown ]
Mar:                       [I didn't do calenders last] year
Joh: Ri:ght
```

Fig. 5.6 A reminiscing conversation about some old Christmas photographs. Speech spoken loudly is shown in capitals

```
Jen: What am I doing making a phone ca:ll? why would I take
     that picture [look at those thi:ghs ]
Cla:             [ O h  t h- o::h ]do you remember what
     you're doing here?
Jen: No.
Cla: Think for a second Valerie's birth[da:::y?]
Jen:                                    [OH: yeh] Valerie got
     her new uhm  pager that she want so badly un I was pagin'
     her
Cla: U[ h u r? ri:]ght? uhu?
Jen:  [that's right]
Jen: That's why you took a picture with her makin' a phone call
```

Fig. 5.7 A reminiscing conversation in which one person prompts another to remember why a photograph was taken

Fig. 5.8. Dan compliments Marjorie's appearance in one of the photographs and Marjorie objects and finally diverts attention to another photograph. This is a common response to compliments, but should not be taken to indicate that the compliments are unwelcome (Pomerantz 1978). Other things people commented on included subjects taken from unusual angles, how people and things have changed, the exposure, focus and composition of a photograph, and the camera equipment used. Again, these subjects are unlikely to make good voiceover clips for photographs.

More promising voiceover material can be seen in the storytelling conversations recorded in the study. These include conversations in which there is also an element of reminiscing (shown as 'Both' in Fig. 5.5). In these cases, some account of the activities taking place in the photographs is often given to those who were not present at the time. However, as we shall see, these accounts were not presented routinely for each photograph as the slide show model of photo sharing might suggest. Instead, they were often triggered by audience interest in a particular photograph or by a turn in the ongoing conversation that made an account relevant at that point. As in storytelling without photographs, accounts or stories of photographs were often marked at their beginning and end, modified according to audience interest and participation, told collaboratively with others who did share the memory, and followed at times by second stories from the audience. The main difference between photo-talk and ordinary conversation seemed to be in the pace and focus of the talk. Photo-stories were embedded in a photo-turning activity that formed the focus of the conversation, and served to prime certain topics associated with each photograph at a particular time and pace. It also meant that participants could pause quietly for several seconds in the conversation to examine a photograph in mutual silence. To appreciate these characteristics, let us consider some further fragments of photo-talk.

Figure 5.9 shows a typical photo-story being told by a wife to her husband. Tracy, the wife, had recently been on vacation with a friend, Annabelle, and was showing photographs of the trip to her husband Simon when she got back. In contrast to the symmetrical kind of reminiscing conversation shown in Fig. 5.6, this conversation is led by Tracy with occasional contributions from Simon. Tracy marks out a photograph of a tea room in Atlanta Springs as worthy of particular attention, by saying "This was an interesting thing" and asking Simon a question about the picture. This

```
    Mar: An who's tha::t.
    Dan: That's a great picture uh you
    Mar: No it's really not=
    Dan: =It is a great pict[ure uh you. ]
    Mar:                     [No it's really] kinda ugly hhh.=
    Tom: =(Na[:r.)]
    Dan:    [No i]t's no[t yer har-]
    Mar:            [ I T I S] UGLY Tom.
    Dan: Let me take a look at it
    Mar: Oh look at Emma...
```

Fig. 5.8 A reminiscing conversation involving an assessment of a photograph

```
Tra: Annabelle and me anyway we ate at the painted lady er:m
     tea room
Sim: Uhum?
Tra: A::nd ee- we were sittin' out [ en this: ]
Sim:                               [Is that one] of thuh
     rooms? in thuh house?
Tra: Uh- no t[his wu]z th[is wuz uh] in still in
Sim:        [oh this]  [ o k a y ]
Tra: Atlanta springs .hh an:d urm
Tra: This? was a real interesting thing can you tell what's
     in that tree? right there?
Sim: They look like pumpkins
Tra: Well those are actually lights.
Tra: An um tch
Tra: Buh- what's hanging in the tree is a cup an she
     an- thuh ther wus liddle tea cups hanging all over this
     tree? an then underneath it in thuh ya:r:d there were
     liddle like dishes to match thuh tea cups so
Sim: Why::?
Tra: I don't know b[u t  a- hh]h.
Sim:              [heh heh heh]
Tra: Annabelle said look at thuh dishes out there they all
     are
Sim: En did you say that's called the tea garden
Tra: I said it was thuh dish garden [hih hih hih]  hh hih
Sim:                                [heh heh heh]
```

Fig. 5.9 A storytelling conversation between a wife and her husband

serves as the preface to a story about the discovery of cups suspended in a tree and dishes (saucers) placed in the ground beneath it. The close of the story is prompted by Simon who asks for clarification of the name of the place and, by implication, some further explanation of the strange decorations. Tracy doesn't provide an explanation but changes her earlier description of the tea room to a "dish garden", underlining the strangeness of the experience. The joint laughter at the end of the story serves to mark its closure for both parties. In this way, the account exhibits many of the properties of storytelling in ordinary conversation, but happens to be triggered by a photograph of the place being spoken about.

The fact that this story is tailored to Simon's interests and intervention, can be seen in Fig. 5.10 where Tracy tells the same story to two other friends, Katie and Lyn. The story elements are similar but the telling is different. It is now more appropriate to the relationship Tracy has with her female friends. Hence, Tracy first picks up on the name of the tea room, the "Painted Lady", to make a joke about its connection to herself and Annabelle. Katies' humourous response and Lyn's laughter indicate that they get the joke, allowing Tracy to continue with the account. This time the story is told in an extended turn which Katie and Lyn refrain from interrupting despite several pauses in the middle. Whereas Tracy's description of the place as a "dish garden" was prompted by Simon's question in the first telling, it has now become the punchline of the story as something she said to Annabelle at the time. In fact these words were most probably spoken to

```
    Tra: An we ate at the painted lady café
    Kat: hmm
    Tra: I thought that was appropriate for thuh two of us
    Lyn: huh huh huh
    Kat: I'm not touchin[g that one]
    Lyn:      [Ha ha ha h]a
    Tra: And they had in thuh tree there were cups hanging in thuh
         tree? an stuck in thuh dir:t were liddle plates haf
         sticking up out thuh dirt an I looked out thuh window an I
         said it's a dish garden
    Kat: Did you ask them why?
    Tra: No  I just got a (picture)
    Kat: Okay interesting
```

Fig. 5.10 A re-telling of the same story as in Fig. 5.9, to two female friends

Simon first, judging from Fig. 5.9. These modifications show that photo-talk is subject to recipient design in the same way that ordinary conversation is, and that this influence may even lead people to modify story elements between tellings.

In addition to the implicit role played by the *identity* of the audience in the storytelling, we observed a much more direct role through their active participation in the conversation. Some indication of this role can be seen in the contribution of Simon to the conversation shown in Fig. 5.9. For example, Simon asked a series of questions that led to the framing and conclusion of the first story told by Tracy. A more dramatic illustration of audience participation is shown in Fig. 5.11. Here Emily and Jack were showing two visiting friends, Grace and Ted, pictures of an old vacation in New Zealand.

The photo-talk in Fig. 5.11 was triggered by the fact that Grace and Ted were soon to visit New Zealand themselves, and were interested in learning more about the country from their hosts' old photographs and experiences. At the beginning of the stretch of conversation shown, Emily and Jack are showing a photograph of a farmstead that contained a craft museum. As it turned out, this was near where Emily bought a spinning wheel, and both Emily and Jack begin to tell the story of the spinning wheel purchase until Grace interrupts with a question. The interruption is done in overlap with the story preface in which the spinning wheel is first mentioned. This serves to reject the story at that point in favour of obtaining further information about a Ferry shown either in the farmstead picture or in another one close by. After a pause, Emily and Jack go on to answer the question together, and a series of related questions about the weather in that region. However they don't give up on the spinning wheel story. A subsequent photograph "back in queensland" is used to set-up the story preface for the second time, using an almost identical form of words. This time the visitors desist from interruption and the story follows.

Although audience participation here only served to delay the onset of a photo-story from the storyteller's point of view, it also elicited useful details of the geography of New Zealand from the audience's point of view. In the context of a pending

```
   Emi: Anyway um (uh that was [that was stunning)]
   Jac:                        [thuh farmstead wa]s beau- I mean
        really intere[sting ]
   Emi:              [But yuh]know it's set up like a museum so
        they demonstrate all thuh different cra:fts uh
   Jac: Un this is thuh
   Emi: That's wh[en (    )] buy the spinning wheel
   Jac:          [spinning wheel]
   Gra:          [Uh- This looks] like a Fe:rry or something
        they're saying goodbye? or welcome or something
     {Pause}
   Emi: Ee- thuh were yuhknow we took a ferry from thuh little
        tow- uh thuh town of queensland over tuh this
        (r[anch over: ) ]
   Jac:   [Down this lake] maybe:: ten mile[s (there)
        -
   {Discussion of mountains and weather for 20 more turns}
        -
   Jac: Oh this is this is back in
   Emi: That's a lovely place
   Jac: This is back in queensland I think you're getting ready
```

Fig. 5.11 A storytelling conversation in which an original attempt to tell a story is rejected and redirected by the audience

vacation to the country this information is likely to have been more interesting to Grace and Ted than the fact that Emily bought a spinning wheel there on a previous trip. This trade-off between the interests of the storytellers and the interests of the audience lies at the heart of this kind of photo-talk and was actively negotiated in the storytelling conversations we analysed.

Where there was more than one storyteller, this conflict of interest and perspective extended to the storytellers themselves. Two storytellers might want to draw out different lessons from the same experience or would simply remember different details of the same event. This led to exactly the kind of collaborative storytelling observed by Edwards and Middleton (1986).

An example of this is shown in Fig. 5.12. In a typical photo-sharing session, Natalie and Edward are showing their photos of a recent camping trip to their parents, Jim and Delia. In this part of the conversation, Natalie and Edward are prompted by Delia to tell a story of their encounter with racoons. Natalie and Edward work together to relate the story from their own recollections, with Natalie uttering the story preface "The first place we stopped was like a zoo". Edward subsequently takes up the story and resists an attempt by Natalie to take over half way through. Natalie completes the story with a more forceful ending than Edward's "So we packed up and left". She uses more dramatic language such as "just freaking along" to indicate the frantic state in which they left, and tie the ending back to her own story preface in which she likened the campsite to a zoo. This shows that photo-stories are not only influenced by the audience, they are also influenced by multiple storytellers who may collaborate to tell the story together.

```
Nat: So this is where we stopped thuh first night
Jim: So yo[u took off from here]
Nat:    ['cuz we had gr(h)eat] tr(h)ouble
    s[o we didn't make      it all]
Jim: [ yer heading for (home)]
Nat: Thuh way
Jim: and Noel? was with yo[u
Del:                       [was this [thuh] next morning?
Nat:                                 [yeh ]
Del: 'Cuz yo[u guy]s stopped in ell aye mm hm
Nat:        [y e h]
Del: That doesn't look ba- I mean it was a bona fide camp site
Nat: Yeh [no it wuz] fine
Del:     [ mm hm ]
Nat:  B[u- thuh first place we sto]pped was like a zoo
Edw:   [ there  were  so  many ]
Edw: Yeah we pulled intuh on[e s i t e] right near thuh warder
Nat:                        [hu-hh. huh]
Edw: Un there is kind a seven racoons en there were skunks
    just walking all over
Del: Oh my goodness
Edw: Then I threw ro:cks at thuh racoons en they just kinda
    brushed um off their eyes en kept comin' after us
Del: Hu-[hh.
Jim:    [Ku-[hh.
Del: So you hadn't un packed yur
Edw: So:: uh[: we're eating ]
Del:        [yuh didn't camp]
Nat: Well we kin[de[r had hh.]
Edw:           [Di[n n e r :]
Del:           [Oh you w]er:e?
Edw: S[o we packed [up an' left] un: went to another one
Nat: [So I (jus-) [ huh  huh ]
Nat: He just drove me I'm like holding on tuh thuh pa:n an' all
    thuh dinner stuff in thuh ca:mper en he's just freaking
    along yuh know
```

Fig. 5.12 Collaborative storytelling by a young couple to their parents

Another feature of the episode shown in Fig. 5.12 is that the story is triggered from a single photograph that does not seem to contain images of the racoons or skunks that the story is about. Delia's original observation is about the campsite Natalie and Edward travelled to *after* they had escaped from the killer racoons! We found it was common for stories to spring off a single photograph in this way, often at the prompting of the audience and sometimes in a direction not illustrated by the photograph itself.

Another example of a story about something *not* shown in the photographs is shown in Fig. 5.13. In this conversation Dan and Mary are being shown photographs by their friends Nicola and Lenny, whose daughter is also present. Just prior to the clip shown, Dan asks if the photographs were taken with Nicola's point and shoot camera. As it turns out, they were taken with another camera that was lost and then

```
Nic: My camera disappeared at thuh Honalulu airport a thought it
     [ wuz stolen ]
Mar: [.hh en that's] why you got thuh new liddle one?
Nic: Yes Len[ny bought me thuh new liddle one] fer my:
Dan:        [ N o : :  k i d d i n g ]
Len: I g[ot her a (elth)]
Nic:    [it's own it wuz] it wuz in a liddle fa:nny pack it wuz
        in a liddle fa:nny pack en it wuz my camera an Dan's
        disposable camera [and my] comb was there
Mar:                      [OH::.]
Nic: An that wu- that wuz like all that wuz in that liddle
     fanny pack .hh en so: I didn't discover it till we arrived
     in San Fransisco so it wuz just lo:st. so a mean it was my
     own stupidity to have [.hh lo:st it s]o
Mar: [That's sad though]
Nic: I was just very upset aboudit end about two weeks ago a
      woman ca:lled ur: from Honalulu baggage clai:m she works
      weekends in thuh baggage claim .hh en she'd found...
   .
   {The story of how the lost possessions were eventually
    returned continues for approx 20 seconds}
   .
Nic: Yeh she- she sent um back so: i[t it's
Dan:                                [isn't that wild
Mar: That's amazing  that's ama[zing
Dan:                           [it's like when Mary lost her
     purse fer some reason that she got it back. a wallet? or
     whatever it was?
```

Fig. 5.13 A first and second story unrelated to the photograph that triggered them

surprisingly found and returned sometime later. Nicola tells this story with help
from Lenny, and Dan responds with a *second* story about a similar incident in which
Mary's purse was stolen and then discovered sometime later.

Sometimes people did unfold a story with reference to a sequence of photos,
but this was always at a pace and manner dictated more by the audience than the
material. More often than not, the turning or passing of photographs within the
group served to make relevant certain topics of conversation, and some of these
would be developed into stories at the whim of the storytellers or the prompting of
the audience. In between these developments, which have been picked out for
illustration in Figs. 5.9, 5.10, 5.11, 5.12, and 5.13, there appeared to be a good deal
of what might be called 'filler talk' by storytellers such as that shown in Fig. 5.14.
This is taken from another portion of the conversation between Tracy and her hus-
band featured in Fig. 5.9. Here phrases such as "an then this", "here", and "that
was", are used to mark the passage onto a new photograph and preface a simple
voice label. Pauses are introduced between each photograph, at which point the
audience gets a chance to make a comment or ask a question. In this exchange
Simon remains silent apart from expressing an objection to Tracy's assertion that
the trees were changing colour.

```
Tra: Lady's things an it was just decorated real real sweet .hh
     an then this was on thuh RO:ad from ur:m
   (1.0)
Tra: Oklahoma to arkansas .hh just some of th- su- thuh scenery
     that we stopped un
   (1.0)
Tra: Took- thuh trees here were just beginning tuh tur:n
   (0.7)
Tra: An: en it just wuh- absolutely gor:geous
   (2.2)
Tra: That was a little f[arm
Sim:                     [can- ee- I can't really tell that much
     of a colour change there  I mean that's not all yellow
     anyway
Tra: mm hm
Sim: That d[oes ] look mottled
Tra:       [koo-]
   (0.9)
Tra: It wuz really priddy
   (1.9)
```

Fig. 5.14 Filler talk by Tracy in between storytelling to her husband Simon. Pauses in seconds are shown in brackets between the lines

5.5 Summary of Findings on Photo-Talk

In short, we have found in this study that photo-talk is composed of two quite distinct types of talk. Reminiscing talk was the most common type, and took place between people who shared the memory represented in the photographs being shown. It was characterised by equal participation in the conversation by all concerned, a good deal of overlapping talk, and a reticence to elaborate the story-of-the-photo beyond identifying where and when it was taken. Storytelling talk was the other kind of talk, and took place between people who shared the memory of the photographs and those who didn't. It was characterised by unequal participation in the conversation, little overlap, and a tendency for the photographer to develop the story behind *particular* photographs in a set. The way in which such stories were told in these conversations, appeared to be similar to the way in which stories are told in other forms of conversation. Hence they involved opening and closing sequences, tailoring to the audience, audience participation, collaboration, and even the telling of second stories in response. The main difference between storytelling in photo-talk and 'ordinary' talk, appeared to be in the way the story was initiated out of a photo-turning activity that made certain kinds of information *topical*.

Returning to the slideshow model of photo sharing we can say that these findings provide further evidence against it. Both kinds of photo-talk are highly interactive and influenced by the participation of an audience. In fact, in reminiscing conversations, the contributions are so equal that it is not appropriate to distinguish between an audience and a presenter. These roles move between participants with each utterance. Even in storytelling conversations where the presenter or presenters are defined

by their knowledge of the photographs, the audience was found to prompt and direct the conversation by timely interventions and responses. Indeed, the presenter's talk was interspersed with pauses to encourage this. The fact that most photo-sharing conversations were found to involve the handling of loose prints may also have contributed to this effect, since it would have allowed the audience to look at images which were out of synch' with the current narrative. Such dislocation of image and sound consumption may have created opportunities for redirecting the narrative to images not currently being talked about. While this would be frustrating in a slide show, it appeared to be the norm in the photo-sharing conversations we examined.

The implications of these findings for capturing conversational photographs are that only certain kinds of conversation are likely to be interesting enough to keep. In particular, only storytelling conversations appear to contain details that generalise beyond the conversational context itself and concern the story-of-the-photo in Berger's sense (1982a). The kind of reminiscing conversations shown in Figs. 5.6, 5.7, and 5.8 are unlikely to have long-term value as a form of voiceover on the corresponding photographs. Furthermore, *within* storytelling conversations, it may only be the stories themselves that are required for capture and storage with the photographs. The intervening filler talk and audience comments may be judged too boring or distracting to retain.

To test these hypotheses, we move onto a second study in which we gave people the chance to decide for themselves, which parts of a storytelling conversation they would like to keep with its associated photographs. This study is similar to those reported in earlier chapters, and involves the generation and review of audiophotos themselves.

5.6   The Storytelling Photograph Study

In order to give people a chance to record their own conversational photographs from storytelling conversations, it was necessary to find a situation in which these conversations could be recorded in full for later review. The spontaneous conversations recorded in the previous study were tape recorded without reference to their associated photographs, and could not easily be followed-up. However, an unpublished laboratory study of photo-sharing, carried out for other purposes, had recorded 14, 20 min sessions of storytelling conversations on video. This study was carried out with Nadja Linketscher and Maurizio Pilu. In this study, each photograph had been scanned and captured in digital form, so that it could be shared onscreen across two rooms through a photo-conferencing application called *PicShare*. An average of 30 photographs were shown and discussed in each session, at a rate of about one every 40 s. One of the values of this system is that it could easily support the capture of photo-sharing conversations. To explore this possibility, an additional follow-up exercise was carried out 12 months later on the recorded conversations. This was done in collaboration with Tony Clancy and allowed us to test reactions to recording storytelling photographs in general.

Ten pairs of work colleagues and four extended family groups were asked to review their photo-sharing sessions at home, by watching themselves on video.

Work colleagues were originally drawn from the HP Bristol site, while families were recruited from the local Bristol area. During this task they were given a printed contact sheet showing thumbnails of all the photographs shared, and asked to mark all photographs for which they would like to keep the corresponding conversation. In addition, they were asked to select a favourite conversational photograph from this set. Participants were then requested to complete a simple questionnaire describing their reasons for keeping the selected conversational photos, for **not** keeping the others, and for choosing their favourite audiophoto. The questionnaire also prompted participants to vote on whether or not they believed the recorded conversation made the photographs more interesting, informative or enjoyable, and to indicate whether they would like to record conversation with photographs again. Most of the conversations related to recent sets of photographs not seen by the photo-sharing partners in each pair or group. However, the interval between the original session and the follow-up research was about 1 year, and gave the conversations and photographs an older feel. This interval also made it difficult for us to track down all the original participants and encourage them to respond. Sixteen out of twenty-two individuals carried out the follow-up task and completed the questionnaire. Twelve people were members of the original pairs of subjects, while four belonged to one family group of two parents and two grandparents.

Despite the missing data, the responses from our 16 respondents contained a wide variety of opinions and experiences with the task, and furnished us with enough data to report a preliminary set of findings on the value of conversational photographs. This is done in the next section through an abbreviated form of analysis used in previous chapters. Statistical data are reported ahead of a discussion of favourite audiophoto selections and comments on the value of conversational audio. However, the statistics refer more to the subjective judgements and votes cast on the medium, rather than to the properties of the conversational photo set, and there is no discussion of the practice surrounding the capture, editing and playback of conversational photo material. We leave these details for a future full-scale trial, and concentrate on determining whether such a trial is justified from the data we have.

5.7   Storytelling Photos Captured

It has already been mentioned that the mean duration of a conversational photograph in this study was about 40 s. The association between sound and image in this set was always one-to-one because of the way a single photograph was presented at nearly full-screen size for discussion at any time. In fact, the duration of an individual conversational photograph was defined by the duration for which the image was displayed at this scale. For the most part, participants tailored their conversation to correspond to the currently displayed photograph, and the speech stream could be segmented quite straightforwardly at image transitions. The exceptions to this rule usually occurred when authors began to introduce the next photograph ahead of the current one. In these cases the entire conversational recording window for any given photograph was displaced to start a few seconds before the display of the photograph itself.

A key question for the study was whether or not participants would choose to keep *any* of the conversation corresponding to particular photographs. We found that on average they opted to keep the conversation on about half of the photographs shared. In those cases they did not express a desire to edit the conversations. This begins to show that there was some perceived value to the storage of storytelling conversation around photographs, but that this did not apply to every single photograph.

Two other factors in the selection of conversational fragments and photographs, were the ownership of the photographs and the relationship between photo-sharing partners. The interaction between these two factors is illustrated in Fig. 5.15. In general, people opted to keep important or forgotten details mentioned in their own talk on their own photographs, or revealing details on the photographs of close partners. They tended to reject most or all conversation on photographs shared by casual acquaintances, as shown by the low percentage of photographs saved by this group in Fig. 5.15. We defined acquaintances as people who did not socialise with each other outside work. In contrast, acquaintances retained nearly as many of their own photographs as did friends. Within the family group, we found that people were more interested in saving conversation on each other's photographs as shown in the figure. In fact, the grandparents in this group opted not to keep any of their own conversation on their own photographs on the grounds that they would not have time to listen to conversation on their own photos in the future. This explains the very low proportion of own photos saved by family members in Fig. 5.15.

The overall ratings for the conversational photographs saved are shown in Fig. 5.16 below. These show that the addition of conversational audio to these pho-
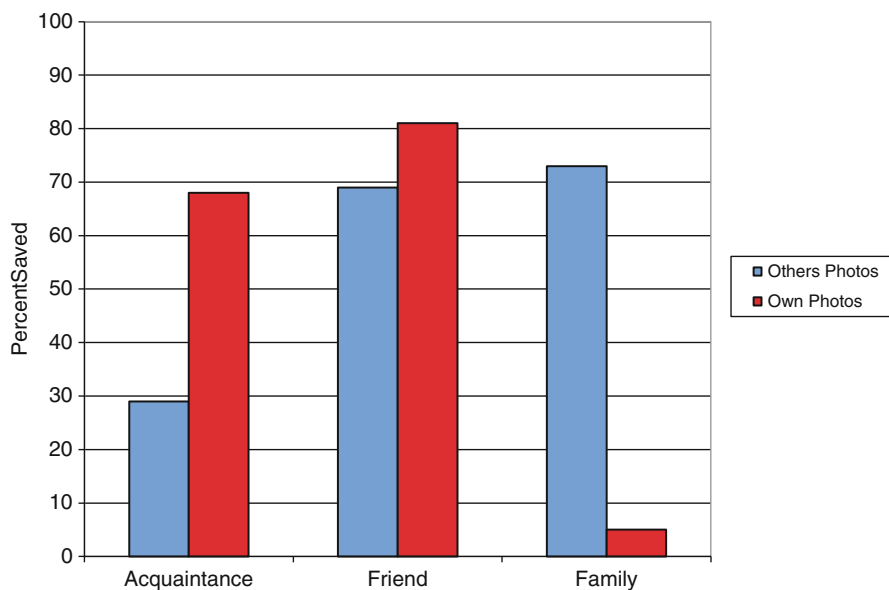
Fig. 5.15 The proportion of conversation saved on other people's photographs and one's own photographs by different groups in the study
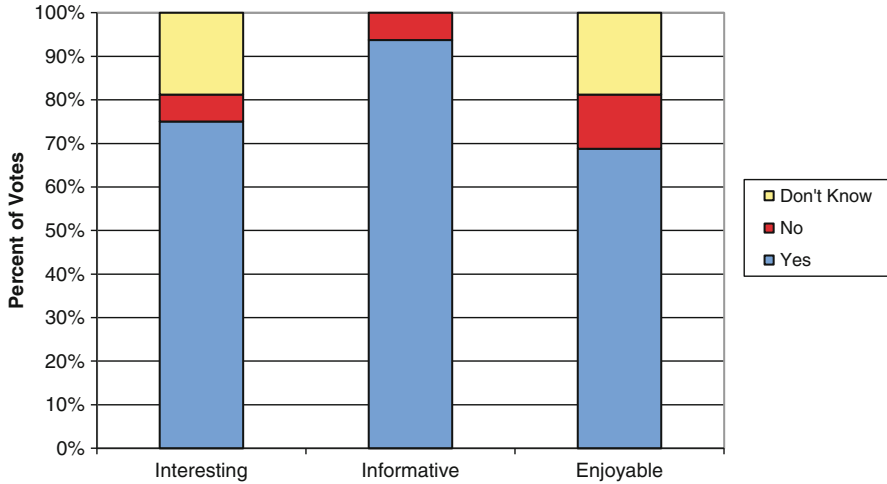
Fig. 5.16 The extent to which participants agreed that recorded conversation enhances a photograph in different ways

tographs was considered to increase their interest, information value and enjoyment by the majority of participants. The biggest consensus was around the 'informative' dimension. When asked at the end of the exercise whether they would like to record conversation on photographs again, three quarters of participants said 'Yes'. The remaining quarter were split equally between saying 'No' and 'Don't Know'. While the reasons for these votes are complex, their importance is clear. There are at least some circumstances in which saving the conversation around a photograph will be considered worth doing, and therefore worthy of further study and support.

5.8 Reactions

5.8.1 Favourites

A number of favourite conversational photographs were selected because they contained well-formed stories. These stories possessed the same kind of clarity and interest found in favourite talking photographs with voiceover. For example, Fig. 5.17 shows two people on a paraglider from the boat below. The conversation is led by one of the paragliders, Christine, who describes her feelings of terror as she took part in the ride. The story therefore provides an evaluation of the experience as well as a broader description of it, and was voted a favourite by Christine's photosharing partner Viv. As Viv points out in her own words, this audiophoto was a favourite because:

The conversation is an interesting and informative narrative on the photograph (Viv)
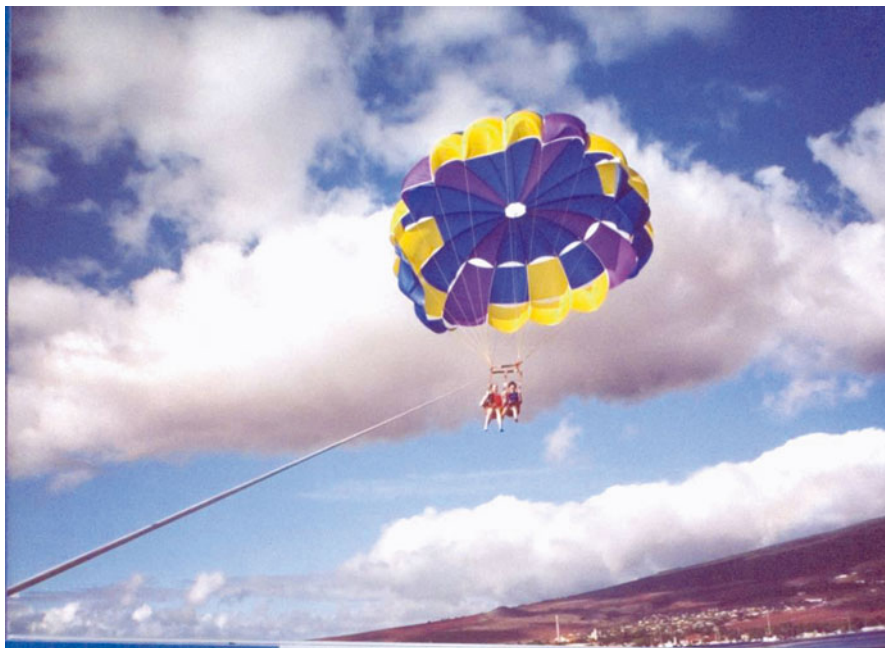
Fig. 5.17 Viv's favourite conversational story told by her partner Christine

However, favourite conversational photos were not always chosen because they were complete stories. Often the talk was considered attractive because it triggered a memory in the author or had some other personal significance. For example, the photographs shown in Fig. 5.18 were chosen as favourites because they reminded the author, Amanda, of the changing appearance of the same lake on consecutive days of her holiday. The accompanying conversation isn't so much a story as a description of the lake on both days with additional questions and observations by the audience, Jess.

A similar conversational photo was selected by Dorothy in the study. She chose a picture of herself in a mountainous landscape, taken by her husband. Her own conversational comments described her memories of being surrounded by wild herbs such as sage, rosemary and thyme, which gave the place a wonderful smell. The conversation itself appeared to remind her of this smell:

The conversation talked mainly about the sights/sounds/smells that went with the photo. This was definitely a bonus that the photo does not capture (Dorothy).

Other conversational photographs were chosen for their personal significance. These included a snapshot of Debbie outside a boutique in New York on her 40th birthday. This was chosen by her mother in the family group, not only because of the content of the conversation but also because of the association with that event:

Fig. 5.18 (**a, b**) Amanda's favourite conversational photographs which reminded her of the changing light on a lake

By hearing the conversation it helped me to feel I was sharing in that special day (Joy).

Both factors appeared to be at work in Fig. 5.19, which was the favourite audio-photo of its author, Lyn. It shows the presentation of a birthday present at a friend's party, and led Lyn to describe what was in the present bag shown. The event was significant in its own right and the chosen moment indicates the kind of relationship Lyn and her friends have with the birthday girl, Tracy. Listening to the conversation later, Lyn was also reminded of details of the event she had forgotten since the conversation:

I'd forgotten Tracy kept asking 'Do you want to know the time?' and kept carrying the bag, until I heard the conversation (Lyn).

Fig. 5.19 Lyn's favourite conversational photograph which reminded her of forgotten details

5.8.2   The Value of Storytelling Conversations

The criteria used for the selection of favourite conversational photographs begin to indicate some of the values of storytelling conversation of this kind. To a large extent it appears to be similar to the value of voiceover on photographs, as identified in the previous chapter. Hence, when people were judging the value of conversation on *other people's photographs* they tended to appreciate talk that told the story-of-the-photograph and added interest to it, or which personalised the photograph with the speakers' voice and opinion. However, there did not seem to be the same sense of intimacy in the conversational records as in the privately recorded voiceovers. Whereas voiceover appeared to convey the speaker's inner thoughts and feelings, and sometimes communicate hidden messages to the recipient, conversation did not. In fact most of the reasons for keeping conversation with photographs were pragmatic rather than sentimental. This brings us to one major difference between the value of voiceover and the value of conversation. Conversation was often selected on people's *own photographs* as a personal reminder of forgotten details of the memory.

The reminding value of one's own storytelling activities was actually the most commonly mentioned. An example of this has already been given in Fig. 5.19, which showed the handing over of a present at a party. Another example of the reminding value of conversation is shown in Fig. 5.20. This is distant photograph of two moose entering a lake in which the photographer, Meg, and her partner are

Fig. 5.20 A conversation photograph about canoeing with a moose

canoeing. Meg's excitement at this encounter is evident from the story she tells of it in the conversation, and is echoed in her comments about why she chose to keep it:

> *I chose to keep this mostly because the moment was pretty cool (seeing a moose) and there was a good story behind it – some of which I had since forgotten about* (Meg).

The other values of interest and personalisation can be illustrated with reference to some of the conversational photographs selected by the family group. Their photographs were unusual in that they contained not only recent photographs taken by the younger 'parents' of the family, but also older photographs of one of the parents, Debbie, when she was about 6 years old. These were brought in by her parents, who we refer to as the 'grandparents' of the family. This set of photographs is interesting because it shows how different members of a family have different views of the value of each others' photographs and conversation.

Regarding the recent photographs brought in by the parents, these were judged by the grandparents as much more attractive than their own photographs. Figure 5.21 shows the grandmother's favourite conversational photo of Debbie mentioned in the previous section. This contained a description of Debbie's 40th birthday trip to Paris which served to connect her mother to the occasion. It was also valued by Debbie's father, who was unfamiliar with Paris and was interested in the comments about it:

> *The Parisian scenes and context was outside my experience and the images were enhanced by the commentary and the added information* (Ron).

Fig. 5.21 This conversation and photograph of a daughter was found to be interesting by her parents

Regarding the older photographs brought in by the grandparents, these were judged differently by Debbie who formed the subject of the photos, than by Debbie's husband Chris. Chris was intrigued by the commentary provided by his in-laws about his wife when she was 6. For example one of his favourite conversational photos was that shown in Fig. 5.22 of Debbie, her brother and two school friends. Debbie appears second from the left. Chris' interest in these kinds of conversations is somewhat understated in the following quote:

It was useful to identify the background to the photograph, i.e. when, where, who, etc.(Chris)

Debbie's interest in her parents' photographs of herself was perhaps the most complex of all. She appreciated the new information provided by her parents in conversation and also liked the fact that the photographs were personalised by the voices of her parents and husband. In addition, the photos and conversation worked together to trigger personal memories she had not had for years. These effects can be appreciated from the audiophoto in Fig. 5.23, and her comments below:

The most valuable part for me was the sharing of historic photos from my parents' collection. I chose to keep the parts where my mum and dad were sharing pictures of me and answering my questions. It forms a good memory for the future. It is good to have a stored memory of mum and dad chatting and explaining their playful responses to Chris' comments....Memories of actually wearing the clothes came to me as I saw the photos (Debbie).

Fig. 5.22 This photograph of a wife with accompanying conversation by her parents was found to be interesting by her husband

Fig. 5.23 A four-way conversation valued for a variety of reasons by Debbie, pictured *left*

Some final comments from participants in the study suggested that there was also value in the original video record of the photo-sharing session. One participant preferred the entire video record to any selection of individual conversational photographs, and pointed out the different feel it had to conventional home video. Another liked the fact that the video happened to contain a split screen showing the face of the 'sender' or main speaker in the conversation.

5.9   Summary of Findings on Storytelling Photographs

By giving people the opportunity to store their storytelling conversations around photographs, we have found that *some* parts of the conversation were considered worthy of keeping because they enhanced the photographs. However, these parts did *not* always correspond to well-formed stories as predicted from the previous study. The selection of conversational segments was related more to the relationship participants had to each other and to the photos at hand.

A key relationship factor was whether the photographs belonged to you or others. This could be seen from Fig. 5.15, which showed that acquaintances were more interested in keeping conversation on their own photographs than on other people's photographs, whereas the reverse was true of family members. Friends were equally interested in keeping conversation on their own and their partner's photographs. Another way of looking at these findings is to say that the value of keeping conversation on someone else's photographs was related to the value of keeping the photographs themselves. Only very close friends or family were interested in doing this, perhaps because they had more shared history and context to maintain with each other.

The choice of which pieces of conversation to keep was dictated by the value they had for each group of people involved. Clear and well-formed stories were selected out of interest in other peoples' photographs. Other remarks and conversational banter were selected for the way each type of talk personalised one's own photographs and reminded the photographer of forgotten details. We also found that a mixture of these motives could be evident in individual selections, such as those made by Debbie in Fig. 5.23. Debbie liked the banter on her parents' pictures of her as a child, but also appreciated her parents stories on these pictures. In fact, Debbie was in the same position as the children in Edwards and Middleton's (1988) study of parent–child photo-sharing. Although she was a subject of the photographs being shared, she did not fully share the memory of the original event. This gave the reminiscing talk a storytelling quality – which was further extended by the presence of her husband as an independent audience.

Despite the provisional nature of these data, the evidence for the value of conversational photographs is clear enough to warrant further research. A future study might explore the use of such facilities in a trial setting, to examine again what conversation would be recorded but also to find out how such recordings would subsequently be used.

5.10   Discussion

We began this chapter with the observation that photographs live in an environment of conversation, and that some of this conversation could be captured in another kind of audiophoto. Two studies were then reported which looked at this environment and the prospects for recording it.

In the first study of photo-talk we found that all the features of conversational storytelling reported in the literature were present around photographs, but only in 40 % of photo-sharing conversations. These conversations were those in which photographers or the subjects of photographs shared the images with an audience who was not present at the time the photographs were taken. We referred to these as storytelling conversations in Fig. 1.2, although the stories they contained were subsequently found to be embedded in other kinds of talk involving questions, answers and general observations. A further 60 % of photo-sharing conversations took place exclusively between people who shared the memory of the events depicted in the photographs. We referred to these as reminiscing conversations in Fig. 1.2, since they involved remembering the event together. Somewhat surprisingly we found that reminiscing talk did not involve direct descriptions of the events or memories represented by the photographs, but rather contained a collection of remarks, questions and judgements about the photos which were highly specific to the participants own recollections. These conversations appeared to be unsuitable for saving, although further work might try to verify this conclusion. In contrast, storytelling conversations appeared to contain just the kind of details recorded in voiceovers, which had already proved to be useful for the interpretation of other people's photographs (in Chap. 4).

In the second study we examined whether storytelling conversation might have the same properties as voiceover, and expected participants to select well-formed stories as a kind of free voiceover. While conversational stories were found to be of greatest interest on other people's photos, other parts of the talk were valued on one's own photographs. Other people's voices and one's own conversation was found to be useful for personalising the photographs and reminding the photographers of forgotten details.

These two uses of conversational photographs are summed up neatly in the opening quote of the chapter, in relation to a wedding album. Conversation over the album might be used as a long-term reminder for the bride and groom, or as an interesting story for their children in the future.

5.10.1   The Role of Conversation in Personal Image Review

The value of recorded conversation on personal photographs for enhancing memory was a surprising outcome of the study. It flies in the face of the rejection of voiceover for this purpose in the audio-annotation study.

One factor may be related to the methods used in the studies. The time interval between recording and listening to personal voiceovers (several weeks in Chap. 4) was much shorter than that involved in recording and listening to personal conversations (several months in this chapter). It is likely that *any* spoken commentary on a photograph will become more useful as a reminder over time.

Another reason may relate to the expected practice of recording voiceover and conversation. Even if the value was perceived to be equivalent, participants may have rejected the idea of recording voiceover on their own photographs because of the effort and time involved. In contrast, participants may have recognised that recording photo-sharing conversation would take little extra effort. This may have led them to view it more positively as a form of indexing their own photographs.

However, a third explanation is particularly compelling. Participants in the audio-annotation trial were most reticent of recording voiceover when they could not imagine an appropriate audience. In fact the most difficult audience to record for was oneself, since participants already knew the story of their photographs and found it hard to anticipate what details they might forget. In conversation, the issue of identifying an audience disappears. The audience is physically present, and from the analysis of photo-talk we know that speakers act with great skill to tailor their conversation accordingly. We also know that audiences themselves act to bring out new details and explanations from speakers. This process results not only in free voiceover for speakers, but *freely-tailored* voiceover full of twists and surprises that speakers could not have anticipated themselves. In this sense the conversational recordings are also personalised, beyond the mere inclusion of other people's voices. They reflect accounts that emerge with the audience, containing facts and details that speakers could not have deliberately recorded for themselves. Such account may work as better reminders of the events than voiceovers, whilst also reminding speakers of the people they shared their photographs with in the past.

The value of recorded conversation on other people's photographs is more easily explained, especially in the light of the earlier findings on voiceover. Essentially, the stories in storytelling conversations work just like voiceovers for a remote audience. They add contextual details and interest that reveals more of the story-of-the-photograph. What is curious in the case of recorded conversation is that these details have already been heard live in conversation. The decision to record them must therefore be related to a desire not to forget them on subsequent reviews of the photographs. This presupposes that the audience expects to keep and review the photographs again and explains why their relationship to the photographer or subject is so important. If the relationship isn't close enough to involve keeping the other person's photographs, then there is no need to remember their interpretation. This account also explains why recorded conversation lacks the intimacy and impact of voiceover heard for the first time. It suggests that the value of recorded conversation for an audience is ultimately linked to a second-hand use in memory, rather than a first-hand use in interpretation.

5.10.2 The Role of Conversation in Social Interaction

The activity of using recorded conversation in live conversation with photographs was not examined in the studies. However, this is unlikely to be valuable, given the strong negative effect of playing voiceovers in live conversations reported in Chap. 4. The activity itself is somewhat bizarre and seems to stretch the reflexive nature of photography too far. While there may be some circumstances in which people want to listen back to themselves or others talking about photographs, I have assumed that they would do this alone and without further conversation. To want to review a conversational photograph with someone else is certainly possible, but would need to involve an interest in the conversation itself rather than the photograph to which it related. A feasible scenario for this might be listening to photo-sharing conversations your ancestors had with their friends and family when they were young. Future work might address such interactions, but I predict that participants will find it difficult to talk over the recorded conversations, and will not want to further record the conversation on a conversation.

References

Atkinson JM, Drew P (1979) Order in court: the organization of verbal interaction in judicial settings. Macmillan, London

Atkinson JM, Heritage J (eds) (1984) Structures of social action: studies in conversation analysis. Cambridge University Press, Cambridge

Balabanovic M, Chu LL, Wolff GJ (2000) Storytelling with digital photographs, Proceedings of CHI 2000. ACM Press, New York, pp 564–571

Berger J (1982) Appearances. In: Berger J, Mohr J (eds) Another way of telling. Butler & Tanner, London

Chalfen R (1987) Snapshot versions of life. Bowling Green State University Press, Bowling Green

Drew P (1985) Analyzing the use of language in courtroom interaction. In: Van Dijk TA (ed) Handbook of discourse analysis. Volume 3: discourse and dialogue. Academic Press, London, pp 133–147, Chapter 10

Drew P (1994) Conversation analysis entry. In: Asher RE et al (eds) Encyclopedia of language and linguistics. Pergamon Press, Oxford, pp 749–754

Duck S (1994) Steady as (S)he goes: relational maintenance as a shared meaning system. In: Canary DJ, Stafford L (eds) Communication and relational maintenance. Academic, London, pp 45–60, Chapter 3

Duck S, Pittman G (1994) Social and personal relationships. In: Knapp ML, Miller GR (eds) Handbook of interpersonal communication, second edition. Sage, London, pp 676–695, Chapter 17

Edwards D (1997) Discourse and cognition. Sage, London

Edwards D, Middleton D (1986) Joint remembering: constructing an account of shared experience through conversational discourse. Discourse Process 9:423–459

Edwards D, Middleton D (1988) Conversational remembering and family relationships: how children learn to remember. J Soc Pers Relat 5:3–25

Edwards D, Potter J, Middleton D (1992) Toward a discursive psychology of remembering. Psychologist 5(10):441–446

Frohlich DM, Chilton K, Drew P (1997) Remote homeplace communication: what is it like and how might we support it? In: Thimbleby H, O'Conaill B, Thomas PJ (eds) People and computers XII: proceedings of HCI'97. Springer, London, pp 133–153

Frohlich DM, Kuchinsky A, Pering C, Don A, Ariss S (2002) Requirements for photoware, Proceedings of CSCW'02. ACM Press, New York, pp 166–175

Goffman E (1981) Forms of talk. Basil Blackwell, Oxford

Grice HP (1975) Logic and conversation. In: Cole P, Morgan JL (eds) Syntax and semantics 3: speech acts. Academic, New York, pp 41–58

Norrick NR (1997) Twice-told tales: collaborative narration of familiar stories. Lang Soc 26:199–220

Polanyi L (1985) Conversational storytelling. In: Van Dijk TA (ed) Handbook of discourse analysis, Volume 3: discourse and dialogue. Academic Press, London, pp 183–201, Chapter 13

Pomerantz A (1978) Compliment responses. Notes on the cooperation of multiple constraints. In: Schenkein J (ed) Studies of the organization of conversational interaction. Academic, London

Sacks H (1968a) Second stories; 'Mm hm'; story prefaces; 'Local news'; tellability. Reprinted, In: Jefferson G (ed) Lectures on conversation volume II, part I. Blackwell, Oxford, pp 3–16

Sacks H (1968b) Second stories. Reprinted, In: Jefferson G (ed) Lectures on conversation volume I, part VII. Blackwell, Oxford, pp 764–772

Sacks H (1974) An analysis of the course of a joke's telling in conversation. In: Bauman R, Sherzer J (eds) Explorations in the ethnography of speaking. Cambridge University Press, Cambridge, pp 337–353

Sacks H, Schegloff EA, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. Language 50:696–735

Schaverien J (1991) The revealing image: analytical art psychotherapy in theory and practice. Jessica Kingsley, London

Sprecher S (1987) The effects of self-disclosure given and received on affection for an intimate partner and stability of the relationship. J Soc Pers Relat 4:115–127

Tversky B, Marsh EJ (2000) Biased retelling of events yield biased memories. Cogn Psychol 40:1–38

Chapter 6
Paper Versus Screen Playback

*You've got flexibility about who's got what and where (Paul,
audiocamera trial)*

Our journey so far has taken us through the properties of different kinds of
audiophoto content. PC-owning families have been encouraged to combine
audio of various kinds with photographs in order to reflect on its potential value
and use. This chapter changes direction somewhat, in order to connect these
perceptions of content to the practicalities and technicalities of its capture and
playback on different devices. In particular, it looks at the way various playback
options affect the experience of reviewing audiophotographs. The interrelation-
ship between content and technology is one of the lessons from the history of
new media development. All successful new media have involved innovation on
both fronts, with content responding to and favouring particular forms of tech-
nology. This relationship is examined here by drawing together reactions from
the same families to the playback of ambient, musical and talking photos on a
range of different devices.

As noted in Chap. 1, there are two main classes of devices for the playback of
audiophotos. Screen-based devices, such as the PC, TV or PDA, already support the
presentation of sounds and images and could be used quite easily to create and pres-
ent audiophoto albums. In addition, paper-based artefacts, such as the conventional
photographic print, album, frame or card, might be augmented with audio using
existing or new technology. These technologies support audio playback from paper,
or *audioprints*. Indeed, technologies for audio-enabled books have been around for
some time, and have led to a number of commercial audioprint products for talking
photo albums, frames and cards.

Both screen and paper forms of playback of the same material were presented to
families in the audio-camera and audio-annotation studies. Their reactions and
comments are reported in detail here. However, before launching into the methods
and findings of this comparison, I briefly review insights from related work to guide
our approach and analysis.

6.1 Related Work

Two sets of prior research are relevant to the question of paper versus screen playback of audiophotos. The first set of work stems from the debate on the future of reading, while the second relates to the importance of tangibility in design.

6.1.1 The Future of Reading

Insights on the use of information on paper and screen come from an extensive literature on the reading of office documents. Developments in document scanning and management technologies in the 1980s led to an active debate about the prospects for paperless offices. It was suggested in the business press of that time that these would be more efficient and environmentally friendly places, in which documents would be routinely authored or captured into the electronic realm, before being distributed, read and processed on-line. Although this prediction has not turned out to be true, it generated and good deal of research into the use of paper and screen-based documents and a clearer understanding of the value of each (Sellen and Harper 2002).

The locus of the debate has now moved to the future of the book, rather than the future of the office (Nunberg 1996). This is because recent developments in computing and screen technology now make it possible to design lightweight reading devices for 'electronic books' (*e-books*). These might be used to carry large amounts of up-to-date information in a form that can be annotated, linked and shared in new ways (Schilit et al. 1999). Because of the high cost of specialised e-book devices, they have been targeted at the professional reading market, and research has continued to focus primarily on office documents and information. However, some of the insights from all this research may carry over to the use of domestic documents and photographs.

For example, it is not just the optical characteristics of a display that affect reading on screen. Reading performance is also affected by text formatting, manipulation and control factors (Dillon 1992). Another insight is that it is important to distinguish between different kinds of paper and screen experiences when making comparisons between them. The history of reading shows that reading (and writing) behaviours have changed with each development in paper technology (Manguel 1996). Hence reading from clay tablets and papyrus scrolls was very different to reading from a handwritten parchment codex or a bound textbook. In the modern age, similar differences exist between the reading of cards, posters, documents, brochures, magazines, books and other forms of paper. Different kinds of screen experience result from reading the same content on screens of different sizes and resolutions; such as on a PC monitor, a TV and a handheld device. Furthermore, extensions and modifications to the *content* of screen-displayed information are also possible on screen-based devices. These include the addition of other media such as audio and video, and other structures such as hyperlinks – including links and communication back to the original author.

These possibilities are creating new forms of computer literacy which have only just begun to be recognised and studied (Bolter 1991; Snyder 1998). A final complication is that people have been found to read material differently, depending on what they are trying to do with it (Adler et al. 1998). For example, people will search specifically for matching information when trying to make comparisons within or between texts.

Leaving aside differences in the *content* and *purpose* of paper and digital documents, a number of studies have begun to reveal important differences in the properties of paper and screen-based media. Thus, in an experimental comparison of reading the same technical document in order to summarise it, O'Hara and Sellen (1997) identified a number of *affordances* of paper that made it easier to use for this task than a computer screen. They define affordances, after Gibson (1979), as physical properties of paper which encourage certain kinds of interactive behaviours. These included:

- Support for annotation while reading
- Ease of manipulation and navigation within a document
- Ability to spread out and compare information across pages

Subsequent work in more naturalistic settings has confirmed the importance of these affordances of paper, especially for supporting the interweaving of reading and writing activities (Sellen and Harper 2002). At the same time it has also shown the value of the following affordances of screen-based material for carrying, finding and exploring information:

- Allows access to a large amount of material
- Supports full-text searching
- Quick links to related materials
- Content can be dynamically modified or updated

The fact that paper and screen media have such different affordances for reading and writing behaviours leads people to move between them for different tasks. For example, the creation of documents tends to be done on-line, with paper in a supporting role. In contrast, the review and discussion of documents tends to be done on paper (Figs. 3.2 and 3.3 in Sellen and Harper 2002).

In reality then, people do not opt for paper *or* screen forms of reading as a general preference, but move between them for different purposes. In a final twist to this conclusion, Luff and Heath (1992, 1997) have shown that this movement can take place at both macro and micro levels. Paper or screen documents may be consulted in different phases of work (macro level) or in the same phase concurrently (micro level). This latter activity involves the co-reading of screen and paper surfaces, and is common not only to conventional offices but to other complex workplaces such as traffic control centres, medical centres and call centres (Luff et al. 2000). In such settings, the *micro-portability* of paper becomes important as a resource for calling attention to information, orienting it to colleagues and bringing it into a conversation. In this respect, paper makes the information itself more tangible.

6.1.2   Tangibility in Design

The notion of tangible interaction with information has been promoted for some time as a desirable property of future personal computing. When computing is 'ubiquitous' and resides in all manner of objects, furniture and architecture, people may be able to interact with it in more natural ways (Weiser 1991). This will not necessarily be through dedicated reading devices. For example, simply moving into a room might trigger the display of ambient sound, or moving an object on a desk might control the display of projected information (Ishii and Ullmer 1997). This approach to the design of future interfaces between people and computers attempts to augment the behaviour of ordinary objects and spaces, in contrast to simulating those objects and spaces within an enclosed computer world. For this reason it aims to create a kind of *augmented reality* rather than a virtual reality (Wellner et al. 1993). Considerable progress is now being made in this area, especially through the use of tags, sensors and wireless networks of information that can be triggered in ad hoc ways (e.g. Hawley et al. 1997; Want et al. 1999).

I mention these developments because they are part of a trend towards tangible computing. They show that the tangible nature of paper is now recognised as an advanced and desirable property of human computer interfaces in general. Such tangibility may give paper some functional advantages over screen-based devices, when it comes to reading, writing and sharing information.

Another reason for mentioning tangibility in design is that it may also confer some non-functional advantages over screen-based technology, especially in the photography domain. This is because people have a tendency to develop sentimental attachments to objects, both as personal possessions and as souvenirs. For example, in a survey of cherished possessions, it was found that the *utility* of the possession was a relatively minor reason for the object being special (Csikszentmihalyi and Rochberg-Halton 1981). More common reasons were that the objects triggered *memories* or *associations* with people or events, gave a certain *style* or *personality* to the home, and provided enjoyable *experiences* or *qualitie*s. Photographs in particular were valued for their memories, qualities and association with others. They turned out to be the most cherished objects cited by grandparents and the second most cherished objects cited by women of all ages. The association of objects and photographs with people was particularly strong. This has been noted in other contexts and cultures where objects become almost magically endowed with the personal characteristics of their owners (Hoskins 1998). In some cases these objects are considered to have power in their own right, while in others they are seen as merely signs or remembrances; as in the catholic and protestant views of the bread and wine used in Christian worship (Schaverien 1991). Schaverien goes on to show that images as well as objects can take on these kinds of meanings, even to the extent of serving as tangible talismans or charms.

Given the importance of tangible objects as both functional and non-functional things, it may not be so easy or desirable to replace them with digital copies. This can be seen from a recent study of digital music, which found that although users enjoyed the greater choice of music available in MP3 format, they still wanted to

purchase and handle it through a form of physical music packaging (Brown et al. 2001). In another context, it has been suggested that physical souvenirs might be brought into the digital world by taking a photograph of them and annotating the photographs with stories (Stevens et al. 2003). However, a photograph of a souvenir is unlikely to be as powerful as the object itself in evoking a connection to the past. A more promising approach might be to augment the souvenirs themselves with stories, so that the objects can continue to be handled in the usual way (Frohlich and Murphy 2000). Essentially the same argument has been applied to audiophotos (Frohlich et al. 2000) and is developed below.

6.1.3   Application to Audiophotography

In short, although the idea of a paperless home has never received the same attention as a paperless office, we can expect many of the same insights to apply. Rather than domestic paper documents and photographs being replaced by screen-based technology, it is more likely that they will continue to co-exist with it, perhaps in a modified form. This is because the affordances of paper and screens are different, and will continue to lead people to move between them for different purposes. The key question for digital photography is therefore not 'Which is best, a printed or screen-displayed photograph?' but rather: *'In what contexts will people prefer certain kinds of paper or screen-based experiences with photographs?'*

This question is also a key research issue for audiophotography. Because of the existence of talking book technology for augmenting physical paper with sound, it is possible to provide a paper alternative to the screen-based audiophoto. The audioprint alternative not only removes the need to use screen-based technology for this particular multimedia format, it also provides continuity with existing tangible methods of handling photographs. As we have seen from previous work on tangibility, this may be an advantage both functionally and non-functionally. Functionally, it may easier or more flexible to handle, retrieve and share printed photographs than screen-based ones. Paper may also be a more accessible and reliable medium for archiving photographic images long term. Non-functionally, prints may be better for forming psychological attachments to photographs. Conversely it may be more difficult to form attachments to ephemeral screen images, or to believe that they will not vanish forever. Viewed from this perspective, we might ask an additional question to the one mentioned above, concerning movement from the status quo: *'What advantages must screen-based photographs offer for people to give up the tangibility and familiarity of prints?'*

6.1.4   Approach

Our approach to these questions is to layout a range of different paper and screen options for playback of the very same audiophoto content. These options were then experienced and discussed by participants in our two main studies of ambient sound,

music and talking photographs. Analysis then focuses on the contexts in which people report using or preferring certain playback options, and especially on the conditions for shifting to screen-based forms of playback. In exploring context, we refer again to the diamond framework in Fig. 1.2. Because the playback experience in these studies was limited by the early technology for audioprints, we rely largely on the participants anticipated uses of audiophotos on paper and screen. This is a shortcoming of the focus group methodology itself. These studies should therefore be seen as precursors to more extensive field trials with fully integrated versions of each playback technology shown.

In the next two sections, we step through the findings of each study separately, before bringing them together in a final discussion

6.2   Playback of Ambient Photographs

Photographs with ambient sound attachments were reviewed on paper and screen in the context of the audio-camera trial described in Chap. 2. The overall methods used in the trial are described in that chapter. In this section we describe the review procedures in further detail.

The first stage of audiophoto review took place in home visits after four families had captured ambient sound photographs using an audio-camera unit on their summer holidays. Within these visits families jointly reviewed the total set of photos and sound clips captured. This was done first without sound from the prints alone, and then with sound played continuously from an *audio-cassette* loaded into each families' home hi-fi unit. This can be seen as a form of audioprint playback from paper, in which the prints and sounds are unsynchronised. After this experience the families selected about 10 favourite audiophotos for incorporation into a digital album that we later assembled for them. The original prints and cassette tapes were left with the families for them to show to others if they wanted to.

About 1 month later, families were invited in pairs to attend two separate 'Feedback Groups' at HP Labs Bristol. Each group session started with a discussion of their overall reaction to audiophotos and their subsequent use of the prints and cassette tape. Their favourite audiophoto albums were then made available for them to show to each other in two forms:

- **PC album**: Participants could click through a screen-based album on a multimedia PC, using a mouse to select on-screen navigation buttons. Albums were authored in a multimedia editing package called *StudioM* from GoldWorks Inc. Sounds played only when the photographs were themselves selected by clicking on them. Simple audiophotos with a one-to-one association of sound and image were shown one screen at a time. More complex audiophotos with a single sound clip for several images were also shown on the same screen, but with individual images revealing themselves at the right time in the sound clip. It was suggested that material from the camera could be uploaded to the PC for this experience.

- **CD album**: Participants could randomly access sound tracks from a hi-fi system using track numbers written on the back of each printed photograph. The hi-fi unit was placed on a table next to a pack of loose prints. It was suggested that track-numbered prints might be provided with a corresponding audio CD from a photofinishing shop, or that the prints and CD could be made with a home PC and printer.

After each family had reviewed each others' albums in these formats, two further options were presented for comparison:

- **TV album**: The PC album experience could be replicated on an interactive television. It was suggested that the camera might be plugged into the television and used to control the on-screen album.
- **Audioprint scanner**: The printed photograph could have sound data on the back, in a form that could be scanned and played back from a handheld device. This capability was simulated using the prototype shown in Fig. 6.1. A handheld box containing a speaker and a slot was plugged into a minidisc player. Four audioprint samples, one from each family, were prepared with corresponding audio tracks stored on minidisc. Automatic playback of sounds from the prints was demonstrated by inserting a print into the slot in the box, pressing 'Play' on the minidisc unit, and passing the print around the group. It was suggested that such prints could be made on a conventional printer.
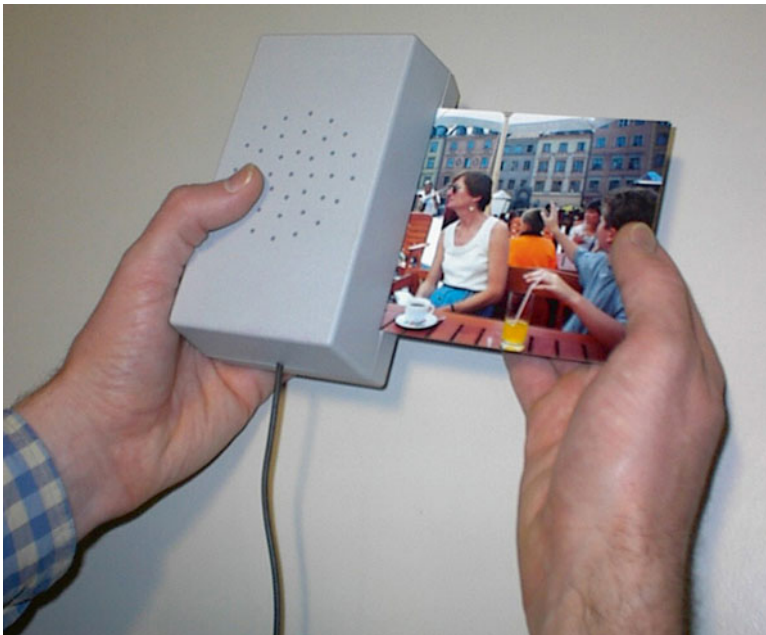
Fig. 6.1 The prototype audioprint scanner (Fig. 1 in Frohlich et al. 2000)

Individual participants were then given forms containing all the above options, including the original audio cassette method experienced at home. They were then asked to place them in order of preference. The results of individual rankings of these options were discussed in relation to the handling of entire sets of audiophotos. Two further options for sending or exhibiting individual audiophotos were presented for final discussion:

- **Audiophoto cards**: A self-playing audiophoto could be created as a gift by customising a recordable greeting card. Three commercial cards of this kind were passed around the group as examples of this. They contained photographs from different families in the trial with sounds that could be played by pressing a corner of the card. It was suggested that these could be made up manually by sticking adhesive prints to a card base and recording the sound through a line in socket or an integral microphone. Alternatively they could be requested as a digital photofinishing service.
- **Audiophoto frame**: A self-playing audiophoto could be displayed in an audio-enabled photo frame. In the absence of any commercial frames of this kind at the time of the study, a prototype was made as shown in Fig. 6.2. This comprised a standard $6'' \times 4''$ picture frame with an analogue sound recording and playback unit stuck to the back. The sound could be started or stopped by pressing a button on the back of the sound unit. It was suggested that this could be made up manually by inserting a printed photo into the frame and recording a corresponding sound through a line in socket or integral microphone.

Fig. 6.2 The prototype audiophoto card and frame

6.3   Results for Playback of Ambient Photographs

6.3.1   Overall Preferences

The results of the playback ranking exercise are shown in Fig. 6.3. The average rank score is shown for each playback type. So higher preferences are shown by lower scores. Essentially, people preferred the two methods of review with the strongest association of sound and image; the PC album and the audioprint scanner. In both these cases the sound and image data are stored and invoked together, rather than kept separately and indexed to each other, as with the audio-cassette or the CD album. These last two options were felt to be too difficult or restricting to handle in the context of discussing and sharing photos face-to-face. The following comments were made about the audio-cassette method, but apply equally to the CD method. This is because in practice the CD was allowed to play through like an audio-cassette tape, without manual intervention from the users. It appears that in the heat of a photo-sharing conversation it is not possible to control images and sounds separately:

> *"It was too fiddly" (Mervyn)*
> *"There's no time to linger over the photos when you are listening to the sounds at the tape pace" (Philip)*
> *"I wanted to control the pace (of the tape) and hold the photo myself" (Sue)*

In the course of discussing the status of these preferences, it became clear that each individual was not voting for a single method of playback in order of preference. In conversation afterwards, most people expressed a desire to use more than one method of playback depending on the context. Significantly for our analysis this often included a screen *and* paper method. This can be seen from Fig. 6.4 which
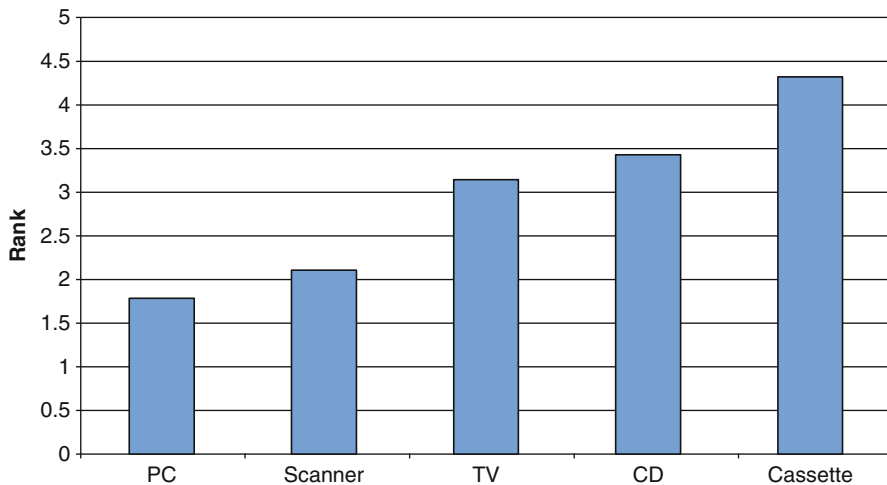
Fig. 6.3 Average ranks for each of five audiophoto review options
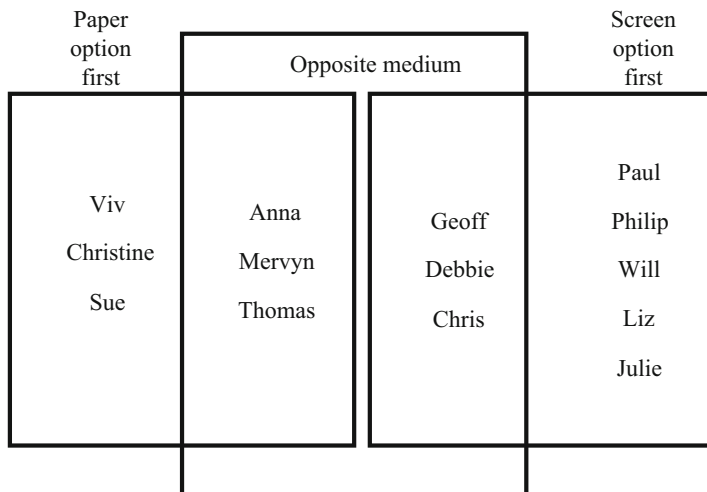
Fig. 6.4 Overlapping preferences for paper and screen playback options

shows the overlap of first and second choices made in the study. The participants were split broadly on whether they voted for a paper-based option or a screen-based option as their first choice. The undisputed paper option was the audioprint scanner while there was heated discussion between the advantages and disadvantages of the PC versus the TV (see below). However, when we look at the *second* choices of each group we find that a subset select the alternative medium. Furthermore, paper is chosen by these people as a *complement* to the screen option or vice versa.

6.3.2   Reactions to Devices

The reasons for the above preferences become clearer when we consider participants' reactions to individual device options. Since the audio-cassette and CD options were quickly rejected by participants, discussion centred around the other main options of the PC, TV and audioprint scanner. We now step through discussion of these devices in turn. Reactions to the card and frame are also covered at the end of this section.

PC Album

The PC album was liked and disliked for a number of reasons. The teenagers in the trial mentioned liking the PC itself, and associated it with playing games in an upstairs study or bedroom. The idea of presenting photographs and sounds on the

PC therefore fitted into this recreational pattern of use. The PC also turned out to be in an appropriate location for sharing audiophotos with their friends, who they tended to entertain upstairs.

"I like computers. I can switch off and go and play games" (Julie)
    "All you have to do is turn on the computer" (Thomas)
    "It's so easy to use, there's an option to have sound or you can just flick through" (Liz)
    "I don't take my friends to the lounge because that's adult. With a computer in my bedroom that would be a social place" (Anna).

The adults were more ambivalent. Their opinions split roughly along gender lines. The fathers in the study recognised the power of the PC for editing audiophoto material and creating albums. They found the demonstration album interface easy to use and enjoyed the professional look of various features such as the automatic transitions associated with multi-photo pages.

"It looked impressive and you can trawl through them easily" (Chris)
    "I'm starting with the editing and I'm thinking you're going to have to sit at a PC and do that anyway so why not have it based around a PC and have it portable to other media?" (Paul)

In contrast, the mothers in the study disliked the PC itself, and considered it too 'technological' a device for presenting photographs. They also argued that it was in the wrong place for sharing audiophotos with family and friends, since visitors tended to be entertained downstairs in the living room.

"I just think the more technological it gets, the more it puts me off. I want something simple, I don't want it complicated" (Sue)
    "The PC is not for us. Its in a small room in our loft conversion" (Debbie)
    "I don't want to get away from prints" (Viv)

One of the objections to the PC album was that it was not portable. This was seen as a disadvantage because it restricted audiophoto access to a particular room of the house and also prevented the sharing of audiophotos with others outside the home. Sharing with grandparents who did not have a PC was thought to be a particular problem in this respect. A number of suggestions were made by teenagers and fathers to overcome these objections. These included the use of a portable (laptop) computer for sharing audiophotos, storage onto floppy disks or CD-ROMs, and printing the photographs without the sounds. In all these cases, the solutions were seen as supplementary to the presentation of audiophotos on a desktop PC, which was still seen by some participants as the best vehicle for showing material to a group of people. Again these findings imply a movement between different devices for different contexts of use.

TV Album

Many of the advantages of the PC were seen to apply also to the TV. Hence the large screen format for photographs and the high quality sound, were appreciated by those who valued such attributes on the PC. Additionally, the TV was sometimes

seen as more suitable than the PC as a device for sharing photographs, because it encouraged a more relaxed and familiar style of interaction in a better location. Since the TV was considered equally at home in a bedroom or living area, this option appealed to both teens and adults who imagined using one in the most suitable location for them. The compatibility of storage format was also seen as an advantage. Because the study pre-dated the launch of DVD players in the UK, participants assumed that the audiophoto albums would be stored on video-cassette which was seen as a pervasive medium. Interestingly, these participants expected the same level of control and interactivity with the video-album as with the PC-based album, and did **not** want to view the album as a static slideshow.

> *"You could have it just like a PC. You could just flick through and have a special thing to put it through. I just thought it would be much better to have the video because more people have got a TV and video than have a PC – so that means you could show it to your friends" (Will)*
>
> *"You would sit on comfy chairs, and if you were showing special occasion photos to friends you'd present them on TV in the lounge" (Debbie)*
>
> *"TV is more sociable and people are used to it" (Viv)*

The similarity of the TV and PC albums led some participants to propose an ideal solution involving both. This was either seen as a single integrated device or two separate devices with connectivity between them. The PC could then be used for editing and creating albums that could be moved to the TV for sharing.

> *"I think ultimately if PCs are in every home and they are going to be Entertainment Centres in people's lounges, then no problem" (Chris)*
>
> *"Well if you have one of these predicted combined PC/home TV things then it would be an ideal combination wouldn't it?" (Geoff)*
>
> *"Well if you had an input that was for both PC and TV then you could do anything you like on the PC, if you've got one, and send it down the line on the internet or fax or whatever, or print and edit, and use the television in a social situation and send it to someone else with a TV. So you've got flexibility around who's got what and where" (Paul)*

Audioprint Scanner

The ability to print audiophotos and play them back in your hand was welcomed by all the women in the trial. They saw instantly that this would allow them to continue using printed photographs as usual but with the added bonus of associated sounds. They praised the 'low tech' feel of the device, the fact that there was little or no set-up required, and the benefit of preserving tangible prints. One of those benefits was the longevity of paper as a storage medium. They simply didn't trust digital forms of storage. Another benefit was portability. Printed photographs can be shown anywhere, and mothers wanted the same flexibility with audiophotographs.

> *"Its small and mobile and the photos and sounds are together" (Christine)*
>
> *"And you've still got prints" (Viv)*
>
> *"I just want to show them casually where I am at the time… With this you could sit out in the garden" (Sue)*

"There is something nice about having that tangible print in your hand. In fact I'd feel very insecure if the images were just stored on the PC" (Debbie)

The use of physical prints also appeared to imply a more intimate form of sharing than displaying photographs on a large screen. For this reason it was seen as particularly suitable for sharing audiophotos within the family, in reminiscing conversations.

I imagine we would use it more than when we are showing photographs to other people (Christine)

A subset of men and teenagers also saw the audioprint scanner as a useful complement to the PC or TV (see again Fig. 6.4). They also mentioned portability as an advantage over a fixed PC or TV, and even suggested that the scanner was more appropriate than a laptop, which seemed too 'high tech' for sharing audiophotos. This sentiment also appeared to extend to the micro-portability of handling individual photographs. This was reflected in a number of comments about the accessibility of prints.

"I don't want to show £5 worth of photos on a £2000 laptop" (Chris)
"It's portable. You can bring it to other places" (Mervyn)
"I flip through photos 'cos they've got that feel to them. If you add equipment it feels like you are imprisoning the thing" (Geoff)
"You can go to the back one (with the scanner) without pressing the button all the time" (Thomas)

A number of concerns about the audioprint scanner emerged in this discussion. These included the ramifications of sending audioprints to others, editing audiophoto material, putting audioprints in albums, and getting better sound quality from a smaller device. Mothers in particular asked about sending material to others and realised that recipients would also need a scanner to playback the sounds. Although the cost of these devices was said to be low, this was still seen as a drawback to the solution. Fathers were concerned about editing material on the camera alone and struggled to understand if this could be done on the scanner. In fact, with this technology it could not. Some participants noticed that the scanner could not be used on photographs in albums where the printed sound data would be inaccessible. Several also objected to the size of the scanner itself and wondered if the external sound quality could be maintained in something more like an MP3 player. All these concerns lowered the perceived value of the audioprint scanner somewhat, especially in the eyes of fathers and teenagers. This led them to ask for printing from the PC, where editing, sending and high quality playback could be achieved more easily.

Audiophoto Cards

The additional options of sending or exhibiting individual audiophotos in self-playing cards or frames were shown at the end of each session. The technology involved in both solutions was similar, comprising the storage of up to 20 s of sound on a voice chip attached to a miniature microphone and speaker. However, the

packaging of this technology within a cheap cardboard greeting card designed for sending, or a solid wooden photo frame with a glass cover, led to quite different reactions from participants. In general, they loved the cards but hated the frame.

Initial reaction to the audiophoto cards was overwhelmingly positive, as people imagined themselves receiving cards from friends and family. They were seen as novel, fun and easy to use. Furthermore, participants recognised that there was no need for the receivers of cards to have additional playback technology, such as a special scanner. The fact that they were self-contained and self-playing was seen as their biggest strength. This train of thought led participants to consider how they would use the cards as senders, and generated a range of creative ideas. Special occasions such as graduations, the birth of a baby or a birthday were mentioned in this context, as well as postcards from holiday. One father in the trial noted that new sounds could be re-recorded over olds ones, so that you could re-use a birthday card. Another father thought of recording holiday postcards from home before departure.

> *"That's brilliant... You could send a photo of the kids to grandma with them singing 'Happy birthday'" (Debbie)*
> *"They'd be great for a special occasion" (Mervyn)*
> *"You could say'Happy birthday' and send something else from your holiday" (Thomas)*
> *"You could get your mother to send her birthday card back!" (Geoff)*
> *"It would be a lot easier than writing (postcards). You could record it before you go!" (Chris).*

To make the cards more suitable for these uses, participants suggested several improvements. They wanted the cards to be slimmer and in postcard form. They asked for a longer duration of sound storage so that they could send a clip of ambient sound with a voice message appended. Others suggested the possibility of adding multiple photographs. A final request was the ability to generate the cards from a PC, since the process of attaching digital photographs and sound clips from a camera was unclear. In short, the cards received considerable attention in the discussion because they were seen to have so much potential for sharing audiophotos with remote family and friends. In contrast, discussion of the audiophoto frame was short and not so sweet.

Audiophoto Frame

Contrary to our expectations, there was an almost universal negative reaction to the audiophoto frame demonstration. In fact the most positive comments related to using it as a more long-lasting card that the recipients might want to keep. For example, one family thought of giving a framed audiophoto of their friend's baby's dedication to them as a present. The idea of keeping and displaying personal audiophotos in this form just didn't work for most of the participants. Whereas people were comfortable with the notion of glancing repeatedly at the same displayed photograph, they reacted against listening repeatedly to the same sound clip.

"Oh my goodness, That is over the top" (Christine)
    "It would sell very well in Bavaria or Russia. They go in for musical boxes there" (Geoff)
    "I wouldn't want to hear it over and over again" (Sue)

Although the audiophoto frame was not liked as a vehicle for displaying personal audiophotos, several people thought of commercial applications such as talking books, travel guides and holiday souvenirs. The most creative application was a kind of autographed audiophoto, explained enthusiastically as follows:

Imagine if you were in a fan club and it was (a photo of) your favourite pop star... David Cassidy in my day... And it had a personal message. ['Hello Debbie I'll always love you'(Chris)]. You could sell buckets. Fantastic! (Debbie)

6.3.3   Summary

In short, we have found that it was rare for individuals to express a preference for a single paper or screen-based method of audiophoto playback. Most participants recognised that one solution did not fit all circumstances of use, and expressed a desire to move between solutions depending on circumstance. A common combination, at least for teenagers and men in the trial, was to use a PC to create audiophoto albums which could then be shown on the TV or printed out as audioprints. Women were more inclined to favour audioprints alone, but also expressed interest in sharing audiophotos on the TV.

Within each paper or screen-based playback category, participants were very sensitive to the form and type of option available. Not all options were considered equal, even when the very same core technology was involved. For example, audiophoto cards and frames elicited opposite reactions as a result of the very different packaging involved. The card packaging implied a model of purchasing and creation designed for sending audiophotos to others. This behaviour was valued by participants and judged to be well-supported by the technology. In contrast, the frame packaging suggested the long-term display of audiophoto material for oneself. This was not valued so highly, and led to rejection of the technology. A similar thing happened with the perception of audioprints whose sounds could be played back from an audio CD or a handheld scanner. Although printed photographs were used in both cases to trigger the sounds, the *method* of triggering was seen as simply more convenient with the handheld scanner and allowed greater freedom of movement. The scanner also simplified the overall process of producing audioprints, since it cut out the necessity for a separate storage device (i.e. the CD itself).

The perception of screen-based options was also sensitive to the option type. Careful attention was paid to the different behavioural ramifications of using the PC or the TV as an audiophoto player. Participants noted that they were located in different rooms in the house and that this would affect how easily they could show audiophotos to different groups of people. The distance from the TV and PC screen

was thought to affect the tone of the playback experience, and the greater flexibility of the PC was felt to be more suitable for editing audiophoto material. Finally, the different storage formats of PCs and TVs and their ownership by others became an issue, as participants thought of taking audiophotos out of the house or sending them to others. All these findings show that participants judged each playback method on a very wide range of criteria. These included the set-up costs of using the equipment, constraints on the contexts of use, and compatibility with existing technology and behaviour – as well as the playback experience involved.

6.4   Playback of Musical and Talking Photographs

A similar approach to that used in the audio-camera trial was used again to assess reaction to paper and screen–based methods of musical and talking photo playback. Participants in the audio-annotation trial were shown the same demonstration material on a variety of paper and screen devices. All of these demonstrations were shown in the Feedback group sessions, after participants had reviewed and discussed their own musical and talking photo albums on a desktop PC (see again Chaps. 3 and 4).

As a result of the playback findings in the audio-camera trial, reported above, we modified the mix of paper and screen-based options used in the audio-annotation trial. The importance of portable screen-based options was acknowledged by the addition of a Tablet and a Palmtop device to the PC and TV. The success of self-playing audioprints and the involvement of album enthusiasts in the trial also led us to include an Audiophoto album option, alongside a handheld Audioprint player and Audiophoto card. Because of the lack of interest in the Audiophoto frame, and the already large number of playback options, we decided to drop the frame from the list of paper options. This left us with four screen-based and three paper-based demonstrations as described below.

The screen-based options for audiophoto playback were shown as follows. In each case it was suggested that participants could associate music or record voiceover on the device itself, or via an editing process on a desktop PC.

- **PC album**: Participants viewed each other's musical and talking photo albums on a flat panel display with stereo speakers, placed on a circular café table. The interaction was controlled by using a mouse to select forward and back buttons which turned over virtual pages of the album (see Fig. 3.1). Sounds played automatically on presentation of each page.
- **TV album**: Demonstration material was presented on a 21 inch TV screen with stereo speakers. The TV album was authored using HP Memories Disk software on a PC to burn a CD-ROM in video CD format. The CD was then played back from a DVD player connected to the TV, and controlled from a DVD remote control device. The album was played through in slideshow mode, but it was suggested that pages might also be advanced manually, as with the PC album.

- **Tablet album**: A demonstration PC album, also created in Macromedia Director, was shown on a 12" Fujitsu 512 Tablet PC. Photographs were shown nearly full-screen at an absolute size of about five inches by seven inches. Participants controlled the turning of screen pages by tapping on screen-displayed buttons with a pen stylus. Again, sound played automatically on presentation of each page.
- **Palmtop album**: A lightweight version of the Tablet album with more compressed image and sound files was created using the HP Pocket PC camera software running on an HP Jornada PDA. This software already supports audiophoto capture and playback; automatically associating sound and image files with the same filenames. Photographs were shown full-screen at a size of about two inches by three inches. They could be controlled using forward and backward movements of a rocker switch on the casing, with sounds playing automatically for each new page.

The paper-based options for audiophoto playback were shown in the following way:

- **Audioprint album**: An audio-enabled photograph album from Brookstone Gifts was filled with demonstration content from the trial. This product supported the association and playback of a 20 s sound clip for each double page spread, from numbered buttons beneath the pages (see Fig. 6.5). The demonstration album was created by placing individual six inch by four inch photographic prints on self-adhesive pages and recording the first 10 s of their associated music or voiceover tracks in memory. Participants could then browse the album, pressing the appropriate buttons for each page at will. It was suggested that participants could create their own albums manually.
- **Audioprint player**: A new improved audioprint player prototype was built in response to feedback on the previous prototype and the results of a separate technical investigation. The second version was based on the use of solid-state storage in the paper itself to hold the audio data associated with the photograph. Hence a voice chip from ISD measuring 5 mm square and just 1 mm thick was mounted on a substrate and embedded in each print sample. Each chip stored up to 30 s of high quality sound. The sound could be played back in real-time via contact with the handheld player shown in Fig. 6.6. Three audioprint samples were created for demonstration in the Feedback session. It was suggested that individual audiophotos could be printed on an adapted HP printer loaded with 'audiopaper'.
- **Audiophoto card**: A two-page opening card similar to that used in the previous trial was demonstrated in the Feedback groups. It was suggested that participants could create their own cards manually.

After seeing these demonstrations, participants were asked to indicate on a form, which device they would like to use to playback musical or talking photographs. They were allowed to tick more than one option for each type of audiophoto, in recognition of the fact that multiple options were requested by individuals in the previous study. Forms were then collected as a basis for discussing selections. The resulting votes and comments will now be described.

Fig. 6.5 The audioprint album
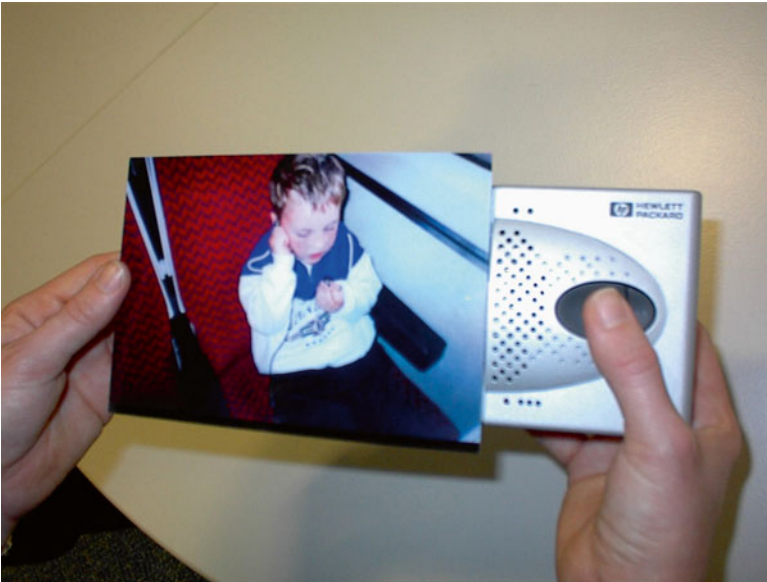
Fig. 6.6 The second prototype audioprint player (Fig. 3 in Frohlich et al. 2000)

6.5   Results for Playback of Musical and Talking Photographs

6.5.1   Overall Preferences

The preferences for different kinds of playback methods are shown in Fig. 6.7. This time, preferences are expressed in terms of the number of votes for each method. Each of the 11 participants was allowed to vote for as many methods as they thought they would use. Votes were cast separately for the playback of music and talking photographs.

   This graph is a dramatic illustration of the fact that no single playback method was preferred over others. The PC and TV options came out top overall, but a large number of people also indicated wanting to use the palmtop, album, card and print methods. The least favoured method was the tablet, although even here, four out of 11 individuals said they would be interested in using it. A further finding, not shown in the table, was that each participant selected an average of 2 screen methods and 1 paper method of playback. Again there was no single dominant playback method, even within the preferences of individual participants.

   One pattern that is evident in Fig. 6.7 is the divergence of playback preferences by content type. Some playback methods received more votes for musical rather than talking photographs or vice versa. The PC and TV were seen as more suitable for playback of musical photographs, while the handheld devices and paper methods were seen as more suitable for talking photographs. The reasons for this pattern and the selection of multiple methods of playback, can be seen from the discussion of preferences that followed the voting activity. It is to this discussion that we now turn.

Fig. 6.7 Preferences for the playback of musical and talking photographs

6.5.2   Reactions to Devices

As before, there was more discussion of favoured than non-favoured options for playback, with comments focusing more on the positive than the negative aspects of each device. However, no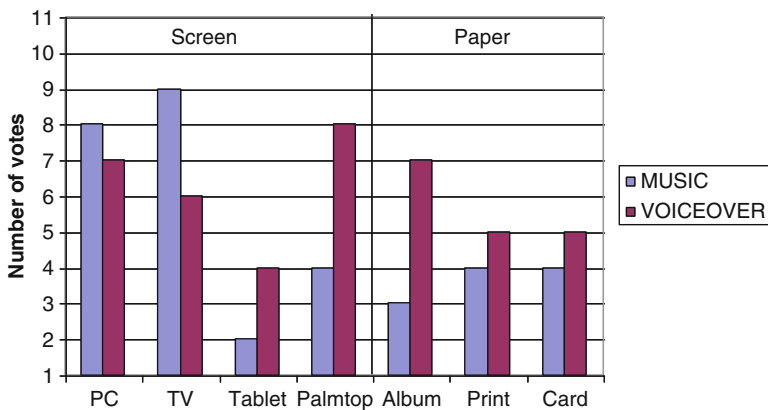 device in this study was ruled out as a potential method of playback, unlike the audio-cassette and CD album in the previous study. There was also more openness to playback of audiophotos on the PC and other screen-based options, particularly by women in the sample. This may have been because the study was carried out in 2003; several years later than the audio-camera trial. Participating families were therefore more familiar with computers and the sharing of photographs via the PC and internet, even if they did not necessary own a digital camera themselves. Hence we begin with comments on the PC as an audiophoto viewer, and go on to consider all the other screen and paper-based methods of playback in turn.

PC Album

The PC was seen as a good way of storing, editing and sharing audiophotos privately within the family, and as a method of sending audiophotos to other people. It was seen as less suitable than the TV for showing audiophotos to extended family and friends. This was partly because the interaction was said to become cramped with more than about 2 people, and because the location of the TV was more appropriate to entertaining guests (see below). PC and TV contrasts of this kind were common in the discussions and the idea of networking one to the other emerged again as a way of getting the 'best of both worlds'. One elderly member of the group pointed out that the PC screen was difficult to read from compared to the TV, but on the whole the consensus was that both devices delivered a high quality audio-visual experience.

> *"I would get a reasonable sized display with the quality and resolution but I'm very biased I think" (Harry)*
> *"Well my PC screen is bigger than your TV screen so its fine for showing photos" (Vicky)*

It was the storage, editing and communication capabilities of the PC that really differentiated it from the TV. People could imagine editing audiophoto material much more easily on the PC, and sending audiophotos to others via email or the web. They also saw the PC as a possible place to archive photographic material. These kinds of uses were cited as often as sharing audiophotos on voting forms, which also asked participants to write down when they would use each playback method selected. Some typical written comments were as follows:

> *"As a record or to send to family or friends" (Traci)*
> *"Sending or placing on the internet" (Rob)*
> *"Editing photos and manipulation" (Gordon)*
> *"Personal use only" (Rachel)*

TV Album

The TV was seen primarily as a way of showing audiophotos to groups of family and friends. This was cited as the most common context of use on the voting forms. The viewing distance from the TV, seating comfort, and the public location of the TV were all said to contribute to its suitability for group sharing. Some typical comments from the discussions were as follows.

> *"Its more relaxed I think sitting around like this than sat on a computer chair…I just find it more relaxing to sit and watch a TV than a PC" (Rachel)*
>
> *"For me the TV is something where a group of people can share it and there's no problem. With the PC once you get more than two people it tends to get a bit crowded" (Jane)*

Other reasons for using a TV album included greater familiarity with watching TV rather than using a PC, especially by visiting parents or grandparents. This reason carried over into a debate about how to send audiophotos to distant relations who did not have a PC. Although the possibility of sending video-cassette tapes was mentioned here, participants were more inclined to send augmented prints of one kind or another (see paper options below).

Tablet Album

Only a small number of participants voted for the tablet as a playback device, and most comments made in discussion were negative. For example, it was considered too heavy and bulky to hold or carry for long, and too difficult to share with others because of its restricted angle of viewing. Essentially, only about two people could comfortably look at its surface at any one time. On top of this the sound quality was thought to be poor compared with the PC and TV.

> *"You've got an angle problem" (Harry)*
>
> *"You've got to get it at the right angle as well otherwise you can't see anything. Sounds awful as well" (Rachel)*
>
> *"Its unrecognisable" (Gordon)*

In spite of these sentiments, several people mentioned its potential for showing audiophotos in a business context. Jane worked in an Art and Design Department of a university and thought that the tablet would be useful for carrying around a multimedia portfolio. Bryon suggested using it as a selling tool, and began to imagine new kinds of audiophoto adverts that might be created for this purpose. These comments show that certain devices are associated with certain contexts of use, and that both these factors affect the kind of content that might be created and shown in them. Unfortunately here, the work-related contexts of use associated with the tablet device appear to be similar to those of a laptop and worked against its use with domestic audiophotographs. A similar effect was encountered in the audio-camera trial when discussing the merits of sharing audiophotos on a laptop computer. Like the laptop, a tablet was felt to be a form of technology overkill for the casual showing of photographs.

Palmtop Album

In contrast to the tablet, the palmtop received a large number of votes and much more positive comments. Its main advantage appeared to be its extreme portability, which made it suitable for showing audiophotos out of the home. This was cited as the main context for use on the voting forms in connection with visiting friends, going on trips and carrying it around to use in unplanned conversations. These factors were reiterated in later discussions. Somewhat surprisingly, the strongest advocate of the palmtop was Betty, a grandmother in the trial. She preferred it to more familiar paper-based solutions because it held a great number of photographs but fitted in her handbag.

> *"I like the idea of that, that you can take it around anywhere if you were visiting somebody you could take it there to show them" (Betty)*
> *"Well that is one for the grandmothers to put all the photos of the grandchildren in, similar to the things they carry around now to the WI (Womens' Institute) or wherever they gather" (Jane).*

Although the viewing angle and sound quality problems were mentioned again for the palmtop album, participants seemed more willing to tolerate them for the convenience of showing audiophotos wherever they happened to be. Two concessions they did make to these problems were to mention one-to-one sharing with individuals rather than groups of people, and to favour the showing of talking rather than musical photographs. In fact twice as many people voted for the palmtop album as a method of playing back talking photographs compared to musical ones (see again Fig. 6.7).

Audioprint Album

In contrast to the palmtop, the self-playing audioprint album was seen as too large to carry around. It was therefore considered to be a piece of technology for archiving and sharing audiophotos in the home. Its two perceived uses both related to 'special' photographs. Participants saw themselves using it as a showcase for archiving and sharing memories of an important event or trip. Alternatively they considered it suitable for creating as a gift to send to family and friends abroad. Given these occasional uses, participants required much higher quality sound capture and reproduction to match the high quality of printed images. This was revealed in complaints about the short duration and poor quality of music and voice clips in the demonstration album. Like other print-based methods of playback, the album was seen as more suitable for voice commentary than music. In this case the commentary was felt to go particularly well with the personalisation of a special set of photographs, as shown in the following quote.

> *I think the album would be nice with the sound to send to people who live a long way away who you don't ever see. So you could do the pesonalised message that went with it. I'm*

thinking about years ago when my mum's brother lived in South Africa and we used to sit at home and make tapes that we sent him with all the news, which wasn't just about news, but family voices. (Jane)

A number of participants were concerned about the potential cost of an audio-enabled album compared to what they usually spend on a conventional one. This led two people to suggest a re-usable album player, similar to the audioprint player, which could be used with any album. Albums themselves would have to have some kind of removable storage medium associated with them that could be inserted into the player when required.

"Could you put something different in there to record on, like a tape or CD, for different albums?…What I mean is, if you've got a lot of albums, if you had something like that which would slot into the album, you could put a different recording into each album to suit" (Byron)

  "You'd have to make it so that the photo album wasn't disposable, i.e. you can take the photos out and put a new photo in and play it" (Vicky)

Audioprint Player

Whereas the audioprint album was seen as a device for special audiophotos, the audioprint player was seen as a device for playing back routine audiophotos. This distinction roughly echoes that between loose prints and albums today, where albums are reserved for a subset of the best prints (Frohlich et al. 2002). Hence, participants spoke about using the player to browse through printed audiophotos themselves, and also to annotate regular prints with voice messages for sending to others. This accounts for the increased number of votes for voice rather than music playback in Fig. 6.7. The issue of recipients of audioprints needing a player to play-back the comments emerged again in this study. However, participants here resolved it by limiting the use of sending to very regular or routine recipients who could be provided with a player with the initial set of audioprints.

I could buy the attachment and send it too, if I was going to send a lot of photos. And the kids could record their voices and photos or whatever and I would just have to mail the photos off. And they can just plug them in and listen, or just look at the photos if they don't want to listen (Vicky)

Audiophoto Card

Audiophoto cards received the same pattern of votes as the audioprint player (see again Fig. 6.7). However, they were judged more similar to the audioprint album in their potential use. They were associated with the sending of special audiophotographs to distant family and friends on occasions such as birthdays, anniversaries

and other causes of celebration. In fact they were treated as extension of conventional greeting cards, which could be used to hold personal audiophoto material instead of a pre-designed message and graphic. Participants appreciated this flexibility, not only for sharing photographs more effectively, but also for designing new kinds of communication on these special occasions.

> *I would certainly be willing to make a card on some specific special occasion. I guess that is taking a picture from the camera, printing it out and making a recording, whether that's music or voice (Jane)*

6.5.3   Summary

Findings on the playback of musical and talking photographs were essentially similar to those on the playback of ambient photographs. No single playback method was preferred above others and participants expressed a desire to use different methods for different purposes. In fact, because of the change of methodology for measuring preferences, we were able to determine that each individual wanted to use about three different playback methods. On average two of these were screen-based methods while one was paper-based. In this study, movement between options was more a matter of context than personality, since there were fewer idiosyncratic choices or age and gender differences. Common contextual factors affecting playback selection were the type and size of the audience for audiophotos, whether they were being shown in or out of the home, and whether the audiophotos were routine or special in some sense. Mention of this latter dimension was a new development, and appeared to be significant for the choice of screen or paper methods of playback.

Routine audiophotos were associated mainly with screen-based playback. Participants expressed a desire to review their own audiophotos together on the PC, or to show them to visitors on the TV. They also favoured use of a Palmtop device for showing routine audiophotos out of the house. This trend extended to the regular sending of audiophotos to distant family and friends through the PC. The only exception to this rule was mentioned in the case of sending audiophotos to people without computers. Participants seemed to favour using audioprints with duplicate sets of audioprint players distributed across households.

In contrast, special audiophotos were usually associated with paper-based playback. Hence, the audioprint album was seen as a vehicle for archiving precious sets of audiophotos for personal use, or for sending to others as a gift. Audiophoto cards were seen as a more disposable kind of photographic communication to others on special occasions. This trend appears to mirror a recent decline in the number of digital photographs being printed at home. Consumers are becoming more comfortable about reviewing photographs on-screen and more selective about which photographs they decide to print out.

6.6   Discussion

Having remained neutral about the technology involved in audiophotography in previous chapters, this chapter has been devoted to its examination for audiophoto playback. A key issue to resolve was the relative value of paper versus screen-based methods of playback, since unlike video clips, audiophotos could technically be played back from a variety of audioprint devices. Previous research on the future of reading and tangibility in design suggested that both paper and screen methods of playback might be preferred for different reasons, although the usefulness and familiarity of photographic prints might inhibit a move away from the paper medium.

In fact we found that participants were impressed by the screen-based audiophoto albums they were shown in each study, and indicated their willingness to use them for playback in a variety of contexts. The PC, TV and Palmtop albums were found to be most appealing, because of a number of features they possessed which paper prints did not. These can be seen as comparable to the affordances of screen-based documents observed by Sellen and Harper (2002) and listed in the introduction to this chapter. For audiophoto review, screens were valued for the following affordances:

- Compact access to a large amount of material
- Content editing
- High quality audio-visual experience
- Viewing comfort
- Efficient transmission of material

These screen properties overlap with those of Sellen and Harper, particularly through the inclusion of editing and access to large amounts of material. The latter property was probably the main attribute underlying the attraction of the Palmtop as a portable method of playback. However, other properties in their list were not mentioned as important in the domestic audiophoto context, such as full-text searching or links to related materials. Instead, other properties were important, such as audio-visual quality, comfort and transmission. All these factors lead us to an answer to the question of *'What advantages screen-based photographs must offer for people to give up prints?'* They must begin to create a more comfortable and compelling viewing (and listening) experience than can be achieved on paper, with associated gains in the efficiency of storage, editing and transmission of material.

This shift towards screen-based methods of playback by no means eradicated the desire for printed audiophotos, but it did mitigate their use. Essentially audioprint methods of playback came to be reserved largely for the review and sharing of *special* audiophotos. In particular, the audioprint album and cards were seen as good ways of packaging selected audiophotos for special occasions and sending them as personalised gifts to other people. Audioprints used with a handheld player were considered for more routine use, alone or with a small number of family members. These methods possessed a number of properties in common that made them

superior to screens for these purposes. These can be seen as affordances of paper that are relevant to audiophoto playback.

- Ease of manipulation and navigation
- Ability to spread out and compare information
- Longevity of archiving
- Lowest common denominator for sharing
- Tangible form

Again there is overlap with the affordances of paper important to the office setting, mainly through the continued relevance of manipulation and the ability to spread out more than one image at a time. However, the ability to write on prints or other paper whilst reviewing photographs was not relevant. Instead, it was more important to know that recipients had a way of playing and seeing what they had been given, that the material was not going to become obsolete over time, and that it had a tangible form that could count as a gift. Other benefits of tangibility may well emerge in practical use, since differences in the dynamics of screen and paper conversations are not immediately apparent in a focus group setting.

In short, we have found that the lesson of the paperless office debate also applies to the debate about the future of digital photographs in the home. Domestic photographs like office documents will not go completely paperless because the properties of paper remain valuable for particular contexts of use. This value is increased in the case of audioprints where ambient, musical and spoken sounds can be invoked from the paper to enhance memory and communication around it. Screen-based photographs and sounds extend these values, when contexts of use change. This leads us to predict the co-existence of paper and screen-based methods of audiophoto playback in and out of the home.

Exactly which methods are appropriate to which contexts is complicated to explain from the data, but some answer is demanded from our other research question about context switching: *'In what contexts will people prefer certain kinds of paper or screen-based experiences with photographs?'* Using the contexts of use identified in the diamond framework of Fig. 1.2, I have attempted to summarise the findings for each context in Table 6.1. This shows the playback methods considered most suitable for each context, across the two studies. An entry in a cell refers to the kind of photo sharing that was thought to be appropriate to that particular method and playback context. Empty cells simply reflect the lack of any comments on the appropriateness of the method and context combination.

Rather than describing these playback choices again for each context, it is worth noting some of the factors that led participants to change their choices within each column of the table. For example, we found that the *location* of the activity affected playback selection especially in social settings. The PC and TV were preferred for sharing audiophotos in the home, while the palmtop or audioprint scanner were preferred for sharing out of the home. *Audio quality* was another factor affecting choice of playback method in several contexts. In general, higher quality was required for music playback and for public sharing of audiophotos, and this favoured the PC and TV methods whenever possible. The *existing infrastructure* was also an

Table 6.1 A summary of audiophoto playback methods preferred for different contexts of use

		Remembering	Interpretation	Reminiscing	Storytelling
Screen	**PC album**	Routine individual downloads and editing	Recipient of routine emails or web postings	Routine group reviews	
	TV album			Routine group reviews	Special group presentations
	Palmtop album	Routine individual reviews		Routine pair reviews	Special pair reviews
Paper	**Audioprint player**	Routine individual reviews	Recipient of routine audioprints	Routine group reviews	
	Audiophoto card		Recipient of special occasion cards		
	Audiorprint album	Special event album	Recipient of special event album		

issue in the selection of methods in the 'Interpretation' column of the Table 6.1. This led some participants to choose self-contained paper-based methods of playback like the audioprint card or album when sending material to other people. Finally, the *effort* of creating or transmitting material was considered when thinking about the overhead of using each method. This worked against some of the paper-based methods such as the audioprint album.

These factors reveal more than the use of a wide variety of criteria for playback selection. They show a certain kind of systems thinking by participants, who judged each method in terms of its overall process of use as well as by the playback experience it delivered. This view is consistent with a way of thinking about media as socio-technical systems of production and consumption rather than simple, technological means of presentation.

Hesse (1996) develops this argument in relation to the future of the book, by contrasting the printing press (technology) with the modern literary system of supporting authorship and readership around fixed writings (system). The real challenge of new technology, according to Hesse, is not in providing new technical methods of reading books, but in changing the relationships between author and reader so as to transform the system of book publishing. In the same way, the real challenge of audiophoto technology is not *only* in providing new ways of reading photographs, but *also* in changing the relationship between photographers, subjects and audiences so as to transform the system of domestic photography.

Participants in these studies appeared to recognise this instinctively for the system of photography represented by the diamond framework. Some presentation methods appeared to priviledge or disadvantage certain actors or activities in the system, compared with others. This led people to optimise the system for particular uses, by switching technical methods. This results in a kind of *domestication* of technology, by selecting and adapting its properties to existing social practices.

References

Adler A, Gujar A, Harrison BL, O'Hara K, Sellen A (1998) A diary study of work-related reading: design implications for digital reading devices, Proceedings of CHI'98. ACM Press, New York, pp 241–248

Bolter JD (1991) Writing space: the computer, hypertext and the history of writing. Lawrence Erlbaum Associates, Hillsdale

Brown BAT, Geelhoed E, Sellen AJ (2001) The use of conventional and new music media: implications for future technologies. In: Hirose M (ed) Proceedings of INTERACT 2001. IOS Press, Tokyo, pp 67–75

Csikszentmihalyi M, Rochberg-Halton E (1981) The meaning of things: domestic symbols and the self. Cambridge University Press, Cambridge

Dillon A (1992) Reading from paper versus reading from screens: a critical review of the empirical literature. Ergonomics 35(10):1297–1326

Frohlich DM, Murphy R (2000) The memory box. Pers Technol 4:238–240

Frohlich DM, Adams G, Tallyn E (2000) Augmenting photographs with audio. Pers Technol 4:205–208

Frohlich D, Kuchinsky A, Pering C, Don A, Ariss S (2002) Requirements for photoware. In: Proceedings of the 2002 ACM conference on computer supported cooperative work, ACM, pp 166–175

Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin, Boston

Hawley M, Poor RD, Tutja M (1997) Things that think. Pers Technol 1:13–20

Hesse C (1996) Books in time. In: Nunberg G (ed) The future of the book. University of California Press, Berkeley, pp 21–33

Hoskins J (1998) Biographical objects: how things tell the stories of people's lives. Routledge, London

Ishii H, Ullmer B (1997) Tangible bits: towards seamless interfaces between people, bits and atoms, Proceedings of CHI'97. ACM Press, New York, pp 234–241

Luff P, Heath C (1992) Tasks in interaction: paper and screen based documentation in collaborative activity, Proceedings of CSCW'92. ACM Press, New York, pp 163–170

Luff P, Heath C (1997) Mobility in collaboration, Proceedings of CSCW'98. ACM Press, New York, pp 305–314

Luff P, Hindmarsh J, Heath C (eds) (2000) Workplace studies: recovering work practice and informing system design. Cambridge University Press, Cambridge

Manguel A (1996) A history of reading. Knopf Canada, Toronto

Nunberg G (ed) (1996) The future of the book. University of California Press, Berkeley

O'Hara K, Sellen A (1997) Comparison of reading paper and on-line documents, Proceedings of CHI'97. ACM Press, New York, pp 335–342

Schaverien J (1991) The revealing image: analytical art psychotherapy in theory and practice. Jessica Kingsley Publishers, London

Schilit BN, Price MN, Golovchinsky G, Tanaka K, Marshall CC (1999) The reading appliance revolution. IEEE Comput 32:65–73

Sellen AJ, Harper RHR (2002) The myth of the paperless office. MIT Press, Cambridge, MA

Snyder I (ed) (1998) Page to screen. Routledge, London

Stevens MM, Abowd GD, Truong KN, Vollmer F (2003) Getting into the living memory box: family archives and holistic design. In: Proceedings of 1 AD, 35–41. First international conference on appliance design, Bristol, 6–8 May 2003

Want R, Fishkin KP, Gujar A, Harrison BL (1999) Bridging physical and virtual worlds with electronic tags, Proceedings of CHI'99. ACM Press, New York, pp 370–377

Weiser M (1991) The computer for the 21st century. Sci Am 265(3):94–104

Wellner P, Mackay W, Gold R (1993) Computer augmented environments: back to the real world. Commun ACM 36(7):87–96

Part II
What Happened Next?

Chapter 7
Research and Development in HP

With some simplification, we can divide the inquiry into two main categories: user oriented and technology oriented. Both are equally important! Without a strong grounding in potential users' desires and needs, it is unlikely that any new idea will gain traction. But at the same time, an invention may be ever so desirable, but if it is not actually within the possibilities of existing or future technology, it will simply never be realised Holmquist (2012, p. 39)

7.1 Introduction

The historical and company context of the original audiophoto work, very much influenced what happened next within HP. And what happened next, unfolded both as an impact on HP products but also on the ongoing research work carried out by other research scientists in the Labs and Divisions. I review these different impacts in the following two Sects. 7.2 and 7.3.

One recurrent difficulty in conducting this analysis has been framing the original audiophotography research and vision, and following its evolution over time. Scientists, designers and inventors naturally want to know what happened to their ideas and tend to adopt an egocentric point of view. This is reinforced by companies who want to track and manage the year-on-year performance of research staff. However, most innovation contexts are populated by multiple people with the same innovation remit. And people routinely influence each other and work in teams to develop their ideas. This means that it becomes harder and harder to track the influence of particular ideas and research as time goes on, and as those ideas take hold within the culture of the research group and larger company. Individual insight and one-way influence soon turn to mutual influence and joint development, as more people get involved and the research ideas drift from their original context. This is actually one of the biggest discoveries of the exercise and something we will return to in the discussion. But first let us trace the uptake of the early collaborative audiophotography work in HP, as described in Frohlich (2004), as best we can.

7.2 Influence on HP Development

The commercialization of audiophotos started some time ahead of the 2004 Audiophotography book, as reported in a recent paper (Frohlich and Sarvas 2011). The first audiophoto product from HP was an application for the Pocket PC camera attachment for the *Jornada* Personal Digital Assistant (PDA) in 1999 (see Fig. 7.1). It supported the annotation of captured images with a short voice message. The audio was saved in WAV format and given the same filename as the image file. The device was targeted at mobile professionals such as surveyors, journalists, estate agents and salespeople, so adding voice labels to photographs was considered useful in supporting their work. Interestingly this was not a hypothesis that came out of the audiophoto research, which was focused on mainstream domestic photography by families. But the interpretation and transfer of the idea in this context came about through a colleague, Guy Adams. Guy happened to be working on a *Pocket PC* camera attachment for the Jornada with Chee-Keat Fong at HP Singapore where the Jornada was made. The transfer of audiophoto support was easy because the Jornada device already had sound capture built into it, for recording voice memos and considerable processing capacity for supporting novel applications. In this respect, it anticipated the smartphone platforms of today, which make very good devices for capturing audiophotos as we shall see later in the book. The audiophoto 'app' later transferred from the Jornada to the iPaq around 2001 when HP bought Compaq and rationalized its production of mobile devices.

Although we had argued for the creation of a specialized 'audiocamera' in the HP range, the company decided to introduce sound capture on all its cameras in 2002. This was heralded a year earlier by its introduction in the HP Photosmart 650 camera which was a high end digital SLR, likely to be used for professional as well as amateur use (see Fig. 7.2a). A particular feature of the sound design was the use of an audiophoto 'mode' which, when on, activated sound recording on half-depression of the shutter button and continued recording as the shutter was fully depressed and

Fig. 7.1 The HP Jornada PDA

Fig. 7.2 Two of HP's audio-enabled cameras from across the camera line: (**a**) HP Photosmart 650 DSLR camera; (**b**) HP Photosmart 425 compact camera

held down after the image-capture click. This allowed users to frame the capture of a sound window around the image capture event, as recommended by our research. However this was considered to be too complex for incorporation in the compact camera line in 2002 (see Fig. 7.2b). It was dropped in favour of a modeless point-shoot-and-hold action, where users simply held down the shutter button after the click to record sound until the shutter button was released. The resulting sound clip (WAV file) was then stored inside the corresponding JPEG image file. The audiophotos could be played from the back of these cameras or transferred to a computer, and onward to the internet and other computers. Unfortunately for the uptake of audio-photography, embedded sound playback from JPEG images was not universally supported by all image viewers of the day, nor even by HP's own photo website. This meant that audiophotographs could not be easily shared with non-HP camera owners. Even more unfortunately, HP pulled out of camera manufacture in 2007.

A final audiophoto product was HP's 2004 software for sharing and authoring photo slideshows in Video CD format on a CD-ROM. This was called HP Memories Disc and conventionally allowed users to set the slideshow to music (see Fig. 7.3). A later version worked with the import of embedded sound-photos from HP cameras and mixed the ambient sounds with the music backing at playback time. A version of this software was eventually taken into a common image management application called HP Image Zone used across all HPs imaging products. Such mixing of different kinds of sounds with single or multiple photographs was a strong recommendation of the audiophoto work and featured at the end of the 2004 Audiophotography book. Some aspects of this recommendation were incorporated into a new file format for CDs and DVD being developed by the Optical Storage Technology Association called MPV (derived from MultiPhotoVideo or MusicPhotoVideo). However MPV now seems to be an inactive format due to the rise of downloadable rather than physical media.

More recent products from HP continue to support elements of the audiophotography vision, especially those including augmented paper. However these become

Fig. 7.3 HP memories disc creator software

more difficult to trace back to the early audiophoto work as time goes on, because of the influence of further related research projects both inside and outside HP. For this reason, we turn now to review the subsequent internal research before describing HPs very latest product innovations in this area as part of the Discussion section.

7.3   Influence on HP Research

In addition to these commercial activities, many of the core ideas and technologies related to audiophotography appeared to influence other people and research projects in HP Labs. Again, the influence of the work reported in the 2004 book, was felt earlier inside HP because most of the findings were available 4 years earlier from the year 2000. The following summary is essentially a snapshot of the way ideas influence other ideas, especially in the vibrant creative culture of HP's research centres. It should be remembered that the audiophoto work had involved about 10 other members of staff directly, each of whom had their own research agendas and projects going on in parallel. Regular seminars, meetings, trips and demonstrations also served to expose the work to many more staff, especially in the Bristol and Palo Alto Labs. Consequently, the ideas became part of the collective 'zeitgeist' of the Labs and appeared to stimulate or influence a number of other projects.

Fig. 7.4 HP memory spot (**a**) and its potential applications, (**b**) a printed audiophoto, (**c**) a talking book, and (**d**) a hospital wristband (Reprinted with permission)

Perhaps the biggest internal consequence of the work was the setting up of a team of engineers led by John Waters and Ed McDonnell to create a tiny wireless chip that could be embedded in paper, with sufficient memory to store an audio clip and perhaps other data. This derived directly from our audioprint demo of 1999 and several related patents (Adams et al. 1998, 1999a, b; Frohlich and Hickey 1999; Frohlich et al. 2000). HP announced the successful development of the resulting chip, called *HP Memory Spot*, 4 years later in July 2006 (see Fig. 7.4). The chip is the size of a grain of rice, holds between 256 k and 4 MB of data, and could cost as little as 10–50 cents. It can also transfer the data to any reading device such as a smart phone at up to 10 megabits per second, via a proprietary wireless link. Despite the promise of the chip for audioprint and other applications, HP have yet to commercialise it; illustrating the many factors needed to take a new technology to market and the time this takes.

Six other augmented paper projects also sprung up in the early 2000s. Four of these were part of my own continuing work in the area, as crystalised in an ongoing project called TANGENT: tangible media interaction. Following our audioprint demonstration in 1999, Guy Adams and I took the simple step of taking the same audiochip we used in the photographic print and embedding it in a paperclip instead. In fact we had to design a new kind of *Audiopaperclip* with the clip at one end and an electronic tab connected to the embedded chip at the other. Then we designed a handheld and desktop audiopaperclip player to playback the sound (see Fig. 7.5). The players had both play and record settings so that new messages or sound record-

Fig. 7.5 Audiopaperclips (**a**) with a handheld (**b**) and desktop (**c**) player

ings could replace or extend old ones. We imagined that such clips could come in different colours for different kinds or durations of sound, and that they could be recorded and attached manually to printed photograph albums. Alternatively the player could be built into an album with wired contacts around each page (Adams and Frohlich 2002 – patent application). As with HP Memory Spot, the audiopaperclip concept opened up new applications outside photography, and was probably best suited to exchanging voice messages with office documents. This use was demonstrated in the UK Design Council's *Great Expectations* exhibition of October 2001 in Grand Central Station New York, and featured as one of ten Best of British designs of the year (Carlowe 2001 – Focus Magazine).

Having separated the audiochip from paper and embedded it in a paperclip, it was only a small additional step to consider embedding it in other objects. Physical memorabilia and souvenirs were an obvious target, since they stood to gain many of the same benefits of sound as photographs. In some ways, memorabilia act like photographs as memory triggers, and can be seen to extend domestic photography as a form of reminiscing (Frohlich and Fennell 2007). A variety of new product concepts in this area were envisioned through personal collaborations with Rachel Murphy and Jacqueline Fennel around 2000–2003, both of whom were hired by HP from the Computer Related Design course at the Royal College of Art.

The earliest concept was a Memory Box, inspired by a woman's jewellery box and the layers of history that can accrue on precious objects that outlive a human lifetime (Frohlich and Murphy 2000). The box was designed to support the capture and playback of spoken stories on objects, which we suggested would be tagged with RFID stickers to identify them. As owners placed a new object in the box for the first time they would be given the chance to record the story, which would be played back automatically as the object was removed (see Fig. 7.6). A mock-up of this idea was shown to 11 US families for feedback and discussion, leading to further concepts and design ideas. For example, women liked the idea in its current form, but children wanted a version that worked for their secret collections of things stored in other containers and drawers. There was also a call for audio-annotation of larger objects such as furniture, and a mobile solution to allocated household possessions in a talking will.

The augmentation of memorabilia represented by the Memory Box work proved so promising that we looked for opportunities to extend it. This materialized in 2002

Fig. 7.6   The memory box

through an inclusive design project sponsored by HP Labs Bristol but run by the Helen Hamlyn Research Centre at the Royal College of Art. This was a managed form of open innovation as discussed above, but is described here because of its' funding by HP. Jacqueline Fennel carried out the project on 'Multisensory memorabilia' which aimed to explore how partially sighted people remember the past without the usual recourse to photographs (Fennell and Frohlich 2005). In partnership with five reference users who were registered blind, Jacqueline worked up a variety of ten novel product concepts for remembering the past in new ways. Interestingly, many partially sighted people go gradually blind through age or disease, so an emerging requirement was for a way of bringing existing photos back to life through touch. This was explored in two concepts for tactile photographs. However, the closest concepts to audiophotography were those featuring the novel use of audio. These included a *History Answering Machine* for keeping conversations and offering them for playback on their birthday each year, a *Memory Shelf* for creating a database of talking objects recognized by their weight, *Best China* for recording memorable conversations around the dinner table, and *Hospital Armband Chatter* for recording conversations around the hospital bed. Memory Shelf and Best China were created as full working demonstrators for exhibition and feedback, and are shown in Fig. 7.7.

My fourth augmented paper project, was a collaboration with John Robinson and Enrico Costanza at York University and the photographer Tony Clancy from University of Central Lancashire. We co-designed and built an *Audiophoto Desk* for playing back sounds from printed photographs thrown onto a desk surface (Frohlich et al. 2004). The desk used an overhead camera to recognize the prints and fetch associated sounds from a hidden PC, and was exhibited at the Royal College of Art

Fig. 7.7 The memory shelf and best China concept envisionments, from Fennell and Frohlich (2005) (Used with permission of Helen Hamlyn centre for design)

in the autumn of 2004 (see Fig. 7.8). An interesting property of the desk was its ability to playback multiple audiphotos simultaneously, and to allow the physical control of their volume and stereo panning by pushing them backwards or forwards and left or right. This led Tony Clancy to explore the creation of small sets of audio-photos with different kinds of sounds and images that worked well as an 'ensemble'. The vision for a commercial version of this was as a domestic desk lamp with wireless connection back to a home PC. The photographic part of a stored audio-photo could be printed out in the usual way, but its associated sound could be played back under the lamp.

A related audioprint exhibition called *Walk in the Wired Woods* was hosted in the Labs itself 2 years earlier, as an outgrowth of the Mobile Bristol project. This was exploring the notion of ubiquitous media that could be triggered anywhere in both indoor and outdoor environments by location sensors. The resulting 'mediascapes' could be authored by artists and content developers to create walking tours, situated plays, mixed reality games and gallery experiences (e.g. Stenton et al. 2007). An early experiment was to playback sounds on an iPaq handheld computer relating to large printed posters suspended from the ceiling, using overhead ultrasonic microphones to identify its indoor location. A photographer and sound artist collaborated to make the audiophoto content of woodland scenes and sounds (Hull and Reid 2003).

A final audioprint technology was developed by a combined Bristol and Palo Alto group of scientists and engineers. This was called *Active Photos* and involved the association of text, audio, image and video data with different regions of a printed photograph (Kindberg et al. 2004). This was done by placing the photo on a passive digitizer tablet and marking up the photo with a stylus to either associate or playback content for specific regions. The associated content was played back on either an integral iPaq screen in the tablet or on a nearby computer.

Fig. 7.8 The audiophoto
desk with John Robinson,
Tony Clancy and David
Frohlich (*from left*)

In addition to the interest in augmented paper and objects around the Labs at this
time, there was equal interest in supporting new kinds of screen-based experiences
with photographs and other personal media. This included various kinds of story-
telling with images. For example, a small group in Bristol had created a wearable
camera and were experimenting with the processing of large amounts of personal
video data, akin to Microsoft's work on *SenseCam* and *MyLifeBits* (Gemmel et al.
2002, 2004). One immediate finding was that much of the video taken passively on
a wearable camera is shaky and difficult to watch without feeling sick. However, it
can be used to extract so called 'trophy shots' (still image frames) and candid audio
clips around them. Furthermore, video can be generated from individual or multiple
images in a series to tell short stories, as in a documentary film. My own collabora-
tion with Dave Grosvenor from this group resulted in numerous patent applications
for *audiophoto storytelling and annotation*, including one for panning over still
images during an ongoing narration (Grosvenor et al. 2002).

Other members of the same Bristol group, including Guy Adams and Andy
Hunter, were working on new CMOS camera hardware and software, and creating
a flexible camera testbed (or 'breadboard') for exploring new functionalities. In fact
this was used to create a reference design for a programmable CMOS camera
docked to a PC. Discussions about adding sound capture to this camera resulted in
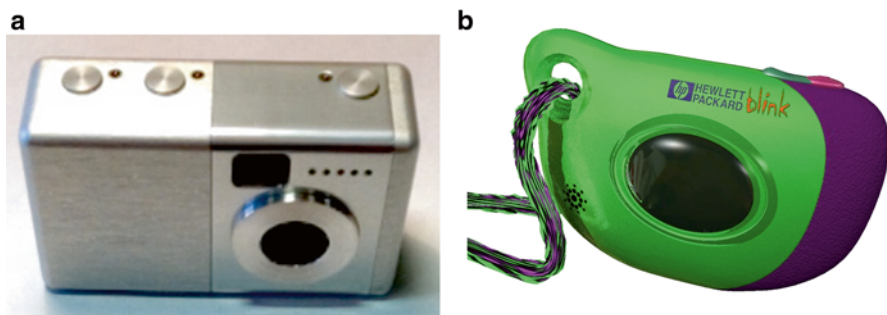
Fig. 7.9 The blink audiocamera (**a**) and envisioned product (**b**)

the design and implementation of *Blink* in 2001: the Labs first digital audiocamera with a separate sound capture button for the left index finger (see Fig. 7.9a). This had been a key recommendation for capturing sound on a camera following the audiocamera trial of 1996/1997, to allow separate sound and image capture from different buttons. This allowed memories to be captured by sound or image alone, but also by flexible combinations including the holding down of the sound capture button while one or more photos were taken. Blink supported this flexibility, and displayed the resulting sound, image, audiophoto and audio-slideshow media in an interactive webpage. While we never got to trial the new audiocamera with members of the public, it was popular for its creative potential, and explored artistically by Tony Clancy in a number of audiophoto works. It also triggered an envisionment for a new audiocamera product, by Jim Gerard, an HP industrial designer on sabbatical in the Bristol Labs. The proposed audiocamera form is based around the circular shape of speaker, but was never commercialized by HP (see Fig. 7.9b).

An additional design exploration of audiophotography products was commissioned by HP Labs as part of a collaboration with the Computer Related Design Department at the Royal College of Art. Bill Gaver and Heather Martin were briefed in 1999 to explore the space of products in this area, by creating a workbook of related concepts. The results of this work were subsequently presented back to HP and documented in a conference paper (Martin and Gaver 2000). A range of new domestic product concepts were developed and linked to a broader investigation (Gaver and Martin 2000). This culminated in two new audiophoto prototypes shown at the RCA in November 1999. One was an *Evesdropping Camera*, whose audio-photographs could be triggered in response to loud sounds such as laughter. The other was a *Digital Shelf*, along which barcoded photographic prints could be moved to play different sounds along the shelf. Alternatively it was suggested that a small postcard-sized display could be moved along the shelf to flip through digital audiophotos stored in the shelf. Although these concepts were never commercialized by HP, they served to illustrate the potential of the approach for new forms of creative expression and communication.

Several years later and further afield in Palo Alto and Cambridge Massachusetts, colleagues were thinking about how to generate and retrieve audiophotos automatically. Hence in Palo Alto, Margaret Fleck created a system for '*evesdropping on storytelling*' in which all conversations around a photograph were captured, as in the notion of conversational photographs in my original book. However, she took this further by doing automatic speech-to-text translation on the recorded speech (Fleck 2004). The resulting text could then be used as keyword tags for later retrieval and browsing. In Cambridge, Beth Logan and colleagues had developed a search engine called *SpeechBot* for finding speech, video and music from the web (van Thong et al. 2002). After Beth heard about audiophotos as a new media type, we filed a joint patent for retrieving and browsing audiophotos from their associated audio alone (Logan and Frohlich 2004). Both inventions suggest new ways of reviewing photographs based on the stories told about them or the similarity of music and ambient sounds associated with them.

A final group of colleagues in Palo Alto were beginning to research the future of photography on cameraphones rather than cameras, in anticipation of the revolution to come. Central to their vision was the notion of multimedia storytelling based on the circulation of personal media and conversation in a social network (Debaty et al 2004; Rajani and Vorbau 2004). Talking photographs were therefore a natural fit to this vision and something they explored in two systems with very different starting points. *MemoryNet Viewer* was essentially a photo sharing web service (like Flickr) supporting both text and voice annotation of photographs. Interestingly this was intended to display the media of networked 'friends' in a continuous slideshow on a situated display. Other friends could comment on individual photos and see or hear the comments of others. Most spoken comments were provided in co-located storytelling around the displays, while most textual comments were provided alone for remote circulation. To encourage the recording of spoken comments for remote audiences, the team created another system called *StoryMail*. This was essentially a voicemail system accessible from a cameraphone that allowed users to record and share voicemails augmented with photographs or video clips. This was later released as an HP product called *StoryCast* in 2005, as reported by Red Herring magazine. Again the integration of audiophoto recording and sharing in another context, both amplified and changed the original idea.

7.4   Discussion

The range and variety of subsequent audiophoto projects in HP following the original work, is testimony to the attraction of the idea within the company. But how much do these projects tell us about the uptake of innovation within large corporations and the prospects for widespread industry support of audiophotography itself? It turns out that the company focus of this analysis and chapter tells us more about innovation than audiophotography as such, because the company perspective is always to maximise the impact of research on products and therefore the return on
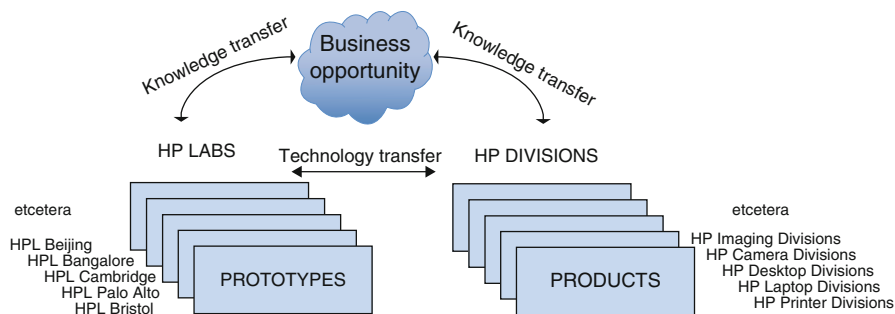
Fig. 7.10 Knowledge exchange within HP

research investment. The company inclination with any new business opportunity such as audiophotography, is to realize any immediate benefits within existing product lines whilst at the same time working up that opportunity for future products. In HP's case, their products span computing, printing and imaging industries of which photography is only a very small part. Hence there was a natural long-term pull away from the original photographic context to explore the benefits of audiovisual media and augmented paper for other products and markets. In general, the shorter term work was carried out by divisional marketing and R&D groups specialized by product category, while the longer term work was carried out by the corporate research labs segregated by geography. This situation is represented in Fig. 7.10 and played out in Sects. 7.2 and 7.3 above.

Hence, immediate commercial transfers affected the capture of photographs with sounds on HP cameras and their playback through HP imaging software on mobile and desktop computers. This was relatively straightforward to implement since some cameras were being developed to support video as well as still image capture at the time, with integral microphones for sound capture and speakers for sound and video playback. Interestingly, these features were inherited 'for free' on the Jornada PDAs and the high-end digital SLR cameras, both of which were targeted more at mobile professionals than at domestic consumers. This led to an initial commercialisation of audiophotography outside the market for which it was originally researched and designed.

The implementation of audiophoto capture support was also different to that recommended by our research. Ironically, the Jornada supported voice annotation of photographs which we argued against, rather than ambient sound capture around the photographic event. The HP 650 DSLR camera tried to rectify this, but with a single button rather than with a separate sound capture button as recommended for the left index finger (see again Fig. 7.9a and the *Blink* prototype camera). The second button was too difficult to implement because of the placement of the optical viewfinder in this top left position on the body of the camera (from rear). So with Divisional colleagues, we co-invented, designed and patented an alternative method of half-button depression for onset and offset of sound recording from a single shutter but-

ton (Battles et al. 2002). These examples show the opportunistic nature of technology and knowledge transfer, and the way that this sometimes leads to modifications of the original ideas on the one hand and further invention on the other.

The transfer of audiophoto support to HP Memories Disc and Image Zone software was also relatively straightforward and opened up other uses of audiophotos outside photography. Because the same software was supported with computers, printers, scanners and all-in-one devices this opened up a debate about the possible role of audio and audiophotos in each of their markets. Greatest interest was probably from the printer divisions intrigued by the possibility of augmented paper, and this emboldened the Labs to propose and eventually fund the HP Memory Spot project. The fact that this technology had applications outside photography in the printing of multimedia documents and labels (Fig. 7.4) was seen as a strength that might eventually offset the cost of additional research investment. The HP corporate Labs was the best place to develop this, because of its research orientation and product independence, although that created the problem of finding a Divisional customer later on.

Smaller scale investments were represented by the many other lab projects which sprang up more spontaneously to incorporate elements of audiophotography in other research contexts. Some of these were opportunistic, making small step changes in other prototypes to test the feasibility and benefits of sonic augmentation. For example, the creation of audiopaperclips from the same chips used to create embedded chip audioprints (Fig. 6.6) was a small step technically, but it opened up a new user experience with photographs and other documents that could be annotated with detachable sounds. This was ultimately more suitable to an office stationary market than a domestic photography one, although HP chose not to commercialize this either, even though it sold office stationary such as paper and ink. Other investigations into voice annotation of online photographs through MemoryNet and StoryMail were small adaptations of the audiophoto idea that led to further commercializations both inside and outside HP. StoryMail resulted in the HP StoryCast multimedia messaging service, while aspects of MemoryNet can be seen in numerous digital storytelling apps that will be discussed in Chap. 9. Again these examples show a diversification of applications of audiophoto *technology* outside the domestic photography domain. They also generate opportunities that may not be particularly well suited to HP's existing technologies, markets, channels or competencies. In this respect they do what all good research labs should do, which is to challenge the company with opportunities to change and grow their business in new directions.

A final observation can be made about the nature of this challenge with respect to the involvement of external collaborators. In general, the research projects involving other organizations and people *outside* HP, generated the most different and challenging opportunities. This could be seen in the collaboration with artists on the Mobile Bristol project for *Walk in the Woods* (Fig. 7.9), with photographers and engineers at the Universities of Central Lancashire and York for the *Audiophoto Desk* (Fig. 7.8), with students and staff at the Royal College of Arts for the *Memory Box, Memory Shelf, Best China, Eavesdropping Camera and Digital Shelf* proto-

types (Figs. 7.6 and 7.7). For the most part these were not done in partnership with Divisional customers and were deliberately provocative with respect to their audience, which included academic and end user groups of interests to the other partners. In effect, these collaborations were controlled forms of open innovation, commissioned by HP in collaboration with HP staff, but subject to some independence from HP goals and thinking. Consequently they were valuable to consider but difficult to commercialize.

The fact that many of these prototypes and ideas did not make it into direct product 'descendents', does not mean they did not influence later HP products. Depending on the timing of their creation and the relationships between researchers, they mutually influenced each other as described in publications and cross citations. And this influence continued over time, as new researchers adapted *those* technologies to new contexts and HP markets, as in the above examples. Within the evolving culture of HP there continued to be interest in multimedia photography and memory capture, and in the augmentation of paper with digital content and life. These interests can be seen in more contemporary product innovations and acquisitions such as the Aurasma augmented reality platform from HP Autonomy (www.aurasma.com), the HP Live Photo smartphone app (http://www8.hp.com/uk/en/ads/printables/live-photo.html) and the recently announced HP Sprout PC with overhead camera and projector (http://sprout.hp.com/). In fact the Audiophoto desk approach of doing image recognition on photographic prints to fetch associated digital content, turned out to be the most powerful way of bringing paper to life, and extensible to objects and hand gestures. HP Live photo is a simple commercialization of this approach through a smart phone app which recognises a printed photograph to play an associated video clip on the phone. And the overhead camera of the Sprout PC echos that of the Audiophoto desk, but for bringing paper and objects to life through a fully-fledged PC screen and speakers. While this *also* echos much earlier experiments with digital desks (e.g. Wellner 1993) and the growing momentum of an internet of things, it would be wrong to overlook the role that *all* the audiophoto prototypes played in these recent developments. As we have seen, ideas are taken up and used rather opportunistically by individual people or teams within particular institutional contexts, and there is evidence of at least early uptake of these ideas within the Zeitgeist of HP Labs and Divisions.

Many of these HP developments exploit the technology and functionality of audiophotographs outside the photography context. In the next two chapters we will consider audiophoto research and product developments in the wider industry to see if they have brought the proposed practice of audiophotography any closer to reality.

References

Adams G, Frohlich DM (2002) Photo album with provision for media playback via surface network. US 20040008209 (15.1.04) US 6975832 (13.12.05) JP 2004034692 (5.2.04), 13th Mar 2002

Adams G, Frohlich DM, Rix MS (1998) Method and system for providing a printed image with related sound. US 20050264657 (12.1.05) US 6930713, 16 Aug 2005

Adams G, Frohlich DM, Rix MS (1999a) Printing of image with related sound. US 20020075464 (20.6.02) US 6563563 (13.5.03) EP1095312 (15.10.03), 8th July 1999

Adams G, Frohlich DM, Rix MS, Waters J (1999b) Selectively attachable device for electronic annotation and methods therefor. US 20030225572 (4.12.03), 15 July 1999

Battles A, Staudacher D, Thorland M, Frohlich DM (2002) System for capturing audio segments in a digital camera. US 20030174218 (18.9.03) JP 2003283904 (3.10.03), 22 Feb 2002

Carlowe J (2001) The best of British. Focus Mag 108(November):36–42

Debaty P, Goddi P, Gossweiler R, Rajani R, Vorbau A, Tyler J (2004) Enabling informal communication of digital stories. HP labs technical report no HPL-2004-180

Fennell J, Frohlich DM (2005) Beyond photographs: a design exploration of multi-sensory memorabilia for visually impaired people. HP labs technical report number HPL-2005-151

Fleck M (2004) Evesdropping on storytelling. HP labs technical report number HPL-2004-44

Frohlich DM (2004) Audiophotography: bringing photos to life with sounds. Kluwer Academic, 220 Dordrecht

Frohlich DM, Fennell J (2007) Sound, paper and memorabilia: resources for a simpler digital photography. Pers Ubiquit Comput 11(2):107–116

Frohlich DM, Hickey M (1999) Digital camera with sound recording. WO/2000/048388 (17.8.00) EP1028583 (16.8.00) EP1151600 (7.11.01), 27 Apr 1999

Frohlich DM, Murphy R (2000) The memory box. Pers Technol 4:238–240

Frohlich DM, Sarvas R (2011) The role of HCI in innovation, Proceedings of CHI 2011. ACM, New York, pp 713–728

Frohlich DM, Adams G, Tallyn E (2000) Augmenting photographs with audio. Pers Technol 4:205–208

Frohlich DM, Clancy T, Robinson J, Costanzo E (2004) The audiophoto desk. In: Proceedings of 2 AD: 2nd international conference on appliance design, p 139

Gaver B, Martin H (2000) Alternatives: exploring information appliances through conceptual design. In: Proceedings of CHI 2000, ACM Press, New York, pp 209–216

Gemmel J, Bell G, Lueder R, Drucker S, Wong C (2002) My life bits: fulfilling the Memex vision, Proceedings of ACM multimedia 2002. ACM, New York, pp 235–238

Gemmell J, Williams L, Wood K, Lueder R, Bell G (2004) Passive capture and ensuing life issues for a personal lifetime store, Proceedings of CARPE 2004. ACM, New York, pp 48–55

Grosvenor DA, Frohlich DM, Hall GP (2002) Method and apparatus for producing video and audio-photos from static digital images. US 20050008343 (13.1.05) JP2005038399 (10.2.05), 2 Dec 2002

Holmquist LE (2012) Grounded innovation: strategies for creating digital products. Elsevier, San Francisco

Hull J, Reid R (2003) Designing engaging experiences with children and artists. In: Blythe MA, Overbeeke K, Monk AF, Wright PC (eds) Funology: from usability to enjoyment. Kluwer, Amsterdam

Kindberg T, Tallyn E, Rajani R, Spasojevic M (2004) Active photos, Proceedings of DIS 2004. ACM, New York, pp 337–340

Logan B, Frohlich DM (2004) Image organization method and system. US 20050281541 (22.12.05), 14 June 2004

Martin H, Gaver B (2000) Beyond the snapshot: from speculation to prototypes in audiophotography, Proceedings of DIS 2000. ACM, New York, pp 55–65

Rajani R, Vorbau A (2004) Viewing and annotating media with memory net, Proceedings of CHI 2004. ACM, New York, pp 1517–1520

Stenton SP, Wee S, Hull R, Goddi PM, Reid JE, Clayton BJC, Melamed TJ (2007) Mediascapes: context-aware multimedia experiences. HP labs technical report HPL-2007-113

Van Thong J-M, Moreno PJ, Logan B, Fidler B, Maffey K, Moores M (2002) Speechbot: an experimental speech-based search engine for multimedia content on the web. IEEE Trans Multimedia 4(1):88–96

Wellner P (1993) Interacting with paper on the digital desk. Commun ACM 36(7):87–96

Chapter 8
External Research

> *Invention is not a matter of a sudden flash of inspiration from which a new device emerges 'ready-made'. Largely it is a matter of minute and painstaking modification of existing technology. It is a creative and imaginative process, but that imagination lies above all in seeing ways in which existing devices can be improved, and in extending the scope of techniques successful in one area into new areas. (Donald and Wajcman 1986, p. 8)*

8.1 Mapping the Territory

Identifying external research in audiophotography after 2004 is a difficult but not impossible task. It is made tractable by the academic practice of citing references related to any current field of study. This makes it possible to trace the forward citation of references to the initial audiophotography vision, as represented in four papers (at 17 November 2014):

1. 47 citations to an early conference paper on the approach (Frohlich and Tallyn 1999)
2. 24 citations to a journal paper describing the audioprint opportunity (Frohlich et al. 2000)
3. 86 citations to the original *Audiophotography* book (Frohlich 2004)
4. 10 citations to a conference paper on the audiophoto desk (Frohlich et al. 2004)

While many of these citations to these four publications overlap, and others are being made continuously, there were probably around 100 unique pieces of work to consider 10 years after publication of the Audiophotography book. These broke down into three categories which further refine the search for external developments in the field. Some references were from *HP authors* whose related work has already been reviewed in Chap. 7. Other citations are really *passing references* to aspects of the original work or its representation of a class of similar works. Hence, the audiophotography work contains a number of contributions which can be referenced separately, such as the method of conducting ethnographic work in the home (cited in Drazin and Frohlich 2007), the fine-grained analysis of 'phototalk' (cited in Swan et al. 2008), linking sounds to printed photographs on the Audiophoto desk (cited in

van Erve et al. 2011), and representing interactions with photographs in a diamond framework (cited in Lindley et al. 2009). The examination of sound with pictures can also be classified more generally as a form of multimedia photography (e.g. cited in Durrant et al. 2011a, b; Whittaker et al. 2010; Wilson et al. 2012) or a new form of multimedia communication on cameraphones (e.g. cited in Kindberg et al. 2005a, b). However, a third set of citations were from authors who were *developing the core ideas* or claims of audiophotography in different ways. It is to these authors that we now turn, first to identify particular strands of research and second to review their findings.

Somewhat surprisingly, it turns out that there are only a small number of research strands and groups who have been actively developing the notion of audiophotography since 2004. In fact there are only six main lines of investigation as shown in Fig. 8.1. The figure shows developments as branches of an audiophotography 'tree', exploring the ideas of audiophoto **messages** and **narratives**, audio **context** for photographs and **triggers** for memory, and the playback of audio from **paper** and **artifacts**. The metaphor of a tree was more productive here than that of the ripple used in Chap. 1, since distinct lines of research can be shown to emanate from the cited references, growing and sometimes splitting over time.[1] Three industrial research groups are actively involved, including Microsoft Research Cambridge, Motorola
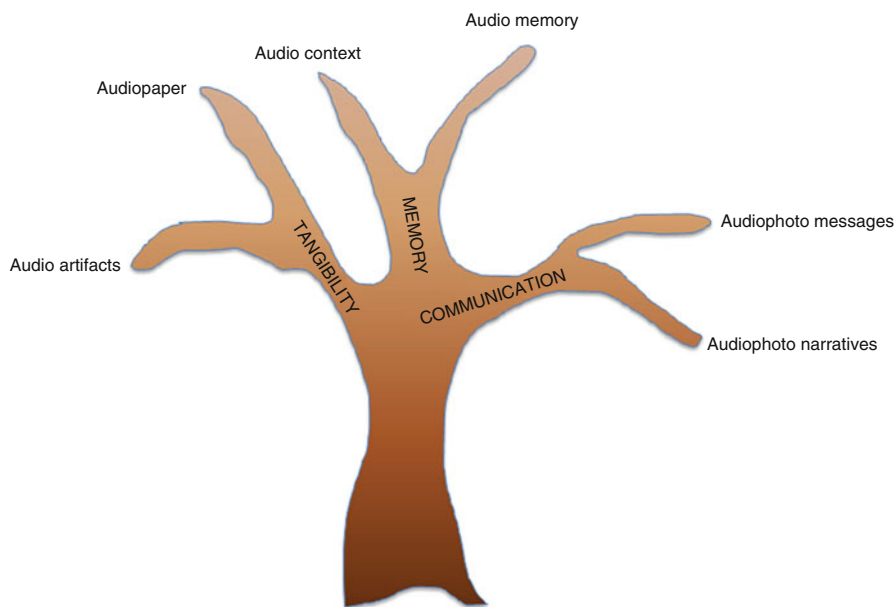

Fig. 8.1 External research developments in audiophotography

---

[1] The fact that coherent rings of wood develop together across the trunk and branches of a tree is a nice connection between the two metaphors.

Labs and IBM Research Tokyo. Eight other university groups are involved, including Aalto, Surrey, Swansea, Victoria Institute, MIT Europe, Sheffield, York, Maryland and UCSD. This is in addition to internal HP developments reviewed in Chap. 7 by HP Labs and their open innovation partners the Royal College of Art, the University of Central Lancashire and the University of the West of England. Many of the individual researchers in these groups know each other, are co-authors across the tree, and indeed have moved between these institutions over the 10 year period from 2004 to 2014. This illustrates again the personal nature of scientific research and technological design, and the way in which ideas circulate and develop within academic and industrial communities.

In the next six sections of this chapter, I review the lessons of each strand of the audiophotography tree in turn. A final section discusses the connections between them. Discussion of the implications of the findings for audiophotography and theories of innovation are reserved for Chap. 10, following a brief review of commercial developments in Chap. 9.

8.2   Audiophoto Messages

The initial context for work on audiophotography, as the name suggests, was photography. More specifically it was the process of capturing images through CCD and CMOS sensors inside digital cameras that brought everyday photographs into the digital domain. This laid the foundation for combining photographs with other digital media such as sound, and for considering the impact of sound on photography itself. However, a second step in the digital photography revolution took place in parallel to this work around 2001–2002, with the embedding of digital cameras inside mobile phones. The resulting 'camera phones' were seen as communication devices (phones) first and foremost, with images as an exciting new way of illustrating conversations or text messages with pictures (c.f. Sarvas and Frohlich 2011). Audio capture was not supported on early versions. Furthermore picture quality was low and did not rival that of contemporary digital cameras of the day. In this context, Koskinen and colleagues (2002, p. 95) referenced our early audiophoto work (reference 2 above) in a design recommendation to incorporate spoken narratives with pictures. This was based on findings from a small-scale trial in which participants sent sequences of picture-and-text messages to each other, telling true or fictional stories over time.

Audio was subsequently incorporated into an early Multimedia Message Service (MMS) supported by Radiolinja, the Finnish mobile phone operator. This allowed users to combine photographs, text and audio clips in multimedia messages that could be sent directly to other camera phones. The service was tested in a further study led by Koskinen in which 25 participants used a Nokia 7650 with integral camera. Subsequent write-ups referenced the 2004 Audiophotography book and explicitly considered the role of sound in 'mobile multimedia' communication (Koskinen 2005, 2007; Koskinen and Battarbee 2006). The uptake of MMS

messaging in the trial, and subsequently the market, was somewhat disappointing. Participants took many photographs to share face-to face with those around them or with others later on. When photographs were sent to others on the phone, they usually incorporated text annotations rather than audio. Text was used in two genres of picture messages referred to as *postcards* or *stories* (Koskinen 2007, Chap. 4). Postcards were simple image + text combinations containing a greeting and some description of the surroundings or context for the photograph. Stories were sequences of postcard-style messages designed to reveal an unfolding experience or storyline over time.

About 10–13 % of messages contained audio, depending on the way they were counted and analysed in different publications (c.f. Koskinen 2007, Chap. 5; Koskinen and Battarbee 2006). Audio was generally used to enhance an image-and-text combination assembled by default through the interface. For this reason it seldom carried the main narrative content of the message but rather embellished it through greetings, singing, laughter, cries or other ambient sounds. Elements of sound could be recorded intentionally or unintentionally by the sender, and recipients appeared to make use of both to imagine the context and location of the message and sender. Various reasons were given for the modest use of sound in relation to text, including the lack of familiarity compared to texting, the mundane nature of images compared to conventional snapshots, and the lack of adequate support for sound management and editing.

These findings on multimedia communication have carried through to the present day in many ways, with text remaining the dominant method of annotating images now shared over the web via social networking sites. In the next section we see that this is culturally specific, and could change in communities with low levels of textual literacy and stronger oral storytelling traditions.

8.3 Audiophoto Narratives

The storied nature of some picture messages was also evident in early audiophoto recordings, such as the surfing narrative of Fig. 1.2 in the original Audiophotography book. This contained a spoof BBC sports commentary by a father while taking a sequence of pictures of his son trying to surf. Quite independently, this kind of audiophoto narrative was also under development at the time by Joe Lambert and colleagues in California, within the digital storytelling movement (Lambert 2002). This emerged from the digitally mediated storytelling performances of Dana Atchley and subsequent community film-making workshops designed to allow ordinary people to tell their personal stories in pictures and sound. Typically, stories would be up to 2 min long comprising a sequence of still images with voiceover, music and ambient sound. They would be assembled in Adobe Premiere on desktop Macintosh computers with expert training and facilitation in film writing and production. Because of the community context in which this was done, the activity was elevated above the level of personal photography to capture the lives and

experiences of communities, and has been applied in numerous forms of community empowerment and activism ever since (e.g. Hartley and McWilliam 2009; Lambert 2013).

In 2006 at the University of Surrey, I was given the opportunity with new colleagues across the UK and beyond, to bring together the audiophoto work with the digital storytelling approach. We did this in an ICT for development project in India. The project was called *StoryBank* and aimed to explore the role of 'mobile digital storytelling' for information exchange within rural developing communities. This addressed a problem of low literacy rates in such regions which was impeding the use of ICT and the web. It did this by allowing people to express news stories and information in audiophoto narratives on the new generation of camera phones that were exploding into the emerging markets of that time (Frohlich and Jones 2008). Situating ourselves within Budikote village in southern India, we studied the success of a local community radio station in order to develop a form of radio-like content with pictures that could be created quickly and easily on a mobile phone. This could be shared using Bluetooth (without internet connection) with other phones or with a public display on the balcony of the village ICT centre. The system was a success in encouraging wider participation in content creation and a virtuous circle of production and consumption within different segments of the community (Frohlich et al. 2009a, b). The approach went beyond conventional digital storytelling by using audiophoto capture techniques on a mobile device rather than filmmaking software on desktop computers. It also went beyond early audiophotography, by developing new techniques to support audiovisual narratives within a new cultural context. Unlike the findings of early MMS use in Finland, voiceover and sound recording was embraced with photography by rural Indian participants, as an inclusive new form of cultural expression in their local language. The fact that we used local Bluetooth communication, also removed the cost of sending MMS messages, albeit at the expense of remote story sharing.

The generalizability of mobile digital storytelling was tested in a follow-up project, carried out in the Mankosi region of eastern South Africa from August 2010. This aimed to explore audiovisual information sharing over wider distances and in a different African culture. It also committed to develop and release an open source toolkit by the end of the project 2 years later. I was co-investigator on the project led by Matt Jones from the University of Swansea, with the Universities of Glasgow, Cape Town and the Meraka Institute involved. The new dynamic of this project was the sparsity of electricity and internet connectivity across the region, and the lack of affordability of mobile internet tariffs. We therefore designed a more robust version of the StoryBank system to be geographically based around solar powered charging stations and community story repositories (Bidwell et al. 2014). As part of this project we designed a new audiophoto narrative app for Android devices called *Com-Phone*, itself an element of the *Com-Me* (community media) toolkit: http://digitaleconomytoolkit.org/com-phone/ This allowed users to string together audiophoto frames with up to three layers of sound, a photograph, and a piece of text on each frame. Any or all of these media items could be inserted into each frame. The interface is shown in Fig. 8.2. Text annotation was also supported because of the
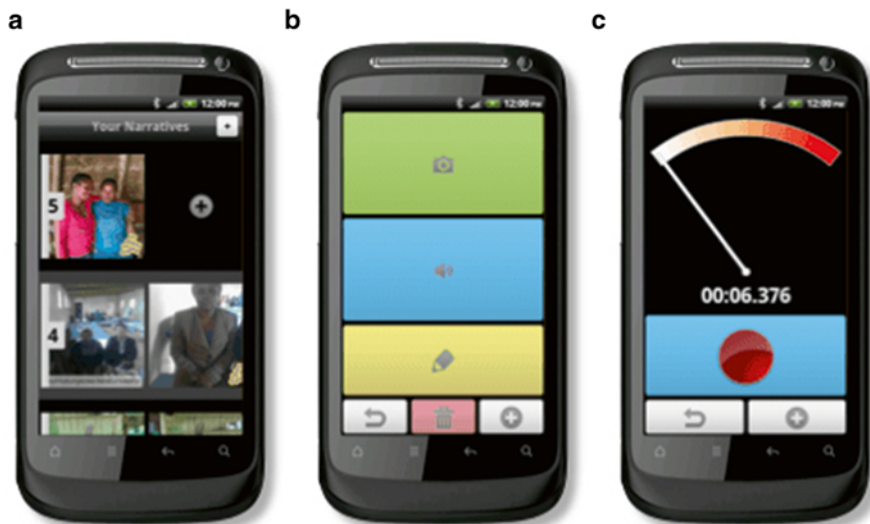
Fig. 8.2 The Com-Phone interface. (**a**) Home screen (**b**) New frame (**c**) Sound recording

higher literacy rates in this region, but was seldom used in practice (Frohlich et al. 2012). The resulting digital stories or narratives could be sent via Bluetooth to nearby devices, but also emailed or uploaded to social media sites like YouTube in wi fi zones or over 3G.

From an audiophotography perspective, the above projects essentially re-imagined the audiocamera as a mobile digital storytelling app, designed for a smart-phone with intermittent connection to the web. Similar apps have since been developed for Apple and Microsoft devices and will be reviewed in Chap. 9. While most of these are designed for youth or working adults in a Western context, one app is designed more inclusively for children and older adults, and has more in common with Com-Phone. This a called *StoryKit* and is described in a research paper by Quinn et al. (2009) from the University of Maryland who cite the original audiphoto work as well as StoryBank. The app runs on an iPhone and was co-designed by children and older adults working with researchers to support the creation of multi-media storybooks. The interface is shown in Fig. 8.3 and allows the creation of complex pages of a book comprising (a) live or (b) stored photographs (c) text (d) audio and (e) drawings. The paper is the story of the design journey rather than the effectiveness of the tool itself, and proposes design guidelines for similar apps. These include recommendations about re-sizing content, attaching sound, adding text and sharing privately.

A subsequent test of StoryKit involved seven intergenerational pairs of users on a day out in a national park, the families of two pairs using it for a further 7 weeks. It also involved an analysis of 245 anonymous stories shared through an on-line portal by people downloading the StoryKit from the app store on iTunes (Bonsignore

Fig. 8.3 The StoryKit page interface (Bonsignore 2010). Letters are explained in the text (Used with permission)

2010). The app was found to be fun to use by children and adults alike, and the mobile device was often passed between them for different contributions to stories. For example children would often record spoken narrative while adults preferred to type. Children seemed to especially enjoy recording sound effects and drawing. Some adults marveled over how it gave the children a new method of expression and a tool for learning and reflection following an experience. Story analysis showed a great variety of forms, from a one page story with two audio clips to a 22 page photo-narrative with 37 photos, one drawing and one sound recording. Audio was used in 40 % of the stories and often accompanied photographs and text (see also Bonsignore 2013 for an expanded analysis with additional data). Again these findings suggest that narrative audiophotography is a creative new medium that may have more attraction outside a typical Western adult audience, in this case for childhood education and intergenerational collaboration.

8.4   Audio Context for Photography

In the sections above we have seen how the use of spoken voiceover with single or multiple photographs can lead to audiophoto messages or narratives of different kinds. In this section, I review three explorations of the use of ambient audio recorded around the photo event. Two of these develop the artistic and aesthetic potential of the audiphoto medium in different directions, while a third uses it in a very practical way to make photography more accessible to people with visual impairment. I begin with the audiophoto arts-based studies.

In a project called RAW, a team from MIT Media Lab Europe developed the audio-photography approach into a distinct form of documentary filmmaking for capturing 'audiovisual impressions of everyday life' (Bitton et al. 2004). Inspired by the mundane nature of scenes within the documentary *Sans Soleil* (1982) by Chris Marker, and the need to promote greater cultural understanding, the team created a new audio-camera system. This recorded 1 min of high quality spatial audio before and after each photograph was taken, through binaural in-ear microphones and software which selected segments from a continuous sound recording. If photos were taken less than 1 min apart, an audiophoto narrative resulted with continuous sound starting and ending 1 min before and after the photo sequence. An additional screen-based playback feature was designed, based on user-selected points of interest in each photograph. These were used as locations from which the image would zoom out to reveal itself up to the photo capture point, before fading out to the end of the audio clip or next photograph. Audiences could also skip to the next photo at will. This system was evolved through user engagements in Ireland, France and Mali, where various user groups were given the chance to record and review the resulting material.

User reactions differed between locations but were generally curious and positive in equal measure. Feedback in Ireland led to the animated and interactive playback scheme, presumably because staring at a static image for 2 min is frustrating. This was about four times as long as the average duration of an audiophotograph in my own work, and even then I found users often skip on before the end of each clip (Frohlich 2004). Feedback from secondary school children in France, revealed idiosyncratic uses for capturing impressions of things inside a home, architectural features outside, or people on the street. Further uses were revealed in a larger study in Mali with 23 participants ranging from 10 to 'over 50', recruited opportunistically across three locations. Analysis of recorded material indicated four distinct uses of the system as described below. Local audience reaction to the audiophotos in a subsequent exhibition reflected satisfaction in what was captured, but some confusion as to the motivations of the researchers.

- *Social glances* – recordings used as a means to strike up conversations with other people
- *Caught in activities* – the capture of everyday activities and skills such a collecting water or performing music
- *Active documentation* – explanation of an activity or issue with narration and/or interview recording

- *Intentional discourses* – direct political messages or commentary to a foreign audience with photo illustration

Three papers from researchers at the Viktoria Institute cite the earliest audiophotography reference (1) in a series of studies in *'context photography'* (Ljungblad et al. 2004; Håkansson et al. 2006; Ljungblad 2007). The first and third papers concentrate on the methodology used in developing the approach and especially the inspiration drawn from *lomography* (Ljungblad et al. 2004, 2007). This is an amateur photography practice using low cost Russian analogue cameras to create particular visual effects. Lomographers 'shoot from the hip' without using a viewfinder and this results in strangely framed and exposed images with a distinctively spontaneous aesthetic. Context photography also attempts to create a new photographic genre by applying image filters based on aspects of the surrounding context. In practice, sound and movement were used in various ways to change the visual appearance of a photograph, as described in Hakansson et al. (2006). Four visual effects were supported on Nokia 660 and 6630 camera phones, through software analyzing the frequency and loudness of ongoing sounds and the amount and location of ongoing movement in relation to the photo capture 'click'. One of these transformations called 'Wave' is shown in Fig. 8.4. In this case, movement creates waves in the image, making it look like a dense liquid; the size of the pixels is proportional to the surrounding sound level.

A user trial with seven Swedish participants ranging in age from 23 to 49 years generated 303 context photographs in total (Håkansson et al. 2006). This revealed interest in the creative possibilities of using movement and sound context to influence photography, but also some frustration in the lack of control and the resulting

Fig. 8.4 A context picture showing distortions due to movement and surrounding sound levels in an office environment (Hakansson et al. 2006) (Used with permission)

images. As in lomography, the context photographers were often surprised by the appearance of images and sought ways of influencing them to create certain effects. Here, this included ways of moving the camera itself as in spinning or panning it or taking photos from a moving car or train, and seeking out noisy environments or screaming whilst the photo was taken. Consequently, the technology affected the kind of photographs taken, rather than reflecting the naturalistic context for conventional camera phone images. Users wanted more control over the calibration of sensor data and its effects, and more subtle influence so that images were not rendered unrecognizable by those effects.

Finally, Harada and colleagues at IBM Research Tokyo and UCSD draw on audiophotography to develop a new kind of camera system for the blind (Harada et al. 2013). Following an observation that many visually impaired people continue to take or share photographs with sighted family and friends, they consider the use of both ambient sound and voice annotation for enhancing and managing photographs for this population. In their '*Accessible photo album'* app for the iPhone they support ongoing capture of ambient sound which can be restarted with a double-tap gesture on the iPhone screen, and stopped on photo capture through a downward swipe gesture. Post capture spoken labels can be recorded on each photo by a tap-and-hold action which stops on release. In this way all photographs can have associated ambient and voice label audio for browsing and playback. In browsing mode, the user can swipe through photo lists, listening to key metadata read out on presentation. On full display the app plays the voice label for a photo followed by the ambient sound. A small scale trial of the system with 5 visually impaired users showed great enthusiasm for the approach in encouraging more photographic capture and better sharing, due to the simple gestural control interface and additional audio information. Voice labels were practically useful for identifying and selecting photos to share while ambient sounds were valuable for conveying atmosphere and additional information to photographers and audiences alike. There was some discussion of **when** each type of sound might be recorded around the image capture event, since the auto-recording of ambient sound before the click often contained set-up talk between the user and their companion. Some users kept this as the voice label, while others wanted more naturalistic sound kept separately. Furthermore, not all users wanted to play voice labels to audiences at sharing time. Notwithstanding these design issues, the approach was highly successful and the authors discuss ways of further combining photographs, sound, video and metadata in the future, and standardizing data formats for sharing.

8.5 Audio Memory

A major finding of the original audiophoto work was that people have a sonic as well as a visual memory for events. That is, they are able to remember the sounds of particular places, events and periods of their lives as well as their visual characteristics. This is not so surprising given that hearing is probably the second major sense

after sight for perception of the world. But we are not so used to the idea of capturing sounds as reminders of the past as we are with the idea of capturing photographs. This led me to coin the term 'audiographs' in the 2004 book (page 2) to refer to the equivalent of a photograph in sound alone, as a way of triggering sonic memory for events.

This notion has been developed in subsequent research citing the audiophotography book. A fascinating introduction to this comes from a serendipitous set of findings across two studies carried out by Bentley and colleagues at Motorola Labs. They noticed similarities in the properties and use of photographs and music emerging from separate studies of each medium carried out between 2002 and 2005 (Bentley et al. 2006). Some of these similarities were in the organization and retrieval of photo and music content. For example, people struggled with an intention to organize both media better, sometimes finding the time to classify, label and file items and sometimes not. This interacted with their ability to browse and find items later, which was typified by a kind of satisficing behavior to settle with something similar enough to what they had in mind. In browsing through both photos and music, people reported getting sidetracked into reviewing material discovered by accident that they hadn't seen or heard for a long time. This reminded them of forgotten people and events and led to protracted stretches of private reminiscing. The authors report strong similarities between the way photographs and music appear to be linked to past events in people's minds, and a propensity to share and talk about this. For example, particular songs were associated with people who had given them the music or listened to it with them, and musical preferences were found to be highly idiosyncratic and indicative of a person's personality. The paper concludes with recommendations for integrating music and photo collections and supporting common activities of managing, combining and sharing media relating to particular people or events.

Two further explorations of the sentimental properties of sound citing the audio-photo work were carried out in my own group at the University of Surrey in collaboration with Microsoft Research Cambridge, and at the University of Sheffield. Both investigations involve multiple studies and explore the role of ambient sound in memory.

In the first published study of the domestic soundscape, we found that certain naturally occurring sounds in the home are evocative of family life and would be worth saving (Oleksik et al. 2008). This was discovered almost by accident through a methodology that asked families to record certain sounds in the home for later discussion. Those recorded in the categories of 'favourite sounds', 'sounds that reminded them of home', and 'sounds they would like to keep' turned out to be particularly attractive and many families asked for a copy of the recordings. We referred to these as *sonic gems* because they stood out from the majority of mundane and forgettable domestic sounds as precious discoveries. They included sounds such as children talking while at play or breathing while asleep, the cat purring, and conversation around a meal. This led to a design proposal for a sonic gem recorder in the form of a wearable recording device on a pendant, and a sonic gem bowl for playing back sounds from physical gems broken off from the pendant (see Fig. 8.5).

Fig. 8.5 The sonic gem recorder and bowl (Oleksik et al. 2008) (Used with permission)
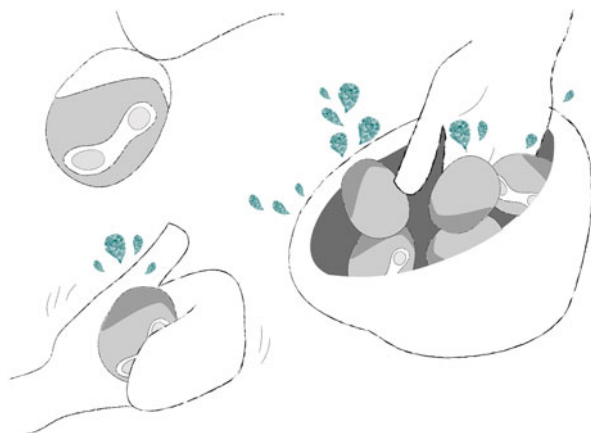
In a follow up study on sonic gems, Oleksik and Brown (2008) at Microsoft Research Cambridge carried out a trial of sentimental sound recordings and gauged reactions to playback from a prototype sonic gem and bowl set. Four families with children under the age of 6 were interviewed at home about their existing sound recording practices before being asked to capture sonic gems on a digital Dictaphone. Some of the recordings were edited and associated with physical 'gems' in the form of plastic eggs containing RFID tags. Families were interviewed in a second home visit about the captured sounds, and gave their reactions to playback from a sonic bowl with integral RFID reader. This played edited sound clips from each gem as it was removed from the bowl.

The findings from phase 1 of the study revealed that families all had prior sound recording experiences but that these had declined with the replacement of audio cassette tapes by CD and MP3 technology, and answering machines by voicemail services. Practices of recording music from the radio had sometimes spilled over into accidental or deliberate ambient sound recording, and many participants reported having ad hoc collections of precious sound recordings in their homes. These included children having fun with audio cassette recording, compilation music tapes, and voice messages on answerphone tapes that had become precious with the death of speakers. Families could not always play back these sounds from old technologies, and in some cases the recordings had been lost or accidentally thrown away.

In phase 2 of the study families were fascinated with the challenge of capturing precious sounds and responded in a variety of creative ways. They recorded children and baby sounds, people singing, conversations, indoor and outdoor ambient sounds, and music. Each of these types appeared to have different sentimental properties and values, but all conjured up memories as they listened back to them with researchers. In many cases the sounds triggered vivid mental images of associated people, events and places, and this led to a discussion of the relationship with photos and video. Where audiences were present at the time the recording was taken, unaccompanied audio was felt to be more immersive and evocative, giving room to the

imagination in the way imagery did not. Playback from the sonic gems and bowl was liked in this regard, especially where audio clips had been shortened to play key phrases rather than long rambling episodes. The paper concluded with recommendations for larger scale studies and for the design of new technology for supporting digital sound capture and playback, with and without other media.

This design challenge was taken up by other researchers at Microsoft Research Cambridge and the Technical University of Eindhoven (Helmes et al. 2009). Taking inspiration from the sonic gems work and other research on lifelogging (e.g. Gemmell et al. 2006), the authors designed and tested *The Other Brother*. This was a semi-autonomous device for the home that captured images and video around sonic events. In fact the device used twin microphones to identify the source of the sound, towards which the device could then turn. A first prototype was tested with three families over Christmas 2008. Although some surprising moments of family life were caught on camera so to speak, families were reluctant to interact directly with the device and somewhat confused about its recording behaviour. A second prototype gave more feedback on its behaviour and invited more interaction and manual control over its operation (see Fig. 8.6). In addition, it was given audiophoto capabilities in the form of recording 10 s of sound around each photograph taken. The resulting sound content was seen as a surprise bonus to still photographs in a second trial with two families. A final prototype retained audiophoto and video capture, removed manual control, and refined the way sound deviations triggered capture. Further testing and development appears to have stopped at this point, but the approach demonstrates some new possibilities for automatic capture of audiophotos and video clips in everyday life.

A final pair of papers describe a new study of family sound recording and the design and testing of new technology to support this. Researchers at the University of Sheffield collaborated with a visiting anthropologist from Rice University to conduct a study of *sonic souvenirs*, defined as sound-only mementos of a family holiday (Dib et al. 2010). These contrast with sonic gems in relating to specific trips

Fig. 8.6 The Other Brother prototype 2 (Helmes et al. 2009) (Used with permission)

away from the family home, where sounds are likely to be less familiar and recorded from a wider diversity of outdoor as well as indoor environments. In fact the study was a replication of the early audiophoto trial reported in references 1, 2 and 3 above, using Dictaphones rather than audiocamera devices. Hence, 10 families with children aged 7–15 were given 30 dictaphones over the summer of 2008 to document 3 days of their holidays in sound alone. They were later interviewed about the captured sounds at home.

Families greatly enjoyed the capture exercise and listening back to the sounds together. In fact, group listening stimulated conversation and speculation about the sounds being played, which itself led to better remembering and spontaneous storytelling in front of the researchers. A great variety of sounds were recorded, including live commentary, distinctive sounds of the areas being visited, and family conversations. As in the Oleksik and Brown (2008) study, families were often surprised by the detail recorded in the sound clips and enjoyed the 'imaginative freedom' of listening back at the pace of the recording itself without accompanying images. In addition, families in this study found the sound clips to be more candid and less flattering than photographs, because they took place over longer stretches of time and were less controllable. This sometimes led to the recording of arguments, crying, and shouting with children, which were either seen as funny in hindsight or deleted before the researchers returned. In other cases, participants used the recording medium as an opportunity for playful performances of various kinds, such as interviewing fellow travellers or speaking to an imagined audience in mock documentary style. This led the authors to discuss differences between sounds and pictures for recording the past, and to recommend more accessible and creative technology for sound management and editing.

One such technology was later developed by the authors themselves, in collaboration with Nicolas Villar from Microsoft Research Cambridge. This was called 'Family Memory Radio' or *FM Radio* for short (Petrelli et al. 2010). The device re-purposed an old Roberts R707 analogue radio with volume and tuning dials and channel selection buttons. It's analogue speakers and interfaces were preserved but re-arranged and driven by hidden digital hardware comprising a CPU, audio chip, USB port and a display for channel labels. Sound clips organized into four folders on a USB stick were accessed from one of four channel buttons and played sequentially in chronological order at a specified volume level (left control knob). Users could also rewind or advance the playback/tracks by rotating the tuning dial anticlockwise or clockwise respectively (right control knob). Channel classification became a subject of discussion, but was set by default to *Time travel* (chronological), *Ambient*, *Voices* and *Favourites*. Pressing a fifth button, marked the current track as a favourite and put it in the Favourites playlist (see Fig. 8.7).

The prototype was taken back to 6 of the original 10 families taking part in the Dib et al. (2010) study, 1 year later in the summer of 2009. Families listened and reminisced to the original sounds which many had forgotten due to lack of access in the intervening period. The radio interface was generally appealing and served to support more convenient shared listening and trigger discussion of future use. Members of the family had different suggestions for channel types but generally

Fig. 8.7 FM Radio prototype (Petrelli et al. 2010) (Used with permission)

agreed with the model of creating and managing these from a computer. A recurring suggestion was for sounds to serve as background to photo sharing activities, with some disagreement as to whether this should be synchronized directly to photo display. This opens up a new design space for more loosely connected audiophoto materials, designed to evoke past memories and emotions in more poetic combinations for a group setting.

8.6   Audiopaper

One of the most novel aspects of the original audiophotography vision was the possibility of playing back sounds from printed photographs or *audioprints*. This led to a number of granted patents on what might be called *audiopaper* for either embedding sound within a photograph or otherwise identifying it so that associated sound can be retrieved from elsewhere. This vision was entirely consistent with the HP context of the original work, since HP were market leaders in printers and could have introduced augmented paper into their printer and consumables range. It was also consistent with early studies of digital photography in which consumers expressed reluctance to move away from the tangibility of printed photographs to screen-based ones. Given the subsequent transition to screen-based photography on smartphones and tablets, at least in the more technologically saturated parts of the world, we might have expected subsequent work on audiopaper outside HP to decline. However, three university research groups as well as my own at Surrey have continued to explore the field, citing connections to the original audioprint work.

These include York, Swansea and the University of California San Diego (UCSD) (see again Fig. 8.1).

The first exploration of audiopaper after 2004 was a user study of the audiophoto desk by collaborators at York. Three groups of four friends created audioprints to share on the desk in a living lab setting called the York Responsive Home (Lindley and Monk 2005). About 8 existing photographs per person were annotated with sound effects, music or voiceover in the knowledge that they would be shared with three particular other people. The selection of audio for this purpose was analysed along with the first 8 min of group photo sharing on the desk, in order to understand the function of audio in reminiscing and storytelling and its influence on the resulting conversation.

Sound did appear to add atmosphere, fun and humour to the activity, but at the cost of authoring associations deliberately for the audience at hand. Music was the most effective category of sound leading to the longest clips, the shortest duration of silent listening, and the longest surrounding conversation. Participants often recognized the track within about 3 s and could begin talking about the photo or its music, while the music continued to play through for about 40 s. In contrast, voiceover clips varied greatly in length and tended to inhibit conversation for an average of 7 s until completed. They then resulted in slightly longer turns at talk than ambient sound photographs, perhaps due to their meaningful content (an average of 37 rather than 33 s). Some delay in the responsiveness of the desk in playing or pausing sounds, together with the tabletop configuration, gave the sessions a more unpredictable and formal character than passing around printed photographs from person to person. This led to recommendations for faster and more reliable audio association and playback with photographs, perhaps from a handheld device.

A further study of the audiophoto desk as a public art installation was carried out by myself and other colleagues from York, originally in Glasgow (Blythe et al. 2008). A visual artist and product designer[2] collaborated to create audiophoto stories of the Govan shipyard area of Glasgow, for public display in the city. The installation was called 'Weegie' which is a slang term for Glasgow residents (Glaswegians) but also a reference to a kind of ouija board experience we tried to create of past voices and sounds from the history of the area. The desk was re-designed for this purpose as a circular table enclosed in a tent-like space (see Fig. 8.8). This allowed users to gather around it from any angle and begin to combine images and sounds from six distinct stories sets, comprising three audioprints each. Individual sets contained ambient, voiceover and musical audiophotos describing past and present life in Govan. Sound volume was controlled by moving photographs between the edge and the centre of the desk, with stereo panning moving to the left or right side of the room as usual.

Exhibition of the desk and its material was cut short in the Glasgow LUV gallery by a technical problem caused by one particular photograph, itself depicting several framed photographs. The exhibition was later restaged in the York Responsive Home and visitors were interviewed about their experience. In addition, a theatre

---

[2] Esme McLeod and Mil Stricevic of Pavillion Design Glagow.

Fig. 8.8 Initial sketch of the 'Weegie' installation by Mil Stricevic (Figure 3 in Blythe et al. (2008) reprinted with permission)

reviewer, film and radio reviewer and comic book reviewer were all invited to write reviews of the exhibition itself. The diversity of audience and reviewer reactions were contrasted with conventional lab-based evaluations of interactive prototypes, such as the Lindley and Monk (2005) study above. Word clouds from audience interviews showed subtle shifts in the perception of Glasgow before and after exposure to the exhibition, while critics discussed the aesthetics of the piece as a whole, the 'magic' of the technology, and its relation to examples and genres of art in other media. The broadly positive feedback served to validate the desk and approach as a new form of artistic expression, and raised questions about the role of interaction critics in interaction design.

An independent line of research by Ann-Marie Piper and colleagues at UCSD examined the role of audiopaper photographs for encouraging social interaction with older people (Piper et al. 2013, 2014). In a detailed case study of an extended family, the authors used the Livescribe digital pen and Anoto paper to augment printed photographs with custom audio recordings. For this they used the TAP & PLAY authoring toolkit created on another project (Piper et al. 2012) to allow users to mark regions of photographs or surrounding areas and associate them with recorded voiceover or other sounds. Sounds could be played back from a speaker in the pen, simply by tapping on active regions in the album. In the study, a printed audiophoto album was created by relatives of Ethel for her 105th birthday, and subsequently enhanced and reviewed by Ethel and her family and friends over a 5 month period. The findings revealed a strong interest in the technology, together with tangible improvements in Ethel's memory for relatives and shared events

through repeated playback of names and stories from pictures. The album also brought disparate members of the family together in its creation and sharing, and gave Ethel a new focus for interaction with carers and other residents in her home.

In quite a different context, Robinson and colleagues at Swansea University built and tested the *AudioCanvas* system for playing back voice annotations from printed materials in an Indian village and South African township (Robinson et al. 2014). This used a voice service on mobile phone to read aloud aspects of the text or play additional instructions and explanations in local language, for participants with low levels of literacy in the printed language. Voice snippets were invoked by touching regions of a *snapshot* of the printed material, which contained QR codes in two corners to identify and register the page boundaries.

A sample of pre-prepared items including a phone bill and map in India, and a newspaper and flier in South Africa, were tested by local people to great acclaim. Despite some problems of framing snapshots and locating active regions on the phone, 58 participants praised the concept and gave it an average usefulness rating of over 8 out of 10. In subsequent work the researchers have developed an authoring toolkit called *PaperChains* (Pearson et al. 2015). This works like the TAP & PLAY toolkit, in allowing audio recordings to be linked to arbitrary regions of printed material for later playback. However, the main focus of *PaperChains* appears to be collaborative storytelling through hand drawn sketches and voice annotation. This application of audiopaper moves away from audio-photo material to illustrate the value of other tangible image and sound combinations for developing communities.

8.7 Audio Artifacts

Two parallel investigations from researchers based mainly at the University of Sheffield and Microsoft Research Cambridge explored the stories people tell about cherished objects in the home. Both cite the audiophoto book as highlighting photographs as a class of cherished objects, but also the related HP research by myself, Rachel Murphy and Jacqueline Fennel (Frohlich and Murphy 2000; Frohlich and Fennell 2007). This is described in Chap. 7 and essentially shows that objects themselves can function as memory triggers in the same way as photographs, and can be augmented with sound or other media to enhance remembering (see again Figs. 7.6 and 7.7 on the *Memory Box*, *Memory Shelf* and *Best China* design concepts). The overarching practice of using physical or digital artefacts as icons of experience and identity, was referred to as a kind of *domestic iconography* in Frohlich and Fennel (2007). The new lines of investigation begin to show how this works for objects.

Petrelli, Whittaker and colleagues from Sheffield approached this question from the perspective of what they call *mementos*: 'an object given or deliberately kept as a reminder of a person, place or event' (Petrelli et al. 2008). They carried out a single study of mementos kept by 12 families with children in the north of England to see how they served to map the memories and identities of individual family members amongst the ecology of other objects within the home. The methodology

involved taking a tour of the home with each participating family member, and asking them to identify public, family and personal rooms in the house and select at least three objects from each type of room that they considered 'special'. Participants explained their choices and reasoning as they went along and were asked further questions about special digital objects in a final phase of the home visits. Three related papers documented different analyses of the data, focusing on physical mementos (Petrelli et al. 2008), the contrast between physical and digital mementos (Petrelli and Whittaker 2010), and photo mementos (Petrelli et al. 2014). While a full summary of this work is beyond the scope of this review, I will now highlight some significant findings on the relationships between photos and other objects as memory triggers, and the contrast between physical and digital mementos.

One of the most striking findings to emerge from the study was the sheer diversity of objects that came to have special meaning to people, and the idiosyncratic nature of those meanings. These included professional and self-made artwork, everyday objects like mugs, clocks, furniture, books and ornaments, functional and non-functional objects which are seldom used such as old children's toys, diaries, receipts, jewellery and shells. The meaning of these objects related to their origin and history with participants, and often represented significant relationships with other key people, or periods of self-discovery, achievement and happiness. The findings were reminiscent of those from a similar study of Chicago families carried out in the 1970s, and demonstrate the importance of sentimental personal possessions in everyday life (Csikszentmuhalyi and Rocheberg-Halton 1981).

Printed photographs around the home made up about 16 % of selected mementos, and themselves took a variety of forms. These included framed photos placed on surfaces or mounted on the wall, as well as loose photos arranged in albums, pinned to boards, carefully made into collages, or stored in shoeboxes and other containers. Significantly, only one participant chose a digital object unprompted in the original tour of the home, revealing the hidden nature of most digital photos, video and other materials. Digital mementos were greatly appreciated according to the prompted discussions, but did not lend themselves to the same kind of distribution and display around the home. This led the authors to propose new technologies for integrating the digital with the physical, as in a *MemoryBook* for storing electronic snippets of media with physical trinkets, or a *ProjectoFrame* for projecting a collage of images on the wall from a physical photo display on a table (see Fig. 8.9). The creative act of assembling and crafting artefacts for remembering was taken further in a related study, also citing the audiophoto work. Petrelli et al. (2009) asked 10 families to make time capsules of their lives to be viewed 25 years in the future. Although this was a departure from the analysis of naturally occurring mementos in the home, the results showed a similar preference for physical artefacts which were actively selected or created to be expressive and personal.

Sellen, Kirk and colleagues at Microsoft Research approached the question of domestic iconography from the slightly different perspective of home archiving (Kirk and Sellen 2010). Skeptical of the argument for life logging technologies to record human experiences in ever increasing detail, the authors look to existing practices of collecting objects and capturing media in the home for clues about the

Fig. 8.9 ProjectoFrame prototype (Petrelli et al. 2014) (Reprinted with permission)

role of such artefacts in everyday life. They also apply these findings to the design of new home-based archiving systems. Taking a more sociological view than Petrelli and colleagues, they explored how and why families keep sentimental objects and media in their homes. A similar methodology was used with 11 UK families at different stages of family life, invited to discuss special objects during a home tour. This time participants were asked to cover every room in the house and mention any items of sentimental value, which were photographed for detailed discussion in a second home visit. They were also prompted to identify special digital artefacts around key electronic devices such as the home computer, and given time in the follow-up visit to reflect explicitly on the differences between their physical and digital collections.

Analysis of the types of artefacts discussed, revealed two new categories lying between three dimensional physical objects and digital media. These included so called '$2^1/_2D$' items such as paper documents, letters, certificates, cards and cuttings, and '*hybrid*' (analogue) media objects such as vinyl records, audio cassettes, CDs and DVDs. Furthermore, a classification of the function of objects across all categories showed six core values of home archiving:

- **Defining the self** – items reflecting the personality and history of individual members of the family (e.g. a Bay City Rollers book or old university work)

- **Honouring those we care about** – items displayed to celebrate the achievements or identity of significant others in the family (e.g. pictures of grandchildren, artwork by children)
- **Connecting with the past** – items demonstrating a link to past events or people (e.g. a grandmother's jam recipe book, a stone from a beach)
- **Framing the family** – items showing what kind of family lives in the home (e.g. artwork in public rooms, family photographs)
- **Fulfilling duty** – items important to family roles (e.g. adoptive children's toys, photographs of distant or deceased relatives)
- **Forgetting** – upsetting items too precious to throw away (e.g. boxes of old letters, old email messages)

Drawing on these findings and their implications for design, Kirk et al. (2010) and Sellen (2011) report the design and use of a *Family Archive* device which aims to allow families to integrate their physical and digital artifacts in the digital realm. In effect they implemented and tested an earlier design concept by Stevens et al. (2003) for a *Living Memory Box* through which to scan and store photographs of 3D and $2^1/_2$D physical objects alongside digitized and digital media. The *Family Archive* was implemented as a waist-height touch-sensitive tabletop display. It also contained an overhead camera for scanning documents and objects into virtual shoeboxes which appear to be floating beneath the table surface. A 1 month trial of the system by three households showed uneven use of digital media uploading and scanning functions by different family members, with mixed reaction to scanning. Some items such as sweet wrappers didn't scan well, producing disappointment, while others such as a 6 year old boy's dinosaur collection did, producing delight. Serious archiving of precious objects was accompanied by playful scanning of unusual objects such as one family's new kittens. The system raised questions about gender roles for archiving and long term compatibility with other devices such as the home PC. Reflecting on the lessons of this research, Sellen (2011) pointed out the importance of iterative development and testing of domestic technology, with attention to family dynamics and relationships. She also went on to promote physical augmentation as an alternative approach to digitizing physical objects. This involves augmenting physical objects with digital data, as in the *Memory Box* (Frohlich and Murphy 2000). Shoebox is one such system explored by the Microsoft group to browse photograph sets as represented in an adapted slide transparency box (Banks and Sellen 2009). Running a finger along the top of the stack displays the corresponding photo set on an integral display on the front face of the box (see Fig. 8.10). Further explorations of this kind are reported in Banks (2011).

Although audio is not explicitly represented in many of the design concepts and prototypes emerging from both strands of work above, physical artefacts are shown to be powerful triggers for memory, and symbols for other social acts within the family. As such they attract comment and conversation around them in a way which goes beyond the original audiophoto medium, into live storytelling and performance.

Fig. 8.10 The shoebox prototype (Banks and Sellen 2009). The *top* of the box lifts up to reveal a stack of pseudo-transparency cases for touch-based selection (Used with permission)

8.8   Connections Between Research Strands

In the last chapter we saw how the original audiophotography work was followed up inside HP, through research and development activities designed to maximise the impact on HP products. This sometimes resulted in the attempted transfer of new technology ideas to domains outside photography, as in demonstrations of location-based services (e.g. Hull and Reid 2003), or in the application of design ideas to new technology platforms such as web services (e.g. Rajani and Vorbau 2004). Conscious attempts to stimulate open innovation through collaborations with academic groups were only partially successful from HPs point of view. They opened up the design space for audiophoto products and services, but in a way which was difficult to map to existing HP products and plans. In this chapter, we have seen this happening in even more stark contrast to HP company interests, through independent research carried out at other universities and companies. A major finding of this review is that this external research has been focused mainly within the domain of domestic photography and its three core values of supporting memory, communication and identity (Van Dijck 2008).

This can be seen in part from the tree diagram of Fig. 8.1 summarising the six strands of new research citing the original work. A deeper understanding of the properties of sound, paper and memorabilia for memory is given by the audiograph and tangible audio research strands. Research into audiophoto messages and narratives is all about communication through audiovisual storytelling, while context photography deals with the stimuli and dynamics of experience capture, and the way it can represent the identity of individuals and communities. Many additional design and technology ideas were developed across these categories, but most deal with support for what we later called 'domestic iconography', broadly defined (Frohlich and Fennell 2007).

Elaborating further, the research generated a range of new insights relating to the role of sound and sound-image combinations, beyond those suggested in the original audiophotography work. In general these were supportive of the audiophotography vision, but served to extend it into new territories of application, markets and technologies.

For example, the audiophoto message and narrative work served to highlight the value of image sequences with accompanying sound and text, as a new form of communication on a mobile phone. This was seen to have great value in non-Western cultures with stronger oral storytelling traditions. Work on the audio context for photography took this emphasis on the sequential environment for photography further. It first showed the value of very high quality binaural recordings before and after a photograph is taken, and the need to animate images through pan and zoom effects when accompanying sound clips are long (Bitton et al. 2004). It also raised new possibilities for triggering image capture from sound events and even modifying the quality of the image itself to reflect aspects of the sonic and movement context of capture (Ljungblad et al. 2004). Audio context was also shown to be useful in making photographs more accessible for sight impaired users (Harada et al. 2013).

Further details on the properties of sound for memory emerged from research on sonic gems and souvenirs, with accompanying design ideas for how to capture and browse sound clips *without* associated photographs (e.g. Oleksik et al. 2008; Petrelli et al. 2010). Finally, new insights on the value of printed photos or other materials as interfaces to sound and memory, emerged from work on audiopaper and audioartifacts. Essentially, physical items can come to have sentimental associations for people, and can serve as accessible triggers for sonic memories and information, especially for those less familiar with modern screen-based technologies (e.g. Kirk and Sellen 2010; Piper et al. 2013; Robinson et al. 2014).

In addition to these substantive new findings on audiophotography and related practices, the research in this chapter also calls attention to the human factors of research and innovation in a scientific context. This is illustrated here by the fact that relatively few research groups are involved in developing this work, and many of the individual researchers have a history of collaboration within or between groups.

At one level this is not so surprising, since members of the scientific community from both industry and academia meet like-minded people at conferences and workshops, and read their work. They may also enter into joint research bids,

projects and PhD supervision, in areas of shared interest. What is perhaps more surprising here is the extent of these collaborations across a single research area, and the level of history that builds up between individuals and groups over time. Hence, in the 10 year span of research reviewed above, many individuals have conducted not one but several projects together as can be seen in the patterns of direct co-authorship shown in the reference list for this chapter. Often this is within groups such as the Microsoft Research Cambridge (MSRC) partnerships of Sellen/Kirk/Banks/Lindley/Taylor, the UCSD team of Piper/Weibel/Hollan, or the Swansea team of Jones/Robinson/Pearson. At other times it is between individuals at different groups such as the UTS-Sheffield collaboration between van den Hoven and Petrelli/Whittaker, or my own collaborations from Surrey with the MSRC and Swansea teams above. Furthermore, most individuals in this community have moved institution at least once between 2004 and 2014/2015, with many travelling to join others with the same interests or continuing their own research strand with new colleagues. This can be seen in the movement of Oleksik from Surrey to MSRC, of Whittaker from Sheffield to UCSC, of Lindley from York to MSRC, and so on.

The timing of these collaborations and movements is deeply significant for the development of work in the field. Most studies are not one-off pieces of research by individuals, but have a history for *each* co-author, and set up trajectories for future research. Therefore a collaboration can merge two or more viewpoints and previous research strands in exciting ways, and a job move can inject new expertise or ideas into a team. To give one example from my own experience, the sonic interventions project between Surrey and MSRC was exciting because it brought to bear Gerard Oleksik's sociological approach to domestic soundscapes together with MSRCs interests and expertise in domestic technology design; resulting in a raft of new design concepts and patent filings. At the end of his contract for this work, Gerard then moved to MSRC to carry out a follow-up study of one of these concepts, Sonic Gems, with Lorna Brown. Again the combination of a sociologist and computer scientist working together on a new media innovation was a great recipe for unique work which could not have been done by either party alone. This work in turn triggered a response and further development of ideas on 'sonic souvenirs' by the Sheffield team of Dib/Petrelli/Whittaker within a couple of years.

The fact that many of these lines of research can pass between groups in this way, and pass back again further down the line, leads to a kind of scientific arms race. However, unlike weapons development, scientific advances are shared openly in annual conferences and publications, fueling faster turnover of studies and design ideas. *Fast design* is therefore a consequence of the free exchange of people and ideas within a scientific culture, and can be seen as a kind of fuel for innovation which seems to proceed at a slower pace. In the next chapter, we turn to the classes of innovation in audiophotography that typify today's market, before reflecting on the interplay between research and development, design and innovation at the end of the book.

References

Banks R (2011) The future of looking back. Microsoft Press, Redmond/Washington, DC

Banks R, Sellen A (2009) Shoebox: mixing storage and display of digital images in the home. In: Proceedings of the 3rd international conference on tangible and embedded interaction. ACM Press, New York, pp 35–40

Bentley F, Metcalf C, Harboe G (2006) Personal vs. commercial content: the similarities between consumer use of photos and music. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 667–676

Bidwell NJ, Robinson S, Vartiainen E, Jones M, Siya MJ, Reitmaier T, Lalmas M (2014) Designing social media for community information sharing in rural South Africa. In: Proceedings of the Southern African Institute for Computer Scientist and Information Technologists annual conference 2014 on SAICSIT 2014 empowered by technology, ACM, p 104

Bitton J, Agamanolis S, Karau M (2004) RAW: conveying minimally-mediated impressions of everyday life with an audio-photographic tool. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 495–502

Blythe M, Robinson J, Frohlich D (2008) Interaction design and the critics: what to make of the Weegie. In: Proceedings of the 5th Nordic conference on human-computer interaction: building bridges, October 2008. ACM Press, New York, pp 53–62

Bonsignore E (2010) The use of StoryKit: design implications for intergenerational mobile storytelling. Tech report HCIL-2010-31, HCIL, College Park

Bonsignore E (2013) Sharing stories "in the wild": a mobile storytelling case study using StoryKit. ACM Trans Comput Hum Interact (TOCHI) 20(3):18–38

Csikszentmihalyi M, Halton E (1981) The meaning of things: domestic symbols and the self. Cambridge University Press, Cambridge

Dib L, Petrelli D, Whittaker S (2010) Sonic souvenirs: exploring the paradoxes of recorded sound for family remembering. In: Proceedings of the 2010 ACM conference on computer supported cooperative work. ACM Press, New York, pp 391–400

Donald M, Wajcman J (1986) The social shaping of technology. Open University Press, Milton Keynes

Drazin A, Frohlich DM (2007) Good intentions: remembering through framing photographs in English homes. Ethnos 72(1):51–76

Durrant A, Frohlich D, Sellen A, Uzzell D (2011a) The secret life of teens: online versus offline photo displays at home. Vis Stud 26(2):113–124

Durrant D, Rowland D, Kirk DS, Benford S (2011b) Automics: souvenir generating photoware for theme parks. In: Proceedings of CHI 2011. ACM SIG-CHI, New York, pp 1767–1776

Frohlich DM (2004) Audiophotography: bringing photos to life with sounds. Kluwer, Dordrecht

Frohlich D, Fennell J (2007) Sound, paper and memorabilia: resources for a simpler digital photography. Pers Ubiquit Comput 11(2):107–116

Frohlich DM, Jones M (2008) Audiophoto narratives for semi-literate communities. Interact Mag 15(6):61–64, Nov/Dec 2008

Frohlich D, Murphy R (2000) The memory box. Pers Technol 4(4):238–240

Frohlich DM, Tallyn E (1999) Audiophotography: practice and prospects. In: CHI'99 extended abstracts. ACM SIG-CHI, New York, pp 296–297

Frohlich DM, Adams G, Tallyn E (2000) Augmenting photographs with audio. Pers Technol 4:205–208

Frohlich DM, Clancy T, Robinson J, Costanzo E (2004) The audiophoto desk. In: Proceedings of the 2 AD: 2nd international conference on appliance design. Bristol, UK, pp 139–140

Frohlich DM, Bhat R, Jones M, Lalmas M, Frank M, Rachovides D, Tucker R, Riga K (2009a) Democracy, design and development in community content creation: lessons from the StoryBank project. Inf Technol Int Dev (ITID) 5(4):19–36, Special issue on Human Computer Interaction for International Development (HCI4D)

Frohlich DM, Rachovides D, Riga K, Frank M, Bhat R, Edirisinghe E, Wikramanayake D, Jones M, Harwood W (2009b) Story Bank: mobile digital storytelling in a development context. In: Proceedings of CHI 2009. ACM Press, New York, pp 1761–1770

Frohlich DM, Robinson S, Eglinton K, Jones M, Vartainen E (2012) Creative cameraphone use in rural developing regions. In: Proceedings of MobileHCI'12. ACM Press, New York, pp 181–190

Gemmell J, Bell G, Lueder R (2006) My life bits: a personal database for everything. Commun ACM 49(1):88–95

Håkansson M, Gaye L, Ljungblad S, Holmquist LE (2006). More than meets the eye: an exploratory study of context photography. In: Proceedings of the 4th Nordic conference on human-computer interaction: changing roles, October 2006. ACM Press, New York, pp 262–271

Harada S, Sato D, Adams DW, Kurniawan S, Takagi H, Asakawa C (2013) Accessible photo album: enhancing the photo sharing experience for people with visual impairment. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 2127–2136

Hartley J, McWilliam K (eds) (2009) Story circle. Wiley, Oxford

Helmes J, Hummels C, Sellen A (2009) The other brother: re-experiencing spontaneous moments from domestic life. In: Proceedings of the 3rd international conference on tangible and embedded interaction. ACM Press, New York, pp 233–240

Hull J, Reid R (2003) Designing engaging experiences with children and artists. In: Blythe MA, Overbeeke K, Monk AF, Wright PC (eds) Funology: from usability to enjoyment. Kluwer, Amsterdam

Kindberg T, Spasojevic M, Fleck R, Sellen A (2005a) I saw this and thought of you: some social uses of camera phones. In: Proceedings of CHI 2005. ACM Press, New York, 1545–1548

Kindberg T, Spasojevic M, Fleck R, Sellen A (2005b) The ubiquitous camera: an in-depth study of camera phone use. Pervasive Comput 4:42–50

Kirk DS, Sellen A (2010) On human remains: values and practice in the home archiving of cherished objects. ACM Trans Comput-Hum Interact (TOCHI) 17(3):10

Kirk DS, Izadi S, Sellen A, Taylor S, Banks R, Hilliges O (2010) Opening up the family archive. In: Proceedings of the 2010 ACM conference on computer supported cooperative work. ACM Press, New York, pp 261–270

Koskinen I (2005) On sound in multimedia. In: Proceedings of designing pleasurable products conference. Eindhoven, NL.

Koskinen IK (2007) Mobile multimedia in action. Transaction Publishers, London

Koskinen I, Battarbee K (2006) Thinking about sound in mobile multimedia. In: Proceedings of Ubicomp 2006 workshop. Irvine, US.

Koskinen I, Kurvinen E, Lehtonen T-K (2002) Mobile image. Edita Publishing Inc, Helsingfors

Lambert J (2002) Digital storytelling: capturing lives, creating community. Digital Diner Press, Berkeley

Lambert J (2013) Seven stages: story and the human experience. Digital Diner Press, Berkeley

Lindley SE, Monk AF (2005) Augmenting photographs with sound for collocated sharing. In: Proceedings of the home-oriented informatics and telematics. Springer, York, UK, pp 155–170

Lindley S, Taylor AS, Durrant A, Kirk D (2009) Colocated social practices surrounding photos. Int J Hum Comput Stud 67:995–1004

Ljungblad S (2007) Designing for new photographic experiences: how the lomographic practice informed context photography. In: Proceedings of the 2007 conference on designing pleasurable products and interfaces. ACM Press, New York, pp 357–374

Ljungblad S, Hakansson M, Gaye L, Holmquist LE (2004) Context photography: modifying the digital camera into a new creative tool. In: CHI'04 extended abstracts on human factors in computing systems. ACM Press, New York, pp 1191–1194

Oleksik G, Brown LM (2008) Sonic gems: exploring the potential of audio recording as a form of sentimental memory capture. In: Proceedings of the 22nd British HCI group annual conference

on people and computers: culture, creativity, interaction, vol 1. British Computer Society. Liverpool, UK, pp 163–172

Oleksik G, Frohlich D, Brown LM, Sellen A (2008) Sonic interventions: understanding and extending the domestic soundscape. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 1419–1428

Pearson J, Robinson S, Jones M (2015) Paper chains: dynamic sketch + voice annotations. In: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing. ACM Press, New York, pp 383–392

Petrelli D, Whittaker S (2010) Family memories in the home: contrasting physical and digital mementos. Pers Ubiquit Comput 14(2):153–169

Petrelli D, Whittaker S, Brockmeier J (2008) Auto topography: what can physical mementos tell us about digital memories? In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 53–62

Petrelli D, Van den Hoven E, Whittaker S (2009) Making history: intentional capture of future memories. In: Proceedings of the SIGCHI conference on human factors in computing systems, April 2009. ACM Press, New York, pp 1723–1732

Petrelli D, Villar N, Kalnikaite V, Dib L, Whittaker S (2010) FM radio: family interplay with sonic mementos. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 2371–2380

Petrelli D, Bowen S, Whittaker S (2014) Photo mementos: designing digital media to represent ourselves at home. Int J Hum Comput Stud 72(3):320–336

Piper AM, Weibel N, Hollan JD (2012) TAP & PLAY: an end-user toolkit for authoring interactive pen and paper language activities. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 149–158

Piper AM, Weibel N, Hollan J (2013) Audio-enhanced paper photos: encouraging social interaction at age 105. In: Proceedings of the 2013 conference on computer supported cooperative work. ACM Press, New York, pp 215–224

Piper AM, Weibel N, Hollan JD (2014) Designing audio-enhanced paper photos for older adult emotional wellbeing in communication therapy. Int J Hum Comput Stud 72(8):629–639

Quinn A, Bederson B, Bonsignore E, Druin A (2009) StoryKit: designing a mobile application for story creation by children and older adults. Tech rep HCIL-2009-22, Human Computer Interaction Lab, University of Maryland, College Park

Rajani R, Vorbau A (2004) Viewing and annotating media with memory net. In: Proceedings of CHI 2004. ACM Press, New York, pp 1517–1520

Robinson S, Pearson JS, Jones M (2014) AudioCanvas: internet-free interactive audio photos. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 3735–3738

Sarvas R, Frohlich DM (2011) From snapshots to social media the changing picture of domestic photography: the changing picture of domestic photography. Springer, New York

Sellen A (2011) Family archiving in the digital age, Chap. 11. In: Harper R (ed) The connected home: the future of domestic life. Springer, London, pp 203–236

Stevens MM, Abowd GD, Truong KN, Vollmer F (2003) Getting into the living memory box: family archives & holistic design. Pers Ubiquit Comput 7(3-4):210–216

Swan L, Taylor AS, Harper R (2008) Making place for clutter and other ideas of home. ACM Trans Comput-Hum Int 15(2):Article 9, 24 p

Van Dijck J (2008) Digital photography: communication, identity, memory. Vis Commun 7(1):57–76

Van Erve D, Vos G-W, Van den Hoven E, Frohlich DM (2011) Cueing the past: designing embodied interaction for everyday remembering. In: Proceedings of DESIRE 2011 conference on creativity and innovation in design. ACM Press, New York, pp 335–345

Whittaker S, Bergman O, Clough P (2010) Easy on that trigger dad: a study of long term family photo retrieval. Pers Ubiquit Comput 14:31–43

Wilson ML, Craggs D, Robinson S, Jones M, Brimble K (2012) Pico-ing into the future of mobile projection and contexts. Pers Ubiquit Comput 16:39–52

Chapter 9
External Development

> *The digital photographer potentially will be so thoroughly linked to a multiplicity of media, both as recipient and producer, that communication of whatever kind becomes more important than the singularity of the photographic vision. (Ritchin 2009, pp. 145–146)*

9.1 Product Design Space

In 2004, the only way to conveniently record audiophotographs was on an HP Photosmart camera, or on a 'feature phone' with MMS. Most other methods involved capturing images and sound clips separately and assembling them into combinations or narrative sequences in video editing software. A few other models of digital cameras or camcorders supported the annotation of photographs with voice labels, and photo browsing software sometimes allowed users to play a slide-show of photographs with accompanying sound. But these uses barely scratched the surface of the aesthetic and affective possibilities afforded by the audiophoto medium, as described in Part I of this book. As we have also seen in Chap. 7, the lack of a widespread standard for encoding audiophotos meant that the embedded sound JPG used by HP cameras was not interoperable with other devices or services. These particular audiophotos could not be sent by email or Bluetooth to other people for playback, or posted on a website in a form that would play.

In July 2015 as I write this chapter, the audiophoto product landscape is very different. Three significant classes of innovation have created new possibilities for capturing and sharing audiophotos in different forms, and new combinations of image and sound are now supported outside the photography domain. Innovations include 3G and 4G smartphones with downloadable multimedia apps, tablets of all shapes and sizes providing a new platform for photo and video browsing, and social networking sites like Facebook and Twitter which support collaborative multimedia file exchange. Furthermore, developments in augmented reality and printed electronics are leading to the maturation of a number of audiopaper products featuring photographs and sound. Finally, the falling costs of data storage together with the increasing rates of data transmission, now mean that audiophotos can be encoded and shared in *video* format without compromising the use of a camera or camera

functions on portable devices. This makes audiophotographs or audiophoto narratives a kind of video genre, which is suddenly compatible with all video enabled devices and services.

In the rest of this Part I briefly review the *kinds* of audiophoto products available in various technology categories, using current products as examples. These products will date quickly or become obsolete in due course, because of the fast turnover of software and hardware in the computer industry. However, the classes of product may change and evolve more slowly, and begin to indicate the different social contexts in which audiophotographs are entering the cultural consciousness. Since most examples are taken from technology-rich Western cultures, this review is biased geographically and culturally towards the global North/West. Given the promising uses of audiovisual expression in developing communities, as described in Sect. 8.3 above, it would be interesting to compare this review with one generated from within the global South/East.

Regarding the technology categories which follow, I begin with camera and camcorder products supporting audiophoto capture, before reviewing two kinds of apps for smartphones and tablets. Audiopaper products are then reviewed before a final discussion of the relationships between categories.

9.2 Cameras and Camcorders

Somewhat ironically, current cameras and camcorders provide the least support of all product categories for audiophotography today. In initial phone enquiries to a number of major dealers in the UK, I was told that they have no products supporting sound and photo combinations. Furthermore, early web searches for cameras with sound resulted in sites containing recordings of mechanical shutter noises. Interestingly, an early discovery from these searches was of surveillance cameras with sound, and the *Boriyuan spy clock* (23 July 2015) which uses motion and sound detection to trigger silent video capture from a hidden camera. This is a commercial application of the context photography paradigm and the Other Brother prototype of Sects. 8.4 and 8.5 above, although not for the aesthetic or sentimental purposes they propose.

However, further feedback from more knowledgeable sales staff, online forums and colleagues, revealed a number of digital cameras with voice annotation features or hybrid photo-video modes. Some of these are reviewed by Chesler (2015) in an online magazine called Opposing Views. They include the following models and features:

- *Casio Exilim* series cameras – several models have Dictaphone capabilities for monoaural voice recording from a rear-facing microphone (e.g. EX-S6SR, EX-G1RD). The intended user model appears to be voice note dictation in between photo and video capture, to remind users of locations they have visited or to capture lectures or meetings in pictures-then-sounds.

- *Nikon COOLPIX* series cameras – support both Dictaphone voice recording and Voice Memo functions for annotating individual photographs with a voice label of up to 20 s (e.g. COOLPIX S9).
- *Ricoh RDC* and *Caplio* series cameras – support Voice Memo features on photographs. These can be played back from images on the camera itself, or sound clips can be exported separately as WAV files. They also support general purpose Voice Recording (e.g. Caplio G3 and G4).
- *Samsung* cameras – support both Voice Recorder and Voice Memo functions with up to 10 h capacity or 10 s/photo (e.g. TL105 and PL200 respectively).

Since it is becoming popular to take short video clips on digital SLRs, a number of cameras now support still image capture from *within* a video recording (e.g. *Canon EOS 750D*). This hybrid video/still mode has the potential to record audio but this has not yet been realised by manufacturers. They appear to take the technical point of view that if you are going to record ambient sound you might as well do it with moving image rather than still image. However, as Chap. 2 shows, there are other aesthetic, psychological and social reasons to record ambient audiophotos rather than video, and this capability could be a new media option that consumers may appreciate.

In short, the support for audiophotography on cameras and camcorders has changed little in the last 10 years. Essentially it is missing, or relegated to specialist sound recording or annotation features which are not seen as central to the capture of photographs or video clips.

9.3   Audiocamera Apps

In stark contrast with the conservatism of digital camera design, the app market is brimming with innovations in multimedia capture and editing. Many of these support audiophotography, either through simple *audiocamera apps* focusing on sound and image capture, or through *audiophoto narrative apps* supporting sound and image combination of various kinds. This distinction is similar to that between simple audiophoto messages and more complex audiophoto narratives described in Chap. 8. There is a tendancy for the audiocamera apps to be designed for smartphones and the narrative apps to be designed for tablets, although many apps can be installed on both platforms. Because of competition between mobile operating systems, there are different types of audiophoto apps for Apple, Android, and Microsoft devices. None have yet been developed for Amazon devices. In the rest of this Part I will review audiocamera apps, leaving audiophoto narrative apps for the next section.

At the time of writing this chapter in July 2015, there are at least 20 audiocamera apps on the market. Their names, platforms and tag-lines are shown in Table 9.1.

Seven apps are implemented for both Android and Apple devices, six are exclusive to Apple, five are exclusive to Android, one is exclusive to Windows and one is for Android and Windows. By far the majority of audiocamera apps are focused on supporting the kind of talking photos described in Chap. 4. These have names like

Table 9.1 Example audio-camera apps (July 2015)

App name	Platform	Description
Audiosnaps	G/A	AudioSnaps takes pictures with sound
Speaking Photo	G/A	It is the simplest way to turn your pictures into stories
picSpeak	G/A	Capture your photos with sound using pikSpeak
PhotoVoice	G/A	Take picture – add voice – share with the world
Talking Photos	G/A	Take a picture, then record and attach audio message
Voisnap	G/A	Camera app that records the sounds of the moment you take the photo
Incerts Snap	G/A	Allows teachers who use the Incerts assessment system to capture photos and video/audio clips to support their judgements
Talking Pictures	A	Record up to 10 touch-activated sounds in a photo
Picle	A	Telling stories with pictures and sound clips
Tunepics	A	Tunepics lets you share your favourite pictures with any song and capture the emotion behind each moment
Picio	A	Picio captures a photo and records sound, so you can remember and share the experiences in a fun way
PixKix	A	Add a soundbite to your photos
Chatter Pix	A	You and your students can turn pictures into talking pictures
Sound photos	G	Listen to your photos
Photo Speaks	G	3D avatars that repeat your every word
Audio Photos Free	G	It adds another dimension to your photos by adding ambient sound
Photo 306 by Sfera	G	Record the environment around you and capture both the image and sound all in one revolution
RIXA	G	Send personal audio greetings with picture to your friends and family!
Foundbite	M	Foundbite combines photo and sound to capture the real atmosphere of a place, event or experience
Eye Me	M/A	Connect and share photos with voice captions on Twitter and Facebook

G Google (Android), *A* Apple (IOS), *M* Microsoft (Windows)

Speaking Photo, PicSpeak, Talking Photos, Talking Pictures, Photo Speaks and so on. They are usually simple 1:1 image-to-sound associations designed to be shared on social media platforms such as YouTube, Twitter or Facebook. The duration of sound recording is usually kept low, and ranges between 5 and 30 s. Two of these apps, *Incerts Snap* and *Chatter Pix* are designed for teachers and young children respectively for justifying assessments or bringing iPad pictures and images to life. The *Talking Pictures* app in this category is unusual in supporting multiple (1: many) image-to-sound associations from up to ten touch regions of a single photo. The *RIXA* app is also unusual in putting emphasis on the 'audio greeting' first, and inviting image annotation to illustrate a spoken message. Finally, the *Photo Speaks* app is particularly innovative in converting a photograph into an animated '3D' avatar, who speaks the voice message associated with the photo.

Four apps are designed primarily for ambient sound capture around an image event, as in Chap. 2. These include *Voisnap, Audio Photos Free, Photo 360* and *Foundbite*. This all allow the image capture event to take place within or alongside the sound recording. For example, in *Voisnap*, the audio recording is started first before the photo is taken, and stopped manually or automatically after 30 s. In *Photo 360*, sound is recorded during the rotation of the smartphone through 360° as a panoramic photo is taken.

Only one app in Table 9.1 is focused on supporting musical photographs for sharing, as in Chap. 3. *Tunepics* recognises the emotional power of music in its' tagline, and allows users of iTunes music to pair tracks with personal photographs as well as provided artwork.

No apps in this category support the recording of conversational photo-talk on photographs, as in Chap. 5. However, an approximation of this is supported through asynchronous sharing of talking photos on social media sites where verbal discussion threads may build up. This is also an explicit aim of certain audiophoto narrative apps, as shown in Table 9.2 and discussed in the next section.

Table 9.2 Example audiophoto narrative apps (July 2015)

App name	Platform	Description
Flipagram	G/A	Flipagram is the easiest way to create and share great video stories using your photos, videos and music
Explain Everything	G/A/M	Lets you annotate, animate, narrate, import, and export
Shuttersong	G/A/M	Create unique photos by adding music, sound or voice
Roxio MediaBook	G/A/M	Share your memories in a multi-media ebook
Story Maker	G/A	Creating great stories is now easier than ever with StoryMaker
Brabble	G/A	Free social sharing app that allows sharing of video, photo, audio and text all in one network
Com-Phone	G	Com-Phone helps you create multimedia narratives, combining photos, audio and text to tell digital stories
U Messenger	G	Experience a new way to connect with friends through photos, text, Audio Photos and stickers
Adobe Voice	A	Tell your story
Memory Box Stories	A	Discover, create and share stories that bring memories to life
Shaddow Puppet	A	Combine photos and video with your voice and favorite song
Pixengo	A	It's time to UNMUTE YOUR PHOTOS!™ and give them your voice
Clipagram	A	Photos + Audio = Amazing Video
Appture	A	Secure photos + audio
Blurb Mobile	A	Create shareable stories using photos, videos and audio
Digisocial	A	Capture the beauty of life with voice and photos together
Educreations	A	Annotate, animate, and narrate nearly any type of content as you explain any concept
Shared Impressions	M	A modern photo and slideshow app

G Google (Android), *A* Apple (IOS), *M* Microsoft (Windows)

9.4   Audiophoto Narrative Apps

Table 9.2 shows 18 example audiophoto narrative apps, found at the time of writing. All of these apps go beyond the early PC and Mac software to set photo or video slideshows to music, and make more or less sophisticated efforts to incorporate voiceover and ambient sound as well. For example, *Flipagram* allows you to set a photo series to a music backing *or* to a voiceover narration, but not both. It does however allow the incorporation of short video clips in the series, along with photographs, resulting in a kind of photo/video slideshow with sound. A typical user model for adding music is to associate the whole image series with a music track and set a duration for the show or each image presentation. A typical user model for adding voiceover is to manually advance the image series during voiceover recording. The insertion of video or other sound clips will usually result in automatic adjustment of the duration of associated visuals. This is easier using a frame-by-frame organisation as in our own *Com-Phone* app, where up to three layers of sound (for ambient, voiceover and music) can be overlaid on a single image (see again Fig. 8.2). These are mixed from a single starting point and play out to the length of the longest clip. In fact *Com-Phone* is a kind of cross-over audiocamera/audiophoto narrative app, which allows the recording of single audiophoto or multiple audiophotos in sequence on a smartphone, so that ambient sound can be recorded at the same time as each image. Otherwise ambient sound has to be recorded separately at image capture time and imported into the narrative apps later.

A number of differing contexts and motivations are evident for the apps in Table 9.2 and these affect the way in which multimedia narratives are supported and shared. Four distinct value propositions can be seen in the description of apps and can be used to roughly classify them as follows:

- Improved memories of life – *Roxio MediaBook, Memory Box Stories, Appture, Shared Impressions*
- Better storytelling over photographs – *Story Maker, Com-Phone, Pixengo, Clipagram, Blurb Mobile*
- Clearer explanations – *Explain Everything, Adobe Voice, Shadow Puppet, Educreations*
- Richer conversations – *Flipagram, Shuttersong, Brabble, U Messenger, Digisocial*

The memory-oriented apps seek to bring photos to life mainly through voice narration. For example, *Memory Box Stories* encourages users to select a sequence of photos and record up to 5 min of narration which is later synchronized to the photos. It also suggests family reminiscing activities in which old photographs are recaptured (digitized) with stories from parents. The book or album metaphor often underlies these apps, as with the *Roxio MediaBook* which uses professional templates for layout of photographs and numerous transition effects. *Appture* also introduces location details as a way of capturing special journeys or travelogues. These apps are intended to generate rich multimedia accounts of private experiences

within a family or small group, as suggested by the description of *Shared Impressions*, and the emphasis on secure photo sharing in *Appture*.

While the memory-oriented apps are centred on reminiscing, the story-oriented apps are centred on storytelling. These apps use similar techniques to tell a multimedia story to a wider audience, and emphasise verbal storytelling as the central component. Hence, *Story Maker* is a journalistic tool designed by The Guardian Project to help mobile/citizen journalists capture secure and compelling stories in pictures, video and sound. It uses story templates and has links to a "55 lesson course in Journalism, Security, Photo, Audio, and Video production". *Com-Phone, Pixengo* and *Clipagram* seek to support successively less sophisticated effects for a consumer market. They do this in simpler ways that still enable users to narrate over single or multiple photo sequences, which may or may not be annotated with text or other sounds. *Blurb Mobile* is an interesting variation on these apps, coming as it does from Blurb's background in the self-publishing of print or e-books (http://www.blurb.co.uk/). The app can be seen as a supporting a kind of multimedia photobook or photo essay with up to eight photos, a 30 s audio clip per photo and a 10 s video clip per story.

Four explanation-oriented apps lie outside the domestic photography market altogether. *Adobe Voice* is directed to the business presentation market while *Explain Everything, Shadow Puppet (Edu)* and *Educreations* are directed to the education market. In fact Adobe Voice was described in one on-line article as "A happy app for making explainer videos", since it is radically simpler than other Adobe software such as Photoshop, and can be used to explain just about anything (Pogue 2014). Users are prompted to start with a verbal explanation of their idea or concept and illustrate it with stock photographs and music before choosing a visual theme. *Shadow Puppet* is a simpler version of this without access to stock photos and music, which tries to handle both music and voiceover across a series of images. Its *Edu* version is promoted to teachers and pupils for creating classroom materials or allowing children to express themselves in multimedia stories. *Educreations* and *Explain Everything* have a different interactive whiteboard metaphor and user model, for allowing concepts to be explained in real time through mash ups of background images, text and handwritten annotations with sound effects and voiceover.

Finally, five audiophoto narrative apps attempt to start audiophoto social networks or introduce audiophotos into existing ones. For example, *Digisocial* bills itself as a new 'revolutionary voice social network'. It has been likened by commentators to be a photo sharing website like Instagram, but with the possibility of audio as well as text annotation (e.g. Davison 2013). This option is carried through to responses on posts to the *Digisocial* website, so that audio discussions and threads can build up on audiophotos themselves. I count these as new kind of audiophoto narrative in which the narrative builds up in audio form on a single photo, in an asynchronous version of the conversational photographs of Chap. 5. *Shuttersong* is similar, while *Brabble* generalizes the approach to support the sharing of standard photo, video, text and audio files with 'brabbleback' commentary in any of these formats. *U Messenger* lies somewhere between *Shuttersong* and *Brabble* in supporting a range of audiophoto posts, but in the context of organizing live photo chats

with small groups of friends or family who can discuss and further annotate online materials. *Flipagram* does not yet support its own social network as such, but promotes the sharing of audio photo/video narratives to other social networks like YouTube, Instagram, Facebook, Twitter, WhatsApp and Vine. However it is promoting the approach on its own website and through *Flipagram* pages and communities on other networks (http://flipagram.com/).

9.5 Audioprint Products

Perhaps one of the biggest changes in domestic photographic practice brought about by the digital revolution is a movement away from printing photographs. Certainly the routine printing of photographs to view them has now gone, since they can be viewed on the originating camera or smartphone immediately and at zero cost, and then sent to a variety of other screen-based viewing devices such as desktop computers, televisions, tablets and photo displays. Initially this was done as a device-to-device (peer-to-peer) communication, but it is now done increasingly as an internet transaction in which photographs are posted on photo websites or social networking sites to be accessed remotely. In writing about this in 2011, Risto Sarvas and I pointed to a transition from the Kodak era of photography ("you press the button and we do the rest") to a Digital era in which printing becomes a more reflective and selective activity (Sarvas and Frohlich 2011). Interestingly, this is now enabled by a new raft of digital print services tied to online collections and albums, including photobook printing by companies such as PhotoBox.

   All this is relevant to the product landscape of what I call 'audioprint products', for playing back sounds from printed photos. It turns out that this not only mirrors the history of analogue to digital photography in many ways, but it also anticipates what may be a new analogue/digital or Post-Digital era. This is because the market contains legacy audiophoto frames, cards and albums designed to allow manual recording of sounds with printed photos, but is also starting to support online specification of audiphotos to similar devices. Furthermore, the rise of augmented reality techniques for tagging real-world objects or locations with digital content is just starting to be applied to printed images to fetch associated sounds or video. In fact, services like *Blippar* described below, directly implement the audiophoto desk technique of recognising a printed image to playback associated content. This begins to realize an 'internet of things' vision for photography, linking ordinary photobooks and other printed photo outputs to audio content and players lying outside the paper itself.

   Different generations of audioprint products can be illustrated by stepping through the history above. These essentially co-exist today, given that analogue photography remains a niche practice. Figures 9.1, 9.2, and 9.3 represent solutions in the Kodak, Digital and Post-Digital eras respectively, using photograph albums as an example print output. Similar activities apply to other audioprint outputs such as talking photo frames and recordable greeting cards. Hence Fig. 9.1 shows the pho-

Fig. 9.1 Photo and
audiophoto display in the
Kodak era
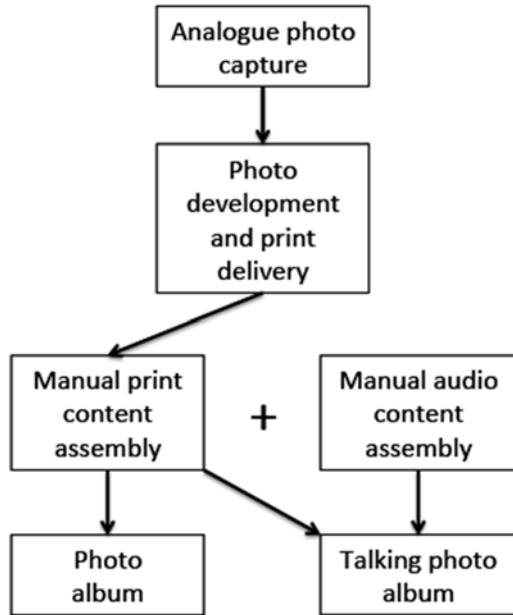
towork activities involved in assembling a conventional and so called 'Talking photo album' based on film photography. Essentially, analogue camera film is developed and printed as loose prints, which are then inserted into the transparent sleeves of a photo album. Talking photo albums simply instruments these albums with electronics to capture and playback sounds from individual pages, recorded manually into the album during assembly. A wide variety of blank *Talking photo albums* of this kind can currently be bought for about £20 or £30, from companies such as Brookstone, Talking products, EKKO, TTS and Talking Albums. Some of these companies such as Alz Products, Liberator and Inclusive Technology promote these products for use with people suffering memory loss or experiencing communication problems. Other instantiations of this technology are designed for single photos to be displayed in *Talking photo frames* or sent as *Recordable greetings cards*. Many of the same companies are involved, or others such as Sound Expressions specializing in the card market. Sound Expressions sell a wide variety of talking paper formats including greeting cards, envelopes, postcards and toys. They also sell adhesive sound module inserts to stick onto existing cards. The variety of products is now great than in 2004 and the quality of solution better at the high end of the range. This is because the digital audio technology embedded in these products has advanced in recent years, but also because the use model still fits with digital photography as shown in Fig. 9.2.

   The use of a digital camera or cameraphone instead of an analogue camera, still allows consumers to print and assemble conventional and talking photo albums,
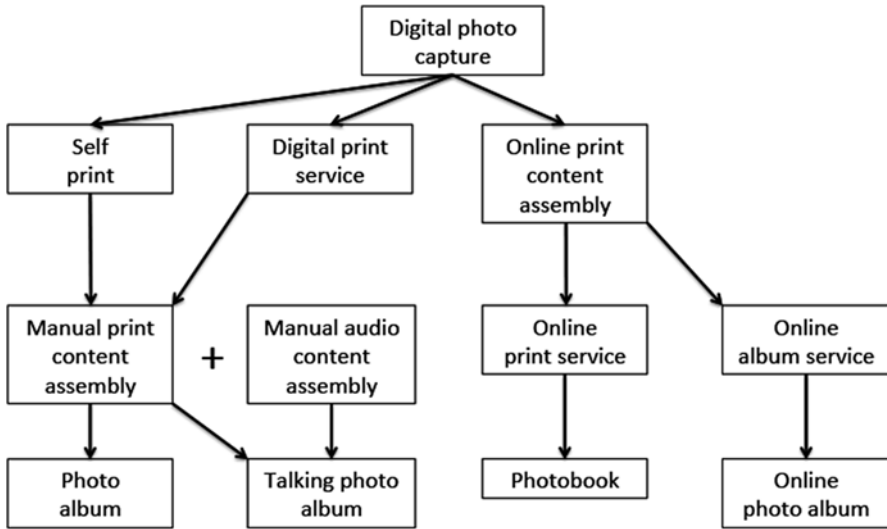
Fig. 9.2 Photo and audiophoto display in the Digital era

frame and cards as before. This is shown on the left hand side of Fig. 9.2, which shows that manual assembly is now enabled by a digital print service or self-printing of photographs at home. The big transformation shown in this diagram is the presence of on-line photo content assembly. This can lead either to on-line posting to a photowebsite or social networking service, or to specification of a printed *photobook* via an on-line print service. A big difference between these two options is cost, with photobooks being expensive and on-line posting being free, or at least cheaper if extended cloud storage is involved. Hence printing has become more selective than online posting. Typical photobook services are provided by companies such as Photobox, Snapfish, Bob Books, PhotoLeaf, Yo Photo and Artesan State. Many of these companies also sell alternative print outputs such as personalized greeting cards, posters, t-shirts, mugs and so on, while others are specialized in a subset of these (e.g. MoonPig for personalized cards and gifts). Note that audioprint products are supported in this era, but not in a way that is compatible with the emerging audiophoto camera apps and audiophoto narrative apps described in previous sections. For this we need to look to Fig. 9.3.

Full audiophoto capture is shown at the top of Fig. 9.3 and is shown to lead directly to on-line audiophoto content assembly. This can result in on-line audiophoto album display on a variety of non-specialist services such as YouTube, Facebook and Twitter, or specialist ones such as Digisocial, Shuttersong, U Messenger or Brabble. It can also result in specification of a printed photobook with associated sound. In principle, this could take the form of a conventional Talking photo album or card populated with content from the online version. Sound Expressions and EKKO seem to have some provision for online design of certain promotional cards in their range. However, the market seems to have moved to an

Fig. 9.3 Photo and
audiophoto display in the
Post-Digital era
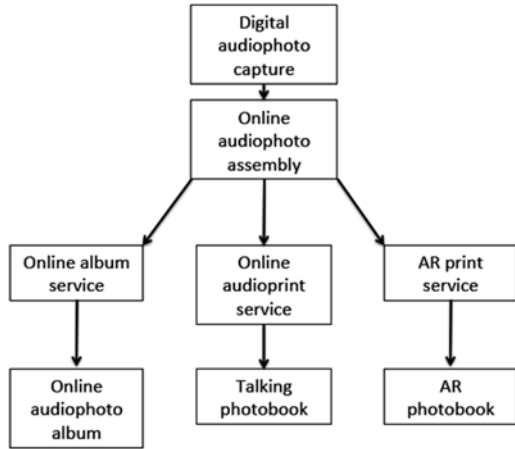
easier solution based on straightforward printing of photobooks or cards with QR codes or other tags to invoke associated sound playback on a smartphone. This augmented reality solution is represented in several services such as Shutterfly's *Photo Story* book. This comes with an iPad app through which to create the album with up to 30 s of narration per page. Users can then send off for the printed photobook version which has QR codes on each page which then play the sound on the scanning device (Dove 2013).

As before, this solution also applies to other print formats such as cards and posters with services such as HP Live Photos mentioned at the end of Chap. 7, which allows consumers to print index shots of video clips which trigger video playback on a smartphone. Moon Pig and Blippar have similar apps for consumer and commercial advertising markets respectively, and work by recognising the image itself rather than a tag on the image. In contrast, *Tap For Message* is a company and service based on the sale of greeting stickers than can be attached to any package or card, and associated with online audio or video content via a smartphone. Stickers contain near field communication (NFC) tags which are recognised by certain smartphones, and legacy printed QR and URLs for non-NFC enabled phones.

An interesting hybrid product that approximates the self-contained photobook of Fig. 9.3, is a new kind of 'Videocard' with integral LCD panel and speakers. This is supplied by a number of companies including Spreengs, Vidioh and TQR cards. For example, the Spreengs video card currently sells for £20–30 and contains an embedded display of between 2.4″ and 7″ in the right hand side of an opened greeting card with 165 MB of embedded storage. It can play either video clips or photo slideshows with music. The content can be transferred to the card via a mini USB cable from a laptop ('You card it!') or specified online for the card to be made and mailed to its recipient ('We card it!'). Although designed for conventional video content, it could display audiophoto narrative content just as well, to bring digital stories to the printed page.

9.6   Discussion

The development of new technology products and services is a vibrant area which often appears to progress in a piecemeal fashion, by constantly seeking novelty in functionality and design. However, by reviewing developments in one particular area over a 10 or 11 year period, we have seen that these experiments are underpinned by bigger changes in computing infrastructure and driven by growing consumer interest in particular types of information or media experiences. In the case of audiophotography products and services, they have been hugely influenced by the development of the world wide web itself and the practice of sharing media on social networking sites like Facebook and Twitter. The posting of photographs for textual comment has become a widespread practice in the West, which audiophoto products must fit into. In addition, the emergence of smartphones and tablets with their corresponding app model of evolution has influenced the way in which sounds and images can be captured and shared, compared with the more camera-oriented model that was the subject of the Audiophotography book of 2004. This led us to three major discoveries about the development of products in this area.

First, we found that innovation in the capture of audiophotos on digital cameras and camcorders has stagnated, compared with the explosion of audiophoto-related apps for smartphones and tablets. The latter products are competing across multiple platforms to establish a new user model for attaching simple voiceovers or ambient sounds to individual images (in audiocamera apps) or assembling audiophoto slideshows to tell visual stories with all kinds of sound (in audiophoto narrative apps). These can usually be posted on the web as a kind of *video*, for textual comment in the usual way. The sophistication and complexity of audiovisual capture and editing distinguishes one app from another, and probably serves a diverse market of users who want more or less control over the process. This appears to mirror the differences between various point-and-shoot, compact zoom and digital SLR cameras themselves, which offer more or less control over the process of image capture. The fact that these cameras cannot support audiophotography in the same way as the apps, constitutes a product opportunity in its own right. Camera and camcorder manufacturers would do well to attend to the developments we have reviewed and consider incorporating some of them in their products. Since they already have a portfolio of products varying in complexity for different user groups, they could support audiophotography at different levels of complexity across this range.

A second discovery of the review is that a battle seems to be raging over the way in which audiophoto materials are shared over the web. Different formats for audiophotos still exist, making them more or less simple to share on existing photowebsites or social networking sites. Those which are stored in video format are the easiest to share, while those stored in specialist multi-file or slideshow formats require their own sites. This has led some companies like Digisocial to position audiophotos as a new social media form with a dedicated network, whereas others like Flipagram are attempting to stimulate community formation within or across other platforms. Whichever model takes hold, the sharing of audiophoto material

over the web is already well underway and is only likely to increase in the future as people realize that they can add sounds to photographs for special effect or create video-like presentations out of still images and sound. As with many other forms of specialist media production, this may lead to an online community of practice in which tips and techniques are shared and inspirational material is exhibited.

A final discovery of the review is that printed audiophotos are enjoying a resurgence through twin developments in printed electronics and augmented reality. These embody hybrid physical/digital interactions within an emerging internet of things. Hence, physical multimedia documents can be specified online and printed as a service like photobooks (digital-to-physical). Alternatively they can be printed with markers, tags or recognised visually to cause associated sound to play on an adjunct device (physical-to-digital). This is why the future era for digital photography might be called Post-Digital, because context and behaviours will not reside in the digital realm alone, but be distributed across digital and physical artifacts, including paper. In this domain it may result in a resurgence of printed photographs with digital 'connections', and an eventual spillover into the world of objects and souvenirs. Experimental activity of this kind has been evident in the research literature for some time, as described in Chaps. 7 and 8, but is now emerging into the commercial market. This is the long nose of innovation in action again (Buxton 2008), bringing audiophotography closer to a mass-market practice, but not in the forms envisaged at earlier stages of its development. This miss-match between early designs and later innovations runs through the entire audiophotography case study, and will now be discussed in the final chapter of the book.

References

Boriyuan spy clock (23 Jul 2015, online search result). http://www.amazon.co.uk/Boriyuan-Recorder-Security-Hidden-Detector/dp/B00OYN7OHQ

Buxton W (2008) The long nose of innovation. Business Week, 2 Jan 2008. http://www.business-week.com/stories/2008-01-02/the-long-nose-of-innovationbusinessweek-business-news-stock-market-and-financial-advice

Chesler M (2015) Digital cameras that record audio. Opposing Views. http://science.opposing-views.com/digital-cameras-record-audio-12308.html

Davison P (2013) Talk to your pictures with Digisocial. Adweek Social Times. http://www.adweek.com/socialtimes/talk-to-your-pictures-with-digisocial/535832

Dove J (2013) Hands on: Shutterfly Photo Story for iPad makes photo books that talk. Tech Hive online magazine. http://www.techhive.com/article/2046150/hands-on-shutterfly-photo-story--for-ipad-dispatches-photo-books-with-narration.html

Pogue D (2014) Adobe voice: a happy app for making explainer videos. Yahoo Tech online magazine. https://www.yahoo.com/tech/you-know-adobe-right-this-is-the-company-that-85081015729.html

Ritchin F (2009) After photography. WW Norton & Company, New York

Sarvas R, Frohlich DM (2011) From snapshots to social media-the changing picture of domestic photography: the changing picture of domestic photography. Springer Science & Business Media, London

Chapter 10
Lessons

10.1 Introduction

> *What counts as an innovation is historically relative because
> they are constituted within particular technocultural formations.
> These formations emerge at certain historical moments in
> response to a wide range of influences – economic, political,
> institutional – through interactions among people, between
> people and the matter of the world. (Balsamo 2011, p. 8)*

In the high tech computer industry, products themselves evolve at a tremendous
rate. Anything more than about 3 years old is probably in need of an 'upgrade', and
the mindset of the industry is ever forward-looking in search of the next big thing.
In this book we have looked back over 10 years to some audio-enabled camera
products from HP, and traced their obsolescence and subsequent re-emergence as
smartphone and tablet apps by other companies. We have also seen the development
of new scientific understandings about the role of sound and pictures in communica-
tion and memory, through numerous research projects and prototypes. This has
been a story of what it means from *inside* the industry to try to research and develop
the next big thing. In stark contrast to the speed with which products come and go,
ideas and values appear remarkably stable. Furthermore, the process of realizing
new values through technology is remarkably long, interconnected with other inno-
vations, and distributed across the globe. Individual products have a *history* and
provenance that usually stretches back many years, into the corporate research labs
of other companies or university groups, and previous products in an evolutionary
tree. This history is illuminating as an insight into how innovation actually works
and what can be done to manage it better.

In this chapter I draw together some of the insights and lessons of the audiopho-
tography innovation story in three categories. First I review the lessons for audio-
photography itself, in terms of its' value, practice and technology. Second, I discuss
the lessons for innovation and how it seems to work at a practical and personal level
in the new media area. Finally, I return to topic of fast design and slow innovation
which framed the investigation, to comment on their relationship as it has been
illustrated here. This involves a consideration of the links between research and
development, and their implications for the research-practice gap as discussed by
Norman (2010). I also consider some connections with movements promoting slow

design and fast innovation. The book finishes with some conclusions for the industry and recommendations for those who work in it.

10.2 Lessons for Audiophotography

The findings of the review have different implications for the value, practice and technology of audiophotography as interrelated facets of the subject. Let us consider these in turn.

Regarding lessons on the value of combining sound with pictures, numerous studies both inside and outside HP have endorsed and extended the original claims. Small scale evaluations of various HP audiophoto systems such as *Active Photos* (Kindberg et al. 2004), *MemoryNet Viewer* (Rajani and Vorbau 2004), *StoryMail* (Debaty et al. 2004) and the *Audiophoto Desk* (Frohlich et al. 2004) were all positive and confirmed many of the values of sound in Part I of this book. For example, we saw that Lindley and Monks' (2005) evaluation of the *Audiophoto Desk* confirmed the value of sound for adding fun, atmosphere and humour to photo sharing, and that music was particularly effective in enriching the activity without inhibiting the talk.

Other studies and systems reinforced the value of specific types of sound for memory and storytelling. For example, studies of sonic gems and souvenirs showed that people have an audio memory for events which is triggered by **ambient** sound recordings in a similar way to the triggering of visual memory by photographs (e.g. Oleksik et al. 2008; Dib et al. 2010). A recent collection of readings on 'Sound Souvenirs' deepens these insights on ambient sound, and also promotes again the nostalgic value of music (Bijsterveld and Van Dijck 2009). Furthermore, the design exploration of Fennell and Frohlich (2005) show how **conversations** of different kinds, around the hospital bed or the dinner table, could be recorded and played back in different ways to remind people of relationships and encounters. Finally, the work on audiophoto narratives described in Sect. 8.3 shows the value of **voiceover** for explaining image sequences in story form (e.g. Frohlich and Jones 2008; Frohlich et al. 2012; Quinn et al. 2009), and connects with a large body of work in digital storytelling.

While much of this work was conducted in the domestic photography domain, other work was not, and serves to open up the application of audiophoto content to other areas and markets. A few of these areas are listed below and show that there may be multiple *audiophotographies* stretching outside the domestic context which was the motivation for the original work:

- Community journalism, e.g. *Com-Me* (Frohlich et al. 2012)
- Cultural heritage, e.g. *RAW* (Bitton et al. 2004)
- Literacy learning, e.g. *StoryKit* (Quinn et al. 2009)
- Photography for the blind, e.g. *Accessible Photo Album* (Harada et al. 2013)
- Public art, e.g. *Weegie* (Blythe et al. 2008)

- Oral history, e.g. *Tap and Play* (Piper et al. 2012)

Regarding lessons on the practice of audiophotography, these stem from a massive expansion of the user model for capturing and sharing sounds and images. My original vision was for the recording of an extensible multilayered audiophoto, partially recorded on an audiocamera, with further sound added or mixed at playback on a variety of screen-based devices (Fig. 9.4 in Frohlich 2004). A loose printed version was also envisaged to play from an associated holder or mobile device. Subsequent research and development work has explored many other user models customised to their particular application contexts and influenced by the technology of the day.

One such model reverses the primacy of audio in relation to photo, and proposes an **audio-only** or **audio-first** approach. Audio-only activities can be seen in the *Sonic Gems*, *Memory Shelf*, *Best China* and *FM Radio* prototypes which support different ways of capturing and playing back sound clips without photographs. Audio-first activities can be seen in the *StoryMail* model of voicemail with pictures, and an option in the *StoryBank and Com-Phone* apps to record voiceover first. At the opposite end of this dimension, an alternative photo-first user model is represented by **audio-annotation** of regions of an individual photograph supported in the *Active Photo* and *Tap And Play* systems.

Another user model is the **audiophoto narrative** model of *StoryBank*, *Com-Phone* and *StoryKit* which imply an image sequence accompanied by voiceover and various other optional sound types or layers. Some of these also support **audio-photo-text** combinations to further extend the expressive power of the medium with written language. A variation of this approach is to record a linear sequence of **audio-photo-audio**, as in the *Accessible Photo Album*. All these systems can play audiophotos back as a kind of animated **slide show**, and some can be encoded and shared as video files on social media websites. Full **integration with video** or video effects is another new user model that is evident in both research and development work. For example, virtual zooming or panning over an audiophoto was evident in the *RAW* system and represents the generation of **video from stills**. In the commercial realm, systems such as *Flipagram* and *Shadow Puppet* support the **insertion of video clips** themselves in audiophoto sequences.

These and other ways of combining sound and image affect the practice of assembling an audiophotograph and how it appears to audiences. The main lesson here is that there are now multiple ways of doing audiophotography, and different user models may be more or less appropriate to different application domains.

Regarding lessons on the technology of audiophotography, these have been partially covered in Chap. 9. There has been a stagnation of support on cameras themselves, but a growth in the number of audiocamera and audiophoto narrative apps on smartphones and tablets. These vary according to the user model they employ from amongst the kinds discussed above. In general, there is an emphasis on supporting talking photographs over other types, and stringing these together in narrative sequences. The prevalence of photo sharing with family and friends over social media, is leading to the integration of audiophoto narratives as a form of video (e.g.

Adobe Voice), or single audiophotographs as talking points in a more verbal social network (e.g. *Digisocial*).

Taken together, these trends appear to be leading to a kind of *hyperphotography* of the kind discussed by Ritchin (2009). This involves the transcendence of single images into linked combinations and sequences accompanied by other media: "*A digital photograph also can be said to represent a single moment, but one that can be easily lengthened and amplified by joining it to many other moments through various forms of hyperlinks and hybrid media, including extending it as a movie*" (Ritchin 2009, p. 142). Modern audiophotography combines photography, sound recording and videography in this kind of way, and opens up a more integrated practice of self-recording and expression.

10.3   Lessons for Innovation

Perhaps the biggest lesson of this review for innovation, is that it never stops. Innovation is a large multifaceted thing, which is best looked at as an ongoing process rather than a bounded activity with a beginning and end. Indeed, the audiophotography lense on innovation, taken in this book, originated itself in a particular time and context, and owes a debt to many preceding innovations in analogue and digital audio capture, photography and communication. And it will be possible and interesting in another 10 years to consider the integration of sound and image in future media technologies and behaviours.

To understand the multifaceted nature of innovation a little better, I would like to return to a framework I have used before in a previous publication about the relationship between innovation and human computer interaction (Sarvas and Frohlich 2011). HCI studies the way in which people interact with digital technology and is involved in the design and testing of new technologies to change the way we live. In the article we used the story of the early HP audiophotography work to argue that HCI tends to overlook business factors in its methods of study, and to stop short of tracking the evolution of prototypes and products over time. The current book takes up that argument by tracking the evolution of work in audiophotography since then, and highlighting its' current manifestation in products. In reflecting on innovation before, we pointed to four research activities that define any new business opportunity and the various lines of connection between them. These included *User research* to understand the domain of application and user value of the technology, *Technology research* to realize the core technology to be applied, *Design research* to package the technology solution in a product and *Business research* to work out who to sell it to at a profit. A diagram of these activities and connections was given in Fig. 7 of the article, and suggested as a way of documenting R & D activity to refine any individual business opportunity. For present purposes, I show an adapted version of this figure with the research activities replaced by their *outputs*, including Applications, Technologies, Designs and Markets respectively (Fig. 10.1).

Fig. 10.1 Elements of a business opportunity defined by user, technology, design and business research (Adapted from Fig. 7 in Frohlich and Sarvas 2011)
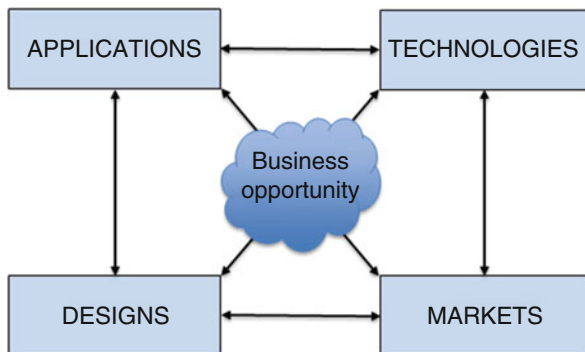
We can use this framework to characterize the starting vision for the audiophotography opportunity as expressed in Frohlich (2004) and Part I of this book. Taking the domestic photography application area as its starting point, the proposal was to use co-indexed sound-and-image software to record audiophotographs on an audiocamera and play them back on the camera and other ICT screen devices. An additional element of the proposal was to use embedded chip technology in paper to playback the sound from printed photographs from a handheld player. The creation of a new paper stock with embedded electronics fitted with HP's business model at the time of making profit from paper and ink consumables. The main market for all this was to be families, which were the biggest purchasers of HP home computers and printers at the time. This combination of elements is shown in Fig. 10.2 below, and was itself the result of a whole series of studies and experiments in each research area dating back to about 1996. Some of these studies are reported in Part I of this book.

Subsequent R & D activity as reviewed in Part II of this book essentially expanded the business opportunities for audiophotography enormously, by exploring new applications, technologies, markets and designs for the idea. Designs are represented by either prototypes or products that appear to embody assumptions about applications and markets in their design rationale (c.f. Moran and Carroll 1996). This can usually be read in research papers describing the motivation for prototypes and who they are tested on, or in the advertising literature describing the benefits of products and their intended customers. Assumptions about technologies can usually be read from their design specifications. Some example new elements to emerge from the review are shown in Fig. 10.3 below. With so many elements listed on the diagram, this is clearly no longer depicting a single coherent business opportunity. Instead, the elements outline an expanded business opportunity space from which individual businesses have been or could be defined. In fact one way of *generating* innovations would be to mix and match elements across categories to evolve designs for new applications, markets and technologies.

One pattern to emerge from the review of these developments has been noted before in Chaps. 7 and 8. Closed innovation within HP appeared to take a classic technology transfer approach of considering the use of sound-and-image software

Fig. 10.2 The original
2004 vision for
audiophotography as a
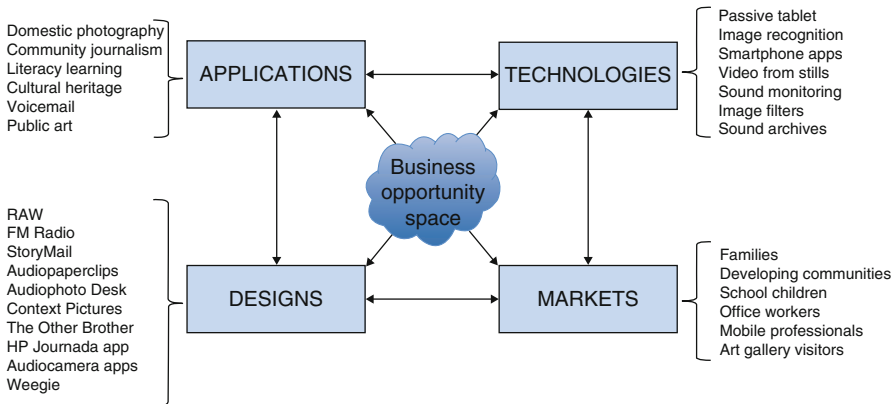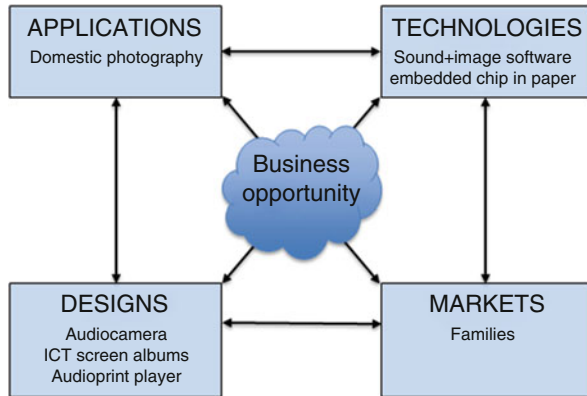business opportunity

Fig. 10.3 Example new developments in audiophotography since 2004

and audioprint technology in other market and application contexts within the company. This resulted in prototypes such as *Audiopaperclips* and *Active Photos* for an office worker market, and products such as *StoryMail* and the *HP Jornada* audiophoto software for a mobile professional market. Open innovation on the other hand tended to look for other technologies, markets and designs for the same domestic photography application. Examples here include prototypes such as *Context pictures*, *The Other Brother* and *FM radio*, or the *audiocamera apps* of Table 9.1. Exceptions apply to both areas, and in time the distinction seems to disappear, especially in open innovation, which is influenced by so many other factors and unconstrained by a single company strategy. This applies to my own work, which was able to consider the application of audiophotography to community journalism in developing communities through *Com-Phone*, and a more artistic implementation of the *Audiophoto Desk* as a public art installation in *Weegie*. Other examples include the *StoryKit* prototype for school children and the *RAW* prototype for capturing cultural life and heritage in developing communities.

One aspect of innovation not captured in these figures is the way in which one idea leads to another over time. This was another strong finding of the review that was illustrated by the story-like write up of examples in the text where I tried to preserve the sequential order of studies and systems. This can be seen most clearly in Chap. 8, which is framed by the tree diagram of Fig. 8.1 showing six related branches of research. One study follows another down each branch of the tree, and often builds on the findings and insights of previous studies. For example, on the Audio Memory branch of Sect. 8.5, Oleksik et al. (2008) study of domestic soundscapes led directly to a follow-up study on sonic gems (Oleksik and Brown 2008), which itself was inspiration for Dib et al. (2010) study of sonic souvenirs and later development of the *FM Radio* prototype (Petrelli et al. 2010). For this sequence, the domestic photography application was initially relaxed to include other aspects of domestic life, only to be returned to later in new solutions for memory recording, in and out of the home.

Because academic research in a university context is not framed by any business imperative to make money, it does not need to align all the elements of Fig. 10.2 into a coherent business opportunity for any prototype. Typically markets are underdefined in character and size. Alignment only becomes necessary for commercialization, as a 'knowledge transfer' task for industry. We consider this dynamic further in the next section in the context of reflecting on the time course of research and development activities and the relation between them. For now we can conclude that innovation involves the development of sequences or portfolios of design proposals over time, in either prototype or product form, and that these proposals make implicit assumptions about other key elements of innovation.

Tracing the arc of multiple designs over time and what they teach about various associated applications, markets or technologies would therefore be a good thing to do in specific innovation areas. This would help to understand what has and has not been explored, and to learn the lessons of design history in a current area of innovation. I have done some of this task here for the audiophotogaphy area, but better tools and techniques are needed for mapping the relationships between prototypes and products and showing their evolution over time. Developing such techniques is beyond the scope of this book, but there are five areas of research which would be worth looking at as starting points for this work.

Thinking and methodology on **design rationale**, mentioned above, was specifically aimed at articulating the implicit assumptions behind the design of interactive systems in the 1990s. A good introduction to the area is given in a special issue of the HCI journal edited by Carroll and Moran (1991) and also in their later book on the subject (Moran and Carroll 1996). Although this subsequently fell out of favour in the area, elements of the same idea have emerged in a recent recommendation for **annotated portfolios** as a way of capturing the lessons of design explorations (Bowers 2012; Gaver and Bowers 2012). Rather than theorise such lessons in scientific papers, the authors suggest the simple textual annotation of design portfolios for related design concepts or prototypes. This can be said to generate a kind of intermediate level design knowledge between artifacts and theory (Lowgren 2013). While design rationales capture the assumptions of individual systems, and

annotated portfolios convey the knowledge and approach of particular designers or studios, my recommendation is to map the evolution of 'naturally occurring' prototypes and products in particular innovation areas. This is much closer to a **design collection** or history project, or the kind of new techniques now used in **patent mapping** to visualise patent families sharing similar claims or citations. Bill Buxton has promoted the tracking of digital technology prototypes and products through his own *Buxton Collection*, now supported by Microsoft (http://research.microsoft.com/en-us/um/people/bibuxton/buxtoncollection/), and a number of patent mapping tools are now available including *PatSnap* which claims to embody the next step in 'IP analytics' (http://www.patsnap.com/home/report). Making sense of such collections or analytics may require the sensibilities of a historian or design critic in a kind of **interaction criticism**, which is another approach currently gaining ground in HCI (Bardzell et al. 2010). All these movements adopt the kind of evolutionary approach to understanding innovation that is strongly suggested by the audiophotography case study. However, none could quite capture the way in which different research and development issues have been explored by industry or academia in the audiophotography area, as reviewed above.

10.4 Fast Design, Slow Innovation

Returning to the overarching theme of the book, we can see in the evolution of audiophotography that design and innovation proceed at different rates over time. Design is represented by both prototypes and products on either side of the Research and Development divide, and these change rapidly in response to research findings and customer demand. We saw this in Chap. 8 through the academic arms race between universities and corporate research labs, as new prototypes were developed along strands of research about every year or two. It could also be seen in Chap. 9 in the vibrancy of the app market and the redundancy of competing offerings of audiocamera and audiophoto narrative apps in Tables 9.1 and 9.2. Innovation on the other hand is represented by the impact of design on everyday practice, and change in the established way something is done. This appears to have changed very slowly in terms of the practice of combining sound clips with photographs, and is only now beginning to emerge as a genre of video which is easy to generate as part of a photo slideshow, or a method of discussing still images on the web.

Related practices may even have declined, as in the case of capturing and sharing music and other recordings on cassette tapes. This is despite the great potential of audio recording demonstrated in the Audio Memory Sect. 8.5 of Chap. 8. What takes hold as a new activity in the consumer market is difficult to predict and highly sensitive to the cost and effort involved, as well as its fit to existing patterns of behaviour. In an age when recording music from the radio was the cheapest and easiest way to share it, cassette technology thrived. But in the current age where music is available digitally over the web, its manual recording in real time is no longer efficient or attractive. The management of music and other kinds of sound recordings are a key 'related practice' for audiophotography to fit with, as are

photography and video recording. In order for audiophotography to become more commonplace, the technologies of sound, image and video need somehow to become more **integrated**, so that audiophoto capture and sharing is just part and parcel of a new multimedia literacy and communication practice. I believe this is what we are starting to see in the digital storytelling area, which supports all kinds of audio/photo/video combinations through various narrative apps.

A final complication for innovation timescales is that they appear to overlap for different classes of user. In Geoffrey Moore's (2014) terms, there may be groups such as innovators, early adopters, early and late majority and laggards who take increasingly longer amounts of time to embrace any new technology. We saw this in the photowork diagrams of Figs. 9.1, 9.2, and 9.3 describing three eras of audioprint solutions. The continuing existence of analogue film cameras with silent and talking photo albums, demonstrates that there is still a market for the activities of Fig. 9.1 despite the range of new digital and 'post-digital' technologies and activities shown in Figs. 9.2 and 9.3. We must therefore be even more careful not to dismiss new technologies and products too early for their failure to reach the mainstream, since Moores' analysis shows that what innovators and early adopters use today, the mainstream may use tomorrow. It is just that 'today' and 'tomorrow' are more likely to be 15 or 20 years apart, rather than 3–5 years apart as folklore has it.

10.5   The Research-Practice Gap

While this message may be less surprising and less concerning to researchers in academia, colleagues and investors in industry may be more concerned. As we have seen, researchers don't need to commercialise their prototypes and see them become mainstream within a short period of time, whereas developers are paid to do just that. Perhaps one way to speed up innovation is for these two communities to work more closely together, and close what Norman has called the research-practice gap (2010). This is a cultural difference between research and development staff with different backgrounds and objectives, often working on similar technologies in different institutions. If only the bright ideas from university or corporate research labs could be handed off to companies and development units quicker and more efficiently, companies could turn them into products faster. Conversely, if academics understood the needs of industry better, then technology innovation might attend to market factors and be more closely tied to business model innovation (c.f. Baden-Fuller and Haefliger 2013).

The lesson of the audiophotography story here is that Norman is right to highlight the cultural divide between R & D. Researchers in either university or industry are part of a scientific community which meets to share findings and ideas on a regular basis through conferences and publications. In Chap. 8 we saw how this led to rapid development of independent research on audiophotography by a small cluster of interacting institutions and individuals, many of whom knew each other. This contrasts sharply with the picture in Chap. 9, of a raft of independent companies all bringing out similar products in the audiophotography area, without any formal

reference to each other or a sense of building on common work. No doubt some developers in these companies know each other from trade shows or forums on the web, but the culture of competition rather than collaboration is one which surely slows them down in relation to their scientific colleagues.

The mixture of commercial competition and scientific collaboration seen in the corporate R & D of Chap. 7, lies somewhere between the two cultural extremes, especially when open innovation is encouraged into the mix. In practical terms, this led to the rapid spread of ideas on audiophotography into a number of products and domains both inside and outside photography itself. At least here, staff on both sides of the research and development divide talk to each other, and collaborate to turn prototypes into products when there is a fit to business strategies and competencies. This then is a model for how research and development communities should interact more closely with each other in a common area, through actual contact, collaboration and the free flow of ideas. Mechanisms for facilitating this include the integration of conferences with exhibitions and product fairs, and publication venues for a range of contributions which acknowledge the different ways in which research and development knowledge is generated and documented. Staff exchange between universities and companies, or between corporate and divisional research labs is another mechanism which appears to happen naturally in high tech clusters like Silicon Valley, or multinational companies like HP. This may itself result in the kind of *translational developers* that Norman recommends, for translating the results of research into development actions, and development challenges into research questions.

10.6 Slow Design, Fast Innovation

Given the conclusions about fast design and slow innovation described above, it is interesting to note the existence of two movements for slow design and fast innovation. I comment on each of these briefly in reverse order, before summarizing the main conclusions of the book.

Fast innovation is an idea especially associated with the software industry and promoted through the agile development and lean startup approaches. Agile software development was a reaction to the experience of large-scale IT system projects of the 1990s which often went over time and over budget, due to over-planning, insufficient communication and 'all-at-once' delivery (Cooke 2012). The agile approach in contrast involves incremental planning and iterative development and testing to evolve a system with the engagement of stakeholders. The lean startup movement is a related approach promoted by Eric Ries (2011) to apply agile development to new companies or ventures. Ries recommends a process of fast innovation through shortening the time taken to carry out iterative Build-Measure-Learn cycles. Interestingly for this book, he advocates the use of a 'Minimal Viable Product' (MVP) early in this process, defined as "that version of the product that enables a full turn of the Build-Measure-Learn loop with a minimum amount of effort and the least amount of development time" (op. cit., p. 77). Examples include

a video of the Dropbox file storage service for users to sign up from, and a food recommendation service done initially in person by the CEO of *Food on the Table* for a single customer.

These ideas are sympathetic in principle to the evolutionary nature of innovation illustrated by the audiophotography story. In fact they use evolution as a driving principle of design. The MVP is particularly interesting in relation to the division we have found between prototypes and products in the computing industry. Prototypes are both an input and output of research, and act to collect new knowledge about user values and behaviours through some kind of conceptual, experimental or field-based test. Products are also a kind of input and output of development. They generate revenue whilst also gauging customer reaction to the associated design and business model. Usually there is a clear distinction between prototype and product, which may often be created by different research and development communities. In the lean startup approach, this distinction is collapsed to allow development companies to experiment more freely and repeatedly with different versions of a product in the market. This addresses the research-practice gap above by bringing more research and knowledge generation into the heart of development, and making the company more responsive (agile) to user needs. The downside of this approach is that customers may feel cheated with the lack of functionality and finish of early products, in the same way that participants react to prototypes. While this may speed up product and company development, it is harder to argue that it speeds up innovation itself. However many prototypes and products are tested on a market at any one time, it will still take a particular confluence of application, technology, market and design factors to cause an innovative shift in established practice worthy of the name. As we have seen, many of these factors are outside the control of individual companies and take years to move into a constellation adopted by a mainstream segment of the population.

Slow design is another movement which has much in common with an evolutionary model of innovation, but for quite different reasons. The slow movement started in the food industry of Italy as a reaction to the fast food revolution of the US. Its original values were to promote engagement with local produce, sustainable farming, and a respect for cultural traditions of preparing and eating meals slowly, in community with others. Slow design is a term coined by Fuad-Luke (2002) to apply these principles to a more sustainable process of design for wellbeing, which recommends fewer iterations of products over time to increase their value and longevity. It also recommends designs which cause us to slow down as human beings, into a less frenetic and more reflective pace of life. Hence his entry in the Dictionary of Design (Erlhoff and Marshall 2008) says this: *"Slow design positions itself again the "fast design" of the current industrial paradigm, which is governed by unsustainable cycles of fashion and over-consumption, business ethics, and an anthropology that defines everyone as customers. The use of "slow" as an adjective, or instructive adverb, deliberately introduces ambiguity in this context; it implies that time is implicit in all facets of (the) design, and that the purpose is to slow down the process, the outcome, and the effects of the outcome"*. In partnership with Carolyn Strauss, Alistair Fuad-Luke has written a set of principles for slow design

(Strauss and Fuad-Luke 2008). Strauss currently runs *SlowLab* as a platform for slow design projects (http://www.slowlab.net/) and the movement has now spread to other domains such as slow cities, travel, schools, books, lifestyles and money (http://www.slowmovement.com/). An example slow design project is the Slow Lloyd project which invited architects, designers, innovators, students and community members to imagine future uses and interactions with the Lloyd Hotel and Cultural Embassy in Amsterdam (http://slowlloyd.slowlab.net/).

Instead of promoting design evolution as a way of speeding up innovation (as in agile/lean development), slow design recommends it as a way of slowing down innovation and our reaction to it. In an age of rising consumerism across the globe, this is a timely and provocative message which may cause us to think differently about the *purpose* of design and its sustainability in a world of finite resources. The lessons of this review offer some comfort to this movement, in showing that despite the fast iterations of designs in any area, consumers are slower and more discerning than we realise in *adopting* technologies that are offered to them. True innovation is already a slow process and not controlled by any one organisation alone. Furthermore, human values remain remarkably stable in the face of technological change, with technology offering new ways of realizing values rather than offering new values themselves. Audiophotography is a good example of this, since it extends the memory, communication and identity values of photography by offering additional auditory cues and content to complement an image, and new ways of playing back the results on paper as well as screen. As a slow design project it has resulted in numerous related designs over many years, and encourages new forms of reflection on experience captured in auditory form.

10.7 Conclusions

What then can we conclude about audiophotography and innovation as a result of reviewing 10 years' worth of research activity in the field and a snapshot of current products?

Regarding audiophotography we can say that it is very much alive in current smartphone and tablet apps, and has been considerably advanced by numerous research studies and prototypes exploring new ways of combining sound and image for different purposes and markets. Promising new application areas have emerged in addition to domestic photography, including community journalism, literacy learning and cultural heritage capture. Furthermore, the initial possibilities for playing back sound from paper are getting closer to a commercial reality because of advances in printed electronics and augmented reality. As a result of these developments, there are business opportunities for cameras and camcorders to provide better support for audiophoto capture alongside photo and video shooting modes, and for social media sites to incorporate sound with pictures for enhanced storytelling and discussion. There is also an opportunity to improve the augmentation of photobooks with sound, through both online specification and offline addition of voiceover

or conversation. A key additional ingredient necessary to the future expansion of audiophotography is the formation of a stronger community of practice. This is beginning to be nurtured online amongst users of various audiophoto apps, but might be brought together across these groups by an overarching online community portal and a series of localized clubs. Synergies with the digital storytelling movement might be exploited here, or other communities such as oral historians, to promote the approach as a contribution to other domains.

Regarding innovation we have found that closed innovation can be used to explore technology transfer to other applications and markets, while open innovation can be used to find other solutions and technologies for the same domain. Although the design of prototypes and products can be accelerated or connected in various ways, researchers and developers should still expect adoption to take many years in any market, at different rates for different segments. In general, we should think of innovation as something which happens across institutions and time, in an evolutionary fashion. Its development can be steered but not controlled by any one individual or institution, and could be documented and mapped more effectively to identify trends, gaps and opportunities. This constitutes a research challenge in its own right, to reveal the history and provenance of prototypes and products in any area, and the connections between them over time.

Finally, the audiophotography story contains a number of personal implications for would-be innovators. I express these here in six recommendations:

1. Be persistent and don't give up on a good idea
2. Be smart and don't hold onto an idea which doesn't work
3. Treat failure as a learning exercise and advance in design
4. Treat success as a short term victory
5. Engage with the scientific community in your area to benefit from and contribute to research activity
6. Engage with the commercial community in your area to benefit from and contribute to development activity

Perhaps the overarching implication is that innovators are individual members of a global learning community. Their contribution is always situated and built on a history of previous contributions, whether they know it or not, and they have little control over the uptake of their ideas by others. The lesson of this book is to accept that, and enjoy the sociability, serendipity and unpredictability of innovation in all its forms.

References

Baden-Fuller C, Haefliger S (2013) Business models and technology innovation. Long Range Plann 46:419–426

Balsamo A (2011) Designing culture: the technological imagination at work. Duke University Press, Durham

Bardzell J, Bolter J, Löwgren J (2010) Interaction criticism: three readings of an interaction design, and what they get us. Interactions 17(2):32–37

Bijsterveld K, van Dijck J (eds) (2009) Sound souvenirs: audio technologies, memory and cultural practices, vol 2. Amsterdam University Press, Amsterdam

Bitton J, Agamanolis S, Karau M (2004) RAW: conveying minimally-mediated impressions of everyday life with an audio-photographic tool. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 495–502

Blythe M, Robinson J, Frohlich D (2008) Interaction design and the critics: what to make of the Weegie. In: Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges. ACM Press, New York, pp 53–62

Bowers J (2012) The logic of annotated portfolios: communicating the value of 'research through design'. In: Proceedings of the designing interactive systems conference. ACM, New York, pp 68–77

Carroll JM, Moran TP (1991) Introduction to this special issue on design rationale. Hum Comput Interact 6(3–4):197–200

Cooke JL (2012) Everything you want to know about Agile: how to get Agile results in a less-than-Agile organization. IT Governance Ltd., Ely, Cambridgeshire

Debaty P, Goddi P, Gossweiler R, Rajani R, Vorbau A, Tyler J (2004) Enabling informal communication of digital stories. HP labs technical report no HPL-2004-180

Dib L, Petrelli D, Whittaker S (2010) Sonic souvenirs: exploring the paradoxes of recorded sound for family remembering. In: Proceedings of the 2010 ACM conference on computer supported cooperative work. ACM Press, New York, pp 391–400

Erlhoff M, Marshall T (eds) (2008) Design dictionary: perspectives on design terminology. Walter de Gruyter, Berlin

Fennell J, Frohlich DM (2005) Beyond photographs: a design exploration of multi-sensory memorabilia for visually impaired people. HP labs technical report no HPL-2005-151

Frohlich DM (2004) Audiophotography: bringing photos to life with sounds. Kluwer, Dordrecht/Boston

Frohlich DM, Jones M (2008) Audiophoto narratives for semi-literate communities. Interact Mag 15:61–64

Frohlich DM, Sarvas R (2011) HCI and innovation. In: CHI'11 extended abstracts on human factors in computing systems. ACM Press, New York, pp 713–728

Frohlich DM, Clancy T, Robinson J, Costanzo E (2004) The audiophoto desk. In: Proceedings of 2 AD: second international conference on appliance design. Bristol, UK, pp 139–140

Frohlich DM, Robinson S, Eglinton K, Jones M, Vartainen E (2012) Creative cameraphone use in rural developing regions. In: Proceedings of MobileHCI '12. ACM Press, New York, pp 181–190

Fuad-Luke A (2002) Slow design' – a paradigm shift in design philosophy. In: Proceedings of the development by design conference, dyd02, Bangalore, 1–2 Dec 2002

Gaver B, Bowers J (2012) Annotated portfolios. Interactions 19(4):40–49

Harada S, Sato D, Adams DW, Kurniawan S, Takagi H, Asakawa C (2013) Accessible photo album: enhancing the photo sharing experience for people with visual impairment. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 2127–2136

Kindberg T, Tallyn E, Rajani R, Spasojevic M (2004) Active photos. In: Proceedings of DIS. ACM Press, New York, pp 337–340

Lindley SE, Monk AF (2005) Augmenting photographs with sound for collocated sharing. In: Home-oriented informatics and telematics. Springer, New York, pp 155–170

Löwgren J (2013) Annotated portfolios and other forms of intermediate-level knowledge. Interactions 20(1):30–34

Moore G (2014) Crossing the chasm: marketing and selling disruptive products to mainstream customers, 3rd edn. Harper Business, New York

Moran TP, Carroll JM (1996) Design rationale: concepts, techniques, and use. L. Erlbaum Associates Inc., Mahwah

Norman DA (2010) The research-Practice Gap: the need for translational developers. Interactions 17(4):9–12

Oleksik G, Brown LM (2008) Sonic gems: exploring the potential of audio recording as a form of sentimental memory capture. In: Proceedings of the 22nd British HCI group annual conference on people and computers: culture, creativity, interaction, vol 1. British Computer Society. Liverpool, UK, pp 163–172

Oleksik G, Frohlich D, Brown LM, Sellen A (2008) Sonic interventions: understanding and extending the domestic soundscape. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 1419–1428

Petrelli D, Villar N, Kalnikaite V, Dib L, Whittaker S (2010) FM radio: family interplay with sonic mementos. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 2371–2380

Piper AM, Weibel N, Hollan JD (2012) TAP & PLAY: an end-user toolkit for authoring interactive pen and paper language activities. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 149–158

Quinn A, Bederson B, Bonsignore E, Druin A (2009) StoryKit: designing a mobile application for story creation by children and older adults. Tech. rep. HCIL-2009-22, Human Computer Interaction Lab, University of Maryland, College Park

Rajani R, Vorbau A (2004) Viewing and annotating media with MemoryNet. In: Proceedings of CHI 2004. ACM Press, New York, pp 1517–1520

Reis E (2011) The lean startup. Crown Business, New York

Ritchin F (2009) After photography. WW Norton & Company, New York

Sarvas R, Frohlich DM (2011) From snapshots to social media-the changing picture of domestic photography: the changing picture of domestic photography. Springer Science & Business Media, London/New York

Strauss C, Fuad-Luke A (2008) The slow design principles-a new interrogative and reflexive tool for design research and practice. Changing the change. International Design Association, Torino