The sound of walking

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

16.11.1995

Rapport no. 1084

The sound of walking

J.H. Eggen

Voor akkoord: Dr.ir. J.H. Eggen

The Sound of Walking

J.H. Eggen

The sound of walking *

J.H. Eggen

Abstract

The effective use of sound as an alternative communication channel in interactive multi-media applications requires a better understanding of the human ability to listen to the sources of sound. In this report we present an exploratory study on the mapping of source events involved in human walking onto the sound of walking. We discuss a model to analyse and synthesize the temporal organisation of human walking sounds. We also introduce a methodology for studying the perception of basic auditory walking events.

1 Introduction

Sound is caused by objects interacting at certain locations in the environment. People can extract information from the sound about the type of interaction, the objects involved, the location, and the environment; people can listen to the sources of sound (Gaver, 1993a). People use this ability extensively when they perform day-to-day tasks. However, it is virtually unused in the interaction of people with systems based on information technology.

Imagine an advanced game (multi-media, multi-agent, multi-user). The user can interact with other users, game characters and objects in the 'virtual' game world. At the same time, other users and autonomous game characters also play their roles in the scenario. In such highly unpredictable complex environments sound can play an important role. By using sound cues to

*The research described in this report was performed when the author was a guest researcher at the MRC Applied Psychology Unit (APU) in Cambridge, England. It was done in collaboration with Roy Patterson of APU and Bill Gaver of the Royal College of Art in London.

monitor what is happening in the environment the auditory channel can be used as an alternative to the visual channel. In this way, sound not only reduces visual work load but also increases the believability of the interactive scenarios.

The European ESPRIT project *Humanoid* aims at creating tools for the entertainment industry for embedding autonomous realistic virtual humans in games, multi-media titles, film animations, and interactive TV. Part of the research effort is directed towards tools for generating soundtracks for interactive Humanoid scenarios.

In animation a sound engineer normally constructs a soundtrack after the graphical work has been completed. This is not possible in the Humanoid case. Humanoids are autonomous characters whose exact behaviour is by definition unpredictable. Also the interaction of the user with the system is to a large extent unpredictable. These factors make it impossible to specify the soundtrack before the actual (inter-)action takes place. The only option is to generate the appropriate sounds in real-time.

Within the Humanoid project we take an event-driven approach. An evolving scenario is seen as the result of concurrent streams of events. Some events produce sound and some do not. To create a high quality soundtrack a good auditory event to sound mapping is needed.

In this report we present research on the sound producing events involved in walking. We start with a detailed look at human walking. Next, we discuss the temporal patterning of basic auditory events as they occur in natural walking sounds. Finally, we discuss the basic auditory events themselves.

2 Human Walking

To understand how the sound of walking is produced, it is important to look more accurately at human walking. Figure 1 shows a sequence of photographs of a walking man (Napier, 1967). We can distinguish two different phases in the walking cycle: the *stance* phase during which the foot is in contact with the ground and the *swing* phase where it moves through the air.

The left leg, that is the leg nearest to the wall, just entered the stance phase

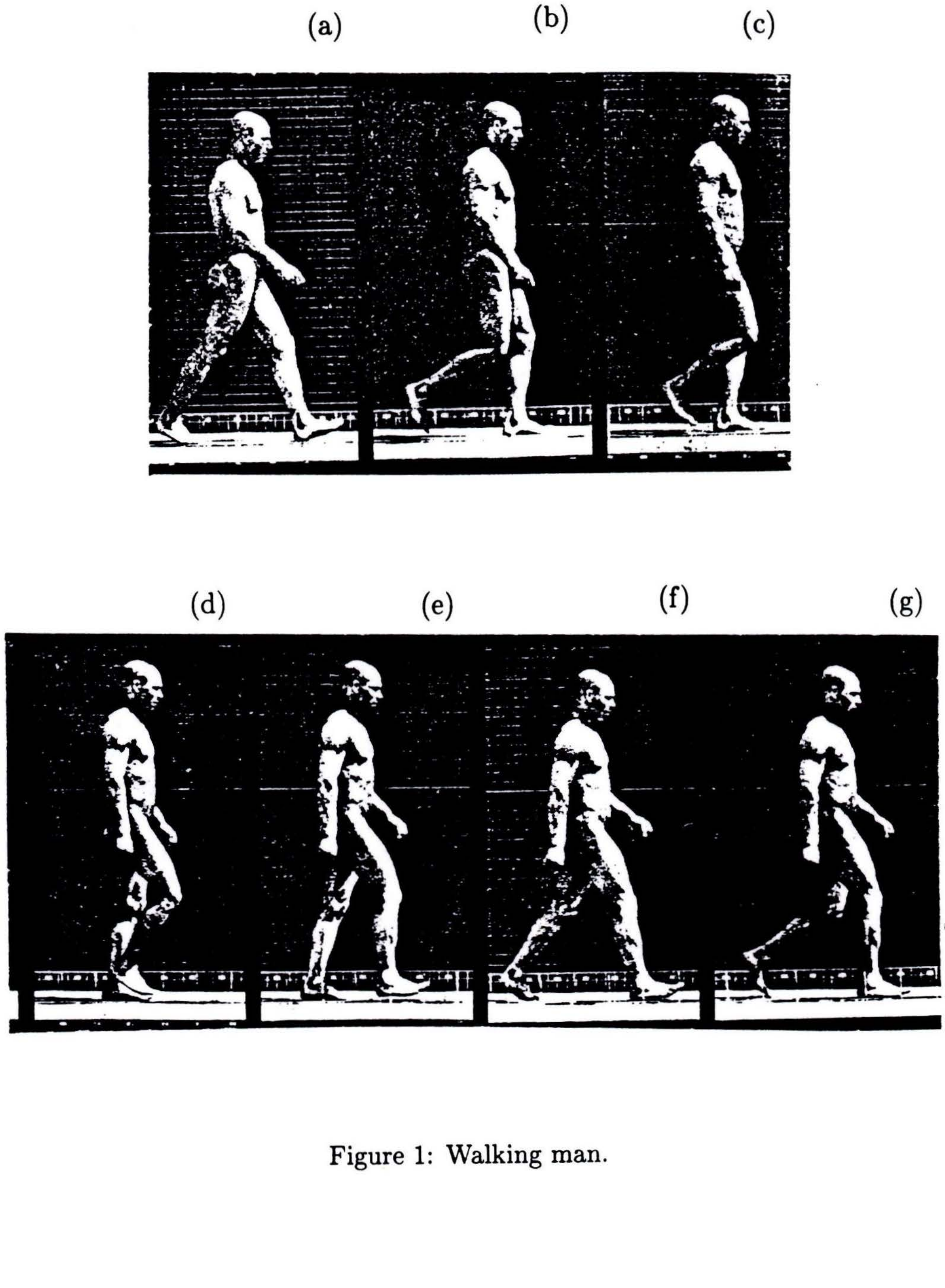


Figure 1: Walking man.

by striking the floor with the heel (figure 1a). The heel strike is followed by the foot-flat strike (figure 1b). In figure 1f, the heel is no longer in contact with the ground, and in figure 1g we see that the toe is about to leave the ground.

The toe-off event initiates the swing phase. The leg in the front of the picture, the right leg, swings through the air until the heel strikes the floor again (figure 1f).

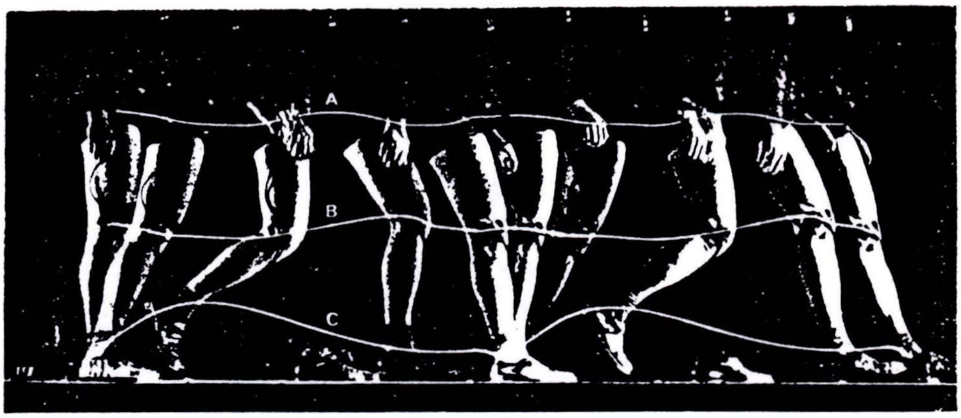
One complete walking cycle is defined as the time interval between the heel strike of one leg and the next heel strike of that same leg. Walking is a rhythmic alternating activity of both legs. As can be seen in figure 1, the stance and swing phases of both legs overlap during normal walking. The time interval during which both feet are on the ground is called the double-support phase as opposed to the single-support phase where only one leg touches the ground.

Events like *heel strike, foot-flat strike, heel off,* and *toe off* are potential candidates for producing sound. We would like to stress that both legs produce these events and that in the actual soundwave the corresponding sounds are mingled.

It is difficult to derive the exact timing of the various auditory events from pictures like figure 1. Therefore, researchers developed a technique in which small lights are attached to the different joints. Figure 2 shows an example of such a recording. Lights were attached to the hip, knee, and ankle. The picture shows the location of the dots at fixed time intervals. By measuring the distance between the dots it is possible to estimate the speed of the various parts.

In this study we used a computer model of human walking to derive the temporal patterning of auditory walking events. This model was developed by Cutting and his collegues in their research on the perception of walkers as dynamic point-light displays (Cutting, 1978a; Cutting, Proffitt, and Kozlowski, 1978). Figure 3 shows how the lights were distributed on the bodies of the walkers. Cutting and his collegues found that subjects instantly recognise that the moving dot displays are produced by walking persons. What is even more amazing is that people can recognize the gender of a walker and whether or not the walker is somebody they know (Cutting

(a)



(b)

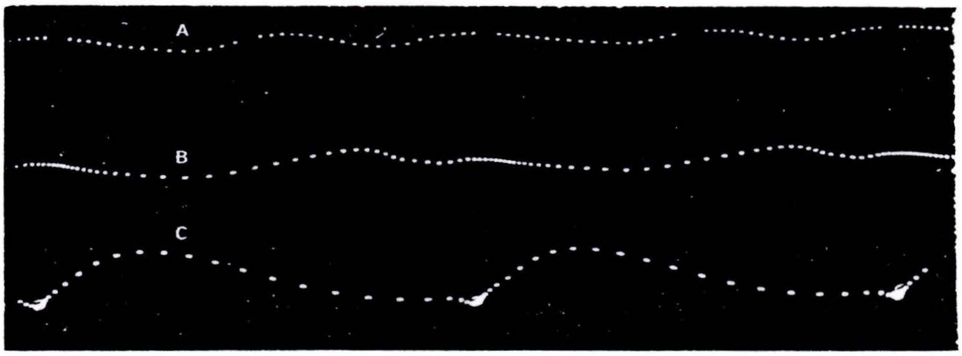


Figure 2: Moving point-lights attached to the hip (A), knee (B), and ankle (C). Two walking cycles are shown.

and Kozlowski, 1977; Kozlowski and Cutting, 1977; Barclay, Cutting, and Kozlowski, 1978). Cutting developed an algorithm for synthesising dynamic point-light displays on a computer screen (Cutting, 1978b). With these synthesised dot patterns he could reproduce the same perceptual effects which had been found with the lights attached to human walkers.

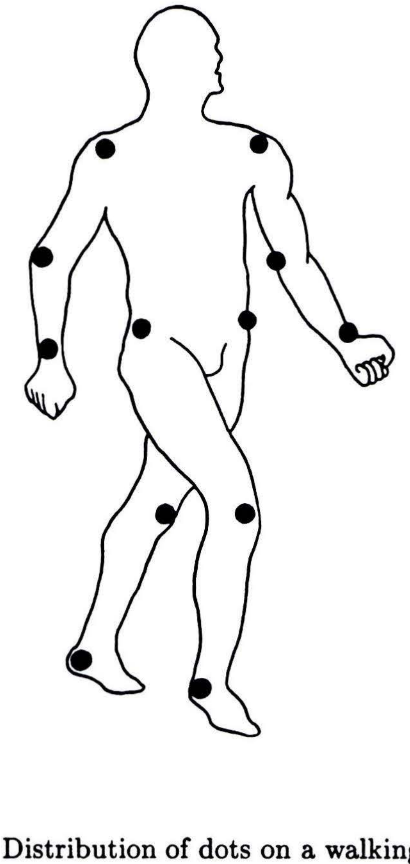


Figure 3: Distribution of dots on a walking man.

We used his model to derive more exact timing information for the various walking events. Figure 4 shows the position, speed, and acceleration of both ankles as a funtion of time. It is possible to estimate the timing of the heel-off, heel-strike, foot-flat strike, and toe-off events from these curves.

The curve which displays the vertical ankle position as a function of time (figure 4a) has two local minima per cycle. The absolute minimum corresponds to the heel-strike event (`hs`) and the other local minimum corresponds to the heel-off (`ho`) event. The vertical ankle speed is zero for both events.

When the heel has struck the ground the vertical speed of the ankle increases until the foot ball hits the ground. From that moment on the vertical ankle speed decreases (figure 4b). This local maximum defines the foot-flat strike event (`ff`).

The stance leg provides the propulsive force that drives the body forward. The walker applies this force by using muscular energy, pushing against the ground with the ball of his foot and then with his big toe (Napier, 1967). In both cases there is an instantaneous increase in the vertical acceleration of the ankle. The right leg in figure 4c (thick line) enters the figure as stance leg. The first discontinuity corresponds to the toe-off event (`to`). The second discontinuity, just before the heel-off event, marks the pushing of the foot ball against the ground.

With the information shown in figure 4 we can now define much more accurately the temporal patterning of auditory walking sounds. In the next section, we will use this knowledge to determine the various acoustic events as they occur in the waveforms of various natural walking sounds.

3 Temporal organisation of walking sounds

In this section we first use the model introduced in the previous section to analyse various human walking sounds. Next, we describe how the model can be used to synthesize walking sounds. In this section we do not analyse or synthesize the isolated auditory events themselves but instead look how these events are temporally organised in natural walking sounds.

The walking sounds used in this study were taken from commercially available compact discs (Aware Inc, 1993; BBC, 1993).

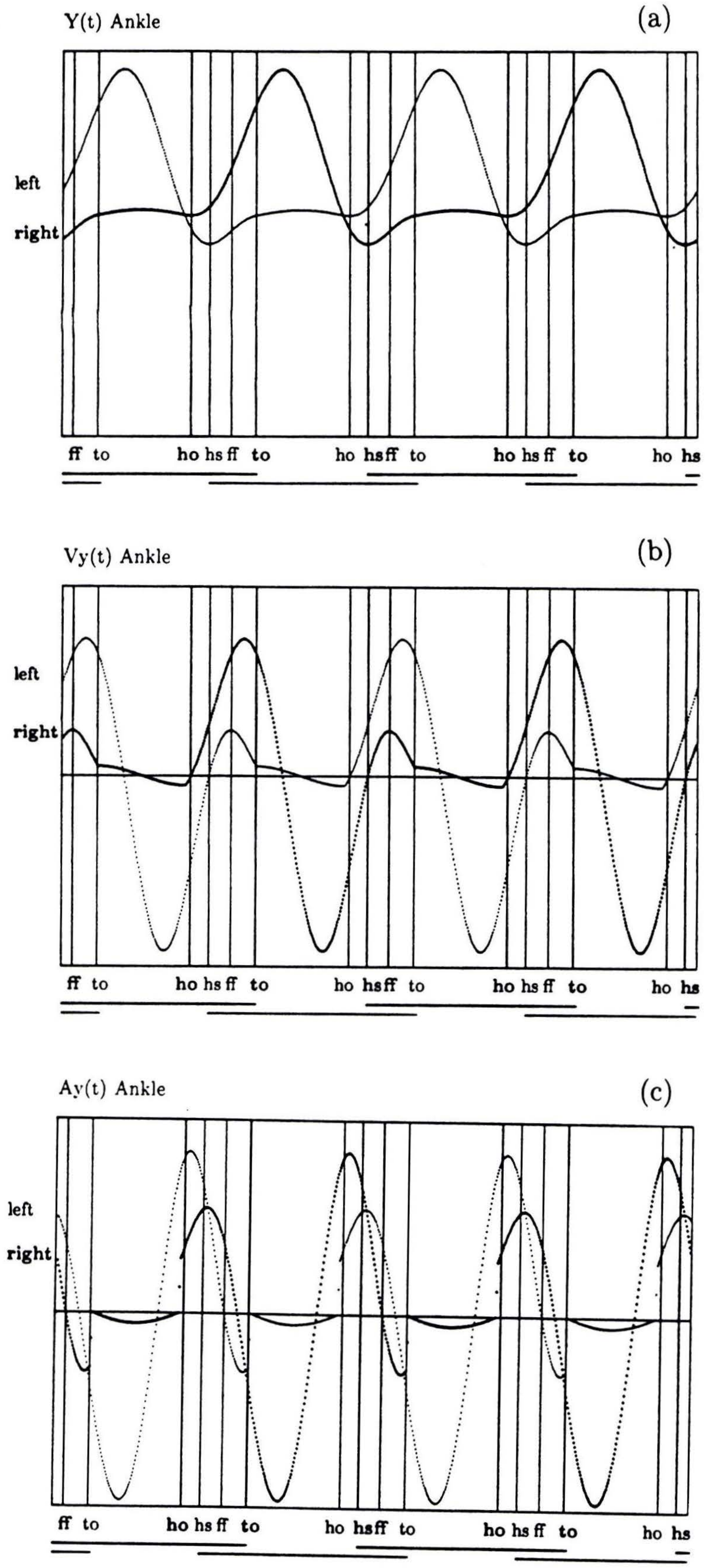


Figure 4: Position (a), speed (b), and acceleration (c) of the ankles as a function of time.

3.1 Analysis

The model predicts the following time order of auditory events during normal walking: heel strike right foot, foot-flat strike right foot, toe off left foot, heel off right foot, heel strike left foot, foot-flat strike left foot, toe off right foot, heel off left foot, heel strike right foot. This timing information, combined with careful listening, enables us to identify the different acoustic structures which occur in the waveforms of different walking sounds.

Figure 5a shows a 5-s excerpt of a 31-s recording of a female walking on a wooden floor with low solid heels (Aware Inc, 1993). The main peaks show the alternating heel strikes of both feet. We can also see the much smaller transients of the foot-flat strikes positioned in time just after the heel strikes. If we zoom in on the time interval around the main peak we can see the heel strike and foot-flat strike in more detail (figure 5b). Both sound events can be characterised as impact sounds. Not only do they reach their maximum magnitude almost instanteneously, they also show fast decay times (about 40 ms for the heel strike and 15 ms for the foot-flat strike). Figure 5c shows another excerpt of the same walking sound. This time we can identify a toe-off scrape. This is a scraping sound which is produced by the foot opposite to the foot which produced the heel and foot-flat strikes. The toe-off scrape is a noisy sound with equal attack and decay times.

Figure 5d shows a 290-ms excerpt of 64-s walking sound produced by a male walking on a country road (BBC, 1993). Although the heel strike again corresponds to the most prominent impact sound, this time we can also see many more small impact sounds. One of them is the foot-flat strike. The other impact sounds are produced by stone crunches.

Figure 5e shows a 307-ms excerpt of a 29-s walking sound of a male walking on a creaky wooden floor with shoes with solid leather soles (Aware Inc, 1993). The sound starting at about 230 ms is produced by the force applied to the creaky floor by the toe of the foot opposite to the foot which produced the heel and foot-flat strikes. Figure 5f shows a 698-ms excerpt of the same walking sound, but now the creak occurs at a different place in time, namely during the single-support phase. During this phase the body completely rests on one leg and this causes the creaky sound. We could also sometimes identify creaky sounds which were positioned just before a heel strike and which overlapped the heel strike itself. These creaky sounds are

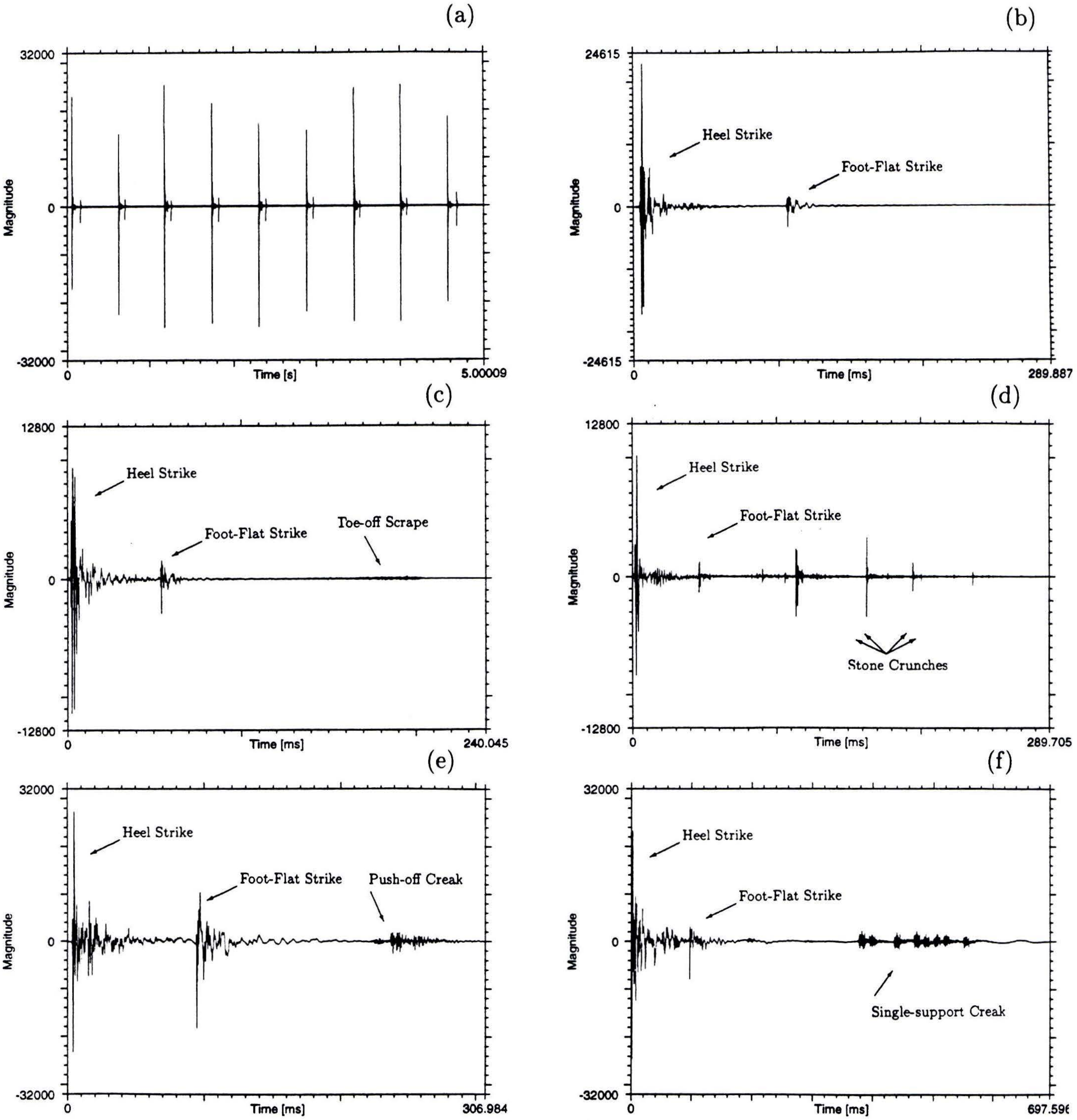


Figure 5: Waveforms of walking sounds showing various auditory walking events (for explanation, see text)

caused by the force applied to the floor by the foot ball of the foot opposite to the foot which produced the neighbouring heel strike.

In the same manner as we identified the auditory events described above we analysed many other walking sounds. Table 1 gives an overview of the sounds which were analysed.

Table 1: Walking sounds used in the present study.

Sound description	Duration (s)	Source
Male, Concrete, Leather Sole	30	Aware
Female, Concrete, Heels	30	Aware
Male, Creaky Floor, Leather Sole	29	Aware
Female, Creaky Floor, Heels	30	Aware
Male, Mud	31	Aware
Female, Mud	31	Aware
Male, Creaky Floor, Rug, Leather Sole	30	Aware
Female, Creaky Floor, Rug, Heels	31	Aware
Male, Wood, Leather Sole	27	Aware
Female, Wood, Low Heels	31	Aware
Male, Pavement	60	BBC
Female, Pavement	59	BBC
Male, Wooden Floor	63	BBC
Female, Wooden Floor	62	BBC
Male, Country Road	64	BBC
Female, Country Road	64	BBC

3.2 Synthesis

Based on the analysis outlined in the previous subsection we can label the various auditory events which occur in the human walking sounds of Table 1. This information can be used to segment the walking sounds and construct a database of auditory walking events. This database of sounds can be used to synthesize walking sounds.

The model used for the analysis of walking sounds was also used as a tool for synthesising them. Based on information about the walking conditions, a selection of the appropriate walking sounds can be made from the database. The selected sounds are positioned in time according to timing information generated by the model. At the moment, the actual synthesizer comprises two modules.

The first module generates an event list. All events in this list have a label and a time stamp. The label describes the type of auditory event. Currently, the following event labels are generated: heel strike right, foot-flat strike right, toe off left, heel off right, heel strike left, foot-flat strike left, toe off right, heel off left, heel strike right. The time stamp of an event describes the position in time of that particular event. In the first module the following parameters can be controlled: the hip excursion (**hex**) parameter influences the relative timing of the events of the walking cycle. Together with the shoulder-to-hip-excursion-ratio (**sher**) it is used to manipulate the perceived gender of the synthetic walking dot patterns. Whereas the **hex** parameter does influence the timing of the auditory events, the **sher** parameter does not. Stepsize and walking speed are the other two parameters which can be used to manipulate the timing of the auditory events.

The second module processes the event list generated by the first module. The output of the module is a sampled data file of a walking sound. This module has the following parameters: selection of basic events, event probability, and timing irregularity. The labels of the event list generated by the first module are expanded. For instance, it can be specified whether the sounds were generated by a male or a female walker. Also, the walking surface can be chosen (for instance, wooden or concrete floor). The probability whether or not a particular auditory event will be used in the actual synthesis of the complete walking sound can be specified. For instance, by increasing the probability of a toe-off event from 0 to 100 % we can manipulate the 'level of perceived creakiness' of a wooden floor. It was found that small irregularities added to the timing information of the original event list improve the naturalness of the synthesised walking sounds. The degree of irregularity is variable.

The two modules, together with the database of auditory events, enabled us to synthesize a wide range of different walking sounds.

3.3 Temporal organisation: suggestions for future research

During our study on the temporal patterning of walking sounds we encountered a number of interesting research questions. Due to time constraints we were not able to pursue these questions any further. We think it would be worthwhile to conduct future resarch on these questions.

One such question relates to the perception of the temporal organisation of the walking sounds. We would like to know, for instance, if and how temporal organisation influences the perceived gender of a walker. It would also be interesting to know how sensitive humans are to the relative timing of the various auditory events of the walking cycle.

It would not be very difficult to develop a segmentation program which could automatically determine the type of auditory event and its position in time in a human walking sound. In fact, we already implemented a preliminary version of a program to automatically identify and segment heel strikes and foot-flat strikes from the waveform of a walking sound. We think such a program could be used to observe and analyse gait disorders. As far as we know, the analysis of walking sounds is not used in clinical practice to investigate neurological damage or other impairments to the human motor system (Rosenbaum, 1991). The analysis of walking sounds could be an easy and economic way to assess gait disorders.

Unfortunately, for practical reasons, we were not able to synchronize the moving dot patterns to the sounds. Ultimately, we would like to build a sound-producing dynamic point-light display walking synthesizer which would allow real-time manipulation of the various parameters we mentioned earlier. We would like to use this synthesizer to study the linkage of visual and auditory modalities in the perception of human walking.

Besides the temporal organisation of the walking sound, the sound quality of the different types of auditory events which make up the sound is also important.

At this stage we still need a database of sampled sounds which correspond to the different walking events. This means that, in practice, we have to record al possible combinations of sound source variations we want or have to support. It would be much more convenient if we could synthesize or modify the sound events in such a way that we can control the sources of the sound directly. In order to do this we have to know much more about the analysis, synthesis, and perception of the auditory events themselves. This will be the topic of the next section.

4 Basic auditory walking events

To study the perception of basic auditory walking events we adopted a methodology which is frequently used in speech research. Figure 6 gives an overview of the methodology (Li, Logan, Pastore, 1991). In a first stage, one investigates which source events can be perceived by human listeners. In a second stage, one tries to find nonarbitrary acoustic structures produced by the auditory events, and in a third analysis step, one tries to evaluate the mapping between acoustic structure and perception.

In this section we will discuss the perception of auditory walking events from these three different angles. But first, we review previous research on the perception of nonspeech sounds which is relevant to the present study.

4.1 Previous research

As we have seen, the most prominent auditory walking events are the heel strike and the foot-flat strike. Both sounds belong to the class of impact sounds. Recently, some studies on the perception of nonspeech impact sounds have been reported in literature.

Repp (1987) studied the sound of two hands clapping. He found that different persons produce different clapping sounds, and that the spectral variability of these sounds is mainly due to differences in hand configuration. The results of Repp's study support the hypothesis that sound produced by a natural source conveys perceptible information about the configuration of that source (Repp, 1987).

Gaver (1988) investigated how well people can judge the material and lengths of bars from the sounds they make when they are hit. He found that it is easier for people to judge the material than it is to judge the length. Gaver developed a physical model to understand the mapping of attributes of the sources and the sound they make. He not only used this model to explain his results, but also to build a synthesizer for generating impact sounds.

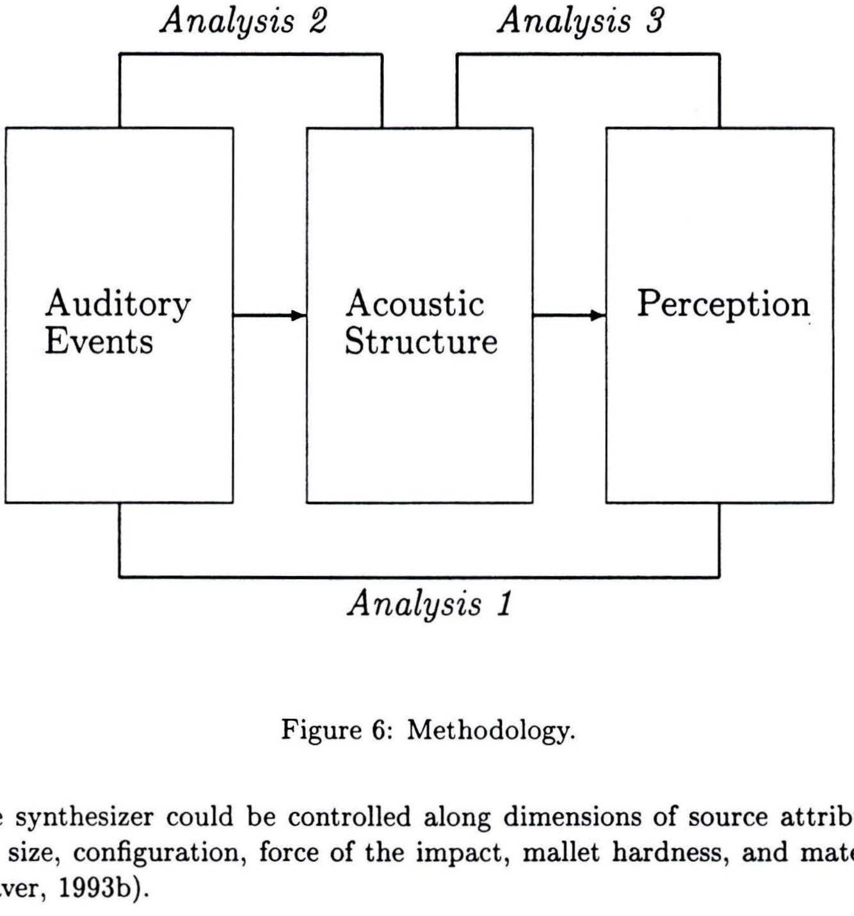


Figure 6: Methodology.

The synthesizer could be controlled along dimensions of source attributes like size, configuration, force of the impact, mallet hardness, and material (Gaver, 1993b).

Freed (1990) studied the auditory correlates of perceived mallet hardness. The sounds were made by striking metal cooking pans with various percussion mallets of different hardness. He found that harder mallets indeed correlate with higher perceived mallet hardness ratings. People could clearly ignore pan-related aspects, such as pitch, and focus on mallet-related aspects of the sounds, or in Freed's words: "listeners are able to separate out interacting dimensions of a stimulus in order to make an environment-oriented judgment" (Freed, 1990, page 319).

The only study, we are aware of, which investigated walking sounds is that of Li, Logan, and Pastore (1991). Li *et al.* (1991) showed that listeners were able to identify the gender of a walker on the basis of acoustic information present in the waveform. They found that identification of walker gender is determined by the integration of multiple factors like anthropomorphic differences among walkers, and differences in shoes. Li *et al.* (1991) concentrated on walker gender. In the next section we discuss which other source characteristics of human walking might be perceived by listeners.

4.2 Auditory Events - Perception

The studies discussed in the previous subsection all demonstrate the human ability to identify properties of acoustic source events. All studies started to examine the relationship between auditory events and perceptual categorisation of the stimuli. For instance, Li *et al.* (1991) studied whether the gender of walking sounds produced by male and female walkers could indeed be identified as male or female. The choice of Li *et al.* (1991) to investigate walker gender was inspired by their observations that we show "an excellent and consistent ability to identify gender, and even the identity of specific colleagues, based solely upon the sound of their walking in the hallway outside our offices" (Li *et al.*, 1991, page 3036). A second motivation to study walking sounds was that these sounds "represent a class of ecologically valid stimuli which impose a nonarbitrary structure on the acoustical waveform. ... this investigation may provide insights for understanding the perception of other complex stimuli" (Li *et al.*, 1991, page 3036).

Although we share the motivations of Li *et al.* (1991), the main reason for us to study walking sounds is the fact that we have to synthesize these sounds in interactive multi-media applications. As explained in the introduction of this report, the use of sound in interactive scenarios calls for a thorough understanding of the mapping of auditory events to sounds. Although walking sounds form only a small subset of the total set of sounds that has to be synthesised in real-time, their importance is demonstrated by the effort usually spend in film industry to record them.

Nowadays, walking sounds for film are still very often recorded life in a so-called Foley theatre[1] (Hartstone and Spath, 1994). In the Foley theatre,

[1]The term Foley is used for sound effects which are recorded in synchronisation with the pictures after production. It is named after Foley who was an effect operator.

besides other props, there are various surfaces for walking, e.g., road, pavement, solid and hollow wooden floors, sand, gravel troughs, and a water tank. Footsteps artists, or 'Foley walkers' are skilled in their art and have experience in dancing. They can very quickly assimilate the rhythm of a walk. They usually work in pairs and bring a suitcase along filled with a variety of shoes (Hartstone and Spath, 1994).

Instead of recording walking sounds in a Foley theatre, they can also be extracted from sound-effects CDs. Table 1 lists some soundfiles taken from sound-effects CDs.

From the way Foley artists create the appropriate walking sounds, and from the descriptions of walking sounds on CDs, we can get an idea of the various source characteristics which might be of importance for the perception of walking sounds. Besides the already mentioned walker gender, walking surface, shoe type, and walking rhythm also seem to be perceptually relevant source characteristics.

The study of Li *et al.* (1991) showed that it is difficult to investigate these characteristics separately. Their results showed that perceived walker gender was not only determined by anthropomorphic measures of walkers, but could also be influenced by shoe characteristics. Unexpectedly, their durational analysis of walking sounds did not show an effect of temporal organisation of auditory walking events on perceived walker gender [2].

In future research we would like to verify and extend the findings of Li *et al.* (1991) and derive more explicit relationships between the various source characteristics involved in walking and their corresponding sound percepts. In particular, we are interested in the perception of walking surface, and whether or not listeners can separate out this interaction dimension from a dimension like shoe type. In the next section, we discuss the possible effects

[2]We think this is in part due to errors in the durational analysis. For instance, Table III in Li *et al.* (1991) on page 3042 shows that for all walkers the temporal proportion for the stance phase is systematically smaller that the temporal proportion for the swing phase. This would mean that during a certain portion of the walking cycle the walker would have both feet of the ground. This is certainly not the case in normal walking. We think this is due to a wrong determination of both phases from the waveform. See, Li *et al.* (1991), page 3041, FIG. 2, and the explanation in the first alinea of section 2. Duration analysis.

the various source characteristics might have on the acoustical waveform.

4.3 Auditory Events - Acoustic structure

There are various ways to study the structure of acoustic waves. The classical Fourier transform is still the most frequently used method for analysing the spectrum of a soundwave. However, several alternative high-resolution techniques have been proposed which perform better than classical estimation methods especially in cases where highly damped signals have to be analysed (Laroche, 1993).

Another approach is to analyse the acoustic wave with algorithms which model the human auditory system (Patterson *et al.*, 1992; Patterson and Holdsworth, 1993). Patterson *et al.* (1995), for instance, developed a set of tools to study the structure of everyday sounds at different stages of the human auditory processing. The first stage models the spectral analysis which takes place in the peripheral auditory system. The output of this stage is a simulation of basilar membrane motion. The second stage models the conversion of basilar membrane motion into a neural activity pattern. The third stage models the central processing required to transform the neural activity pattern into the auditory sensations we hear when presented with the sound. The output of all three stages is a representation of the sound which can be used to study acoustic structure.

The four panels of figure 7 show the simulated basilar membrane motion of four different sounds. Each sound consists of five different 50-ms heel strikes which were extracted from one particular human walking sound (BBC, 1993). The five heel strikes were concatenated to increase the amount of information in the pictures. Each panel of figure 7 corresponds to a different walking condition: a. male walker on wooden floor; b. female walker on wooden floor; c. male walker on concrete floor; d. female walker on concrete floor [3]. Figure 7 suggests some systematic differences between the wood/concrete and male/female conditions: the frequency regions of exci-

[3]These sounds produce an auditory sensation of a drum role on a surface with a characteristic material-like quality. For instance, for the wooden-floor condition it sounds like the stimulus is produced by a woodpecker. Inspired by Patterson (1994a; 1994b), we played the sounds backwards. Careful listening suggested that the wooden timbre of the 'woodpecker' sound disappeared when played backwards and that it was much harder to distinguish it from a reversed concrete drum role sound.

tation seem to be related to the walking surface, and the force of the impact seems to be stronger for males compared to females. In general, however, it is difficult to define acoustical structures which correlate with these source characteristics, especially if we take other recordings of the same conditions into consideration.

The nonspeech studies mentioned above all derived timbral predictors (statistical measures) from acoustical representations of the sounds (Repp, 1989; Freed, 1990; Li *et al*, 1991). Although it was shown that these measures can be used to classify the stimuli along the various sound-source dimensions involved in the sound production, they give no insight into the relation between the physical characteristics of the sound sources and the acoustic structure of the soundwave. To obtain this kind of knowledge we need a physical analysis of the sound production process itself.

Gaver argued that physical analyses can be useful both in suggesting perceptible source attributes and in indicating which acoustic information present in the mass of data produced by acoustic analyses is related to these attributes (Gaver, 1993b). Gaver developed a physical model to understand the effect various source characteristics have on the sound of struck bars (Gaver, 1988). In the remainder of this section we discuss some physical aspects of the source characteristics of auditory walking events.

The walking sounds we studied were all produced by shoes interacting with a walking surface. The type of interaction between shoe and surface characterises the auditory event. For instance, we saw that scraping the surface during a toe-off event produces a sound which is very different from the impact sounds caused by heel strikes or foot-flat strikes. The force of the interaction determines the overall amplitude of the sound and the bandwidth of vibration. Based on this knowledge, Figure 7 seems to suggest that the forces of interaction involved in male walking are stronger than those involved in female walking.

From musical acoustics we know that the excitation spectrum of the impact between a mallet and a percussion instrument is determined by mallet head velocity, mallet material (mainly hardness), and mallet configuration (mainly contact area) (Fletcher and Rossing, 1991). Figure 8 shows an example of this dependency. By selecting different sticks or mallets, percussionists can influence the timbre of the instrument considerably. Hardness

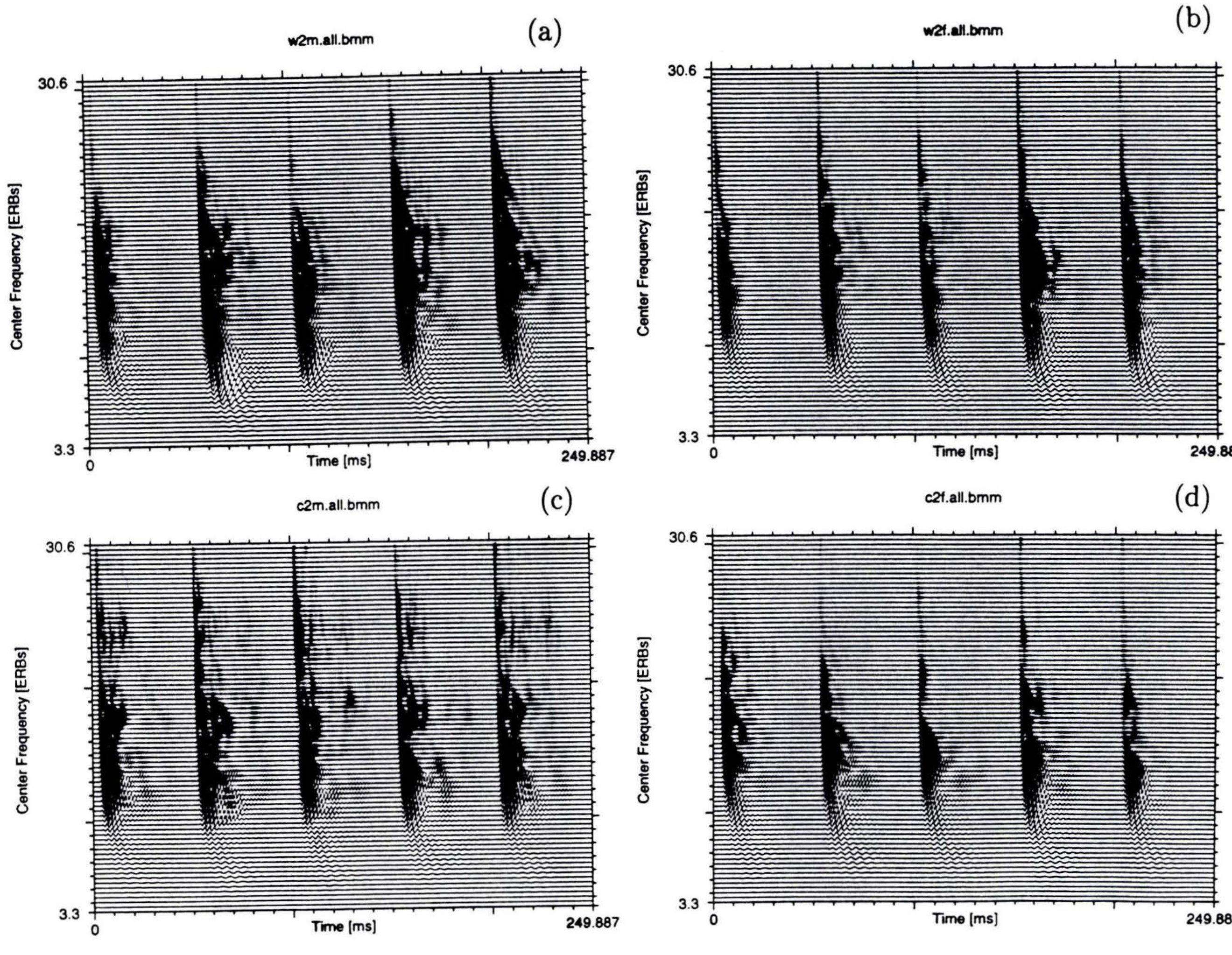


Figure 7: Basilar membrane motion of four different 'five-concatenated-heel-strikes' walking sounds (for explanation, see text).

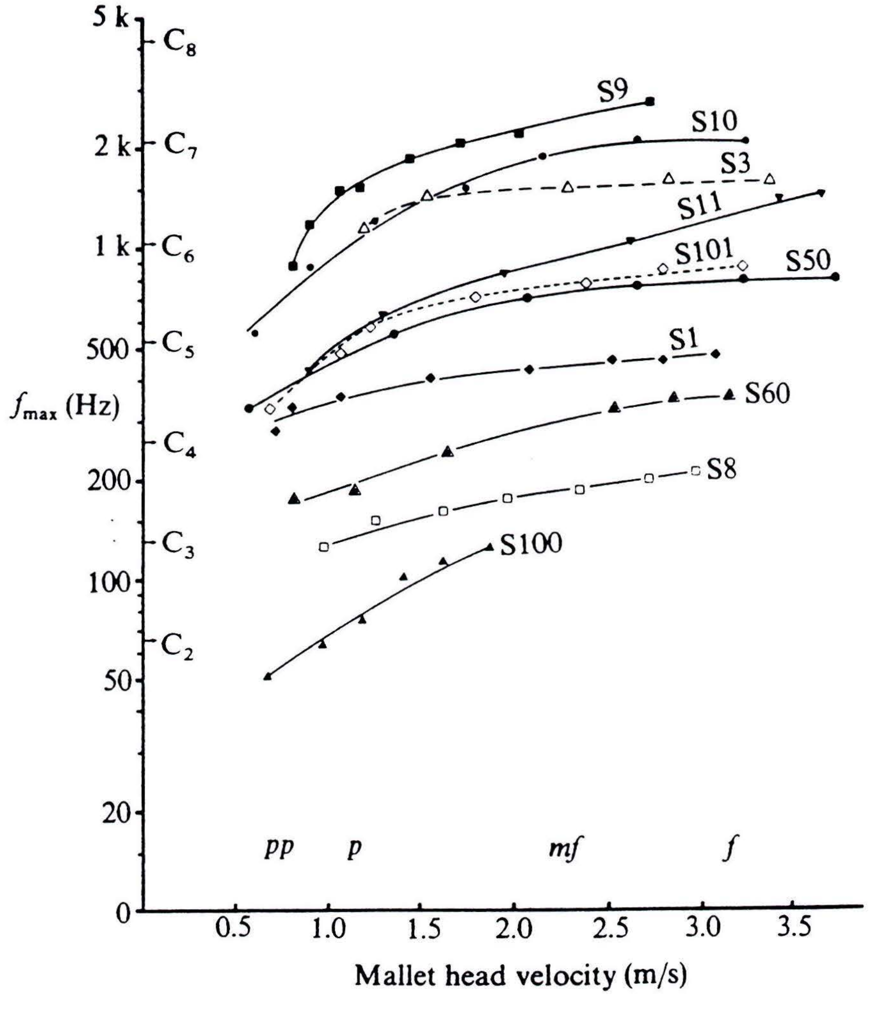


Figure 8: Dependency of the frequency of maximum excitation on mallet head velocity of different mallets (Fletcher and Rossing, 1991). S1 is an all-rubber mallet, S9 is a wooden mallet with rubber ring, and mallet S10 has a hard core wrapped with yarn.

and contact area are the two properties of the beater that have most effect on timbre, but also contact time and beater spot play a role (Rossing, 1982). For walking sounds this implies that when the shoe hits the surface the frequency of maximum excitation is also determined by the velocity of the shoe and the material and configuration of the shoe. The hardness of shoes (e.g., solid heels versus sneakers) and the contact area (e.g., low heels versus high heels) can vary considerably and as a consequence such source characteristics can greatly influence the walking sound.

Besides the effects of shoe material on the sound, the material of the walking surface also plays an important role. The material density and the strength of the restoring forces both effect the frequencies of vibration of the surface. The damping characterisitcs of the material will influence the overall time-course of vibration whereas the internal structure of the material effects the particular fashion the vibrations damp out (Gaver, 1993a).

As we have seen, the contact area between the shoe and the walking surface, which is mainly determined by the shape of the shoe, is important for the walking sound. Other configuration characteristics like size and resonating cavaties also play a role. For instance, shoe size is related to anthropomorphic characteristics of the walker. Bigger shoes normally mean bigger walkers which produce more pronounced low-frequency excitations. Both heels and walking surfaces can have resonating cavaties. Both from architectural acoustics and musical acoustics we know that the way the walking surface is supported can significantly effect the timbre of the walking sound.

In this section we have argued that both acoustic and physical analyses of walking sounds can be used to identify acoustic structures in auditory walking events. In the next section, we will discuss the relationship between acoustic structures and the way they are perceived.

4.4 Acoustic structure - Perception

Once the important acoustic structures of the walking sounds have been identified, we can investigate their perceptual relevance in different ways. Most of the studies previously mentioned establish the relevance of the identified structures by correlating them with the data of the perceptual experiments originally designed to investigate the relationship between auditory events and perception (Repp, 1987; Freed, 1990; Li *et al.*, 1991).

Another possibility is to manipulate the structures in the soundwave with advanced digital signal processing techniques and to present these manipulated stimuli to listeners. By manipulating only one source characteristic in an otherwise constant stimulus waveform it should be possible to alter perceptual classification in a predicted manner.

In a third approach, which is often called analysis-by-synthesis, the sounds are synthesized directly. The analysis-by-synthesis method is frequently used in speech research and provides a way to check whether or not the existing knowledge about a particular sound-producing event is sufficient to synthesize it. At the same time, the method has proven to be a source for new insights.

The analysis-by-synthesis approach was used by Gaver (1988). Based on the physical modelling of impact sounds, Gaver built a synthesizer which can be used to synthesize impact sounds (Gaver, 1993b). The parameters of this synthesizer are directly linked to perceptually relevant source characteristics. Shape, size and material of the object, hardness of the beater, and type of interaction (hitting, scraping) can be specified. Gaver showed that synthesized stimuli could reproduce the same perceptual effects found for natural recordings of impact sounds (Gaver, 1988).

Although Gaver originally developed his synthesizer for generating the sound of struck bars, we think it is flexible enough to synthesize basic auditory walking events as well. We therefore implemented Gaver's synthesizer, of which an overview is shown in figure 9. The synthetic sound is generated by exciting a filter bank with the appropriate input signal. In figure 9, the input signal is a short impuls. By shaping the impuls one can control the perceived mallet hardness: pulse width reflects low-frequency energy, angularity of the pulse reflects high-frequency components. Gaver showed that scraping noises can be synthesized by exciting the filter bank with a band-limited noise, where the center frequency of the noise corresponds to dragging speed, and the bandwidth to the roughness of the texture (Gaver, 1993b). The number of filters in the filterbank and their center frequencies specify the shape of the object (e.g., string versus bar), the center frequency of the lowest filter is used to manipulate the size of the object (e.g., small versus big), and the damping of the filters is used to control the material of the object (e.g., wood versus metal).

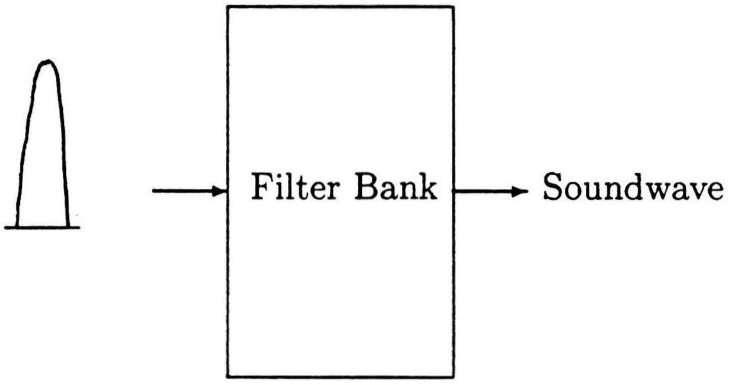


Figure 9: Overview of Gaver's impact-sound synthesizer.

Although informal experiments demonstrated the flexibility of the synthesizer, due to time constraints, we were not yet able to synthesize convincing walking sounds. Nevertheless, we think that after some fine tuning this synthesizer can be used to study and synthesize basic auditory walking events.

5 Concluding remarks

At the end of this report we briefly want to summerize our main findings.

The model of Cutting proved to be a flexible tool both for the analysis and synthesis of walking sounds. It gave insight into the temporal organisation of basic auditory events in human walking sounds.

We proposed a methodology to study the basic auditory events which occur in human walking. Within the proposed framework we discussed the analysis, synthesis, and perception of walking sounds.

Future research conducted along the lines described in this report will eventually lead to a full-fledged walking sound synthesizer. Such a synthesizer will be of direct practical use in applications like the ones pursued in the Humanoid project. The research methods used in the present study on walking sounds can also be applied to study the more general problem of using sound in interactive multi-media applications.

Acknowledgments

I would like to thank Roy Patterson and Bill Gaver who for a long time had wished to work together and invitated me to come to Cambridge to act as a go-between. Thanks to all at APU, PRL, NatLab, and IPO for making my stay at APU possible. Special thanks to Roy, Jay, Mike, Michael, Bob, and all other people at the Unit for their hospitality and for making working in Cambridge, productive, exciting, stimulating, and fun.

References

Aware Inc. (**1993**). *CD-ROM: Speed of Sound* (Aware Inc, One Memorial Drive, Cambridge MA 02142, United States of America).

Barclay, C. D., Cutting, J. E., Kozlowski, L. T. (**1978**). "Temporal and spatial factors in gait perception that influence gender recognition," Perception & Psychophysics, **23**, 145–152.

BBC (**1993**). *BBC SFX Audio CD: Essential People Sound Effects* (BBC CD 863, BBC Enterprises Ltd, 80 Wood Lane, London W12 0TT).

Cutting, J. E., Kozlowski, L. T. (**1977**). "Recognizing friends by their walk: Gait perception without familiarity cues," Bulletin of the Psychonomic Society, **9**, 353–356.

Cutting. J. E. (**1978a**). "Generation of synthetic male and female walkers through manipulation of a biomechanical invariant," Perception, **7**, 393–405.

Cutting, J. E. (**1978b**). "A program to generate synthetic walkers as dynamic point-light displays," Behaviour Research Methods & Instrumentation, **10**, 91–94.

Cutting. J. E., Proffitt, D. R., Kozlowski, L. T. (**1978**). "A biomechanical invariant for gait perception," Journal of Experimental Psychology: Human Perception and Performance, **4**, 357–372.

Fletcher, . H., Rossing, T. D. (**1991**). *The physics of musical instruments* (Springer-Verlag, New York).

Freed, D. J. (**1990**). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," Journal of the Acoustical Society of America, **87**, 311–322.

Gaver, W. W. (**1988**). *Everyday listening and auditory icons* (Doctoral Dissertation, University of California, San Diego).

Gaver, W. W. (**1993a**). "What in the world do we hear? An ecological approach to auditory event perception," Ecological Psychology, **5**, 1–29.

Gaver, W. W. (**1993b**). "Synthesizing auditory icons," Proceedings INTERCHI'93, 228–235.

Hartstone, G., Spath, T. (**1994**). "Chapter 28: Film," in *Sound Recording Practice, Fourth Edition*, edited by J. Borwick (Oxford University Press, Oxford).

Kozlowski, L. T., Cutting, J. E. (**1977**). "Recognizing the sex of a walker from a dynamic point-light display," Perception & Psychophysics, **21**, 575–580.

Laroche, J. (**1993**). "The use of the matrix pencil method for the spectrum analysis of musical signals," Journal of the Acoustical Society of America, **94**, 1958–1965.

Li, X., Logan R. J., and Pastore R. E. (**1991**). "Perception of acoustic source characteristics: walking sounds," Journal of the Acoustical Society of America, **90**, 3036–3049.

Napier, J. (**1967**). "The antiquity of human walking," Scientific American, **216**, 56–66.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and perception* edited by Y. Cazals, L. Demany, K. Horner (Pergamon, Oxford).

Patterson, R. D., Holdsworth (**1993**). "A functional model of neural activity patterns and auditory images," in *Advances in Speech, Hearing and Language Processing* edited by W. A. Ainsworth (JAI Presss, London).

Patterson, R. D. (**1994a**). "The sound of a sinusoid: Spectral models," Journal of the Acoustical Society of America, **96**, 1409–1418.

Patterson, R. D. (**1994b**). "The sound of a sinusoid: Time-interval models," Journal of the Acoustical Society of America, **96**, 1409–1418.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "A software package for time-domain modelling of peripheral auditory processing," Journal of the Acoustical Society of America (to be published)

Repp, B. H. (**1987**). "The sound of two hands clapping: An exploratory study," Journal of the Acoustical Society of America, **81**, 1100–1109.

Rosenbaum, D. A. (**1991**). *Human Motor Control* (Academic Press, Inc, London).

Rossing, T.D. (*1982*). *The science of sound* (Addison-Wesley, Philippines).