CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
AUGUST 1995

DEPARTMENT OF MUSIC
REPORT NO. STAN-M-94

DISTANCE OF SOUND IN REVERBERANT FIELDS

JAN CHOMYSZYN

CCRMA
DEPARTMENT OF MUSIC
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-8180

© Copyright 1995 by Jan Chomyszyn
All Rights Reserved

DISTANCE OF SOUND IN REVERBERANT FIELDS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MUSIC
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jan Chomyszyn
August 1995

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John Chowning (Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Earl D. Schubert

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Julius O. Smith

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Perry R. Cook

Approved for the University Committee on Graduate Studies:

Dean of Graduate Studies

Abstract

Interrelationships between the perception of sound distance and the perception of apparent loudness were examined in a set of three experiments. In experiment 1, the hypothesis that loudness is not a necessary cue for distance was tested using instrumental sounds and natural room environment. Results revealed that relative judgment of sound distance between two sounds could be successfully achieved, even when their loudness was eliminated, if the sounds were spaced by more than two meters. Moreover, the subjects' performance in such conditions was surprisingly high and could not be predicted from their performance with the loudness cue present. Multidimensional scaling of dissimilarities between the sounds in respect to distance indicated that no single physical attribute of the sounds was well correlated with the perceptual distances in the absence of loudness.

Previous experiments suggested that the distance of 'familiar' sounds was perceived more accurately than the distance of 'unfamiliar sounds'. In experiment 2, this hypothesis was tested for sounds which were acoustically nonfamiliar (new) for the subjects, yet were a simulation of the excitation pattern of a familiar (violin) sound. The results showed a very similar performance in the judgment of the familiar and unfamiliar sounds. Remarkably, in some cases distance perception of the non-familiar sounds was better than that of the original sounds. The experiment did not support the assumption that familiarity of sound is an important factor in distance perception.

'Sound constancy' hypothesis was the subject of experiment 3. Artificial reverberation was employed to produce sixteen sounds modelling a player in a room, performing with varied effort at four different distances from the listener's position. Direct-to-reverberant energy ratio was changed according to acoustical principles in order to produce the sensation of different distances of the players. In three subexperiments subjects were instructed: (1) to imagine a room and two players, one close and one remote, and match the loudness of the close player to the loudness of the distant player, (2) given the same sounds as in (1), to not imagine any room and just match the apparent loudness of the reference sound to the loudness of the other sound as it appeared 'in their ears', (3) to imagine the players, like in part (1), allowing that the players may be inside different rooms for each trial. The results showed significant deviations between the loudness judgements of reverberant sounds and the judgments of the corresponding dry prototypes. These deviations formed a logarithmic relationship as a function of the direct-to-reverberant energy ratio, and clearly revealed that reverberant sounds were perceived as louder. Moreover, the subjects were not influenced by the different instructions of the test, which shows that they were unable to concentrate their attention at the required aspect of loudness.

Acknowledgments

I wish to thank my main advisor Prof. John Chowning for his support throughout the course of this research.

Inestimable thanks are due also to Prof. Earl Schubert whose guidance through the research area of psychoacoustics was invaluable to me.

I was very fortunate to have the opportunity to collaborate on occasion with Prof. Albert Bregman, and want to express my gratitude for his encouragement and aid in designing the loudness experiment.

I have to acknowledge Dr. Perry Cook, and Prof. Julius Smith for always being ready to help with the many diverse problems that I have encountered along the way.

Finally, I would like to express many thanks to all my subjects, who voluntarily participated in the difficult tests without any financial reward.

Contents

Abstract iv
Acknowledgments v
List of Tables ix
List of Figures x

1 Introduction and Review of Related Literature 1

1.1 Scaling Techniques in Psychophysics 3
1.2 What is Loudness? 7
  1.2.1 Physical Correlates 7
  1.2.2 Psychological Factors 10
1.3 Distance Perception in Free Fields 11
1.4 Distance Perception in Reverberant Fields 13
  1.4.1 Intensity of Sound and Distance Perception 13
  1.4.2 Reverberation and Distance Perception 15
    1.4.2.1 Direct-to-reverberant Energy Ratio (d/r) 16
    1.4.2.2 Frequency Content 18
    1.4.2.3 Early Reflections 19
    1.4.2.4 Reverberation Time 21
    1.4.2.5 Cross-correlation Coefficient 21
1.5 Other Factors in Distance Perception 22
  1.5.1 Familiarity 22
  1.5.2 Head Movements 23
1.6 Related Sound Processing Techniques 23
  1.6.1 Impulse Response 24
  1.6.2 Auditory Filter 25
  1.6.3 Overlap-add Sound Synthesis Method 27

2 Experiment 1:
Loudness of Sound and Distance Perception 30

2.1 Rationale 30
2.2 Method 32
2.2.1 Subjects 32
  2.2.2 Stimuli and Apparatus 32
    2.2.2.1 The Recording Room 32
    2.2.2.2 Stimuli 38
  2.2.3 Procedure 46
2.3 Results 47
  2.3.1 Reliability 47
  2.3.2 Performance 53
  2.3.3 Scaling 64
2.4 Discussion 68

3 Experiment 2:
Familiarity of Sound and Distance Perception 75

3.1 Rationale 75
3.2 Method 77
  3.2.1 Subjects 77
  3.2.2 Stimuli and Apparatus 77
  3.2.3 Procedure 79
3.3 Results 81
3.4 Discussion 89

4 Experiment 3:
Reverberation and Loudness Perception 91

4.1 Rationale 91
4.2 Method 94
  4.2.1 Subjects 94
  4.2.2 Stimuli and Apparatus 94
  4.2.3 Procedure 101
    4.2.3.1 Simulation of the 'Close' Players 101
    4.2.3.2 Loudness of Sound Sources in Room Conditions 103
    4.2.3.3 Apparent Loudness of Sounds in Reverberant Conditions (in the Listeners' Ears) 104
    4.2.3.4 Loudness of Sounds in Control (Dry) Conditions 104
    4.2.3.5 Loudness of Sound Sources in Abnormal Reverberant Conditions 105
4.3 Results 105
  4.3.1 Loudness of Stimuli in Dry Conditions 107
  4.3.2 Loudness of Stimuli in Reverberant Conditions 107
  4.3.3 Instruction as a Factor of Loudness Judgment 114
4.4 Discussion 116

5 Conclusions 119

List of References

Distance Perception 122
Loudness Perception - Selected References 125
Related References 126

List of Tables

2.1 Loudness match. Comparison of the individual performances 48
2.2 Distance estimation. Comparison of the individual performances 52
2.3 The significance of the differences between the performance on the original and equalized sounds 61
2.4 Dissimilarities between the stimuli in terms of distance 65

3.1 Distance scores for familiar and unfamiliar sounds 84
3.2 Frequencies of correct and incorrect distance judgments of the familiar and unfamiliar sounds 86
3.3 The significance of the proportion of incorrect distance estimates for familiar and unfamiliar sounds 87
3.4 Distance preferences of the equally distant familiar and unfamiliar sounds 89

4.1. Distance ratios between the stimuli (simulating the players) 99

List of Figures

2.1 Crosssections of the recording room 33
2.2 Impulse responses in the recording room 36
2.3 Reverberation parameters in the recording room 37
2.4 Spectrograms of selected stimuli 40
2.5 Power spectra and formants of selected stimuli 41
2.6 The equivalent rectangular bandwidth of the stimuli 42
2.7 The power of the stimuli as a function of distance 43
2.8 The user interface for controlling the experiment 45
2.9 The results of the reliability check 51
2.10 The comparison of individual distance performances before and after the loudness match 51
2.11 Distance judgment as a function of physical distances between sounds 57
2.12 Mean performance of distance judgment as a function of distance between sounds 58
2.13 Multidimensional scaling of distances 67

3.1 An approximation of the familiar sound excitation pattern by the excitation pattern of the unfamiliar sound 80
3.2 Comparison between the familiar sound distance performance and unfamiliar sound performance as a function of the physical distance between sounds 83
3.3 Distance preferences of equally distant familiar and unfamiliar sounds 88

4.1 Demonstration of the 'loudness constancy' hypothesis
and auditory perspective 93
4.2 Spectrograms of the dry, percussion-like sound prototypes of
the stimuli 96
4.3 Distance ratios of the various test positions of the players and
and the listener in a room 98
4.4 The user interface for controlling the presentation 100
4.5 Comparison of mean gain adjustments for the reverberated sounds
and the dry sounds 110
4.6 Mean gain adjustment differences between the reverberated and
dry sounds as a function of d/r or distance ratio 113

Chapter 1

1 Introduction and Review of Related Literature

Despite a long research history our knowledge about the perception of sound distance is deficient in comparison to knowledge about sound localization. This deficiency is especially great in regards to the perception of sound distance in reverberant environments (in rooms). Various aspects of distance perception have been investigated, but since experiments are difficult to design their results are sometimes confusing or incomplete.

Our curiosity about the psychological factors in distance perception has recently become more focused. The exploding development of electroacoustic devices and computerized sound production has turned the attention of the designers to the creation of an appropriate auditory perspective for the produced sound. Such auditory perspective is intended to trigger the imagination of a concert hall acoustical environment in a headphone or a studio loudspeaker presentation of sound, analogous to the way visual perspective enforces the illusion of three-dimensional space from two-dimensional drawing. The ability to invoke the perception of sound source distance is one of the important dimensions of auditory perspective.

The creation of auditory perspective merits special attention in computer music, because in many compositions the location of sound in 'real' or intentionally distorted acoustical space is considered to be an element of the composition, and not simply a secondary addition to the produced sound. To be able to create the sonic illusion of sound distance in a convincing way, we first have to understand how this percept works in both natural room environment and when generated by electroacoustic devices designed for this purpose.

This research is mostly focused on the mutual relationships between distance perception and loudness perception in a reverberant environment. The intensity of sound and the reverberation of a room have been the most frequently studied physical correlates of sound distance. In many papers reverberation has been treated as a uniform entity and implicitly understood as all the changes introduced to the sound by a room except intensity, until recently. In accordance with our intuition both sound intensity and reverberation seem to be responsible for distance sensation. Yet sound intensity is apparently also correlated with another perceptual dimension, the loudness. Therefore, it is conceivable that the presence of intensity may, in some instances, be misleading in distance perception, as it may be inconsistent or even in collision with the reverberation cue. With its function unclear, it is also not known if loudness estimates are necessary for estimates of sound distance.

While the loudness of sound has usually been regarded as an obvious cue to distance perception, it has also long been observed that the perceived distance of sound may be an important factor in loudness judgment, so that they seem to form equivalent concepts for some listeners (Gamble, 1909). The influence on loudness judgments by reverberation was confirmed in an experiment by Warren (1973) on speech examples. A similar effect has been noticed in computer music (Chowning, 1990), but no known systematic study on how the amount of reverberation affects loudness perception has been carried out so far.

A variety of scaling methods of distance and loudness has been applied in experiments on distance perception and loudness perception. These methods involve different requirements from the subjects (instructions), and specific procedures of data collection. These factors are important for the interpretation of the results, and sometimes raise questions about the validity or scope of the experimental conclusions. Therefore, the most important methods will be briefly reviewed, before we move on to the survey of the relevant references.

1.1 Scaling Techniques in Psychophysics

Psychophysics is based on the assumption that the human perceptual system can be used as a measuring device yielding results corresponding to the magnitude of stimulus attributes (Baird and Noma, 1978). These results, reflecting the subjective magnitude of sensation invoked by the stimulus, can then be systematically analyzed and compared to the changes of the physical attribute in question. It is widely accepted that the degree of sensation invoked by the stimulus can be expressed in numbers. Two methods are the most popular: ratio estimation and category estimation (Torgerson, 1958).

In the ratio estimation methods, the subject is told to concentrate on ratios of stimuli. The subject is supposed to provide a number expressing the ratio of the sound attribute under investigation between the stimuli. For example, in the magnitude estimation method, the subject adjusts a stimulus to maintain a constant ratio of the attribute and a constant reference sound provided by the experimenter. In another variation of the method, the free ratio estimation, no standard is presented, and the listener is expected to assign any numbers which, in his opinion, reflect judged ratios among stimuli.

In the most extreme cases, it is assumed that a listener is able to reproduce the *absolute* magnitude of sensation, i.e. to generate a number which reflects a 'true' physical magnitude of the stimulus in question. For example, when judging sound distance such a number should reflect the physical distance between the listener and the sound source or virtual sound source in meters or feet. In addition, it is sometimes assumed that subjects acquire the skill to reproduce such an absolute magnitude corresponding to a sensation during their life experience and are able to provide it when first exposed to the sound in an unknown situation. In such a case, the parameter which causes the sensation is called an absolute cue (Mershon and King, 1975). In contrast, relative cues do not have the potential for providing such information, but may nevertheless still be effective after repeated presentation.

In the category estimation method, the subject is told to assign a whole number, from a certain range of numbers, which enumerate categories of a constant width, to the sensation evoked by the stimulus. The number of categories is arbitrary, is usually kept small (between 5 and 10), and can be much smaller than the number of stimuli. The subject's task can be described as partition of the stimulus continuum (loudness, distance, and so on) into a set of equally spaced bins. In a variation of the above method, the subject estimates a subjective distance between paired stimuli, or their difference according to the attribute of interest. In the simplest case there are only two categories, and the subject judges the relative position of the stimuli on the ordinal scale.

In some experiments many physical properties of the stimuli vary simultaneously. In such a complex situation, an insight into the way subjects perceive the changes can be acquired by modeling the stimuli as points in a geometric space (for example in the Euclidean n-dimensional space). The points correspond to the stimuli in such a way that perceived similarity is represented by the spatial proximity of the points. The task of multidimensional scaling is to provide a configuration of points, which best represents the similarities between the experimental objects. An idea of the nonmetric scheme developed by Kruskal (1964) will be briefly summarized below. For more details, the reader is referred to the review by Baird and Noma (1978), and works by Torgerson (1958), Shepard (1962 a,b), and Young (1968) among many.

The basic goal of nonmetric multidimensional scaling is to match the rank orders of the input proximities with those of the distances between a set of points in a space of a desired dimensionality. Given N (N-1)/2 proximities of a set of N experimental objects we seek to form N points in the space so that monotonicity is preserved:

$$\delta_{ij} < \delta_{kj} \Rightarrow d_{ij} < d_{kl}$$

where δ_{ij} is the ranked proximity between objects i and j and d_{ij} is the derived distance between the corresponding configuration points. Departures from the desired order of points can be easily inspected on the Shepard diagram, a scatter diagram on which all the proximities and distances are paired. Ideally, it should be possible to join all the points of the diagram by moving up and to the right. Points which violate the rule contribute to the variance from this ideal line. Thus, the measure of the 'goodness' of the approximation, the stress is based on this variance:

$$\text{Stress} = \sqrt{\frac{\sum\sum (d_{ij} - d'_{ij})^2}{\sum d_{ij}^2}}$$

where d'_{ij} is the distance when the monotonicity is satisfied.

Starting with a random position of the configuration's points, the MDS procedure attempts to move them so as to minimize the value of the stress. Usually the steepest descent method or the method of gradients is employed to move the points until a minimum is found. Unfortunately, there is a possibility that merely a local minimum will be discovered, with a mathematically non-satisfactory match to the proximities, which is especially likely in one dimension (Shepard, 1974). Interpretability of the spatial dimensions and their number, as well as the general pattern of the final configuration are the main interest in perception. Generally, a low-dimensional solution is preferred, if possible. Any interpretable features of the spatial configuration should be taken into account, including clusters and circular ordering. Projections of the points of the configuration are particularly important, whether on the orthogonal or non-orthogonal axes, if the point coordinates can be correlated with the physical parameters of the stimuli.

While it is not possible to discuss all the scaling methods in detail here, a fundamental problem should be emphasized: mutual understanding has to be achieved between the experimenter and his/her subjects about the task of the listening procedure. As the listeners are a measurement tool, the experimenter has to first explain the procedure, be convinced that the requirements are clearly understood, and that the listeners are able to consciously respond to the stimuli in the desired way. Psychoacoustic experiments often yield a substantial variance in the subjects' responses, which sometimes may make it impossible to draw valid conclusions. If, in addition, the task is unreasonably difficult or obscure, one may doubt whether in such conditions subjects can provide any valid measurement. Reliability checks seem to be, therefore, a necessary element of every perceptual experiment. Each subject should repeat the test at least twice (differently randomized), and a correlation coefficient should be considered between the two sets of paired responses. A low correlation would clearly point to the subjects' inability to perform the experimental procedure, because of its difficulty or poor explanation. Even if the correlation is quite large, the differences between the results of the two test runs may also be large. This may reveal that an extraneous factor played a role during one of the test presentations, such as background noise or de-calibration of the equipment. It may also point to large just noticeable differences (jnd's) of the perceptual attribute being researched. Therefore, the mean deviation and standard deviation of the absolute differences between the paired data from the two test runs should also be calculated. Needless to say, in most perception experiments on sound distance the reliability seemed to be of no concern, and standard deviation (between subjects) was usually large. On the top of it, many of them were conducted in less than realistic conditions from the point of view of an average listener, which raises questions about the validity of the results.

1.2 What is loudness?

Loudness is not only determined by the acoustic intensity of sound. This issue has to be clarified before the literature survey, because it often seems to be misunderstood. Loudness can be best characterized as the subjective intensity of sound (Scharf, 1978). This subjective intensity is dependent on many physical attributes of sound and its context, such as: the physical intensity (or sound pressure), its distribution in frequency, the length of the sound, and a sound relation to background sounds. Loudness is also dependent on the listener. Although people hear in much the same way, large individual differences occur between loudness estimates from one listener to another. It is not within the scope of this work to treat all aspects and details of loudness perception, nevertheless the basic facts will be briefly summarized below. More in-depth descriptions can be found in (Scharf, 1978), (Moore, 1989), (Marks, 1979), and others. In the reminder, the term *intensity* will always mean physical intensity throughout this work, unless otherwise noted.

1.2.1 Physical Correlates

The primary physical factor of determining loudness is the sound pressure (or intensity). For a given frequency, say 1000 Hz, loudness variation with intensity (in sones) can be expressed as a power function of sound pressure:

$$L = k P^{x}$$

where k is a constant depending on the subject and units used. This relationship holds for pressure levels higher than 30 phones (see below for the phone definition). There is an overall agreement as to the general shape of the loudness curve, but there is lack of consensus about the proper value of the exponent. The value of x=0.6 was adopted as the international standard, but the experimental values for the exponent vary. Moreover, it depends on the features of the experiment which theoretically should not affect it, such as, the method employed for collecting data (i.e. magnitude or category estimation), or the range of physical intensities explored (Schneider and Parker, 1990; Marks, 1979).

Loudness depends on frequency. This dependence for pure tones can be shown by the so called *equal loudness contours* (Fletcher and Munson, 1933; Scharf, 1978). The contours show how the sound pressure level of equally loud tones is dependent on their frequency. In general, more sound pressure is necessary in the low and high frequency regions to produce the same loudness as in the middle range. This relationship also depends on the sound absolute level, especially for headphone presentation, becoming flatter at higher intensities. The level (in dB SPL) of a 1000 Hz tone to which a sound is equal in loudness is called the loudness level and the unit of loudness level is the phon.

The loudness of a complex sound depends not only on the sound pressure level of its components, but also on the bandwidth the sound covers. In general, the wider the bandwidth, the louder the sound is. However, if we start to increase the bandwidth of the sound while keeping the total intensity of its components unchanged, the loudness remains constant until the bandwidth exceeds a minimum value, the so called *critical band* (Zwicker, Flottorp, and Stevens, 1957). Beyond the critical band, loudness increases with the frequency spread of the sound, except at intensities near the threshold.

The loudness of sounds longer than 200 ms does not depend on duration. Sounds shorter than 200 ms, especially shorter than 80 ms, are perceived as softer than equivalent intensity long sounds. The *critical duration*, i.e. the time beyond which the intensity becomes constant and independent of duration, varies considerably from one experiment to another. The differences are likely to be attributed to the difficulty of matching the loudness of short tones to the loudness of long tones, hence to the different criteria various listeners use for this task (Scharf, 1978).

Background noise also has an effect on the loudness of sounds. A loud sound can become soft in the presence of masking noise. Partial masking can be thought of as a local rise of the hearing threshold near the frequency of the tone, in order to band pass filter the background noise, and to provide the best signal to noise ratio. Partial masking depends not only on the intensity of the masker, but also on its bandwidth and relative frequency location with respect to the frequency of the masked signal. For more information see 1.6.2, and (Moore, 1989).

A sound heard binaurally is louder than one heard in one ear alone. On suprathreshold levels, loudness of a pure tone or a narrow-band noise measured binaurally is, on average, twice as loud as a monaural sound of the same sound pressure level (Algom, 1989). The level of a monaural tone has to be raised by 10 dB to equate the loudness of a binaural tone. This fact indicates, that it is loudness in two ears which is summed, not just the corresponding energies.

The loudness of a complex sound is usually unequal to the sum of its components' loudness. The Zwicker-Scharf loudness summation method (1965) will be described briefly here, because it is probably the most well known example. Caution is advised when using a 'loudness summation' method to calculate the loudness of complex sounds, because such a method is usually based on simplified assumptions, which are unrealistic for most real sounds. The method may be unsuitable for transient or modulated sounds (most often encountered in speech or music) as it is based on the facts which are explained for steady state long sounds. The method does not account for the temporal summation phenomena either. In the Zwicker-Scharf summation method, an excitation pattern of the complex sound is calculated first (see also 1.6.2). The masking pattern of the sound is calculated to this end, and expressed as a function of tonalness (Zwicker,1961). The tonalness function transforms the critical band scale (the stimulus measure) to the Bark scale (a sensory measure). Stevens's power law is subsequently applied and the excitation pattern is converted to the *specific loudness* expressed in sone/bark units. Finally, the total loudness is produced as an integral of the area under the specific loudness pattern.

1.2.2 Psychological Factors

Individual differences in loudness judgements can sometimes be ascribed to personality differences, as measured by a standard test of anxiety (Stephens, 1970). The author found that subjects with high-anxiety scores produced steeper loudness functions (as measured in sones) than low-anxiety listeners. The way the auditory system works does not change much from one individual to another, but people with different personalities may use various strategies in estimating loudness. Other psychological factors, such as motivation, attention, experience in dealing with sound, or the capacity to understand the task also contribute to the great variability of experiment results. Therefore, "It is not at all clear that methods of calculating loudness using 'psychological' loudness scales give better agreement with loudness judgments than methods based on a physical analysis of the stimuli" (Moore, 1989).

Finally, it has to be noted that there is disagreement on whether loudness is a primary percept, or if it is derived (learned) from distance perception. It was noticed long ago (Gamble, 1909), and demonstrated later (Stevens and Guirao, 1962) that loudness and distance can form equivalent concepts. For Stevens and Guirao distance is secondary: distance estimates are based on a loudness scale. Warren (1963) argues against this thesis. According to his physical correlate theory (Warren, 1958), judgments of sensory magnitudes are disguised estimates of physical magnitudes. In the case of loudness this means that listeners use their experience about the manner in which stimulation varies depending on the distance of the sound source. According to Warren's postulates: (1) loudness judgments are the reciprocal of distance estimates, (2) indicators of the relative distance of a source (reverberance) influence loudness judgements, and (3) loudness judgments can be calculated from the physical principles governing sound. In favor of Warren's hypothesis attention should be paid to the fact that in everyday life we evaluate loudness as one of many properties of the sound source itself. The context in which the sound is heard seems to be important as well, and we certainly take into consideration the distance and size of a sound source when estimating loudness. It is as unlikely to hear a huge, loud, mosquito from a long distance, as it is to hear a soft and close thunderstorm. For more discussion of this subject, please see the description of the third experiment (chapter 4).

1.3 Distance Perception in Free Fields

The distance of sound in free fields has long been one of the major problems in psychoacustics research. Investigation, which began as early as the beginning of the century, concentrated on the exploration of stimulus (physical) correlates producing the sensation of sound distance. From the extensive experimental work we have selected research papers, which represent the related problems and difficulties. Coleman (1963) points to the intensity of sound, frequency spectrum (at near and far distances), binaural intensity ratio, and interaural phase (or time) differences as the physical factors with the potential to create auditory depth.

Intensity as a cue for sound distance has probably been the most frequent research topic. However, a variety of incompatible experimental results have aroused controversy as to their validity. A typical problem in this field can be formulated in the following way: what intensity changes are required to produce the sensation of sound at a multiple distance (i.e. 'twice as far'), relative to a given reference. In Gardner's (1969) experiment, a level reduction of as much as 20 dB led to a doubling of the auditory event distance. Gardner, however, used a speaker identification method. He discovered that a proximity effect occurred, i.e.

subjects tended to assign the apparent source to the nearest rational location (speaker) when the level was kept constant. The estimates were also influenced by the sound type. Distance was overestimated when a shouting voice was used, and a whispered voice resulted in underestimated distances. Overall, the author concluded that, "...the ability of an observer to judge the distance to a loudspeaker source of speech at 0^{o} in anechoic space was found to be, at best, extremely small."

Gardners's finding disagrees with the experimental results of Stevens and Guirao (1962). In their experiment the authors used tones and noises with a headphone presentation. Three methods: magnitude estimation, magnitude production, and category production, were used in the test. About a 10 dB change was sufficient to double the distance, and the resulting relationship was inversely related to the loudness estimates of the stimuli.

9 and 12 dB increases were favored over 6 dB changes for creating an illusion of sound being "twice as close [as the one heard] from an initial position," in an experiment conducted by Begault (1991). Dry speech, piano, and click sounds were used in a two-alternative choice method, with a headphone presentation. In the conclusions, the author postulated that: "... an inverse square law is inadequate for producing a sensation of half distance."

From these results it is rather difficult to come up with any useable way of controlling the sound distance based on a loudness scale, and its importance is questionable. However, a generalization can be derived that, "... a trend exists for this distance [of sound event] to increase less rapidly than the [physical] distance of the sound source." (Blauert, 1983)

As a result of attenuation by air absorption frequency spectrum changes in an anechoic environment. This effect becomes perceptually important for distances exceeding 15 m (Coleman,1968). Low-pass sounds are perceived as more distant than high-pass sounds (Butler et al.,1980). Frequency spectrum may also play a role in distance perception at short distances, from 0.10 to 2 m (Haustein, 1969). His subjects could judge the distance of the sound source near the head very accurately even when the sound level was kept constant.

1.4 Distance Perception in Reverberant Fields

1.4.1 Intensity of Sound and Distance Perception

The intensity of sound manifesting itself as loudness is often regarded as the most obvious cue in distance perception. As we have already seen, it is certainly the case in a free field situation. In a reverberant environments opinions differ, although researchers generally recognize intensity to be the most important factor.

We start our review with Bekesy (1960), who summarizes his articles from 1938. The author does not report on any formal tests conducted on listeners other than himself, nor any statistical results are included, therefore we have to assume that all his conclusions are based on speculation about the idealized models of reverberant fields. In his opinion, "...it seems that loudness has an effect upon the perceived distance only in the absence of other more determinate physical cues." The statement stems from an observation that loudness does not change much the distance of a click, whereas increasing loudness reduces the distance of a tone.

Mershon and King (1975) studied the effect of loudness and reverberation on the perception of apparent distance. They were interested in a specific aspect of distance perception, the 'absolute' cues, as opposed to the 'relative' cues (see 1.1). They found that: "... although a change in auditory intensity may be a good relative cue for auditory depth, it is ineffective as an absolute cue...".

A 'reverberation tunnel' and two speaker positions (at 2.74 and 5.49 meters) were used in the experiment. The stimuli were 5 seconds of white noise.

This finding was later confirmed by Mershon and Bowers (1979) regardless of the listeners' orientation (0 degrees and 90 degrees.) The experiment was conducted in a semireverberant room (10.6m x 7.2 m, 2.7-3.5m high) at five distances: 0.55, 1, 2, 4, and 8 meters. The sound pressure level of five second white noise was normalized at the listeners' position to 60 dBA. In addition, the authors noticed that greater physical distances tended to produce greater group reports of loudness from the subjects. The results will be discussed further in 2.1, and 4.1.

Artificial reverberation and loudspeaker presentation were used in an experiment by Sheeline (1983). He attempted to find the relation between the energy of the direct and reverberant parts of sound when doubling apparent distance. According to his conclusions: "Loudness differences, in all cases, are demonstrated to be the cues that most effectively suggest the distance differences." Although he also stated "The magnitude of the loudness difference required to generate a certain relationship between two sources is not at all predictable, unless the room conditions are known in advance." The last statement refers to his results, which showed that the amount of attenuation of the direct sound is dependent on the reverberation conditions (3-4 dB for low reverberance, 6 dB for medium, and 10 dB for heavy). These results agree with the 6 dB attenuation factor found by Warren, Sersen, and Pores (1958). Speech and tone sounds were used in the test in a speaker presentation in a small reverberant conference room (subjects blindfolded).

Multidimensional scaling (the INDSCAL program) was used by Jullien, Lavandier, and Warusfel (1989) to asses distance judgments in an artificially created environment. The authors simultaneously varied the sound level, LEV (-4, 0, +4 dB), and reverberation time, RT (1.1, 1.7, and 2.5 sec). Changes of the parameters were clearly perceived as varying distance: "Here is an example where one can consider that the criteria (LEV,RT) characterize two perceptive factors in their utilization range".

1.4.2 Reverberation and Distance Perception

Reverberation environments provide a variety of cues to distance as several acoustic variables are altered when the physical distance between sound source and listener is changed. Among the factors regarded as the most important for distance perception in a room environment are the direct-to-reverberant sound energy ratio, the number and distribution of early reflections, and spectral changes, especially 'roll-off' of high frequencies. Binaural differences are sometimes cited, but no serious experimental work has been done along these lines. We should be aware that all of the variables, including the intensity of the direct sound, are influenced by room acoustics, and some of them seem to be correlated. Moreover, in many cases there is no straighforward correspondence between physical parameters in question and distance sensation, because our perception involves many physical dimensions rather than a single one.

Reverberation was usually treated as an indivisible physical phenomenon, and many conclusions are based on the theoretical assumptions of what must physically change along with the distance in the room, rather than on actual measures of the parameters of interest. Recently an attempt was undertaken to vary reverberation parameters separately in a systematic way either by adding artificial reverberation to dry sounds, or, more importantly, by changing the acoustics of experimental studios. Since the early days of radio, broadcasting and recording industry, professionals noted that a distance feeling can be achieved by changing the balance between the direct and reverberant content of the sound. In the experiments pertaining to judging sound distance in reverberant fields experimenters are aware of this fact, even though this ratio is neither measured nor calculated. Therefore, we decided to discuss such work in the following section as well.

1.4.2.1 Direct-to-reverberant Energy Ratio (d/r)

One of the first known places in which the d/r ratio is considered is the work of Bekesy (1960) (see also 1.3). In his opinion: " The perception of the distance of a sound in a room thus seems to depend upon this ratio". However, as he is convinced that sound distance in a reverberant field should be based on the same principles as sound distance in a free space, he denies that d/r ratio is a fundamental percept: "Though this alteration in the ratio between direct and reverberant fields can indeed be used to produce the perception of a moving sound image, this ratio is not the basis of auditory distance". We should be aware that his conclusion is not based on any formal tests.

This point of view was challenged in an experiment by Mershon and King (1975). They found that reverberation could serve as an absolute cue to distance, thus "...provide the basis for a perceptual scale of distance in terms of absolute values (feet, inches, meters, etc.)." Only two distances were used (2.74 and 5.49 meters) in a 'reverberant tunnel', and the d/r ratio was not explicitly discussed in the work.

These results were later replicated by Mershon and Bowers (1979) in a semireverberant room. Five distances were investigated (0.55, 1, 2,4, and 8 meters), as well as two listeners' orientations (0, and 90 degrees). The sound pressure level of the stimuli (5 second white noise) was normalized to 60 dBA at the listener's position. "For both orientations, near distances were overestimated and far distances were underestimated." The experiment confirmed that reverberation can be an absolute factor in determining egocentric auditory distance. Moreover, prior knowledge of the experimental room was not necessary for the operation of the reverberation cue and neither was knowledge of the sounds.

The direct-to-reverberant energy ratio was thoroughly investigated by Sheeline (1983). In addition to relating the direct energy to the energy of reverberant part of the sound in order to place it twice as far (see also 1.3), he explored the just-noticeable differences associated with distance perception. He detected approximately 3 jnd's between the double distance positions, but only for the first 4 or 8 multiples of apparent distance. As the range of distances increased the jnd became larger, and listeners' acuity even less precise.

In yet another experiment Sheeline investigated the usefulness of the ratio for achieving sound depth. In the experiment, he added nine levels of artificial reverberation to dry trumpet tones, varying the reverberant-to-direct sound energy from 0% to 24%, with the step of 3%. The ratio of intensities were kept constant (0, -6, and -12 dB). Subjects estimated the apparent distance difference between the sounds at 0, and -12 dB using a free magnitude estimation method with the range limited to the numbers between 1 and 50, for each of the nine different ratios. According to his results: "...the greatest area of penetration occurred in the reverberation range 3% to 12%, with apparent peak at 6%". He hypothesized that: "... this masking effect [i.e. masking due to high reverberation] is the greatest contributing factor to the limitation of effectiveness of reverberation as a distance cue."

Begault (1991) pursued the importance of implementing the head-related transfer function in auditory localization. Artificial reverberation was added to speech stimuli. Early reflections were modeled as the output of a two-dimensional ray tracing room simulation program. Late reverberation was modeled by using exponentially decaying noise (pseudo random sequences). Next, the direct sound and early reflections were modified by a set of filters corresponding to a head-related transfer function. Finally, subjects were requested to scale distance in inches according to the degree of sound externalization in a headphone presentation. Added reverberation increased the perceived distance in comparison with dry sounds, which was a known fact, but absolute distance judgments showed great variability. More importantly, reverberation helped the externalization of HRTF stimuli: "... 25% of the dry stimuli were not externalized, compared to only about 3% of the reverberant stimuli."

1.4.2.2 Frequency Content

Research on the importance of sounds' spectral content in distance perception has concentrated on the bandwidth of the stimuli as the distance cue. This research is discussed here in the reverberation section because the loss of high frequencies quoted as a reason for increasing distance is mostly caused by room absorption (the frequency dependent absorption coefficient of the walls, furniture, people), and only marginally by the attenuation of the sound in the air (Blauert, 1983; Kutruff, 1991) in the range of distances normally useful in rooms.

Butler et al. (1980) exposed their subjects to 5 sec broadband noise bursts, which were either low-pass with cut-off at 0.5, 1, and 2 kHz, or high-pass with cut-off at 2, 4, and 6 kHz. The sounds were recorded in the ear canal of a model and reproduced in a headphone presentation. Subjects recorded the apparent distance in a free magnitude estimation method. According to Butler: "What did influence distance estimates dramatically was the frequency composition of the stimuli. Low passed sounds recorded in either acoustic environment [echoic, anechoic] were consistently judged to be further removed than high-pass sounds recorded in the same setting". The authors tried to explain this phenomenon suggesting that lifetime auditory experience had taught listeners that distant sounds generally have less acoustical energy in the higher audio frequencies, and that this experience was invoked by the stimuli.

Little in his writing (1992) contradicted this standpoint and showed that "...a decrease in high-frequency content (as might physically be produced by passage through a greater amount of air) led to increases in perceived auditory distance, but only when compared with similar sounds having a somewhat different high-frequency content." In his experiment the author used three low-pass noise stimuli, with cut-off at 5, 6, and 6.7 kHz. The stimuli were matched for loudness before presentation. They were then reproduced at a distance of 3 meters by a loudspeaker in a low-reverberant studio (reverberation time ~0.6 sec). Little's conclusions point to the fact, that: "...spectral information can serve as a relative cue for auditory distance, independent of changes in overall sound level."

1.4.2.3 Early Reflections

Experiments on distance perception in an artificial reverberant environment consisting of only early reflections were performed to prove the usefulness of such an environment for sound spatialization. While we have to show restraint as to the applicability of the results to cases when there are also late reflection cues, the experiments have demonstrated that distance sensation can be achieved by the early reflection field alone.

Gotoh et al. (1977) were interested in factors which would enable subjects to increase subjective distance, while keeping the spatial impression to a minimum. They investigated an effect of delayed sounds (both the number of echoes and their timing) reproduced by a set of loudspeakers, along with the direct sound in an anechoic chamber. A male voice was used as the stimulus. Subjects were asked which sound was farther away in paired comparison. In the case of a single reflection (delay) authors showed that subjective distance was proportional to the delay time. They also discovered that: "... six successive reflections give more feeling of distance than the single or double reflections, and a sound structure simulating the reflections from each surface of an actual rectangular room can create a feeling of distance independent of a spatial or spread impression of sound." Another important fact was demonstrated, namely that sound distance depended on the spatial distribution of the reflections. When the reflections were restricted to the horizontal plane, or to the front direction, the subjective distance was decreased.

These results are generally concordant with the Begault 's conclusions (1987), who applied the image method (Borish, 1984) to reverberate his stimuli. Four and forty eight early reflections from the boundaries of a rectangularly shaped room, and a seven-sided polyhedron enclosure, were produced by a computer program. The four relative positions of the sound source and listener in the rooms were simulated. The modelled distance between the sound source and listener was equal to two meters. Pinnae filtering effects were included in the model. Speech and piano sounds were used as stimuli in a forced choice, paired comparison method, with a headphone presentation. Begault observed that, when intensity was eliminated as a cue, "... a sound convolved with 48 reflections was perceived as more distant than a sound convolved with 4 reflections, as a result of the relative attenuation of higher frequencies in the 48 reflection version." However, in some instances a spatial cue took precedence in his experiment. Since the spaciousness of the sound reverberated with the 48 reflections was greater than those with 4 reflections, sound distance decreased when more of the reflected energy was heard from the side than front of the listener. The author concludes by making an important distinction: "Early reflections affect perceived distance not as a function of the magnitude of their intensity, but rather as a function of their relative spectral weighting of the sound and sometimes as a function of the angle of incidence of the reflections to the listener."

Gotoh's results, demonstrating the possibility to create distance sensation with a single reflection seem to oppose the conclusions of Guski (1990). Guski investigated the effect of a single reflecting surface placed in an anechoic chamber on localization and distance. Unfortunately, the speaker identification method was used with the array of twenty seven speakers visible to the subjects. Three distances were explored: 2, 2.8, 3.52 m. Speech stimuli were presented with natural loudness. Guski concludes: "In the conditions tested here, reflected sounds do not seem to influence distance estimations, and the main source of information for distance localization is supplied by the simple sound level of the source."

1.4.2.4 Reverberation Time

The influence of distance perception by reverberation time was the scope of research by Mershon et. al. (1989). Reverberation time was varied between 2.17 sec (at 500 Hz) in a lively room and 0.58 sec in 'dead' room conditions. Four distances of 0.75, 1.5, 3.0, and 6.0 meters were investigated. Subjects were requested to provide an absolute distance (in feet or meters) to the target sounds. "Reported distance was generally proportional to real distance, but considerably underestimated when room reflectance was low. When room reflectance was high (T60 ~ 1.7s for the range of frequencies used), initial reports of distance were often overestimates; upon repeated presentation, judgments in the high reflectance room became more nearly veridical." Background noise level was demonstrated to be another important element for distance: "The effect of increasing the background noise level was to decrease the perceived distance." We should note here that this result is not consistent with, but rather contradictory to an interpretation based on a simple sound level cue.

1.4.2.5 Cross-correlation Coefficient

The interaural cross-correlation coefficient has been of interest to the concert hall designers (Kutruff 1991). The data indicate a negative correlation between the magnitude of the IACC and subjective preference (Ando 1985). It also contributes to subjective diffuseness or spatial impression.

Kurozumi and Ohgushi (1983) investigated the influence of the IACC on the perception of sound image width and distance. White noises, cross-correlated with seven different levels of the IACC, were used in the experiment in anechoic condition. Four subjects were used. The experiment was carried out in an anechoic chamber and was then repeated in a conference room (reverberation time 0.8 sec), with a similar outcome. The results indicated that: "... as the cross-correlation of white noise increases, the distance of the sound increases." The authors did not propose any hypothesis explaining the physical or psychological reasons for this phenomenon.

1.5 Other Factors in Distance Perception

1.5.1 Familiarity

Familiarity of sound was often reported to be important for distance perception. There were only three known experiments in which this issue was studied, and only one of them attempted to control familiarity in an empirical way. In the three studies, different aspects of sound familiarity were under investigation.

In Coleman's research (1962), familiarity was understood as prior knowledge of the stimuli and acoustical conditions. Thus the learned ability to judge distance was investigated. The experiment was conducted in anechoic conditions (outdoors, with a snow covering), with one second bursts of wide-band random noise. The speaker identification method was used. During the first exposure to the stimulus the judged distance of the source was unrelated to the actual location of the source. However, "... with further trials valid distance judgments became possible."

Gardner's experiment (1968) was not specifically designed to explore familiarity of sound. Neither did he explicitly discuss his results in terms of familiarity. Yet he is sometimes quoted in such context. He used speech as the stimulus, including the whispering and shouting voice. The distance estimates depended primarily on the type of vocal output employed. In this experiment, perhaps the non-familiarity of sound can be comprehended as the atypical articulation of sound (speech).

Familiarity was understood as the lack of meaning in the experiment of McGregor et al. (1985). The sentence "How far away do you think I am?" was used as the familiar stimulus. The same sentence played backwards was the unfamiliar stimulus. The experiment was carried out at a biological field station. It is not clear whether the conditions were anechoic or reverberant. "The result further supports the role of familiarity in relative distance estimation", according to the author.

1.5.2 Head Movements

"Head movement did not improve performance on either task [distance judgment]." This statement best summarizes the results of Simpson and Stanton (1973). The experiment was performed in an IAC room. It is not clear if the conditions were anechoic, but it is known that care was taken to reduce sound reflections. A train of sine-wave pulses was played as the stimulus from a speaker at various distances. Both the magnitude estimation and threshold detection for changes in distance were explored.

1.6 Related Sound Processing Techniques

A number of techniques has been used in the experimental part of this work. The maximum-length sequence measurement technique was used to obtain impulse responses of the recording room in the first experiment. The auditory filter was employed to approximate an instrumental sound excitation pattern in the second experiment. The stimuli for this test were produced as the output of the overlap-add synthesis method. A short description of these techniques is thus included to help with a basic understanding of the applied tools, and to provide appropriate references for the more involved reader.

1.6.1 Impulse Response

The properties of a linear system can be completely characterized by its response to an impulse excitation, i.e. the impulse response (Oppenheim and Schafer, 1989). Sound in a reverberant room can be conveniently analysed as an input to a linear system (Kutruff, 1991). For this reason it is clear that experimental measurement of an impulse response is one of the fundamental tasks in room acoustics.

The simple method of producing and recording an impulse sound in a room has proven to be inadequate, because a sufficient signal-to-noise ratio is difficult to obtain in this way. More advantageous is the use of a sufficiently dense signal, such as white noise, whose autocorrelation function approximates the delta function. After recording a room response to the noise, the impulse response can be reconstructed through deconvolution. Maximum-length sequences turned out to have very useful properties for such measurement.

The essential property of the maximum-length sequence of length $n=2^m-1$, or pseudo-random noise, which is exploited in the impulse response measurement, is that its periodic autocorrelation function is given by (MacWilliams and Sloane, 1976):

$$\rho(0) = 1, \ \rho(i) = -\frac{1}{n}, \quad \text{for } 1 <= i <= n-1$$

Suppose, a few periods of a pseudo-random sequence ρ' was played and recorded in a room. Then, the noise response (periodic) of the room is equal to the convolution:

$$nr'(k) = \rho'(k) * h(k),$$

and the desired impulse response can be recovered as:

$$h(k) = \frac{1}{n} \sum_{j=0}^{n-1} \rho'(k-j)\, nr(j)$$

Fast Hadamard transform turned out to be useful in an efficient algorithm for performing the decorrelation (Borish, 1984). As ρ' in the above expression can be written as a matrix containing the circularly delayed versions of the maximum-length sequence, it can be converted to a Hadamard matrix. Then, taking advantage of its speed, the Hadamard transform can decorrelate the noise response without using multiplications. It should be also noted that the maximum-length sequences can be generated efficiently by using a shift register with feedback.

Consideration has to be taken in planning the experimental procedure, however, because the deconvolution method is very sensitive to the speed de-synchronization of the playback and recording devices (Borish, 1984), and to nonlinear distortions produced by the sequence during the playback (Rife and Vanderkooy, 1989) which may cause substantial errors.

1.6.2 Auditory Filter

The threshold of a signal, in the presence of a masking noise centered around the signal, increases as the bandwidth of the noise increases. After a certain frequency range is crossed it flattens off, however, and further bandwidth spread does not cause a significant change of the threshold. To account for this result the notion of an auditory filter bank was introduced (Fletcher, 1940; Moore and Glasberg, 1987). According to this postulate, the auditory system acts as if it contained a bank of band-pass filters. There is evidence that the center frequencies of these filters correspond to specific locations on the basilar membrane, which is the basis for the filters. When listening to a tone in the presence of masking noise, the listener is assumed to make use of a filter with a center frequency close to the frequency of the tone. Except for a limited range around the center frequency, this filter attenuates the noise, but passes the tone and limited range of the noise. The tone's raised threshold is determined by the amount of noise passing through the filter, so that "...the threshold is assumed to correspond to a certain signal-to-noise ratio at the output of the filter" (Moore, 1989). Clearly, an important task of the auditory filter is to provide the best signal-to-noise ratio, and to help detectability of the tone.

The bandwidth of a single auditory filter (or rather a hypothetical ideal band pass filter of the same center frequency) is called the *critical bandwidth*. Equivalent rectangular bandwidth (Bracewell, 1986) has been adopted as another appropriate measure of the bandwidth of real auditory filters (Moore and Glasberg, 1987). The bandwidth of an auditory filter depends on the center frequency, regardless of which estimate is used, in a remarkably predictable way. Moreover, it has been shown in many different experiments that subjects' responses to complex sounds depend on whether the bandwidth of these sounds is narrower or wider than the critical bandwidth.

Complex sounds give rise to many single auditory filters, so that a filterbank is formed with center frequencies close to the prominent sound components. A procedure for the estimation of the resulting filter shape has been proposed by Glasberg and Moore (1990). The shape of a single auditory filter is derived first by using the notched noise method (Patterson, 1976). In this method, the filter shape is approximated by the *rounded exponential* function

(Patterson et al., 1982), with the assumption of the power spectrum model (short spectrum fluctuations are averaged). The shape is frequency dependent and level dependent. Therefore, a scheme is also provided to efficiently sum the individual filters for a given complex sound to form the filterbank or an *excitation pattern* (Moore and Glasberg, 1987).

The concept of an excitation pattern seems to be essential in many aspects of sound perception. Loudness of a complex sound can be calculated from an excitation pattern (Zwicker and Scharf, 1965) by transforming it to specific loudness domain, and then integrating the area under the pattern. Such a model can be particularly useful in intensity discrimination problems (Florentine and Buus, 1981). The knowledge of the excitation pattern allows one to solve the problem of the frequency selectivity of complex tones, i.e. to asses how well individual partials will be resolved. The transformation of a spectral frame of a complex sound to the corresponding excitation pattern brings out its perceptually relevant features, such as corresponding formant distribution. Therefore, the notion of the excitation pattern may potentially be useful in such areas as speech recognition (Karjalainen, 1984), or the objective assessment of subjective quality of sound (Karjalainen, 1985).

1.6.3 Overlap-add Sound Synthesis Method

The overlap-add sound synthesis method is based on the short time spectral decomposition of an original sound (Allen and Rabiner, 1977; Portnoff, 1980). The following principle is exploited in the method. If X_n is the Fourier transform of the n-th frame windowed signal x,

$$x_n(m) = x(m)\, w(n-m), \qquad X_n(e^{j\omega_k}) = \sum_m x_n(m)\, e^{-j\omega_k m}$$

then x(n) can be reconstructed by taking the inverse Fourier transform of $X_n(e^{j\omega_k})$, dividing out the window, and adding the overlaping frames. In the overlap-add method $X_n(e^{j\omega_k})$ is inverse transformed for each n at which the analysis was performed:

$$y(m) = \sum_n \left(\sum_k X_n(e^{j\omega_k})\, e^{j\omega_k m} \right) = \sum_n y_n(m) = N\, x(m) \sum_n w(n-m)$$

It can be proven (Allen, 1977) that if

$$\sum_n w(n-m) = \text{constant},$$

then perfect reconstruction of x(m) is achieved.

The overlap-add analysis/synthesis method has found applications in speech coding (McAulay and Quatieri, 1986), and computer music (Smith and Serra, 1987; Serra and Smith, 1990). After the analysis part, modifications of the sound are usually made in order to compress the original sound, or to achieve high quality transformations for a variety of musical sounds. For this reason an analysis of the original sound is performed frame by frame.

During the analysis part, frequencies, amplitudes, and phases of the sound components (its partials) are analysed. The analysis involves tracking the trajectories of the partials (amplitudes and frequencies). The basic idea lies in the assumption that the trajectories can be found by tracing prominent spectral peaks, i.e. that the amplitudes and frequencies belonging to the trajectories change slowly and continuously from one spectral frame to another. The set of trajectories so derived is then used to synthesize the deterministic part of the sound. To avoid discontinuities at the frame borders, phase unwrapping and interpolating has to be done for a given frequency track. The cubic phase interpolation function can be used (McAulay and Quatieri, 1986), or simply linear interpolation (Serra and Smith, 1990) to calculate the instantaneous phase from the set of frequencies. Finally, additive synthesis can be employed to obtain the sound result. The stochastic part of the signal can be treated in the same way as the deterministic part (consonants in McAulay and Quatieri work are also treated as sine components), or it can be modeled as the generation of a noise signal, which has the residual characteristics of the original sound (Serra and Smith, 1990).

The analyzed signal can be transformed by time-scale modifications, pitch, and timbral modifications. A variety of interesting effects can be created in this way, including cross-synthesis between two sounds with different characteristics, such as 'talking bell', 'singing flute', and so on.

Chapter 2

Experiment 1

2.1 Rationale

Since the earliest research on sound distance, the variance of intensity along physical distance has been regarded as the strongest discriminable factor. This line of thinking had probably its origin in the research on distance perception in a free field, because in such conditions loudness is the most obvious cue.

In the reverberant field, the overall picture becomes much more complicated, however. Other strong cues provided by the room reflections of the sound source become significant, allowing the listener to distinguish between the loudness and the distance of the sound. In particular, the direct-to-reverberant sound energy ratio (d/r) has been shown to be another important cue to distance perception. Although successful experiments were performed to show that, in a reverberant environment, d/r is a major absolute distance cue whereas loudness is not (Mershon and King, 1975), loudness was present in most of the experiments on distance perception. Except the work by Mershon and Bowers (1979), there has been no systematic research to show the effect of reverberation when the loudness cue is removed.

Yet, an intuition coming from everyday experience, brings about the hypothesis, that loudness may actually be a confusing factor for judging the distance of sound in a room. Loudness can be a conflicting cue, since distant loud sounds may be of the same loudness as closer soft sounds. Although we normally do not easily get confused because there are also other cues suggesting the distance, this experience poses a question about the relative independence of the sound distance from its loudness.

To this end, an experiment was designed addressing the following hypothesis: is loudness a necessary cue to estimate the apparent distance of sound? More specifically, it was to check whether distance judgements of sounds in a reverberant (room) environment would still be precise after their loudness had been matched and therefore eliminated as a cue.

The intensity of sound is only one of the factors correlated with the perception of loudness. It works as a function of frequency, and a few loudness summation models have been proposed for complex sounds (Zwicker & Scharf, 1965). Although these models work in simple cases, there is no general agreement about loudness of complex sounds. In particular, they do not provide a summation method in time (Scharf, 1978), because the integration constant of the ear is not yet known. Moreover, in the estimation of loudness, there is a considerable variance between subjects. To eliminate these difficulties, the task of matching the loudness of the stimuli was left to the individual subjects.

We wanted to use natural sounds, and natural reverberant conditions as well. While it was subject to discussion, if 'reverberation' could be well controlled in a natural environment, we believed that subjects would perform the best in such conditions. In psychoacoustical experiments simple sounds, such as pure tones and broadband pulses, were traditionally used as stimuli. Repeatability was probably among the strongest reasons for their use. These stimuli sometimes worked in a different way than more complicated sounds, in particular, speech, in experiments on distance perception (Gardner, 1969), and loudness perception (Warren, 1973). Our study was intended to be useful for composers, and other musicians interested in distance judgement, therefore we decided to use natural instrumental sounds as stimuli instead. We also devoted an experiment in section three to see if there was an advantage to using familiar stimuli over non-familiar sounds.

2.2 Method

2.2.1 Subjects

Thirty one subjects participated in experiment 1. They were all musicians with extensive musical training and experience, most of them being either active musicians or researchers at the Stanford Center for Computer Research in Music and Acoustics (CCRMA). All subjects had normal hearing.

2.2.2 Stimuli and Apparatus

2.2.2.1 The Recording Room

The stimuli for the experiment were recorded in a typical rehearsal hall (the Braun Rehearsal Hall at Stanford). Horizontal and vertical crossections of this room are depicted in figure 2.1 a, b. Positions of the microphone (dummy head), and loudspeaker are marked in the figure. The Braun Rehearsal Hall is a medium size room (16.6 m x 10.5 m x 7.5 m) designed for music rehearsals, including small student orchestra practice sessions. Curved gypsum structures were suspended below the ceiling along the longer axis of the room to reduce flutter echo and to faster reflect sound to the audience. The reverberance of the room could be adjusted by using three curtains marked by the dotted lines in figure 2.1. The curtains were drawn during the recording of the stimuli. The sole window had a shutter, which was closed. There were no chairs or furniture in the room, except for a grand piano at the far end.

A room can be best characterized by its impulse responses. To measure them in the recording room, white noise responses were recorded at the loudspeaker positions equal to 2, 3, ..., 8 meters from the dummy. A maximum-

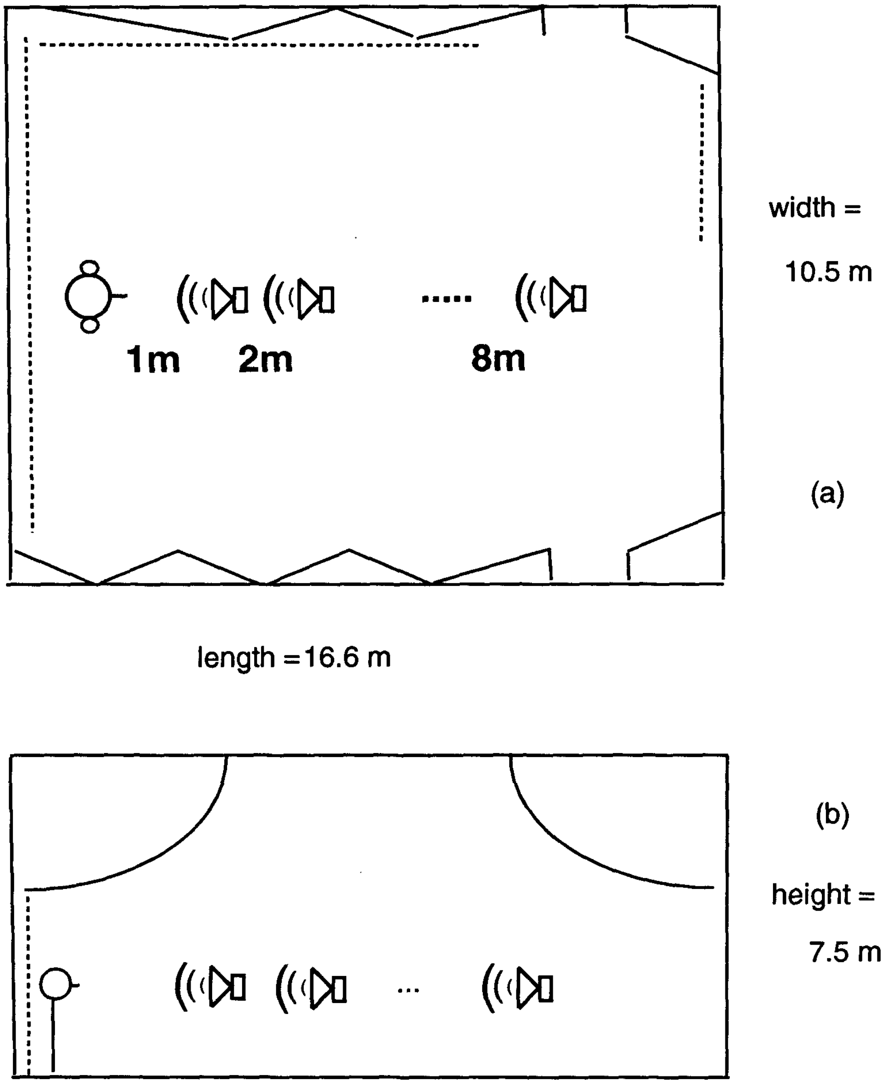
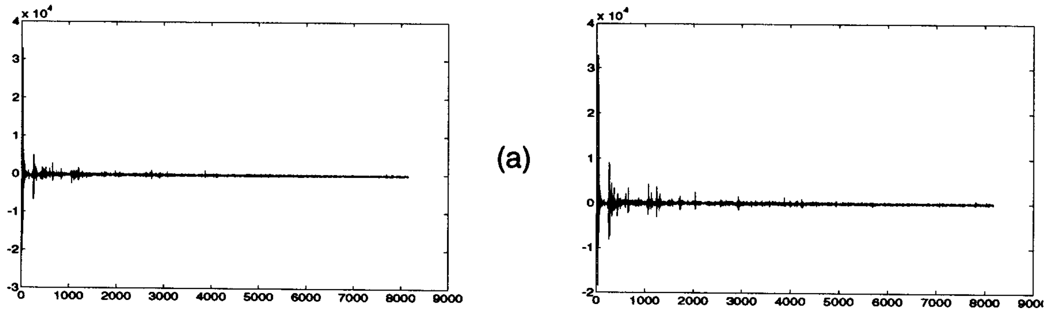

width =
10.5 m
1m
2m
.....
8m
(a)
length =16.6 m
(b)
height =
7.5 m
...

Figure 2.1. Crosssections of the Braun Rehearsal Hall.
(a) horizontal,
(b) vertical

length sequence was reproduced for this purpose (4 periods, 32767 samples = 743 ms each) at the above distances with a constant playback level. An HD-1 Meyer Sound High Definition Audio Monitor was used for the playback (along with a YAMAHA P-2200 professional series amplifier, and a Panasonic sv-3700 professional Digital Audio Tape Deck). A KEMAR artificial head and another Panasonic sv-3700 digital tape deck were used for the recording. The recording level was kept constant regardless the position of the loudspeaker. A 44100 kHz sampling rate was used.

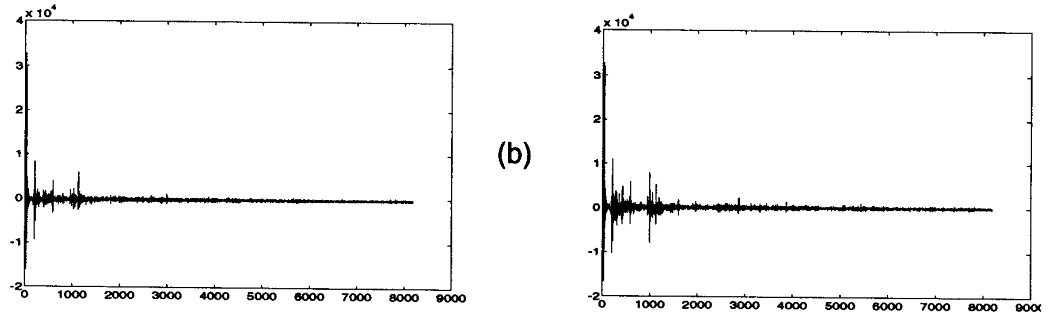
The impulse responses of the room were reconstructed next from the recorded noise responses by using the fast Hadamard transform (Borish, 1984). The first 200 ms of the impulse responses are shown in figure 2.2 a to g. The asymmetrical distribution of the reflected energy in both ears of the dummy is noticeable. This effect can be attributed to the attenuation of the sound by the large, sidewall, curtain. The first reflection of the impulse responses always comes from the floor. Therefore, the time gap between the direct sound and the first reflection becomes shorter as the distance grows (from 5 ms at 2 m to 1.7 at 8 m). The reverberation time T_{60} , calculated from *the method of integrated impulse response* (Schroeder, 1965), equals about 1.05 sec. The early reverberation time equals approximately 30 ms.

Attributes considered to be relevant for distance perception were next extracted from the impulse responses. Figure 2.3a shows the direct-to-reverberant sound energy ratio as a function of the loudspeaker distance. '+' denotes the values in the left ear of the dummy, 'o' – the values in the right ear, and '*' – the arithmetic mean. The horizontal line marks the reverberation distance, i.e. the distance at which the reflected sound energy begins to exceed the direct sound energy. The deviation from monotonicity at about 4 meters in one of the channels is probably caused by the absorption by the non-symmetrically placed curtains. Figure 2.3b shows the clarity index C50, i.e. the



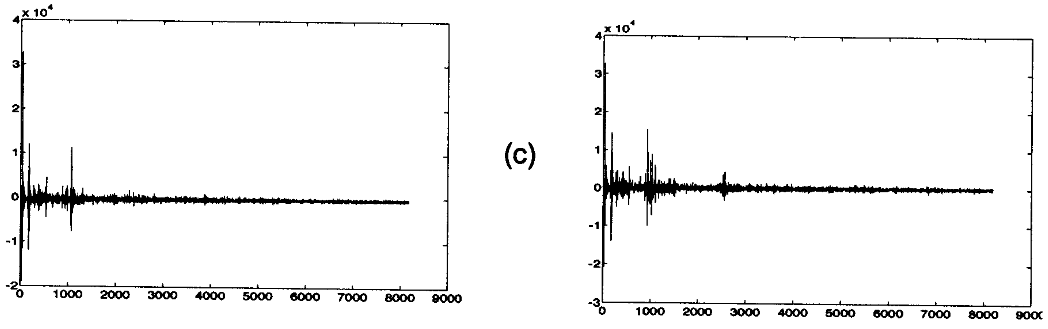
x 10^4
4
3
2
1
0
-1
-2
-3
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(a)



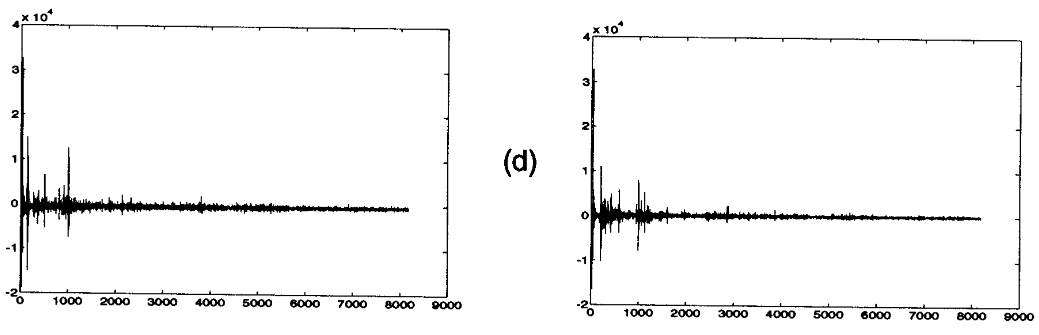
x 10^4
4
3
2
1
0
-1
-2
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(b)



x 10^4
4
3
2
1
0
-1
-2
-3
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(c)



x 10^4
4
3
2
1
0
-1
-2
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(d)

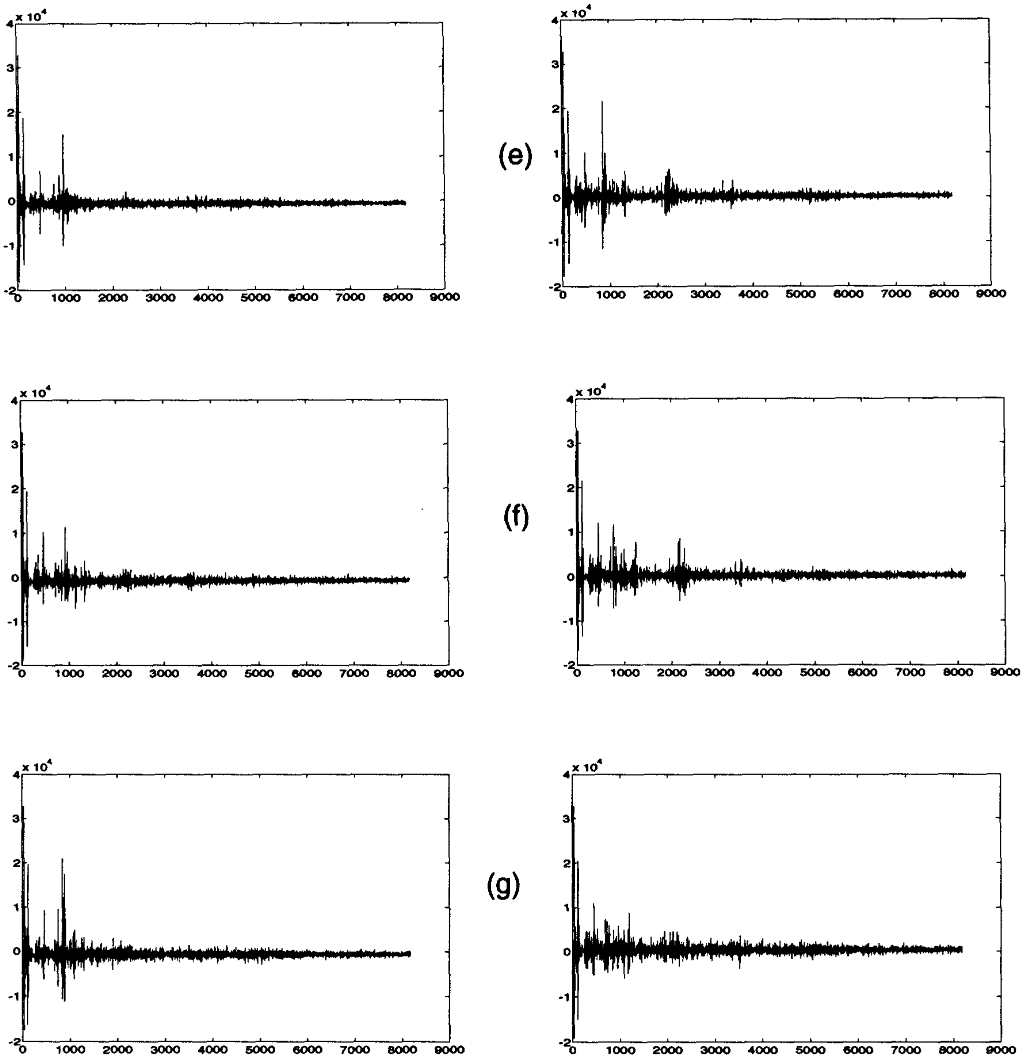


Figure 2.2 (a) - (g). Impulse responses of the Braun Rehearsal Hall at the distances from the microphone equal to 2,3,4,5,6,7, and 8 meters respectively. The first 200 ms is shown (x-axis is scaled in samples, 44.1 kHz sampling rate). The left column shows the left, and right column the right channel.

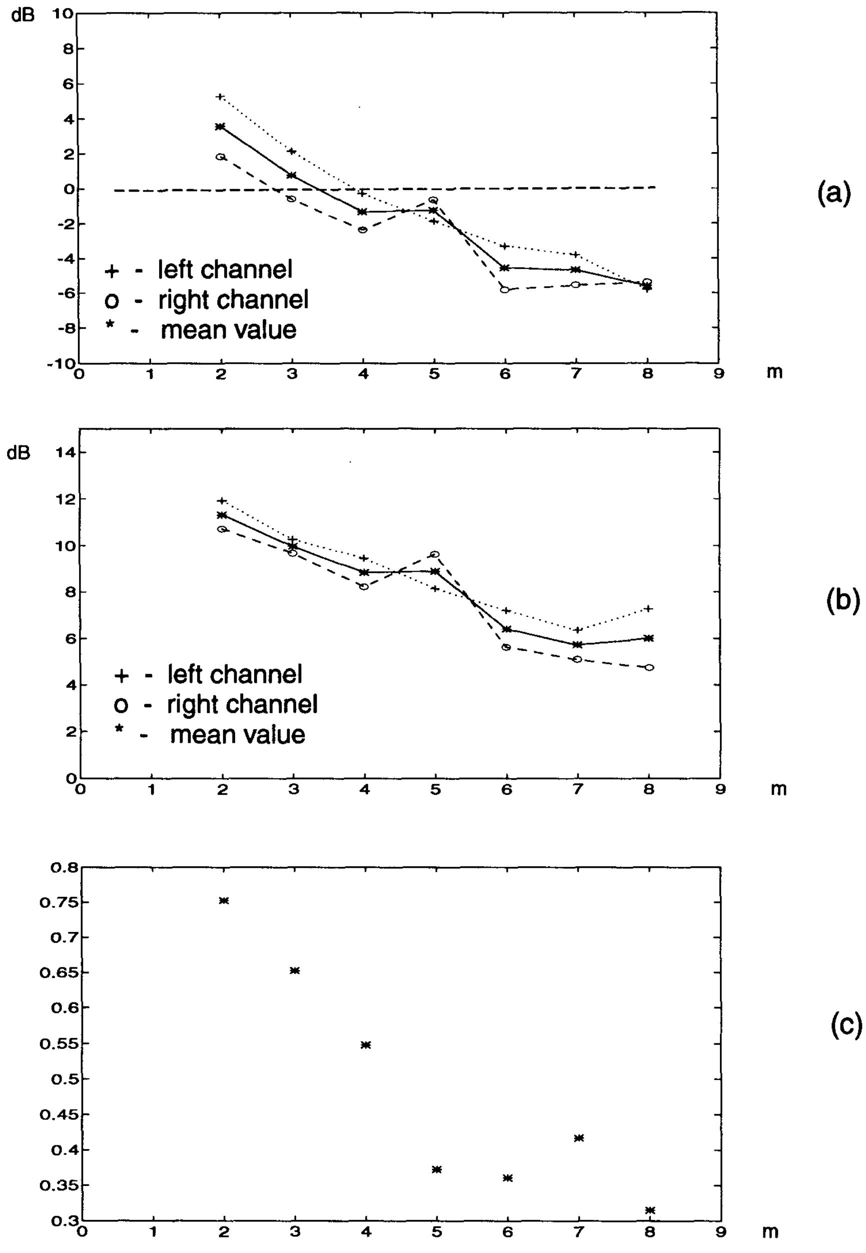


Figure 2.3. Reverberation parameters in the Braun Rehearsal Hall.
(a) Direct-to-reverberant sound energy ratio. The horizontal dashed line marks the reverberation distance
(b) Clarity index C50
(c) Interaural cross-correlation coefficient (first 100 ms) between the ears of the dummy.

energy of the first 50 ms early reflection part (including the direct sound) to the energy of the remaining reflections.

The interaural cross-correlation coefficient between the ears of the dummy was also calculated. The formula and detailed discussion of its importance for the feeling of spaciousness can be found in Kutruff (1991), page 201, or in Ando (1985). Only the first 100 ms of the impulse responses were used. The coefficient decreases pretty rapidly with distance (see figure 2.3c). This can be seen as the indication of the growing energy of the lateral reflections in relation to the energy coming from the front in the median plane.

2.2.2.2 Stimuli

A violin sound (a440, with medium vibrato rate, and duration of about 1.8 sec) was recorded in an anechoic chamber. This sound was then reproduced in the Braun Rehearsal Hall at distances of 1,2,3,...,8 meters with a constant playback level and recorded with an artificial head at a constant gain of recording input. During the recording the speaker was moved along the longer axis of the room, whereas the head remained in the same position. The same positions of the speaker and the dummy head were used for the recording of the noise responses (except an additional recording at 1 meter)–see figure 2.1.

A MDM-4 near-field monitor by California Standard Instruments was used for the playback along with a YAMAHA P-2200 professional series amplifier, and a Panasonic sv-3700 professional Digital Audio Tape Deck. The KEMAR artificial head and another Panasonic sv-3700 Tape Deck were used for recording (with the sampling rate equal to 44100 kHz).

The recording room (the Braun Rehearsal Hall) was specially designed for concerts and rehearsals, but it was not completely acoustically isolated for recording purposes. There was some outside background noise present in the room at the time of the stimuli recording. As the background noise was uniform in frequency, we decided to filter it out. The following procedure was applied (Cook 1990). First, the noise floor was estimated, i.e. given the noise recording alone, its power spectrum was calculated. Next, the spectra of recorded stimuli were compared to the noise spectrum. For each FFT transform bin, if FFT amplitude fell below the noise threshold it was attenuated by a given (constant) ratio. This procedure was then repeated for successive time frames, and an envelope applied to smooth the result. After the filtering, an informal listening test was done to detect possible artifacts and timbral changes introduced by the filtering. As the changes were small in comparison to a great improvement of the sound quality, we decided the sounds were representative samples and reflected what takes place acoustically in a 'real' room environment.

Short-time spectrograms of selected stimuli are shown in figure 2.4 to better characterize the properties of the sounds. The calculation was performed with the Hanning window at a 43 Hz frequency resolution. The time resolution was equal to 3 ms. 8 kHz frequency range and 40 dB dynamics were displayed in figure 2.4. Power spectra of selected frames were also estimated by using the Welch method (Oppenheim and Schafer, 1989). The sound for the analysis was taken at 300 ms after the beginning of each stimulus. Each power spectrum was derived as the result of averaging over six successive FFT frames (frequency resolution equal to 21.5 Hz) with the overlap factor of half a frame (23 ms). The left column of figure 2.5 shows the spectra of the stimuli recorded at 1,3,5, and 8 meters. Only the left channel is shown, because the right channel spectra are very similar. The right column of figure 2.5 shows the spectral envelope and formants of the power spectra. The homomorphic deconvolution method (Oppenheim and Schafer, 1989) was applied to smooth the spectra. The spectra show three distinct formant regions.The lowest region narrows, and all the formants fade out when the distance increases. The bandwidth of the spectra can be formally expressed as the equivalent rectangular bandwidth (Bracewell, 1986). The equivalent rectangular bandwidth was calculated from the power

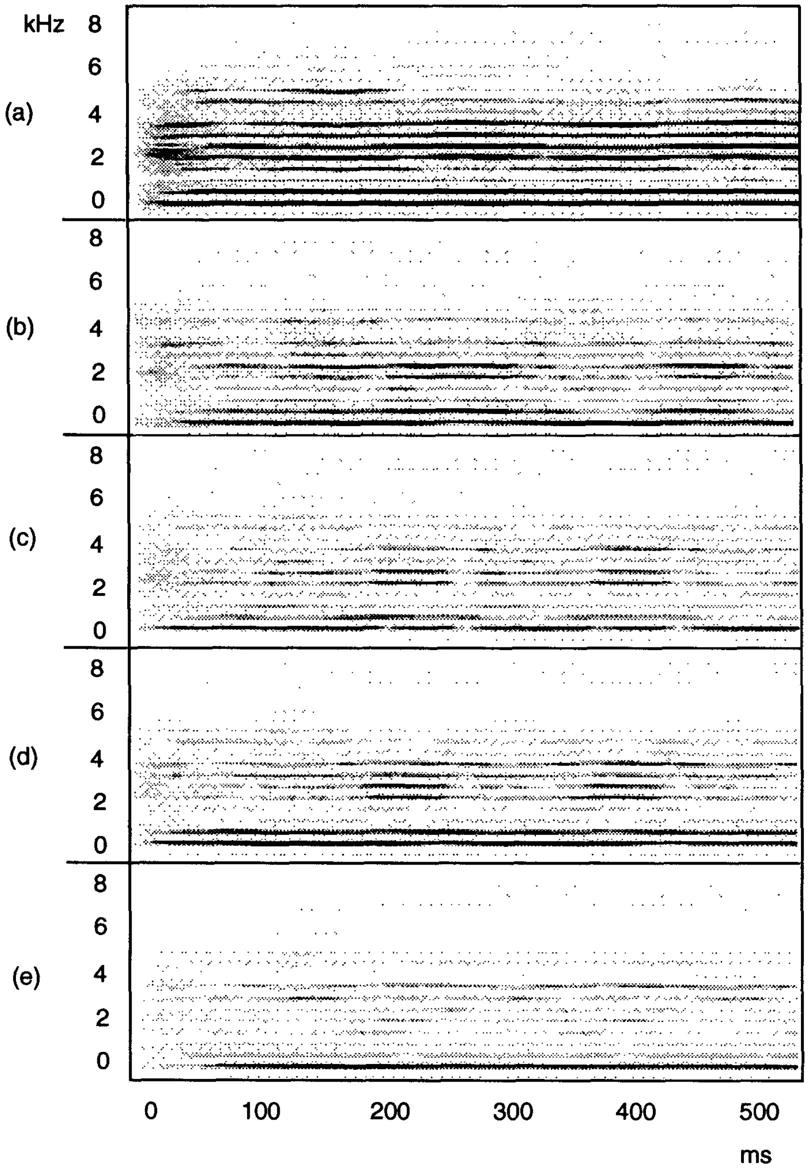


Figure 2.4 (a)-(e). Spectrograms of selected stimuli at 1, 3, 4, 5, and 8 meters from the microphone,
Parameters of the analysis were set as follows:
Hanning window, frequency resolution: 43 Hz,
time resolution: 3 ms.
Frequency range 0-8 kHz, and dynamic range: 40 dB are displayed.

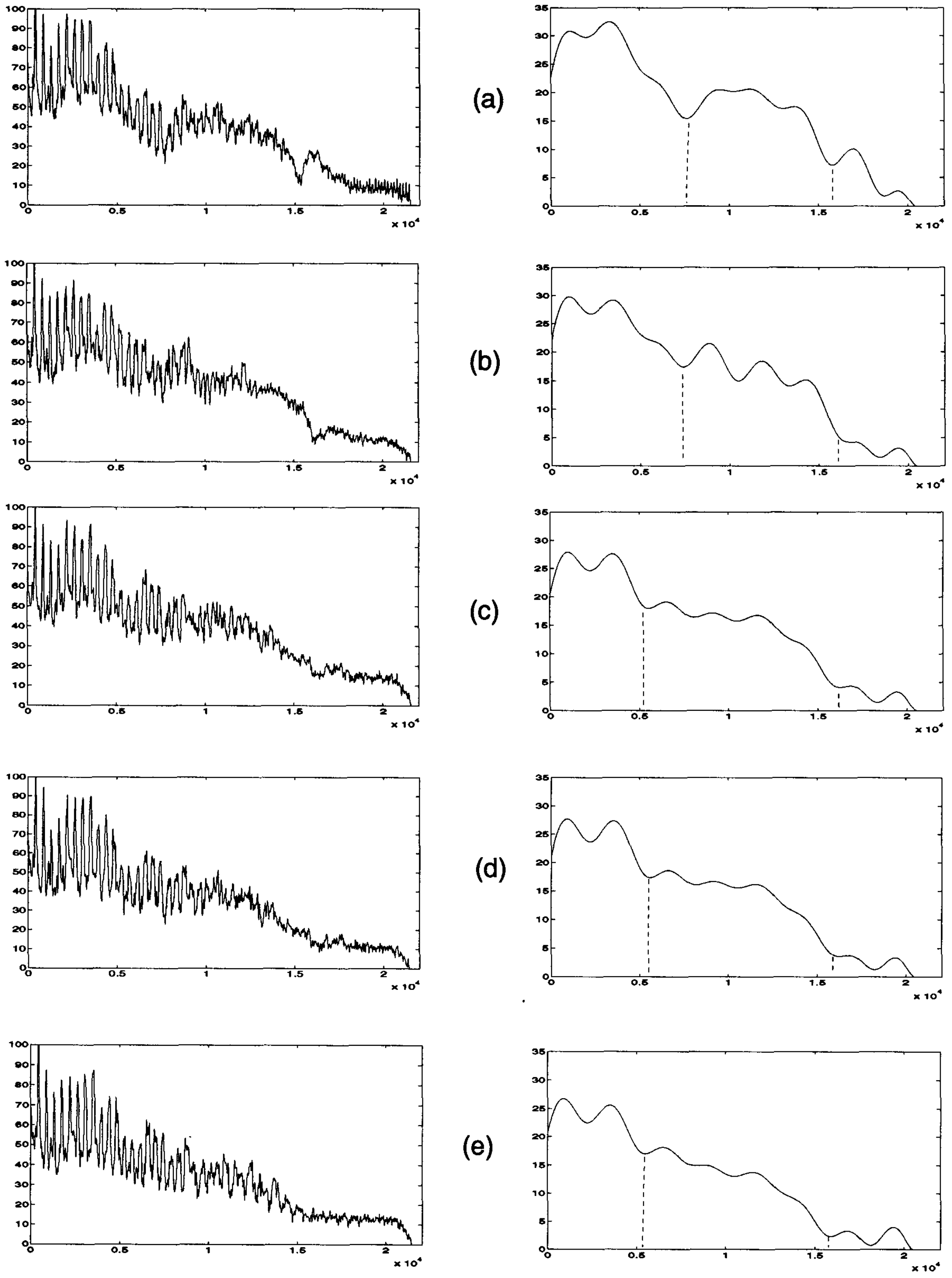


Figure 2.5 (a)-(e). Power spectra of the stimuli at 1, 3, 4, 5, and 8 meters from the microphone (the left channel only). The right column shows the formants. The frequency axis is scaled in kHz.

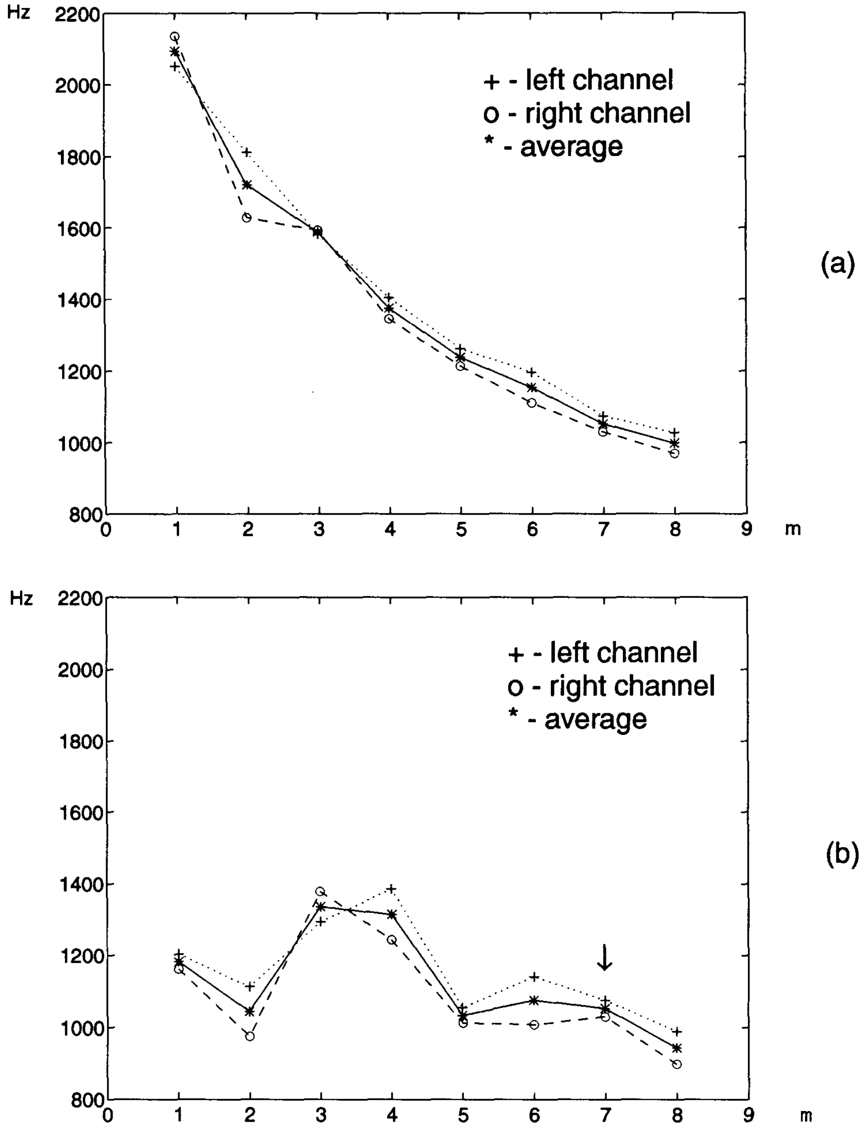


Figure 2.6. The equivalent rectangular bandwidth of the stimuli as a function of distance. 65 dB dynamics range was assumed to be perceptually relevant.

(a) original sounds
(b) the normalized stimuli after they were matched for loudness. The sound at 7 m (pointed to by the arrow) was the reference for the loudness normalization. See text for details.

spectra (non-smoothed) with the dynamics of 65 dB, because the spectral energy below this level is of little perceptual importance. The bandwidth of recorded sounds shown on figure 2.6a clearly decreases with the increasing distance. Points in figure 2.6b mark the bandwidth of the stimuli after the loudness match (for the results of the loudness match see section 2.3.1). The power of the stimuli was matched according to the results of the loudness match with respect to the power of the sound at 7 meters, pointed to by the arrow on figure 2.6b. The bandwidth remains quite stable between 900 and 1200 Hz, except for the sounds at 3 and 4 meters. However, its dependence on distance is no longer monotonic.

Power ratios between the sounds recorded at 2, 3, ..., 8 meters and the sound recorded at 1 meter were also calculated to express its loss along the distance in the room (figure 2.7). Each point represents the average of the left and right channel.

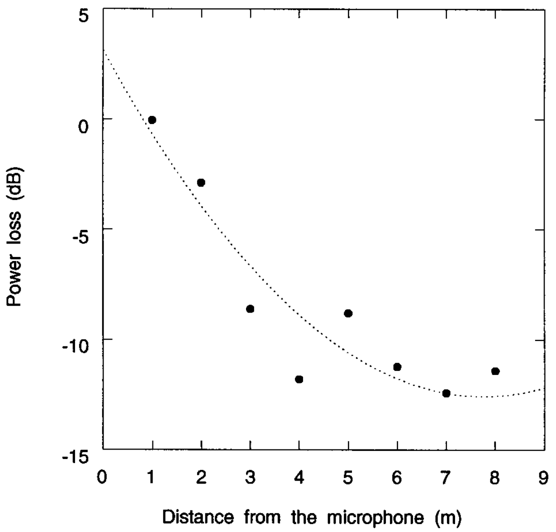


Figure 2.7 Power loss of the stimuli along the distance in the recording room

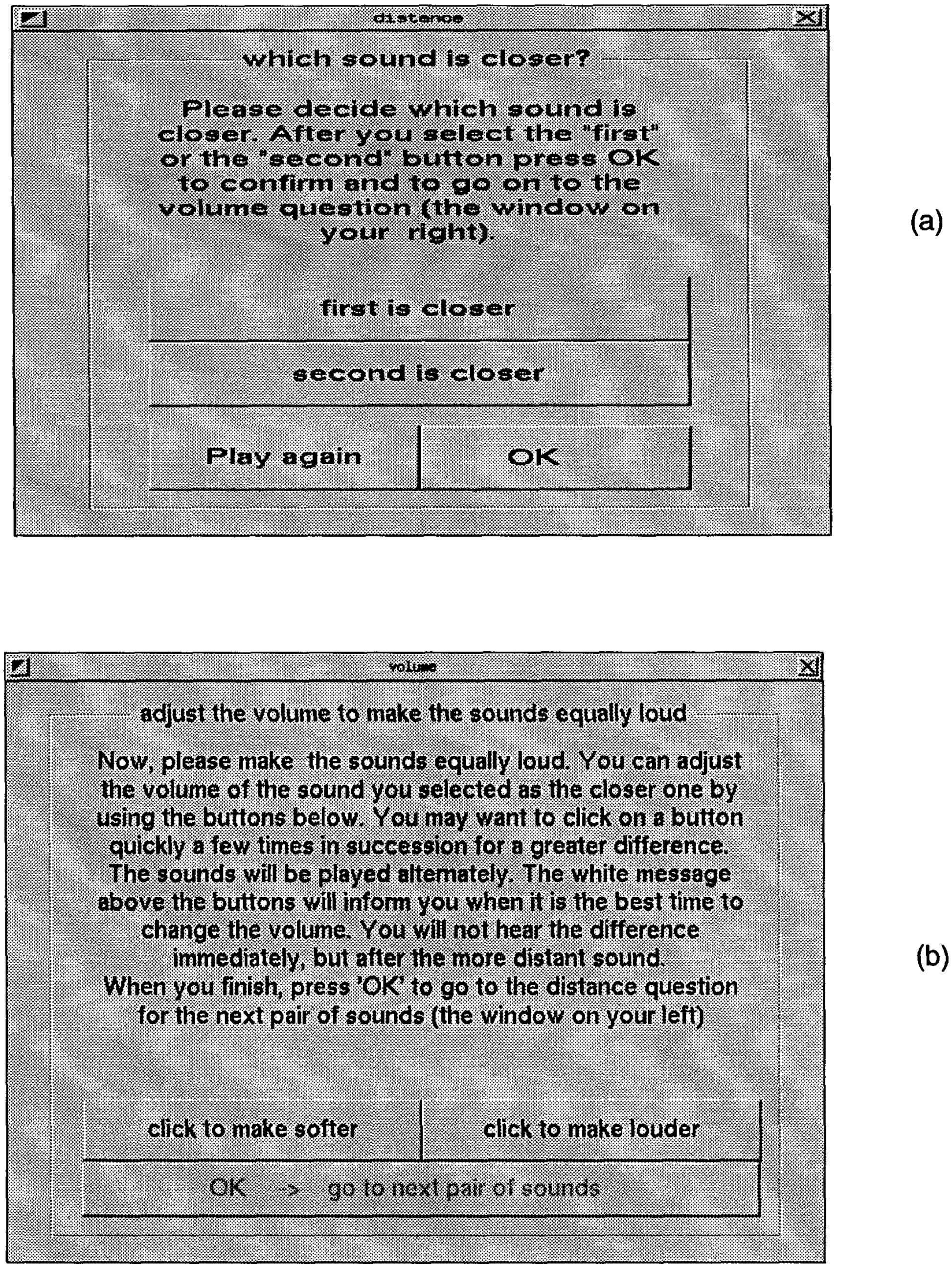
A quadratic regression:

$$\text{power loss (dB)} = 0.26 * \text{distance}^2 - 4.04 * \text{distance} + 3.12,$$

where 'distance' denotes the distance between the sound source and the microphone, shows a rather good fit, with corrected r squared equal to 0.873. In general, power loss is inversely proportional to the squared distance, but there are some anomalous places in the room, in particular at 4m from the microphone. The excessively large energy loss at 4 m position was presumably caused by the ceiling structure. A substantial part of the sound might have been reflected back from the dummy, and then it was absorbed by the curtains and the walls. Some node cancellation came into play at this distance as well. While it is not apparent from the averaged power spectra (figure 2.5), the time evolution of the spectrum (figure 2.4c) shows that the energy of the fundamental and the first harmonics became weak at repeated time instances. The energy gets stronger again at the 5 m sound position.

As described above, the recorded stimuli differed as to several physical parameters because of the effect introduced by the room, i.e. the amount of reverberation, bandwidth, envelope (especially in the attack) and the effective duration (energy of more distant sounds was smaller and decreased faster). They also differed in respect to the power (or intensity).

The sounds were next presented in a formal listening test through headphones. SONY MDR V-6 headphones were used, and a NeXT computer, to control the presentation. A custom designed Objective-C program allowed the subjects to press buttons on the monitor screen (i.e. to click a mouse button on a graphical item representing a button on the monitor screen) that permitted self-control of the presentation. The user interface for this experiment is depicted in figure 2.8.



distance
which sound is closer?
first is closer
second is closer
Play again
OK
(a)

Please decide which sound is closer. After you select the "first" or the "second" button press OK to confirm and to go on to the volume question (the window on your right).

volume
adjust the volume to make the sounds equally loud
click to make softer
click to make louder
OK -> go to next pair of sounds
(b)

Now, please make the sounds equally loud. You can adjust the volume of the sound you selected as the closer one by using the buttons below. You may want to click on a button quickly a few times in succession for a greater difference. The sounds will be played alternately. The white message above the buttons will inform you when it is the best time to change the volume. You will not hear the difference immediately, but after the more distant sound.
When you finish, press 'OK' to go to the distance question for the next pair of sounds (the window on your left)

Figure 2.8. The user interface
(a) for the distance estimation
(b) for the loudness match

2.2.3 Procedure

A two alternative forced choice pair comparison experiment was administered. Sounds recorded at different distances were matched in pairs. All the sound combinations recorded at 8 different distances were included, except those that differed only by the sound order within the pair. Therefore, instead of all 64 pairs of the eight stimuli, only 8x7/2 = 28 pairs were presented to the subjects. The paired sounds were played in a randomized order, which was also different for each run of the test. For a given pair of sounds, the order of the sounds in the pair was randomized as well, and was different for each trial. Each pair was presented once, but the whole test was repeated twice for each subject (with a different randomization) to check reliability of the answers.

In the first part of the experiment, for each reproduced pair, subjects had to decide which sound of the pair was closer to them when compared to the other sound. They were allowed to repeat the pair of sounds an unlimited number of times, until making their final decision. The result was then scored as one if it corresponded with the order of the physical distance of the sounds, or as zero if it was a failure in this respect.

Afterwards, the subjects were requested to match the loudness of the sounds in the pair. In particular, they were instructed to match the loudness of the sound they had decided was closer to the loudness of the other (more distant) sound, which remained constant. Each pair of sounds was played in a loop as many times as desired to allow a precise match. Subjects were permitted to adjust the intensity of the sound by pressing one of two buttons ('softer' or 'louder') with a resolution of 0.5 dB for each click (see figure 2.8b). The power of the closer sound was then adjusted by the corresponding decibel factor and the sound appeared with the new volume during the next repetition in the loop. This amount, i.e. the factor by which subject attenuated the stronger sound to achieve a satisfactory loudness match, expressed in decibels, was eventually stored to a disk file. Subjects had to answer the questions in alternation: i.e. for each pair the loudness match followed the distance question.

The attenuation factors were next used during the second stage of the experiment to reproduce equalized sounds the subjects had made. The equalization was done for each subject individually, by using the means of the two results from the set the subject provided for each pair during the first stage. Given the equally loud sounds, the subjects were asked to answer the distance question again, that is they were to estimate which sound in each pair was closer this time. The sounds in the pairs were randomized differently than during the first stage. Performance at this stage should reflect the result of attempting to eliminate the loudness cue.

2.3 Results

2.3.1 Reliability

The reliability of the loudness match can be based on the consistency between the two measurements each subject provided, i.e. on the ability of each person to repeat the measurement. This consistency can be formally expressed in terms of the correlation coefficient between these measurements. Pearson's correlation coefficient (paired) was calculated for each subject. The results are presented in the second column of table 2.1. All the subjects showed a high consistency between the two measurements. 29 of 31 subjects (93%) completed the loudness match with a very high correlation, greater than 0.9. The worst result, 0.792, entitled us to include all the subjects for further analysis.

The mean adjustment differences of the loudness match, along with the standard deviations, were then estimated from the two sets of attenuation factors provided by each subject. They are also displayed in table 2.1. Because the differences are expected to be primarily unidirectional, a natural question that

Subject number	Pearson's correlation coefficient	Mean difference in dB	Standard deviation in dB
1	0.981	-0.38	0.73
2	0.932	-0.80	1.45
3	0.926	0.14	1.22
4	0.951	-0.13	1.19
5	0.930	-0.29	1.49
6	0.961	0.18	1.23
7	0.937	0.02	1.40
8	0.792	-3.05	3.92
9	0.962	-0.18	1.28
10	0.947	-0.66	1.35
11	0.924	-0.68	1.70
12	0.965	-0.59	1.01
13	0.945	-1.21	1.43
14	0.961	-0.05	1.12
15	0.981	0.14	0.79
16	0.912	-1.32	1.47
17	0.957	-0.18	1.40
18	0.965	0.18	1.04
19	0.939	0.71	1.25
20	0.946	-0.30	1.37
21	0.912	0.21	1.56
22	0.914	-0.43	1.33
23	0.949	-0.43	1.38
24	0.921	0.86	1.64
25	0.935	-0.11	1.58
26	0.947	-0.16	1.27
27	0.959	0.45	1.15
28	0.969	-0.27	1.21
29	0.938	0.16	1.23
30	0.948	-0.59	1.50
31	0.865	-1.02	1.67

Table 2.1. Loudness match. Comparison of the individual performances during the first and the second runs of the test.

arises is whether the mean differences are perceptible. We are convinced that the attenuation factors in the experiment approximate the smallest detectible loudness changes for our sounds, because all our subjects were highly sophisticated listeners accustomed to listening to electroacoustic music. We cannot prove that the adjustments define such changes for our stimuli though, because of the small number of repeated observations (two) for each sound pair per subject. Nevertheless, we will attempt to compare the means with the smallest detectable changes in loudness derived for different stimuli in other experiments. They have been shown to be reasonably stable, regardless of the absolute level (from 20 dB and 100 dB above the threshold), and have values between 0.5 dB and 1 dB for wideband noise, or between 1.5 dB and 0.3 dB for pure 1000 Hz tones (Moore, 1989). For 20 of our subjects the mean difference is smaller than 0.5 dB, and for other 7 it lies between 0.59 and 1.02. We can argue, therefore, that the loudness match provided by these 27 subjects was representative of such loudness matches, especially when we realize that the adjustment button resolution was not continuous, but varied with the step of 0.5 dB. For 26 subjects (84%) the standard deviation varies between 0.73 and 1.5 dB, which is acceptable also. Only subject number 8 shows relatively larger deviations, but on the whole the results prove that the subjects were able to repeat the results with a very high degree of accuracy.

A further check of the quality of the loudness match (attenuator) is depicted in figure 2.9. Mean gain adjustments are juxtaposed to the power losses across all the 28 test combinations on the scatter diagram. As would be expected in the ideal case, the adjustments closely follow power changes along the distance with a linear relationship. The calculation of the correlation coefficient equal to 0.970 between the powers and the gain adjustments reveals this very close correspondence. A regression straight line was fitted to the points as:

mean gain adjustment (dB) = 0.893 * power ratio (dB) - 1.370

where mean gain adjustment (attenuation) is averaged over the thirty one subjects. If the loss of power in the room were the only factor of the match, the coefficient of the line would be equal to 1, and would be identical to the dotted line on the picture. The somewhat smaller value of 0.893 may show an experimental bias, or may be the result of a relationship between loudness and distance perception. Further discussion of this possibility is delayed until section 2.4.

In working out the reliability of distance perception, we used a measure suitable for dichotomous answers of the test (Gower, 1985). Correlation coefficients displayed in table 2.2 are calculated between the binary answers for the corresponding pairs of the two runs of the test. Instead of using the Pearson's coefficient, a proportion of pairs was calculated, in which the values of both measurements were in agreement. All the proportions for the pairs of recorded sounds are above 0.85. The proportions for the pairs of equalized sounds also show a good consistency.

To estimate individual performances, we were interested in whether they were higher than those that would result from chance performance. In other words, the 95% confidence intervals of the binomial distributions based on the proportion of correct answers averaged for the 28 test conditions should not contain 0.5, if the performances are higher than chance. From the tables of the cumulative binomial distribution (N = 28) we can find that the above relationship holds for distributions based on proportions greater than 0.71. Thirty subjects achieved a better result, i.e. they were above the chance level. However, the reliability of distance estimates of the subject number 6 (r = 0.679) can be questioned.

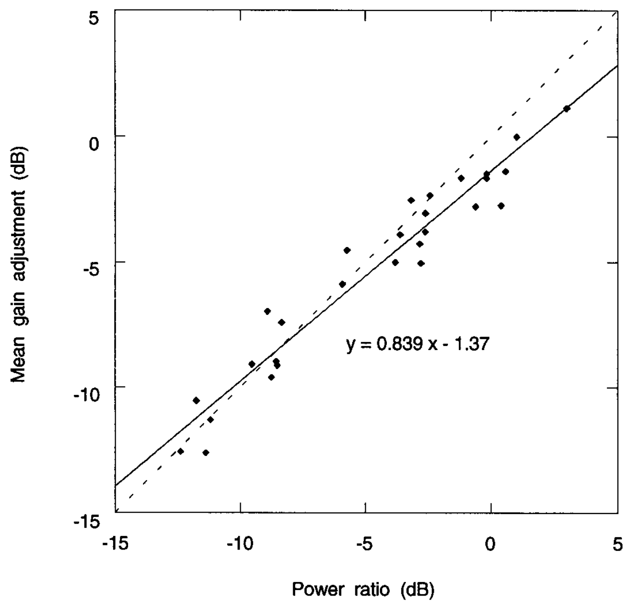


Figure 2.9. The mean gain adjustments as a function of the power ratio for each pair of sounds. The regression line differs slightly from the ideal loudness (the dashed line).

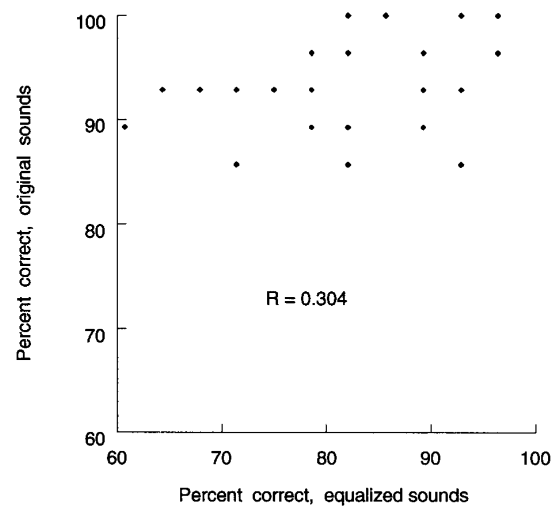


Figure 2.10. The scatter diagram of individual performances. For each point, the mean performances across the 28 test conditions on original sounds were paired with the corresponding performances on the equalized sounds. Points of the graph are uncorrelated, with the Pearson's coefficient equal to 0.304.

Subject number	Recorded sound			Equalized sounds		
	Correlation coefficient (dichotomous)	% correct		Correlation coefficient (dichotomous)	% correct	
		I	II		I	II
1	0.929	92.9	100.0	0.857	75.0	82.1
2	0.893	92.9	96.4	0.714	67.9	60.7
3	0.929	96.4	96.4	0.821	78.6	89.3
4	1.000	92.9	92.9	0.857	89.3	82.1
5	0.893	85.7	89.3	0.821	92.9	89.3
6	0.964	92.9	96.4	0.679	75.0	78.6
7	0.964	92.9	89.3	0.893	92.7	82.1
8	0.857	92.9	85.7	0.821	75.0	92.9
9	0.964	100.0	96.4	0.821	96.4	78.6
10	0.964	92.9	96.4	0.786	92.9	78.6
11	1.000	96.4	96.4	0.857	82.1	75.0
12	0.964	96.4	92.9	0.786	89.3	82.1
13	0.893	100.0	89.3	0.821	82.1	78.6
14	0.964	89.3	92.9	0.821	89.3	85.7
15	0.929	96.4	96.4	0.857	89.3	89.3
16	0.964	92.9	96.4	0.714	75.0	60.7
17	0.821	89.3	85.7	0.964	89.3	85.7
18	0.929	100.0	92.9	0.786	85.7	71.4
19	0.893	89.3	92.9	0.857	78.6	85.7
20	0.964	96.4	92.9	0.964	89.3	85.7
21	0.857	89.3	96.4	0.714	60.7	82.1
22	0.857	89.3	89.3	0.714	82.1	82.1
23	1.000	96.4	96.4	0.786	82.1	89.3
24	0.893	92.9	96.4	0.786	71.4	78.6
25	0.964	92.9	96.4	0.786	64.3	71.4
26	0.929	96.4	96.4	0.929	96.4	96.4
27	0.929	85.7	92.9	0.964	82.1	78.6
28	0.964	100.0	96.4	0.929	92.9	100.0
29	0.964	92.9	96.4	0.929	78.6	78.6
30	0.929	89.3	89.3	0.893	82.1	71.4
31	0.964	85.7	89.3	0.786	71.4	78.6

Table 2.2. Distance estimation. Comparison of the individual performances of the first and the second run of the test. The correlation coefficient is based on the proportion of the correct answers for the same pairs to the total number of pairs in the test (1st repetition of the test only)

The performance during the first stage was extremely high, usually above 85%. The relative distance of the recorded sounds was perceived clearly by the subjects. The performance during the second stage was also very high, with 19 out of 31 subjects better than 80%. The performance of the subject number 2, though very close to the 71% cutoff level, was problematic, but was not excluded because of the very high scores obtained by this person during the first part of the test.

Next, we checked for learning or fatigue by running the t-test (paired). A higher proportion of good answers during the second run of each part of test would indicate a learning effect, for example. The test calculated on the scores of the equalized sounds (second stage) showed no significant differences (t=0.543, df=30, p=0.591), and a very small mean difference of 0.922 (in percent correct). There was no evidence of a learning factor, nor of fatigue between the two repetitions of the test. Similar calculations on the results of the recorded sounds showed no significant differences either (t=-0.740, df=30, p=0.465). Taking advantage of the very close correspondence, we decided to select the results of the first run of each stage only for further analysis.

2.3.2 Performance

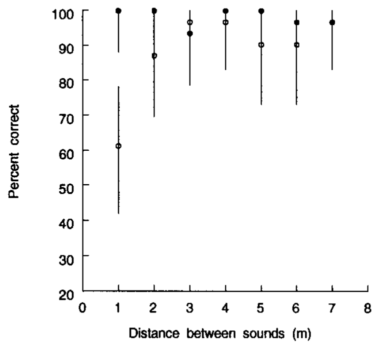
A two-way repeated measure variance analysis (presence of loudness x repetition of the test) was performed on the scores from table 2.2 to check whether the equalization of loudness was a significant factor for the perception of distance. The results showed loudness equalization as a significant effect (F=62.934, df=1/30, p<0.001), since the distances between original sounds were better perceived before loudness was removed. The analysis also confirmed that there was no learning or fatigue effect between the two repetitions of each test (F=0.029, df=1/30, p=0.866). No interaction between the loudness equalization and repetition of the test was discovered (F=0.781, df=1/30, p=0.384).

Figure 2.10 shows a scatter diagram of all the subjects' individual performances. For each subject, the score of the mean distance estimate on the original sounds across the 28 test conditions was calculated (on the ordinate) and paired with the mean performance on the equalized sounds (on the abscissa). The points are distributed quite randomly with a small correlation (R=0.304). Clearly, there is very little correlation between the subjects' average scores of the first and second parts of the test. Although the subjects generally performed better when loudness cue was present, one cannot predict from these results how they would perform in its absence. We have to be careful, because of the generally high performance on the original sounds. However, the plot suggests that all available cues were used for the distance estimate. Unexpectedly, loudness was not the primary one for all the subjects. When loudness was removed, the other cues were sufficient to provide the sensation of distance and the performance remained very high, in a few cases equal to or even greater than in the presence of loudness.

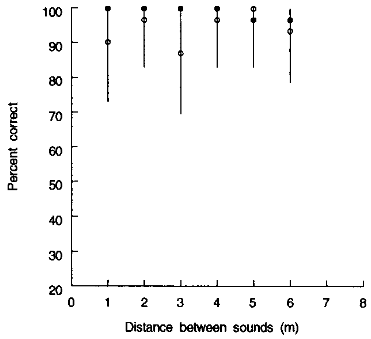
Two physical distances played a role in the discriminability of sound distance. For each pair of sounds in the experiment, the distance between the sounds in the pair provided differential cues, whereas the distance between the closer sound and the microphone defined a boundary to the listening condition (direct to reverberant energy ratio, and sound pressure in case of the original sounds, among most obvious parameters). As a result the subjects might perceive differently the distance of sounds separated by, say, one meter when they were located close to the listener, and when they were both positioned farther in the room. As we assumed that both of the distances affected the performance at the 28 combinations (pairs of sounds) of the experiment, we next intended to analyze how distance perception depended on them after loudness had been removed. In the first place, we expected that if loudness were essential for distance sensation a substantial performance loss would be observed for each of the 28 listening combinations.

In figures 2.11 a to g the performance was arranged as a function of the distance between the sounds for the closer sound positions equal to: 1,2, ..., and 8 meters from the listener, respectively. Black dots represent performance with the loudness cue present, and white dots represent performance with the loudness cue absent. As binomial distributions based on 31 observations are substantially skewed for proportions greater than 83 % (Mason et al., 1989), and in our cases the performance was usually higher, depicting the standard errors on the plot may be misleading. Instead, 95 % confidence intervals are marked to better show the differences in performance for individual conditions. Generally, loudness is significant for distance perception for each two dots whose confidence intervals do not overlap. To test how significant the performance differences are for the dots whose confidence intervals do overlap, we introduced the notion of the performance loss (and performance gain) after the loudness equalization.

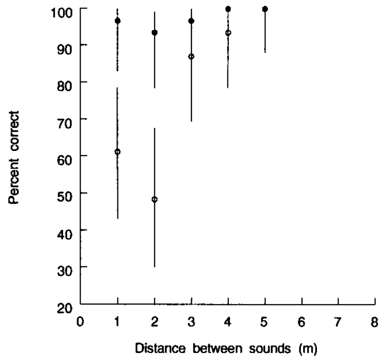
The *performance loss* for each of the 28 test combinations can be described as follows. For each subject, there was a performance loss of distance perception for a particular pair of sounds (j,k), if the person correctly judged the distance order of the two sounds with the loudness cue present, but failed to recognize which sound was closer after the sounds (j,k) were equalized. Similarly, we had a *performance gain*, if a subject wrongly judged a given pair of unequalized sounds and correctly after equalizing. Subjects who failed to recognize the difference in distance for the sounds (j,k) before and after equalization, or succeeded in both cases, did not show performance loss or gain according to this definition. An average performance loss for a particular one of the 28 sound combinations at different distances can thus be modeled by a binomial distribution. It may become necessary, however, to base the distribution on a different number of observations for each of the conditions, because only correct distance judgments of the original sounds are taken into consideration, and they may vary from one combination of distances to another. The variation in the number of observations is usually small in our experiment, however, because



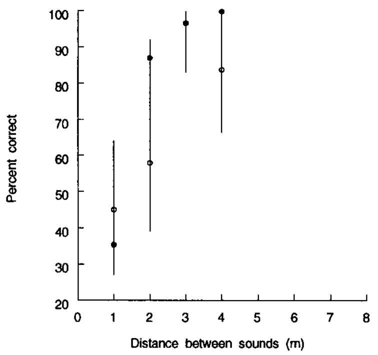
(a) at 1 meter from the microphone



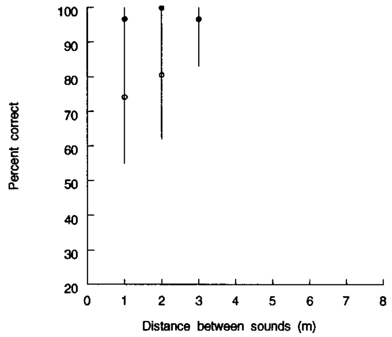
(b) at 2 meters from the microphone



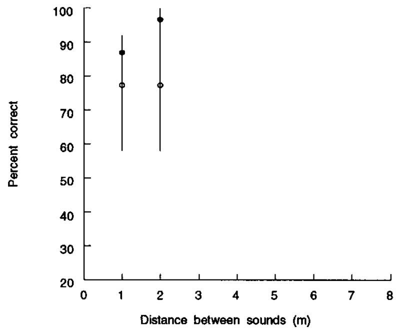
(c) at 3 meters from the microphone



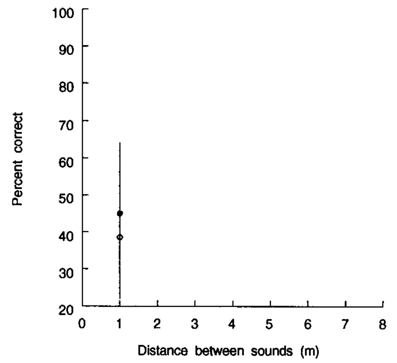
(d) at 4 meters from the microphone



(e) at 5 meters from the microphone



(f) at 6 meters from the microphone



(g) at 7 meters from the microphone

Figure 2.11 (a)-(g) Distance judgments as a function of the physical distance between sounds at 1,2,...,7 meter positions from the microphone. 95% confidence intervals are based on 31 observations. Performance on the original sounds is marked by black dots, and performance on the equalized sounds by white dots.

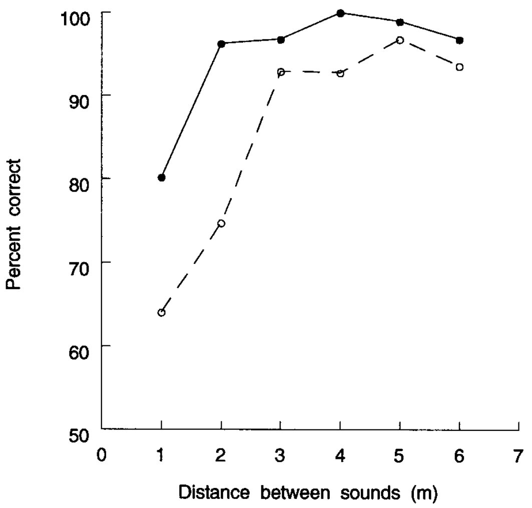
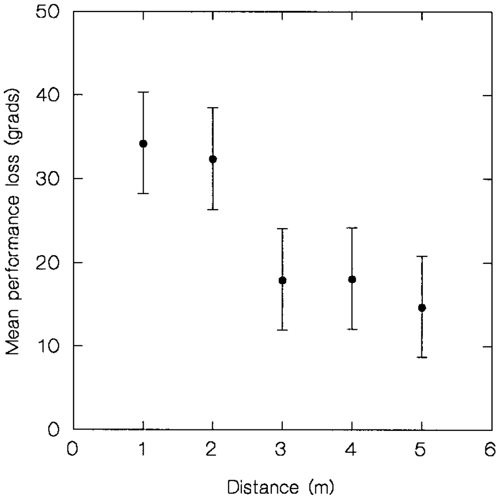
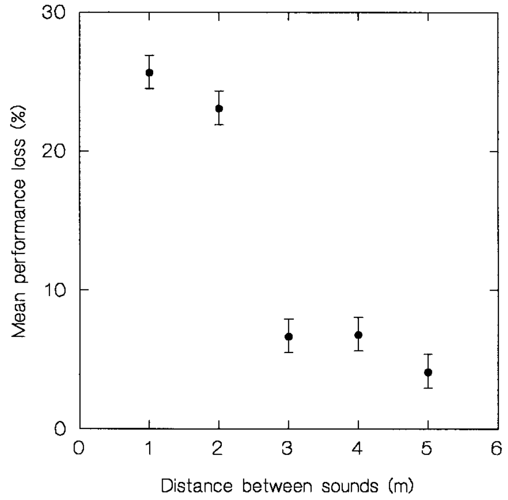


Figure 2.12 (a) Mean distance performance as a function of the physical distance between sounds. Performance on recorded sounds is marked by black dots, and performance on the equalized sounds by white dots. The point at 5 meters represents a pooled mean for sounds differing by 5, 6, or 7 meters.



(b) in the grads domain, after the arcsin transformation,



(c) converted back to the percent correct.

Figure 2.12 (b) - (c). Mean performance loss as a function of the physical distance between sounds. Point at 5 meters represents a pooled mean for sounds differing by 5, 6, or 7 meters. Fisher's LSD 95% confidence intervals are also marked.

of the generally excellent performance during the first part of the test. For the sake of easy comparison between the performance losses in different test combinations, we excluded two of them from further analysis, because for these combinations of sounds (1 meter apart at 4 and at 7 meters from the microphone) the number of incorrect answers on original sounds was substantial. For all the other combinations, this number varies between 27 and 31, with the mean value of 30.2 and median 30 (see denominator in the third column of table 2.3a). While in the analysis of a given condition the significance of performance loss was tested from a binomial distribution based on the true number of correct distance estimates on the original sounds, a reasonably good approximation can also be achieved by using a binomial distribution based on 30 observations.

In the next step, we were interested to point to the test conditions in which the performance loss was *small.* The notion of a small performance loss needs an explanation. If our subjects estimated a distance difference between two sounds (with or without the loudness cue) we had expected a certain performance loss, after shortening the physical distance between the sounds by more than the just noticeable difference. Let us in our experiment define the *small* performance loss as the number expressing the performance loss, which would be obtained as the result of a decrease in perceived distance by one jnd. We have some evidence about the acuity of distance perception from the work of Sheeline (1983). It is generally low, and greatly inferior to the precision of the discriminability of intensity or pitch perception. We assumed, therefore, that the smallest testable performance loss in our experiment (the difference in result corresponding to an error of 1 subject out of 31), equal to 3.2%, is a sufficiently low value to be smaller then the performance loss expected by the decrease of perceived distance between the sounds by one jnd. Therefore, we wanted to separate the test conditions, in which performance loss was greater than the 3.2% minimum obtainable difference from those conditions, in which performance loss was perceptually negligible, i.e. less or equal than 3.2%. Then, we hoped to determine the physical circumstances which contributed to the categorization.

Thus we compared the mean performance loss to the 3.2% limit in each of the test conditions. The tests were performed from binomial distributions based on the proportions of the observed performance loss, which are included in the 3rd column of table 2.3a (in percent). Tables of the cumulative binomial probabilities (Simon and Grubbs, 1952) were used to find the p-values. 'yes' in the 4th column, along with a small p-value, marks the cases in which the hypothesis had to be rejected. The results are collected in table 2.3a. For the sounds 1 m distant from one another at all distances, but 2 meters from the microphone, the performance loss was significantly greater than 3.2%. Similarly, for sounds separated by 2 meters, at all distances except at 1 and 2 meters, the performance loss was greater than 3.2%. 'yes' in the 5th column shows the cases in which we may be confident that the performance loss was really smaller than 3.2% or equal. It also shows the corresponding probabilities (the 'beta' values). For sounds separated by more than 2 meters, in 10 cases out of 15 the performance loss was smaller than 3.2%. For these sounds, in only one case the performance loss was greater than 3.2%. In other cases (marked by '?'), we did not have a sufficient test power to make any reasonable statement.

There are only two cases in which it is justified to test the significance of performance gain in our experiment. These are the conditions excluded from the above analysis, at 4, and at 7 meters, with the sounds separated by 1 meter. A surprisingly low discriminability of the distance on original sounds shows the confusion of the subjects (table 2.3b). If sound loudness was the conflicting cue, we should observe a performance gain in this group of subjects, after they equalized the sounds. As before, we categorized the test conditions into those in which the performance gain was substantial, i.e. larger than 3.2%, and those in which it was small.

distance to the closer sound in m	distance between the sounds in m	estimated performance loss in % correct / no.of observ.	perfor-mance loss greater than 3.2% / p-value	perfor-mance loss smaller than 3.2% / β - value
1	1	38.71 / 31	yes / 0.000	no / 0.008
	2	12.90 / 31	no / 0.077	? / 0.638
	3	3.45 / 29	no / 0.736	yes / 1.000
	4	3.27 / 31	no / 0.736	yes / 1.000
	5	9.68 / 31	no / 0.184	? / 0.638
	6	10.00 / 30	no / 0.184	? / 0.638
	7	3.33 / 30	no / 0.736	yes / 1.000
2	1	9.68 / 31	no / 0.184	? / 0.638
	2	3.27 / 31	no / 0.736	yes / 1.000
	3	12.90 / 31	no / 0.077	? / 0.638
	4	3.27 / 31	no / 0.736	yes / 1.000
	5	0.00 / 30	no / 1.000	yes / 1.000
	6	6.67 / 30	no / 0.397	yes / 1.000
3	1	40.00 / 30	yes / 0.000	no / 0.000
	2	51.72 / 29	yes / 0.000	no / 0.000
	3	13.33 / 30	no / 0.077	? / 0.638
	4	6.45 / 31	no / 0.397	yes / 1.000
	5	0.00 / 31	no / 1.000	yes / 1.000
4	1	72.73 / 11	excluded	
	2	44.44 / 27	yes / 0.000	no / 0.000
	3	3.33 / 30	no / 0.736	yes / 1.000
	4	16.13 / 31	yes / 0.030	no / 0.264
5	1	26.67 / 30	yes / 0.001	no / 0.017
	2	19.36 / 31	yes / 0.011	no / 0.077
	3	3.33 / 30	no / 0.736	yes / 1.000
6	1	18.52 / 27	yes / 0.028	no / 0.264
	2	23.33 / 30	yes / 0.003	no / 0.077
7	1	64.29 / 14	excluded	

Table 2.3 (a). The significance of the differences between the performances on the recorded and equalized sounds. The hypothesis of the performance loss being greater than 3.2% was tested.

distance to the closer sound in m	distance between the sounds in m	estimated performance gain in % correct / no.of observ.	performance gain / p-value
4	1	40.00/ 20	yes / 0.000
7	1	52.94/ 17	yes / 0.002

Table 2.3 (b). The significance of the differences between the performances on the recorded and equalized sounds. The hypothesis of the performance gain being smaller than 3.2% was tested.

distance to the closer sound in m	distance between the sounds in m	estimated performance loss in % correct / no.of observ.	performance loss / p-value
4	1	72.73/ 11	yes / 0.002
7	1	64.29/ 14	yes / 0.000

Table 2.3 (c). The significance of the differences between the performances on the recorded and equalized sounds. The hypothesis of the performance loss being greater than 3.2% was tested

The results of the comparison of the mean performance gain to the 3.2% limit are depicted in table 2.3b. As the p-values are very small, we are entitled to conclude that for the two conditions a significant performance gain indeed took place. For selected combinations of distances of the test (see figures 2.11d, and g) subjects were dealing better with distance after loudness was matched. This also suggests that in some instances reverberation cue was preferred over loudness cue. The remaining group of subjects, i.e. those who were correct in these test conditions, was tested for performance loss (table 2.3c). Interestingly, they showed a significant performance loss after equalizing the stimuli, which means that they used mostly loudness as a cue and were lost after it was removed.

The influence of the distance between the sounds on the performance loss was tested in greater detail. Figure 2.12a shows mean performance on the unequalized sounds (black dots), and the equalized sounds (white dots) as a function of the distance between sounds. Means of sounds distant from each other by 5, 6, and 7 meters were pooled and depicted as one black/white dot. Instead of testing this plot directly, the corresponding performance losses were collected to form a table for the univariate analysis of variance. As there were too few observations available of performance losses corresponding to the sounds distant by 5, 6, and 7 meters from one another, they were grouped together. To make it possible to compare the proportions (performance loss) as a function of distance in an analysis of variance, 30 was adopted as a common number of observations at each of the conditions. As before, the conditions at 4 and at 7 meters with the sounds 1 meter apart, were excluded from the analysis because the number of correct observations of the distances on recorded sounds was much different from 30 (equal to 11, and 14). At some of the conditions, performance losses were close to 0%, therefore the corresponding binomial distributions were skewed. For this reason, the frequencies (counts) of the performance loss were first expressed in degrees by the arcus sinus transformation according to the formula given below (Mason et al.,1989), to make them more normally distributed, and to stabilize the variance.

$$200 * \text{arc sin}\left(\text{sqrt}\left((y+0.375) / 30.75 \right)\right) / \pi \qquad \text{(grads)}$$

Next, a univariate analysis of variance was run, to estimate the means of performance loss as a function of distance between sounds. The results show that distance between sounds is a highly significant factor of the performance loss (F=5.366, df=4/21, p=0.004). Fisher's least significant difference pairwise comparison test detected a significant difference in performance loss between two groups of sounds (p= 0.024). Sounds separated by 1 and 2 meters belong to one group, and sounds separated by more than 3 meters to the other. Figure 2.12b shows the 95% Fisher's LSD confidence intervals marked on the estimated mean performance loss in the grads domain versus distance between the sounds. Figure 2.12c shows the mean performance loss converted back to the percent correct. It can be then concluded that the performance loss is significantly larger for the sounds separated by 1 and 2 meters than for the others.

2.3.3 Scaling

In the metric estimation of the performance parameters done in section 2.3.2 more weight was given to the pairs in which the sounds were close to one another in comparison to those in which they were far apart, because there were more direct observations for such sounds (7 for the sounds distant by 1 m, but only one per subject for the sounds distant by 7 m). However, if we assume that monotonicity and transitivity are in effect in distance perception, we have some indirect estimates of the performances. For example, performance for the pair of sounds at (1m,8m) can be obtained from the estimates for the pairs of sounds at (1m, 7m), and (7m,8m), or from the pairs at (1m,6m), and (6m,8m), and so on. In this approach it is assumed that the performance reflects the closeness of the sounds: a high number for the percent correct judgement of the distance between the sounds in a pair would mean that the sound components are perceived as very distinct, hence distant from one another, but a small number would suggest that they are close on a perceptual scale. Thus, the percent correct numbers scaled by 100, express the probability of discrimination between the sounds. Put another way, they can represent dissimilarities between the equalized sounds with respect to their relative distances. The dissimilarities are included in table 2.4.

Distance between the stimulus and the dummy (m)	1 A	2 B	3 C	4 D	5 E	6 F	7 G	8 H
1 A	0							
2 B	.613	0						
3 C	.871	.903	0					
4 D	.968	.968	.613	0				
5 E	.968	.871	.483	.452	0			
6 F	.903	.968	.871	.581	.742	0		
7 G	.903	1.000	.935	.968	.806	.774	0	
8 H	.968	.935	1.000	.839	.968	.774	.387	0

Table 2.4. Dissimilarities between the stimuli in terms of distance

Next, nonmetric multidimensional scaling (a version of the KYST program) was performed on the data. Sounds were fitted in two dimensions in the Euclidean space. The scaling was stopped when the value of stress did not differ from the value of the previous iteration. The final stress equal to 6.42 % indicates a very good match of the points (Kruskal, 1964). Figure 2.13a shows the Shepard diagram, the plot of distances between the points of the final configuration versus corresponding dissimilarities between the sounds. The points adhere cleanly to a straight line (which would be the smoothed distances). Figure 2.13b shows the final configuration of the points/sounds in the two dimensions. Each letter shows the position of one sound: A corresponds to the sound at 1m, B - the sound at 2 m, and so on. The '2' on the plot means that points G and H are too closely spaced to be marked separately. The points indicate a general semicircular pattern, broken only by the sounds D, and E (at 4m, and at 5m). The arrow on the plot shows the path of the point E from its previous position E' on the circular structure. The stress of that configuration was equal to 7.55%, which is also a quite good approximation. All the sounds, except D and E, appear in their 'natural' order, i.e. they are ordered with respect to the physical distance. A closer inspection of the dissimilarities in table 2.4, however, shows that their entries increase monotonically with distance from the principal diagonal (with the maximum values for the sounds spaced by 5 meters). All the above facts suggest that a one dimensional solution should be sufficient to represent the data (Shepard, 1974).

Figures 2.13 c-d show the result of scaling in one dimension. The Shepard diagram seems to be acceptably smooth, although it shows a larger spread than that in two dimensions. This is related to the fact that the stress value of the final configuration is rather large, 22.79%. A possibility that simply a local minimum was found by the MDS program cannot be excluded, because the available MDS program gives only a limited control on the procedure; for example direct adjustment on the step size is not allowed (Wilkinson, 1992). Nevertheless, the stress value is large because the best distance discriminability occurs for the sounds spaced apart by 5 meters. The discriminability of the equally loud sounds separated by 6 and 7 meters is excellent indeed, but reverberation cues work optimally for the sounds spaced by 5 meters: in 2 cases out of 3 with 100% correct! However, the scaled percent correct might have not been the best measure of dissimilarity for the sounds separated by a large distance because they were too easy to distinguish from one another. When the task is tough, as in case of closely spaced sounds, the recognition errors made by the subjects can be assumed to be a measure of their resemblance, although it will always have a random component. When the task becomes easy and the recognition almost perfect, as it was in the case of largely separated sounds, the error is probably more random, and does not indicate the real distances any more. This is especially true in light of the physical data analysis. Figures 2.3 a-c, and figures 2.6a,b clearly show, that d/r ratios vary very little between the sounds at 6,7, and 8 meters, the cross-correlation coefficient becomes stable at these positions (except a jump at 7 m, which might be in conflict with other attributes), so does the bandwidth.

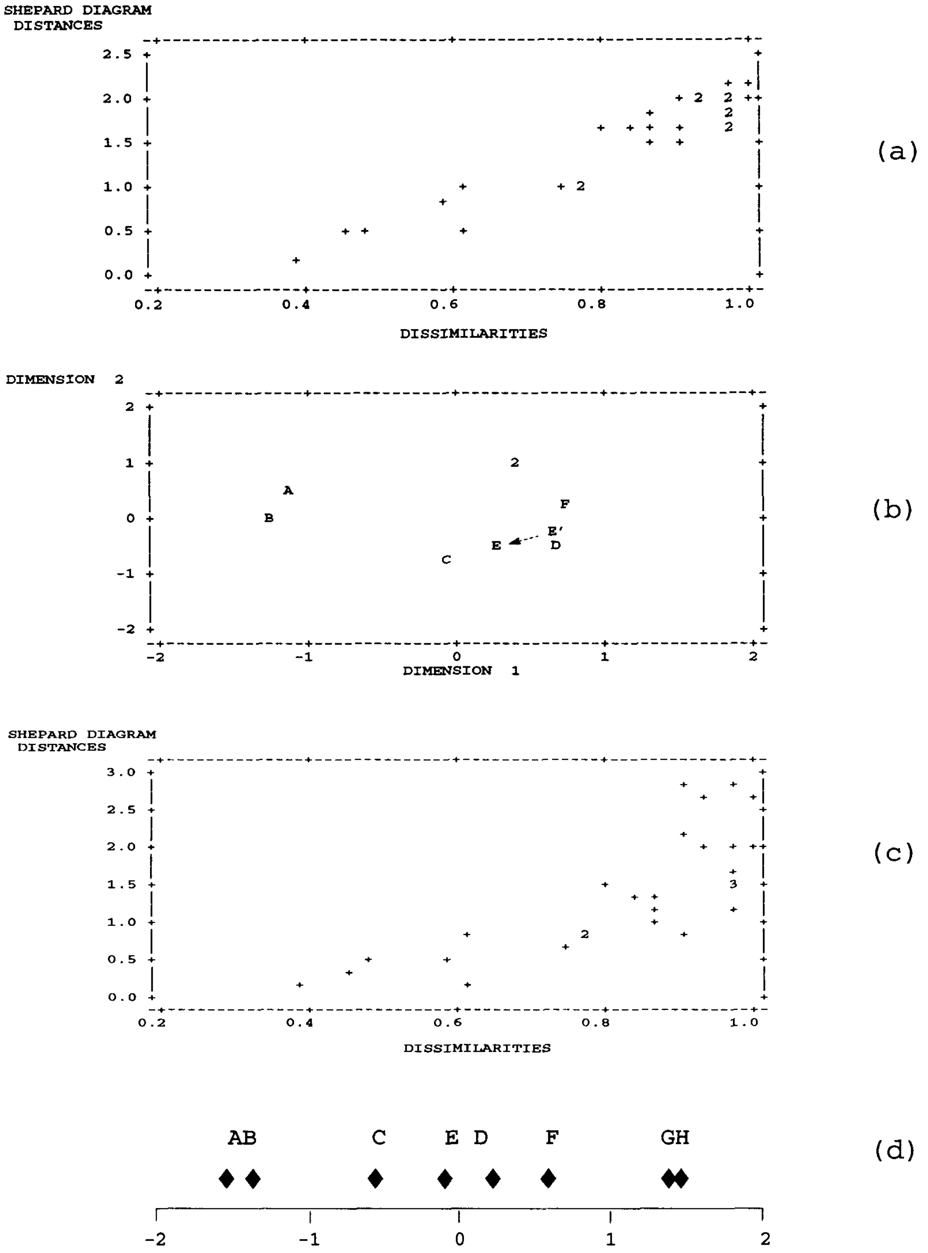


Figure 2.13. The result of the MDS scaling of distances between the sounds after the loudness match in 2-dimensional (a)-(b), and 1-dimensional (c)-(d) Euclidean space. The characters: A,B,C,D,E,F,2 correspond to sounds at 1,2,3,4,5,6 and 7,8 meters from the microphone.

The multidimensional scaling result seems to be in agreement with the thesis that, as in vision, the distance of sound is a one dimensional percept (Sheeline, 1983). What is more important, in the absence of the loudness cue, sound distance is still a clear percept, and remains possible to be modeled in one dimension.

2.4 Discussion

In most of the known experiments that include distance perception in reverberant conditions, intensity was also varied along the physical distance, providing a loudness cue. Yet, since work by Mershon and King (1975) it is known that, contrary to Bekesy (1936), there is little support for the possibility that intensity can serve as an absolute cue to auditory distance in a reverberant environment. Moreover, from the same source, we learn that reverberation can be such an absolute cue. Even given such an extremely demanding task as generating a number for distance on the first exposure to a new sound in unknown acoustical conditions, subjects can be successful in room conditions. These results suggest that other factors, such as reverberation, may be sufficient for distance estimation. Therefore, in the experiment by Mershon, and Bowers (1979), an attempt was undertaken to eliminate loudness as a cue by setting the power of the stimulus (broadband noise) to 60 dBA SPL at the listeners' position, regardless of the stimulus (loudspeaker) distance. In addition to reporting the apparent distances to the sound at five different positions, subjects were required to assign a number between 1 and 1000 to estimate the apparent loudness of the sound.

Numbers assigned to loudness by the subjects showed a large variability in both mean and standard error. As there was a tendency for greater physical distances to produce greater group reports of loudness (both mean and median), the authors hypothesized a possible relationship between perceived loudness and distance. With such a large standard error (between 54 %, and 83 % of the value of the mean scores), and this 'free' scoring system impossible to relate to the stimuli sound pressure, an objection could be raised that the loudness cue might not be completely eliminated, even though the sound pressure level was equated.

As noted before, loudness is a subjective phenomenon, invoking a subjective intensity of sound. By no means should loudness be mistakenly identified with the sound pressure, as it often is. Loudness 'resides' in the subjects, and is only related to–not equal to–the external stimulus sound pressure. For this reason, large individual differences exist, even in a relatively simple task of pure tone magnitude estimation (Scharf, 1978). To eliminate such variability in our experiment, we decided to focus on the individual differences in distance estimation for the experimental condition, i.e. distances of the sounds with loudness eliminated, and compare them to the control condition, in which the same subjects estimated the distances of sounds as they appeared unadjusted in the room. This was the main reason that in our experiment the equalization process was left to the subjects, who did it individually. Only for the purpose of analysis, the differences were averaged across the subjects.

Moreover, the estimation of loudness requires a conscious response on the part of the subject, and therefore a good understanding of the task. While it may be somewhat difficult to comprehend for somebody how to set sound loudness to make it 'twice as loud as the reference', everybody is well accustomed to adjusting the desired volume level on a stereo set, even though sound events may be very complex. Therefore, our procedure for matching loudness, applied in the experiment, was well defined and relatively easy to perform for the subjects.

In this experiment a headphone presentation was used. While it has been proven to be a good way to transmit an impression of a room to the listener, distance perception in such a presentation may be distorted by the presence of the so called 'inside the head locatedness' (Blauert, 1983). Precautions were taken to avoid this effect by using a KEMAR dummy head for recording. Nevertheless, we had to assume that distance estimation could be more difficult for the subjects than in real studio conditions. For this reason, we asked them only for a relative distance judgement in the paired comparison test. Even though this procedure naturally leads to nonmetric scaling, the advantage of the relative comparison method lies in decreasing confusion and being more obvious, than magnitude scaling. In this way we also wanted to avoid great variability in the results, which is present in the data of Mershon and Bowers (1979) (and other works on distance and loudness), or inconsistencies in the results, such as the discrepancies between Sheeline (1983), and Beagult (1991). Moreover, the traditional presupposition that a single stimulus intensity gives rise to a quantitatively unique subjective magnitude without reference to any comparison stimulus seems to be somewhat naive, and may itself be wrong (Shepard, 1981). In the 'absolute' magnitude scaling experiments, subjects seem to be strangely incapable of using the skill, and the results usually show a very large variability. Needless to say, no reliability check is possible there because learning skills are to be avoided by the very definition. There seems to be little advantage in using the magnitude estimation (or production) methods over the above binary relative judgment. The analysis presented by Shepard (1981) shows us that these methods do not determine any more than simply an ordinal structure of the corresponding psychological magnitudes. According to the author: "... these operations do not in themselves permit us to measure inner sensations in any quantitative sense." Therefore, the relative binary comparison of the stimuli presented in pairs is sufficient to derive relevant data, and is in fact beneficial, because of the well defined, simple task.

The success of the present experiment, therefore, depended to a great extent on the ability to remove the loudness cue from the recorded sounds. Though easy to understand, matching the loudness of complex sounds is still a difficult task, especially if the sounds differ considerably in other parameters such as timbre. Moreover, the ability of a person to match loudness can only be estimated as a relative measure, because as in all the experiments on loudness, there is no 'objective' criterion to fulfill (Scharf, 1978). At this point, the reliability of the subjects was extremely important. The results have shown that most of the subjects were able to repeat closely the loudness match. They show a surprisingly good repeatability of the loudness match, as well as rather small standard errors. Therefore we can conclude that the subjects had a very well established notion of what the loudness of sounds at different distances was, could distinguish it from the distance phenomenon, and their attempt at matching loudness was successful. A good correlation between the mean gain adjustments and the powers of the stimuli, as interpreted from figure 2.9, seems to confirm the known fact that, on the whole, intensity is the main factor of the loudness judgement. From the departure of the regression line from the dotted line denoting the ideal match, we can calculate that some of the subjects had to attenuate the stimuli by more than would result from pure power ratios. This relation is roughly inversely dependent on the distance between the sounds, but the evidence is insufficient to claim that this effect is associated with the reverberation cue, because the effect is also very small (~1.5 dB relative difference between sounds distant by 1 m and those 8 m apart), and because other phenomena, such as timbral changes, might have also contributed. Nevertheless, it clearly demonstrates, that a simple equalization of the sound pressure level of the stimuli, such as was done by Mershon and Bowers (1979), might not be sufficient to make the sounds equally loud.

The results of distance perception confirm the very high reliability of our subjects. They indicate a direct relationship between the presence of the loudness cue and the ability of distance recognition. The performance on the sounds without the loudness cue was worse than the performance on unadjusted sounds. This is not surprising, because the conditions of the test were set in such a way that the performance on the unadjusted sounds was usually at the very top of the scale. However, the distance recognition after the loudness cue has been removed was also very high, for most of the cases far above the chance level. Distance perception has got significantly worse only for the sounds close to one another, i.e. sounds different by 1m, or 2m. For such sounds, the difference in reverberation cues was not large enough to provide sufficient cues for the estimation of the distances. This is clear from the analysis of figures 2.11 a, e, and f, and from the results of scaling. Additional problems were caused by the existence of an acoustically 'dead' place in the room, at about 4m from the microphone. Sounds played from this point could not sufficiently excite the room to allow a proper identification of distance at the dummy position. The performance got significantly worse when one of the sounds was placed there, which could be easily seen from figures 2.11 b, c, and d. Placing the closer sound at this place caused a radically worse performance even on recorded sounds (figure 2.11d). This effect can possibly be explained as sound absorption by the unevenly distributed curtains and sound cancelation in some frequency regions. This anomaly also contributes to the large mean performance loss (figure 2.12) at 1, and 2 m. Nevertheless, the overall pattern is clear: small differences in the reverberation cues prevent good distance estimation for sounds with the loudness cue removed, when the sounds are close to one another. For the sounds spaced by more than 2 m, reverberation cues alone are strong enough to allow distance recognition. For such sounds, performance loss in comparison to the performance on unequalized sounds is insignificant.

An inspection of the scatter diagram in figure 2.10 indicates that individual performances in the case when the loudness cue was removed is, to a great extent, independent of the performances when the loudness cue was present. One cannot predict subjects' performance on loudness matched sounds from their performance on original sounds, except that it remains remarkably high. The overall image seems to be clear: loudness is not a necessary cue in estimating distance except for closely spaced sounds.

Although the scaling suggests that the distance of sound when loudness is missing is a unidimensional percept, association of a single physical sound attribute corresponding to the observed distance changes poses certain problems. The only known parameter which changes monotonically with physical distance over the total range of explored distances, and is invariable under the loudness match, is the direct-to-reverberant sound energy ratio. However, it is not at all clear, how subjects are able to use it. The d/r ratio concept is easy to grasp, estimate theoretically, and calculate in the computer. Yet it is difficult to comprehend how possibly the auditory system evaluates this ratio. We are very rarely exposed to either completely dry or extremely reverberant sounds; therefore, it seems feasible that we can only tell the difference between two sounds of different d/r ratios, and not be able to legitimately give any quantitative estimate of this parameter. The very large variance in the Mershon's experiments (1975, 1979) seems to support this claim. In addition to this argument, it is also difficult to imagine that the auditory system is able to derive the dry from the reverberant part of the sound, even for comparative judgment. We know, however, that reverberation can act through the timbral changes. Early reflections are especially likely to produce coloration (Kutruff, 1991) and possibly modulate the amplitude of sound. Yet, after the loudness match performed in the experiment, the bandwidth changes of the stimuli (figure 2.6b) are very small, and seem not to be related to distance. Nonetheless, the interaural cross-correlation coefficient is also not affected by the loudness match, and it could, to a certain extent, account for the distance change. The IACC, as shown in figure 2.3c, monotonically decreases with the increasing physical distance, except the fluctuation at the 7 meter sound position. Moreover, there is some evidence that the binaural system acts as a cross correlator in estimating interaural delay (Licklider, 1959), and that the IACC also, at least in part, accounts for the precedence effect (Zurek, 1980).

This hypothesis is in disagreement with the results of Kurozumi and Ohgushi (1983). They concluded that the increase of distance sensation was associated with an increase of the IACC. What seems to be strange in these results is that, in an average room, the IACC should become smaller with increasing physical distance, not larger, at least in the common range of distances. This is due to the fact, that the sound arriving from the front (with a limited angle range) makes the IACC magnitude larger. Conversely, the energy of lateral reflections contributes to the smaller magnitude of the IACC. If we keep the direct sound energy in the median plane in front of the listener constant, and move the source back, the energy of the lateral reflections usually makes the IACC smaller at the larger distances. This is what in fact can be reported after the analysis of the impulse responses in the above experiment.

Chapter 3

Experiment 2

3.1 Rationale

There seems to be a lot of confusion about the importance of sound familiarity for distance perception. Incidental experiments suggest that the distances of 'familiar' sounds are perceived more accurately than those of 'unfamiliar' sounds.

A possible 'ecological' explanation of this phenomenon could be based on our everyday experience. The ability to estimate the distance of sound is certainly a learned skill. Although it is probably not as crucial nowadays as the ability to estimate the distance of visual objects, it was inevitably of utmost importance in the early days of humanity, and still is in the animal world, where survival often depends on such acuity. Nevertheless, it is important for us as well, as we should not forget that we live in the world of signals, for example on the streets, where our safety continuously depends on the quick evaluation of visual and aural stimuli.

Do we learn how to perceive the distance of sound objects along with other sound features of the objects, such as their timbre? It seems feasible that we unconsciously learn how these features change with the physical distance, and in this way, how to estimate the actual distance. If this is the case, the familiarity of the objects helps us imagining and judging the distance because we know the objects well, and we are experienced in dealing with them. In some situations, however, we are unable to attend to the acoustical features of sound separately. Speech perception can serve here as an example, because we perceive the phonemes of the language, which normally form the meaning of speech, not the associated formant frequencies, which we are unable to hear out. In other cases, we are very good at selecting individual features from even a large complexity, like a conductor, who can simultaneously control the orchestra and listen to a single melodic line. As a rule, we also know how to abstract from room acoustics, when listening to music, or speech. Therefore, we are able to visualize room size given its acoustical image. On the other hand, we might not even recognize some of the common percussive instruments in unusual, for example anechoic conditions, because we have heard them only in rooms.

The question of familiarity can be reduced to the above dilemma: are we able to 'strip' our knowledge about the sound object from the context? In particular, are we able to generalize our ability to estimate the distance of familiar sound, and use this knowledge in a nontypical situation, in the distance judgement of a non-familiar sound?

In our experiment we were interested in a possible answer to this question. We assumed that a violin sound (a440) was familiar to musicians. We created another sound, which very closely approximated the violin sound with respect to many physical parameters. In fact, the approximations of the excitation pattern in the auditory nerve evoked by this sound and the violin prototype sound, calculated from the Glasberg and Moore model (1990), were nearly the same. The unfamiliar stimulus had also the same time evolution as the violin sound. Yet it was a very dense, noise-like sound, constituting an unfamiliar quality even for composers of computer music. In a formal test, we wanted to employ such expert listeners to check, whether the unfamiliar sounds would cause confusion in distance perception.

3.2 Method

3.2.1 Subjects

Twenty one subjects participated in the second experiment. They were all experts in hearing: musicians or music lovers, with extensive musical training and experience. Some subjects were either active musicians or researchers at the Stanford Center for Computer Research in Music and Acoustics (CCRMA). Others were participants in the summer course of computer music, and also active composers, or engineers from the fields of electroacoustics and computer music. All subjects reported normal hearing.

3.2.2 Stimuli and Apparatus

A dry violin tone (A 440) was used in the experiment (the same as in experiment one) as the familiar sound. The tone was 1.8 sec. long, played with a medium vibrato rate. The sound was reproduced through a loudspeaker in a medium size rehearsal hall (the Braun Rehearsal Hall at the Department of Music at Stanford University), and recorded at 2, 4, 6, and 8 meters from the microphone. A Meyer Sound HD-1 loudspeaker was used for the playback, and a KEMAR artificial head for the recording. In this way we created four familiar sounds.

The dry violin sound was further processed to obtain a dry unfamiliar sound. We wanted to create a sound which would preserve many physical attributes of the original, familiar sound, yet be transformed in such a way as to become a new experience even for composers of computer music. Initially, we wanted to create a noisy sound, which would have the same time evolution, as the violin sound, and the same energy distribution in the critical bands. Limited computer power prevented us from achieving this goal, however, we obtained a good approximation.

Synthesis by analysis in the form of the overlap-add method was used to generate the sound. An inharmonic analysis of the original violin sound was first performed by using the SMS program (Serra, 1989; Serra and Smith, 1990) with frame rate of 150 windows/sec (at 44100 Hz sampling rate). The Blackman-Harris 62 dB window was used, with an adaptive length equal to 3.5 periods of the fundamental frequency. Only harmonic analysis of the sound was performed. The power spectrum was then calculated. The excitation pattern was next calculated on the power spectrum for each frame by using a procedure adopted from Glasberg and Moore (1990), and Moore and Glasberg (1987). In this procedure, it is first assumed that each analyzed partial gives rise to a critical band excitation region. Based on the individual filter shapes around the partials within the critical regions, the procedure provides a summation method to generate the corresponding excitation pattern.

Next, in every spectral frame, each partial of the power spectrum was substituted with spectral components according to the following algorithm:

1. Frequency ranges f_l , f_h of the critical band (in Hz) corresponding to this partial were calculated from the experimental relations between the equivalent rectangular bandwidth of the partial and its frequency (Moore and Glasberg, 1987)

2. The critical band (f_l , f_h) was next subdivided into a number of subintervals, each 5 Hz wide.

3. In each of the 5 Hz intervals a spectral component was placed, with frequency randomly selected in this interval,

4. The amplitudes of the new spectral components were made n-times lower than the amplitude of the substituting partial, where $n = (f_h-f_l) / 5$,

5. Phases of the new spectral components were randomized.

Finally, the overlap-add method was used to synthesize the new, noise-like sound frame by frame.

The discrepancy between the excitation patterns of the original and generated sounds was also measured. Figure 3.1a shows the means of the absolute differences between the original and approximated excitation patterns in each spectral frame versus frame number. Figure 3.1b shows corresponding standard deviations. Figure 3.1c shows an example of the approximation for a steady state frame (number 120). The excitation pattern of the generated noise-like sound (dotted line) is more shallow in the peaks and valleys than the excitation pattern of the violin sound (continuous line), but in general, the approximation is very good.

The unfamiliar dry prototype was reproduced and recorded in the Braun Rehearsal Hall in the same way as was the violin sound, at 2,4,6, and 8 meters, to obtain four unfamiliar sounds.

Subjects were exposed to the sounds through headphones. Sony MDR V600 headphones were used, and the mouse-driven NeXT computer program (the same as in experiment 1) controlled the presentation.

3.2.3 Procedure

The procedure was basically identical to the scheme applied in the experiment 1. Only half of all the four familiar sound combinations were presented in pairs, so that each sound was played in combination with every sound except itself. The pairs were presented in a randomized order, which was also different for each run of the test. For a given pair of sounds, the order of sounds in the pair was randomized as well, and was different for each trial.

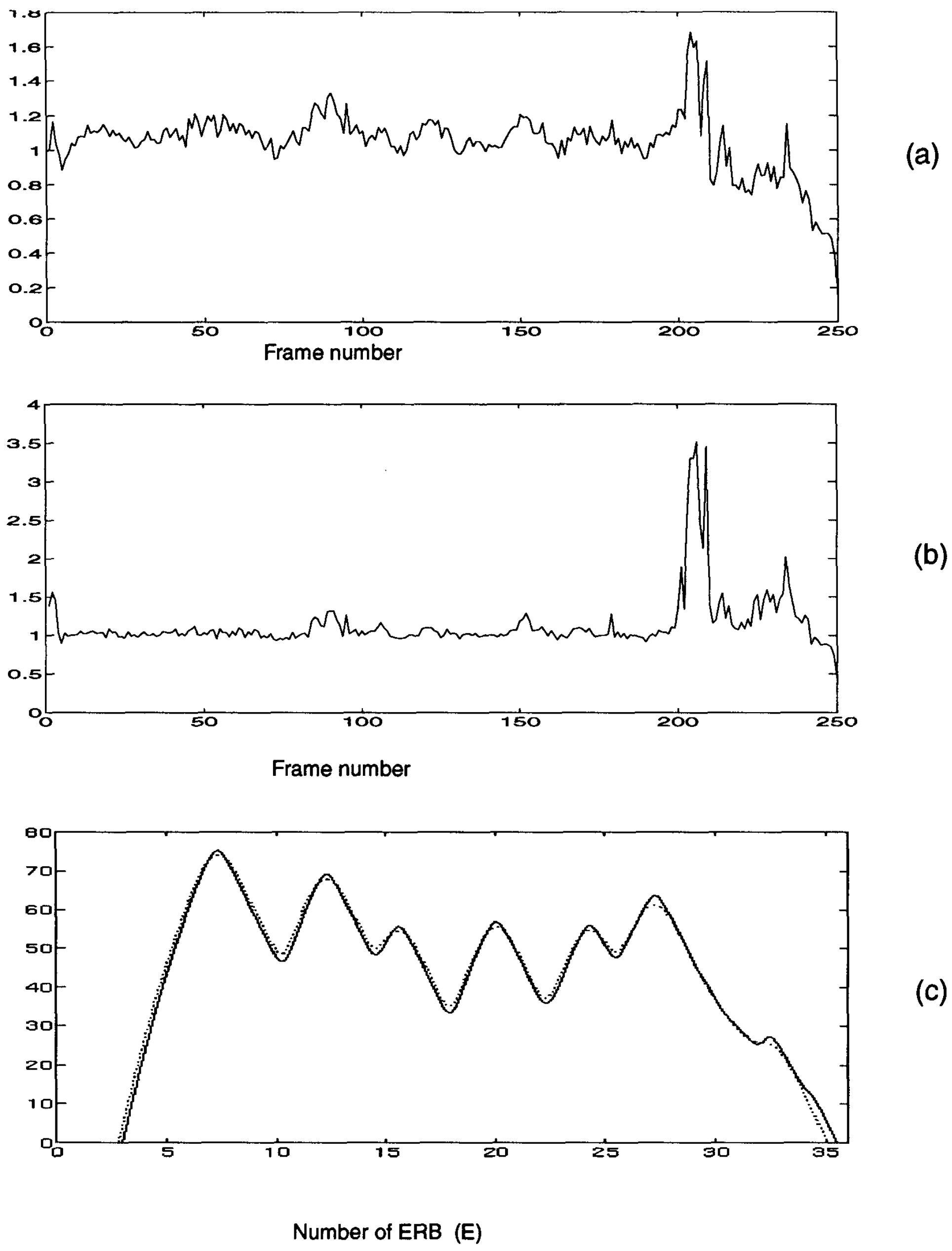


Figure 3.1 An approximation of the familiar sound excitation pattern:
(a) mean absolute differences, and
(b) standard deviations of absolute differences between the excitation patterns in each spectral frame versus frame number
(c) an example of the approximation in a single frame (frame number 120)

Non-familiar sounds were combined with other non-familiar sounds in pairs in a similar fashion. Each combination of a familiar and an unfamiliar sound, including equally distant sounds, was also put together. The order of the presentation was randomized, so that pairs of sounds from different groups were mixed together.

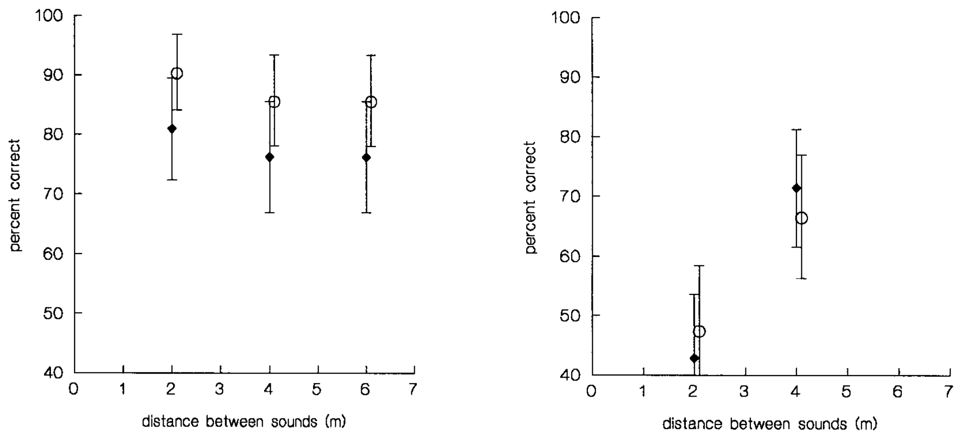
The test consisted of two parts. During the first part, the subjects estimated the relative distances of the paired sounds, and attempted to match their loudness. First, they had to decide which sound in the pair was closer to them–the first or the second. They were allowed to play the sounds back as many times as they wished. After they accepted the decision, the subjects were asked to match the loudness of the sounds, which were played again in alternation, in a loop, as long as desired. They could adjust the volume of the sound they decided was closer by clicking on either 'softer' or 'louder' buttons on the computer screen. With a single click on the button, the subjects could change the power of the sound by +0.5 dB, or -0.5 dB, and then hear the difference. After they decided the sounds were equally loud, the subjects could finalize this decision by pressing the 'ok' button. The gain adjustment (attenuation factor) set during the equalization was stored in a disk file, and the next pair of sounds was presented for distance estimation. The distance and loudness questions were interleaved.

The gain adjustment factors were used during the second part of the test to reproduce the equally loud sounds resulting from adjustments the subjects had made in the first part. The equalized sounds were randomized, and presented in pairs. During the second part, subjects had to decide again, which of the sounds in a pair was closer.

3.3 Results

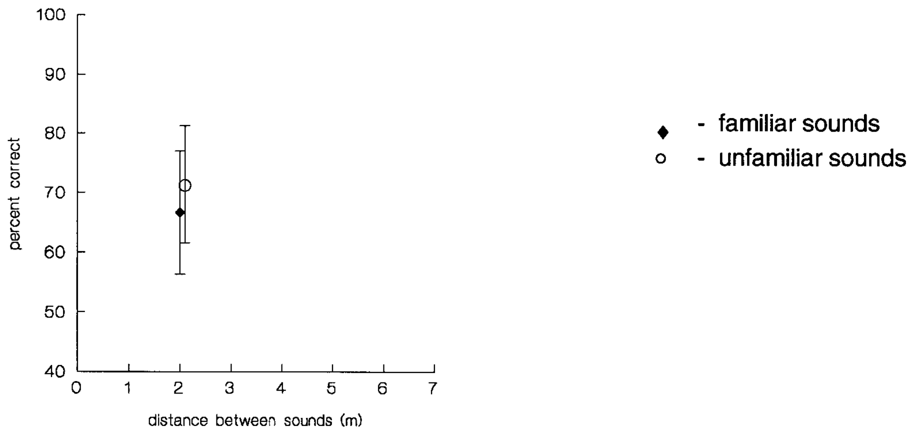
The non-familiarity of the sound should have obstructed distance judgements if it was perceptually relevant. In particular we would expect a notable performance decrease. To explore whether there was such a decrease, judgements of distances between the familiar and unfamiliar sounds were plotted for the six test conditions. Points in figures 3.2a-c represent the subjects' combined performance as a function of distance between the sounds at the position of the closer sound equal to 2, 4, and 6 meters from the listener respectively. The subjects had equalized the stimuli for loudness before the distance judgement was made. Mean performance is shown along with standard error bars based on 21 observations. Black dots mark distance judgements between familiar sounds whereas white dots denote distance estimates between unfamiliar sounds. As can be seen from the plotted distance estimates, sounds of the two categories lie very close to one another, within about a 10% range. Unexpectedly, in most of the cases the performance with the non-familiar sounds is slightly better than the performance with the familiar sounds.

Estimates of distance between familiar and unfamiliar sounds lie very close to one another in figures 3.2 a-c, implying that the familiarity factor is perceived similarly in all six test conditions. To check it formally we collected the observed frequencies of correct distance estimates in a contingency table (table 3.1). Two rows of this table expressed the classification according to the familiarity of sound, whereas the columns showed the results at the 6 combinations of distances from the microphone. The results of a chi^2 test of independence (χ^2=0.198, df=5, p<0.999) showed that we could not reject the hypothesis that the perception of familiarity was independent of our test conditions. The extremely large p-value for this test is remarkable, and confirms our supposition that the performances at the six test conditions are practically independent of the sound category. Although the proportions of the correct distance judgements between the familiar sounds, and the proportions of the right distance estimates between the unfamiliar sounds are pretty much the same for all the test conditions, we should stress that the results show a better performance on the unfamiliar sounds.



(a) closer sound at 2 m from the microphone

(b) closer sound at 4 m from the microphone



(c) closer sound at 6 m from the microphone

Figure 3.2 (a)-(c) Comparison of distance performance in the two categories of sounds (familiar and unfamiliar) as a function of the physical distance between the sounds, at the position of the closer sound equal to 2, 4, and 6 m from the microphone. Standard errors are based on 21 observations. Loudness was equalized before the distance judgment. The positions of the unfamiliar sounds are shifted to the right for clarity.

distance between sounds (m)	2	2	2	4	4	6
distance from microphone (m)	2	4	6	2	4	6
familiar sounds	17	9	14	16	15	16
unfamiliar sounds	19	10	15	18	14	18

Table 3.1 The distance scores for familiar and unfamiliar sounds across the 6 distance combinations

Knowing that the experimental factor, the familiarity of sound, was perceived similarly regardless the distance between the sounds and distance from the microphone, we wanted to investigate whether the introdced non-familiarity significantly decreased (or possibly increased) the performance. For this reason we placed the frequencies of correct and incorrect distance estimates for familiar and unfamiliar sounds at the six test conditions in tables 3.2a-f. The above question can be answered by testing if the difference in proportions of correct distance estimates between the familiar sounds and the unfamiliar sounds equals to zero. This is equivalent to testing if the proportion of incorrect distance estimates between the familiar sounds is equal to the proportion of the incorrect estimates between the unfamiliar sounds (Dixon and Massey, 1983). We will call the proportion of incorrect distance estimates of unfamiliar sounds to the total number of observations the *performance loss*, and the proportion of the incorrect distance estimates of familiar sounds to the total number of observations the *incorrect performance*. Differently stated, we wanted to check if the performance loss caused by non-familiarity could be equal to the incorrect performance, which normally occurs when the sounds are familiar. Our sample was not large enough to successfully calculate the McNemar symmetry chi^2 test, which is usually run to answer such a question. Instead, we tested the proportions directly from the binomial distributions. We ran one sided tests (one for each of the test conditions), with the hypothesis to be checked that the performance loss equals the incorrect performance.

Familiar sounds	Unfamiliar sounds OK	Incorrect
OK	16	1
Incorrect	3	1

(a) closer sound at 2 meters, farther sound at 4 meters.

Familiar sounds	Unfamiliar sounds OK	Incorrect
OK	4	5
Incorrect	6	6

(b) closer sound at 2 meters, farther sound at 6 meters.

Familiar sounds	Unfamiliar sounds OK	Incorrect
OK	12	2
Incorrect	3	4

(c) closer sound at 2 meters, farther sound at 8 meters.

Familiar sounds	Unfamiliar sounds OK	Incorrect
OK	14	2
Incorrect	4	1

(d) closer sound at 4 meters, farther sound at 6 meters.

Familiar sounds	Unfamiliar sounds OK	Incorrect
OK	10	5
Incorrect	4	2

(e) closer sound at 2 meters, farther sound at 8 meters.

Familiar sounds	Unfamiliar sounds OK	Incorrect
OK	15	1
Incorrect	3	2

(f) closer sound at 6 meters, farther sound at 8 meters.

Table 3.2 (a)-(f). Frequencies of correct and incorrect distance judgments of the familiar versus unfamiliar sounds. The results of the test were matched for familiarity. The table contains the number of correct and incorrect distance estimates under the experimental condition (unfamiliar sounds), and control condition (familiar sounds) for the six distance combinations.

The results are assembled in table 3.3. The significance level of the test varies because the binomial distributions used are discrete, but it is always smaller or equal to 0.052. The p-values, and the beta values are also included. All the tests show that differences are non-significant, therefore we cannot reject the hypothesis that the performance loss because of non-familiarity equals the wrong performance. Large beta values in cases b, c, and e confirm the strong evidence about the equality of the incorrect performance proportions of the familiar and unfamiliar sounds. Therefore, the proportion of correct distance estimates under the test condition (non-familiarity of the sounds) is also approximately equal to the proportion of correct distance estimates under the control condition (distance estimates of the familiar sounds). Relatively smaller beta values for the remaining cases a, d, and f do not allow us to derive this conclusion with such a high certainty. However, a large p-value in case d, and a low significance level α=0.019 in the other instances, also show that the differences in proportions in these cases are not significant, and suggest that they are rather small. Concluding, we cannot assume the familiarity of sound as an important factor in distance perception. The results, along with the independence test, show the opposite, i.e. that the effect of the non-familiarity of sound is negligible for distance perception.

	Distance from the microphone		Proportion of incorrect estimates for		Significance level	p-value	β-value
	closer	farther	familiar	unfamiliar			
a	2	4	0.048	0.143	0.019	0.085	0.662
b		6	0.238	0.286	0.044	0.391	0.875
c		8	0.095	0.143	0.052	0.352	0.839
d	4	6	0.095	0.191	0.052	0.152	0.632
e		8	0.238	0.191	0.033	0.368	0.802
f	6	8	0.048	0.143	0.019	0.085	0.662

Table 3.3. The significance of the proportions of incorrect distance estimates for familiar and unfamiliar sounds. The results of the symmetry tests for tables 3.2a-f. Distance expressed in meters.

Pairs of equally distant sounds belonging to different categories (i.e. one of the sounds familiar and the other unfamiliar) were also included in the test. In such cases the answers were scored as 1 if the familiar sound was judged to be the closer one. Figure 3.3 shows the mean distance estimates between such

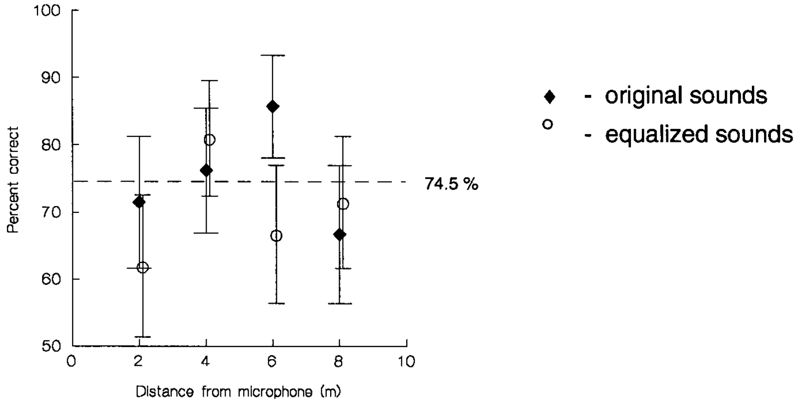


Figure 3.3 Distance preference of equally distant sounds when one of the sounds was familiar and the other unfamiliar. A result was scored as a 'success', when the familiar sound was judged as closer. All the points above the horizontal line are significantly different from chance performance.

sounds versus the distance from the microphone on original sounds (black dots), and after equalization (white dots). As there is no difference in distance between the sounds, we have only four points on the plot for each category. 97% upper confidence interval of a binomial distribution based on 0.5, and 21 observations equals to 0.762, and defines the chance level for the experiment. The corresponding level in percent correct is marked by a horizontal line in figure 3.3, and separates performances which could be obtained by chance (points below the line). The results of the formal test are presented in table 3.4. Two scores on original sounds, and one on equalized sounds lie above the chance level. Noticeably, all the means are higher than 50%, which would not happen if the answers were given by chance. Apparently, when forced to estimate which of the equally distant sounds were closer, subjects chose the familiar sound over the unfamiliar one.

Distance from microphone (m)	Type of sounds Original % correct	p-value	signif.	Equalized % correct	p-value	signif.
2	71.4	0.039	no	61.9	0.192	no
4	76.2	0.013	yes	81.0	0.004	yes
6	90.5	<0.001	yes	66.7	0.095	no
8	66.7	0.095	no	71.4	0.039	no

Table 3.4 The distance preference on equally distant sounds. 'yes' marks significant cases in which familiar sounds were judged as closer. Significance level $\alpha=0.026$.

3.4 Discussion

Based on the results of the first experiment, the step of 2 m between the positions of the speaker in the room was purposefully chosen to be large enough to give a good impression of the change in distance when the sounds were equalized. In such conditions familiarity should work as an experimental (discriminant) factor for distance perception.

There is more belief and speculation about the role of sound familiarity in distance perception than experiments done concerning the topic. Coleman's work (1962) raises the issue of accurate distance judgement (in a free field) upon initial exposure to unfamiliar sounds (wide-band noise bursts). According to his conclusions, subjects learn during the test, obtaining the worst results on the first trials. No test was performed to compare this result with possible subjects' responses to the first exposures to a familiar stimulus. In the second (and the last) known reference (McGregor et al., 1985), responses of the subjects to a familiar speech stimulus played from two distances were compared to the responses to an unfamiliar stimulus (the same speech, played backwards) in an open space. The conclusions of the experiment, that familiar sound distance was perceived more accurately can be questioned, however, for many reasons. First of all, in the acoustical sense McGregor's stimuli were not unfamiliar. Speech played backwards is familiar, but simply does not have meaning. Secondly, his analysis is questionable. If we assign all his subjects who judged the sounds as equally distant to the 'incorrect' category (which would be more appropriate because the sounds were actually not equally distant) and we do the same with the 'undecided' people (i.e. those who were unable to do the test right), the conclusions of his experiment no longer hold.

It seems unlikely from the analysis of our data that the distance of unfamiliar sounds is perceived differently from the distance of familiar sounds. The results of the symmetry test show that introducing the experimental effect (familiarity) is not important. The 'no interaction' relationship between the scores on the familiar and unfamiliar sounds might actually suggest a week effect of a better performance on the unfamiliar sounds than the familiar ones, because the means consistently lie above that of the familiar sounds! The extremely high p-values of the independence test also show that subjects, at least musically trained subjects, can judge the distance of unfamiliar sounds as accurately as the distance of familiar sounds. This conclusion may not be valid for the initial exposure to the unfamiliar sound, though. A very interesting inference can be drawn from the distance performance on the equally distant sounds. Exposed to such sounds, subjects perceived familiar sounds as closer, which would support a hypothesis about a general preference of such sounds in distance perception.

Chapter 4

Experiment 3

4.1 Rationale

As has been suggested by the study of Warren (1973), reverberation may be an important cue for the estimation of sound loudness. This fact has also been noticed in the practice of computer music. It was recently formulated by Chowning (1990), who observed that loudness constancy might take place in a room environment in an analogous way to size constancy in vision. In visual perspective, to preserve the impression of an object's constant size, its physical size has to be diminished in fact, in proportion to the provided perspective. Is this also the case in auditory perspective?

Imagine two sounds in a room, varying as to the playing effort induced by the player and the amount of reverberation, proportional to the distance from the listener. Such reverberation invokes the sensation of distance and auditory perspective. In these conditions, in a loudness judgement of such sounds, if 'size constancy' appears in the auditory world, the loudness of the sound source will be perceived rather than the loudness of the sound wave in the listener's ears. According to the hypothesis, for the two sounds of equal physical intensity in the listener's ears (and similar timbre), the sound played with a greater effort and carrying a higher amount of reverberation would be perceived as louder. This is graphically expressed in figure 4.1 (after Chowning, 1990). The closest singer in this picture sings softly, but the intensity of the sound in the listener's ears is equal to the intensity of the distant singer. The remote sound is also more reverberated allowing the listener to recognize the distance even if she actually does not see the singer. The listener may therefore believe her singing to be louder than that of the closest performer, because the singer must have compensated for the longer distance when she uttered the sound. In a suggested explanation of this phenomenon we assume that for the listener, reverberation cues provide the information that the sound was played from a greater distance, hence it must have been louder at the source.

The primary goal of this experiment was to asses the validity of the above hypothesis on musical material, and by using sophisticated listeners–musicians–as subjects.

The problem can be split conceptually into three parts:

1. Does the presence of reverberation influence loudness judgment?
2. Do factors other than physical intensity influence subjects' judgments of loudness?
3. If this happens, are subjects affected by the request to estimate the loudness of the sound source rather than just the apparent loudness?

An attempt was made in the third experiment to simulate the auditory perspective of a room, and a player in the room in four distant positions playing with a given amount of effort. Listeners were then asked to adjust the loudness of a sound produced by a close player, to match the loudness of each of the remote ones. Four playing efforts, and sixteen different player positions were employed in the test. The task was split into five sub-experiments. In the first one, the subjects produced four equally loud reference sounds, one for each effort, modelling the close positions of the players. In the second one, they were asked to imagine a room, and the four distant players at the different distances in the room. We wanted to check if loudness estimates of sound sources depended on the auditory perception provided by the distance, hence on the varying reverberation. For this reason the listeners were asked to regulate the loudness of the close player so as to match the loudness of the remote players. In the third sub-experiment we wanted to examine whether the listeners were able to distinguish between the loudness of the sound source and just the apparent loudness. In this part of the

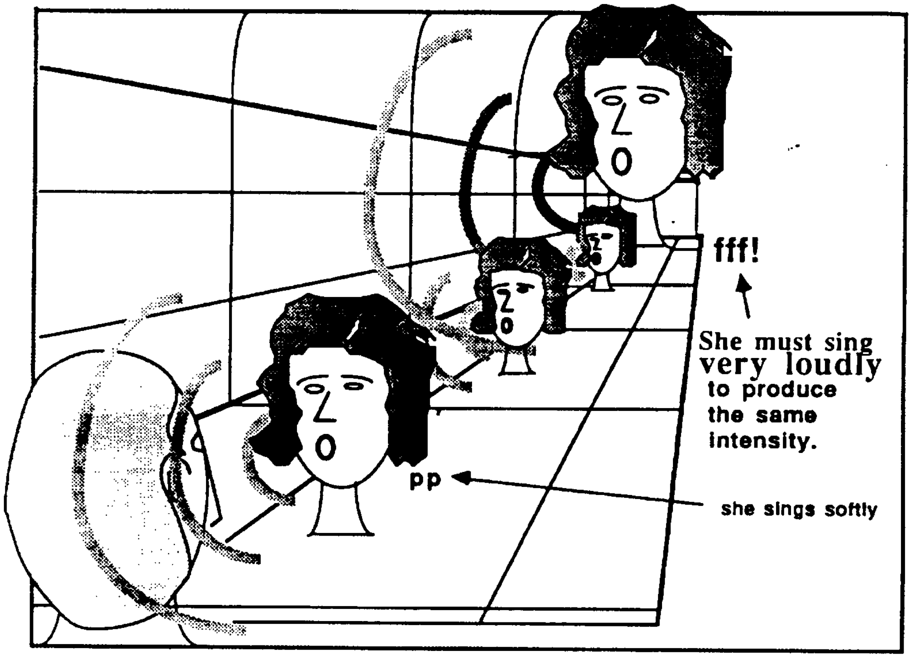

fff!
She must sing
very loudly
to produce
the same
intensity.
pp
she sings softly

Figure 4.1. Demonstration of the 'loudness constancy' hypothesis, and auditory perspective depicted as an analogy to the size constancy and visual perspective

" ... the distant singer must sing very much louder to produce equivalent intensity as the nearest singer, so must she also become bigger in order to produce the same size image at the retina of listener."

" ... it is the constant intensity of the reverberant energy which provides this effect of loudness constancy when there are no spectral cues. A similiar phenomenon occurs in the visual system.": Chowning (1990).

test, the same sounds as in the second sub-experiment were reproduced, but subjects were instructed not to think about any room, but to match the apparent loudness as it appeared in their ears. The fourth part was meant as a control test. Dry prototypes of the 'players' from the other sub-experiments were replayed at the same intensity levels as the reverberated sounds from the other parts, but without any reference to a room. In this part the subjects' task was to provide gain adjustments reflecting the relative loudness of the sixteen sounds, which could then be regarded as the reference loudness levels. Any deviations from these references in the other parts of the test could be attributed to reverberation, and an explicit instruction. Finally, in the fifth sub-experiment, sixteen reverberated sounds were produced in a way that violated acoustical principles, and therefore they could not be regarded as performed in one room. We wanted to check whether the impression of sound sources (players) in the second sub-experiment would differ from the perception of sound sources in such 'abnormal' conditions.

4.2 Method

4.2.1 Subjects

Twenty six subjects participated in the third experiment. They were all musicians with extensive musical training and experience, most of them being either active musicians or researchers at Stanford's Center for Computer Research in Music and Acoustics (CCRMA). All subjects reported normal hearing.

4.2.2 Stimuli and Apparatus

Four short, percussive-like sounds were generated as the output of the computer physical model of the wave digital hammer striking a membrane with different forces (Van Duyne et al., 1994). Variations in force produced an increase in power by 3 dB, spectral changes (an increase of spectral bandwidth shown in figure 4.2), and sharpening of the attack, among other things. In this way the playing effort was simulated, so that the sounds corresponded to four different playing efforts. The length of the produced sounds, as measured from the short-time spectrograms at the level of -60 dB dynamics range, was equal to about 150 ms (figure 4.2).

Next we wanted to model the four different player positions in a room for each of the four playing efforts. Four sounds were created from each of the four dry prototypes in such a way that, for a constant effort, the powers were decreased by approximately 6 dB. In anechoic conditions such sounds would correspond to four physical distances from the listener, with twofold distance ratios between successive sounds. Reverberation was then added to the sounds by using the Yamaha SPX1000 digital room model to simulate the sixteen different positions of the player. The direct-to-reverberant sound energy ratios for the sounds at different distances were calculated accordingly by using an inverse square law formula (Beranek,1954):

$$|p_r|^2 = W \rho_0 c \left(\frac{1}{4 \pi r^2} + \frac{4}{R'} \right)$$

where:

$|p_r|^2$ – mean-square pressure at any point a distance r away from the source,

W – power emitted by the source,

$\rho_0 c$ – characteristic impedance of air,

R' – room constant defining the 'liveliness' of the room.

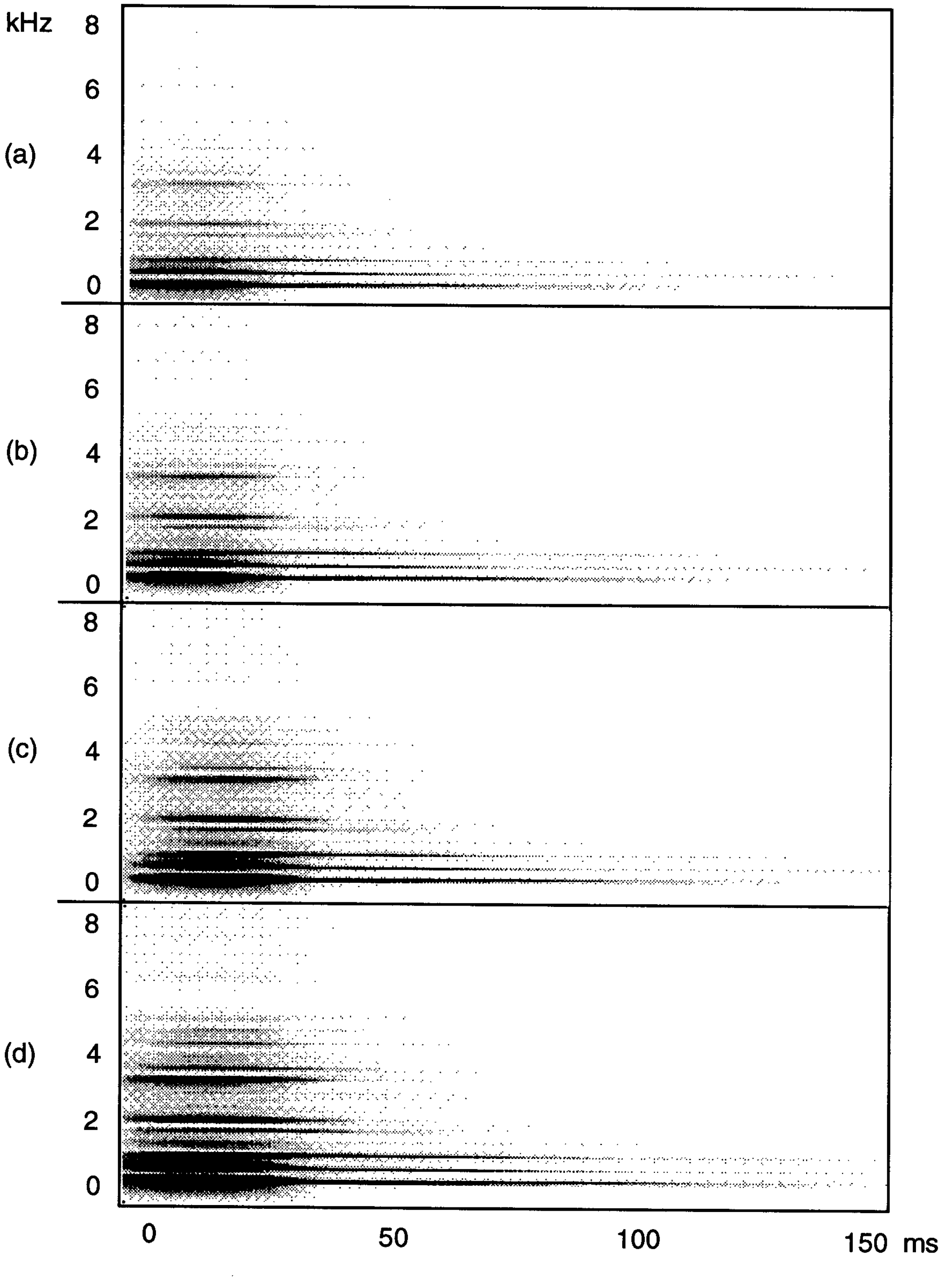


Figure 4.2. Spectrograms of the four dry percussion sounds. (a)-(d) show the increasing effort. First 150 ms is shown with the dynamics of 60 dB. Frequency resulution is 21.5 Hz, and time resolution is 1 ms.

Since R' is constant for a given room it is useful here to express it as a percentage. If we let $^{4}/_{R'}$ be equal to p %, the energy reflected by the room is $p * W\rho_0 c / 100$, and the direct-to-reverberant energy ratio becomes:

$$D / R = \frac{100}{4 \pi r^2 p}$$

where r is the distance from the sound source.

In a room, sound sources of different power will raise the same percentage of reverberant energy, but the reverberant energy power at a fixed distance from the source will be greater for the more energetic sounds. By similar reasoning, if we observe two sounds of the same power, but with different direct-to-reverberant energy ratios in a given place in the room, we can conclude that they originated at two different distances. We should remember that two different playing efforts can be simulated in the model by varying playing force, hence a greater playing effort will yield a more energetic sound source. For this reason, a sound of a given effort will produce unique D / R ratios along the distance axis, which are different from the ratios produced by a sound of another effort.

In our experiment, we wanted to create stimuli at the four fixed power levels (with the ratio of two), out of each of the four playing efforts. In this way all the sixteen sounds could be thought of as being played at sixteen different distances from the listener in the room. This is graphically shown in figure 4.3. The direct-to-reverberant sound energy ratios corresponding to the distances were then calculated according to the above formulas. They are shown in table 4.1. For the calculation, it was assumed that the initial power of the sound sources was equal to the powers of the four dry prototypes of different efforts, hence it was changed with the effort by 3 dB, as shown in the table. As a consequence, for a given power, the sound source distances increased with

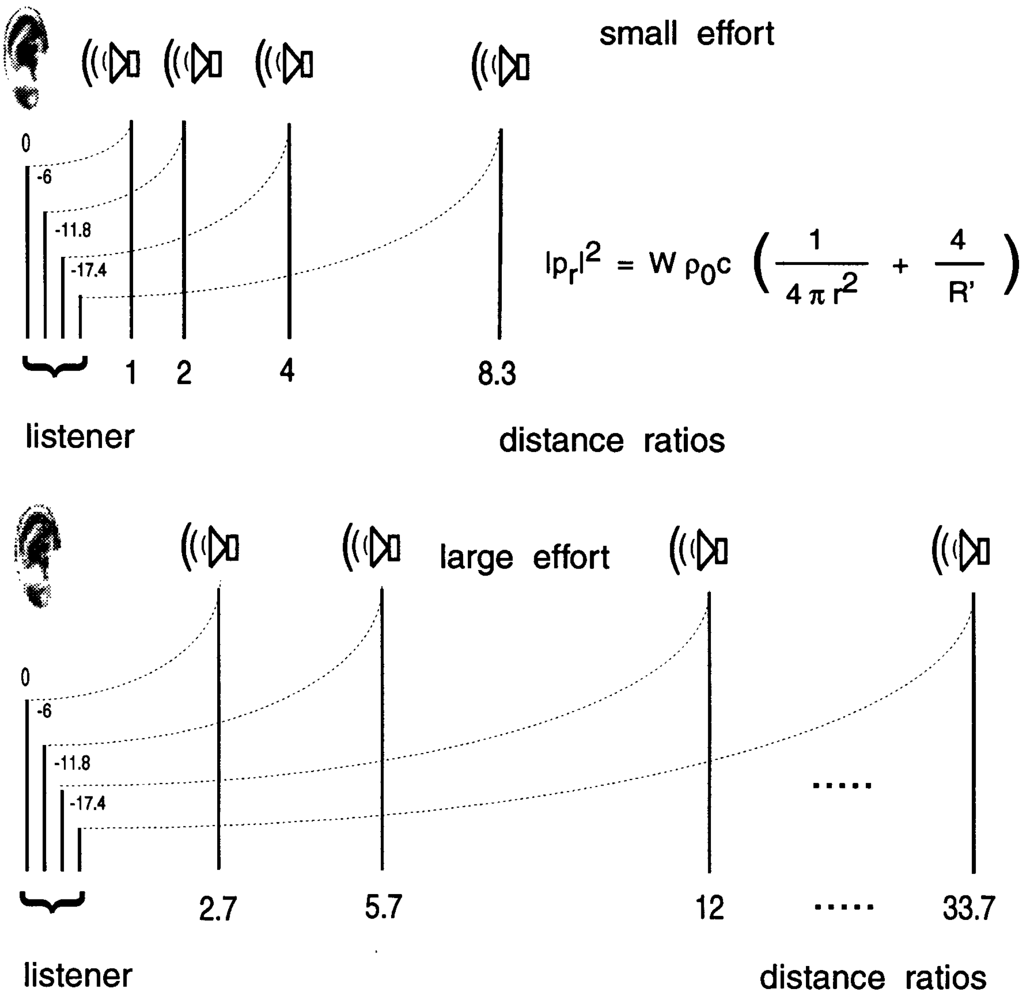

small effort
0
-6
-11.8
-17.4
1
2
4
8.3
listener
distance ratios
$|p_r|^2 = W \rho_0 c \left(\frac{1}{4\pi r^2} + \frac{4}{R'} \right)$
large effort
0
-6
-11.8
-17.4
2.7
5.7
12
.....
33.7
listener
distance ratios

Figure 4.3. A player (sound source) performing with a small effort from different distances will produce four different sound pressures at the listener's ear, with relative ratios: 0, -6, -11.8, and -17.4 dB (top graph). The same player performing with a large effort in the same room has to be further removed to cause the same stimulation at the listener's position (bottom graph). Therefore, if the relative distance ratios of the soft performer are 1:2:4:8.3, then the corresponding relative distance ratios of the loud performer (9 dB louder) will be equal to: 2.7:5.7:12:33.7.
The ratios were calculated by using the included formula.
See also table 4.1.

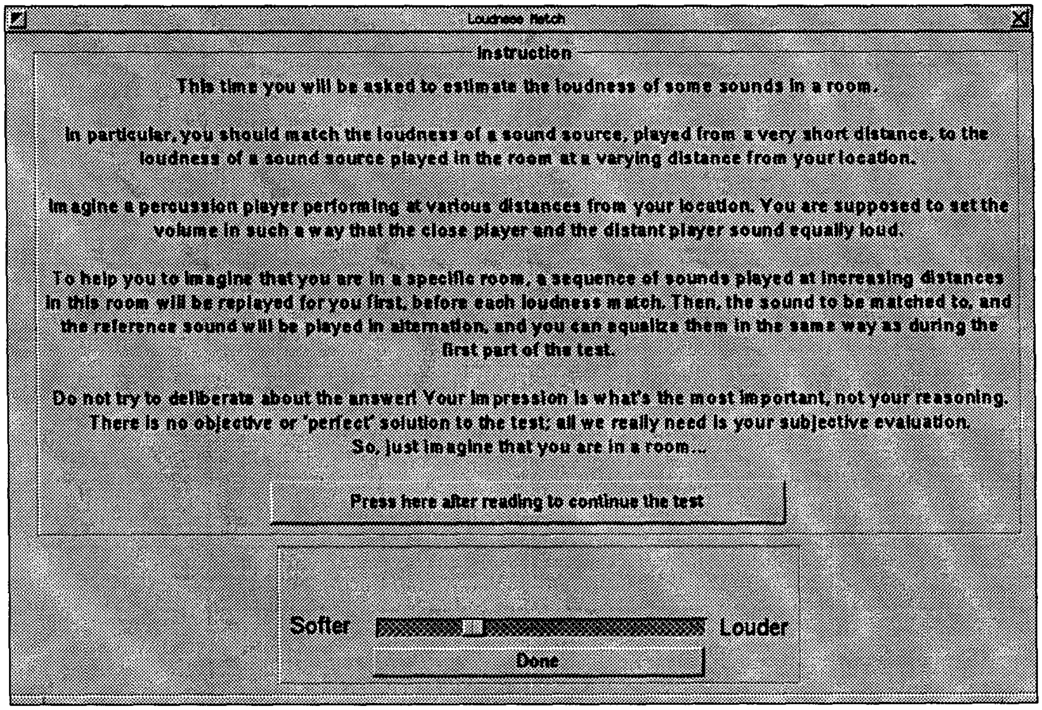
increased effort. Similarly, for a sound of a constant effort, the sound source distances increased together with the power decrease. Therefore, the most powerful of the reverberated sounds was also the closest one (that is, its direct-to-reverberant energy ratio was the largest one). Since the closest distance was achieved for the dry sound of the least effort, it was convenient to regard all the four power levels of the reverberated sounds in relation to the power of this sound. Consequently, the sounds of the 'zero-th' power level were set equal to the power of the least effort dry sound.

The 100% reverb signal was taken from the digital output of the reverberation device and mixed with the digital dry prototypes according to the appropriate ratios. The distance ratios between the sound source and the listener's position for all the sounds depended on both effort and power level, as can be seen in table 4.1. All reverberated sounds in a column had the same power level, marked on top. Distance ratios between the power levels differed by a factor of two (except for the two most distant sounds). It can be also noted that, as a result of mixing, cancelation of some frequencies must have occurred. For this reason, the relative power differences (expressed in dB) deviated from initial multiples of -6 dB. Mixing also increased the length of the stimuli.

relative power level in dB		0	-6	-11.8	-17.4	-17.4
	effort					
0	0	1.0	2.0	4.0	8.3	8.3
+3	1	1.3	2.6	5.7	12.0	9.2
+6	2	2.0	4.0	8.3	18.7	9.4
+9	3	2.7	5.7	12.0	33.7	12.5
dry prototypes		reverberated sounds				

Table 4.1. Distance ratios of the stimuli (simulating the players) in relation to the closest sound (effort=0, power level=0). The last column contains the ratios between the farthest and the closest sounds.

Subjects were tested individually in a small listening studio. The room was well isolated, lined with cushion, so that the reverberation time was very short. Stimuli were played through a pair of loudspeakers, the Meyer Sound 833 Studio Reference, in stereo presentation. The loudest sound was played at the level of 86 dBC. Sound was transmitted through the Motorola 56000 DSP chip, and stimuli presentation was controlled by a custom NeXT computer program. The program allowed the subjects real-time continuous gain control over the sounds. Figure 4.4 shows the user interface of this program.



This time you will be asked to estimate the loudness of some sounds in a room.

In particular, you should match the loudness of a sound source, played from a very short distance, to the loudness of a sound source played in the room at a varying distance from your location.

Imagine a percussion player performing at various distances from your location. You are supposed to set the volume in such a way that the close player and the distant player sound equally loud.

To help you to imagine that you are in a specific room, a sequence of sounds played at increasing distances in this room will be replayed for you first, before each loudness match. Then, the sound to be matched to, and the reference sound will be played in alternation, and you can equalize them in the same way as during the first part of the test.

Do not try to deliberate about the answer! Your impression is what's the most important, not your reasoning. There is no objective or 'perfect' solution to the test; all we really need is your subjective evaluation. So, just imagine that you are in a room...

Figure 4.4. The user interface for the loudness match

The sixteen stimuli were presented in pairs with four corresponding reference sounds. The meaning of the reference sounds will be described later. Subjects were supposed to equalize the loudness of the sounds in each pair. In the pair, the sounds were played in alternation in a loop. A one second delay introduced between each repetition of the pair allowed the subjects easy recognition of the sound order. The volume of the reference sound in each pair (always the second sound) could be continuously varied by moving a slider icon knob on the computer screen with the mouse. Generally, subjects were requested to match the loudness of the first sound by changing the loudness of the second sound (more details in 4.2.3). By pressing a button on the computer screen they could confirm a decision about the loudness match in the pair, and trigger the presentation of the next one.

4.2.3 Procedure

The test consisted of five parts, which differed from each other by the stimuli type, the instruction to the subjects, and the order of presentation. Each part of the test was repeated twice (except the fifth one), to provide enough data for checking the reliability of the subjects. Half of the subjects took the third part first, and the remaining half took the second part as the first one. The ordering of the other parts of the test followed their numbers (i.e. the first was first, and so on). A short practice session was administered before the test to allow the subjects to familiarize themselves with the sounds and the procedure.

4.2.3.1 Simulation of the 'Close' Players

In the first part of the experiment we wanted to create four equally loud dry reference sounds out of the four dry prototypes, corresponding to the four different playing efforts. The reason for using four references instead of a single one is the following. The variation of the effort produced not only differences in intensity, but also changes in spectrum and timing. Therefore, reported loudness differences between sounds of different efforts could be affected by the timbral differences as well as the reverberation. We wanted to minimize this effect and decided that loudness equalization between reverberated sounds and the dry prototypes should be done only between the sounds of a common effort. We also wanted the four equalized references to be as loud as the least powerful sound.

To obtain the four reference sounds, three pairs of sounds were presented to the subjects sixteen times (differently randomized) for the loudness match. In the three pairs, the first sound was always the non-adjustable dry sound of the least power (and effort) coupled with one of the remaining three dry prototypes in the second position. Subjects were instructed to adjust the volume of the second sound to make the sounds equally loud. Corresponding changes to the sounds can be described as follows.

Initially, a constant gain g_0 was applied to the amplitude of all of the sounds, therefore the power of the output reflected their power ratios (0 dB, +3 dB, +6 dB, +9 dB) resulting from the varied attack force (effort). An attenuation of the second sound in the equalization process was achieved by a subject by decreasing the gain factor of this sound, while the gain of the other sound was kept unchanged. Therefore, the change required to achieve equal loudness could be estimated as:

$$\Delta L_{k\,0} \sim 20 * \log_{10} (g_{dry\ k} / g_{k_0})$$

where:

$\Delta L_{k\,0}$ – decibel difference to achieve equal loudness between sounds k and 0
$g_{dry\ k}$ – final gain (amplitude multiplier) of the k-th sound being equalized,
g_{k_0} – initial gain (amplitude multiplier) of the k-th sound being equalized,
k – enumerates effort, k=1,2,3.

If the decibel differences required for a loudness match closely followed the power differences between the sounds, values of ΔL were expected to be located around: -3 dB, -6 dB, and -9 dB.

Three mean gain adjustments were then calculated, based on the sixteen matches for each of the three dry sounds. They were stored in a disk file, and were subsequently used to re-create the four equally loud reference sounds during the remaining parts of the test, individually for each subject.

4.2.3.2 Loudness of Sound Sources in Room Conditions

During the second subexperiment, the subjects were expected to equalize the loudness of sound *sources* in a room. The sixteen reverberated sounds, reflecting the various relative positions between the player and the listener in a room, were reproduced with each one paired with the corresponding dry reference sounds equalized during the first part of the test. In each pair both component sounds had a common dry prototype, therefore they could be thought of as being played with the same effort at two different distances from the listener. The first sound always simulated a player further from the listener, and the second, the adjustable dry reference, a very close one. The reverberated sounds were played with the constant gain g_0, therefore for each sound at a given effort, four sounds at the selected power levels: 0 dB, -6 dB, -11.8 dB, and -17.4 dB in comparison to the least effort dry reference (see table 4.1) were reproduced at the loudspeakers. Gains $[\, g_{dry} \,]_k$ (k=1,2,3) obtained during the loudness match of the first part of the test were initially applied to the dry references. They were subsequently adjusted by the subject in the loudness match procedure, so that an array of gains $[\, g_{rev} \,]_{kj}$ (k,j =0,1,2,3) was produced as a result.

In this part of the test subjects were instructed to imagine a room and a player in the room, playing at two distances. To help them in this task and invoke auditory perspective, a sequence of four reverberated sounds receding in distance was presented before each pair of stimuli. In this sequence, the sounds corresponded to the four different positions of the player available for the given effort. The subjects were then exposed to the actual pair of sounds imitating the players. They were supposed to adjust the loudness of the reference sound to the loudness of the other sound to make the two *players* sound equally loud. They were asked explicitly to match the loudness of the two sound sources. The order of presentation was completely randomized.

4.2.3.3 Apparent Loudness of Sounds in Reverberant Conditions (in the Listeners' Ears)

The same sounds as in 4.2.3.2 were presented in the third subexperiment. For each pair of sounds, the subjects were asked to match the loudness of the two sounds, one reverberated and the other dry, just as it appeared *in their ears.* They were also asked to try not to think about any room or player during the procedure. No leading sequence of sounds was included in this part of the test presentation. The order of presentation was random, and different from the order in the second subexperiment.

4.2.3.4 Loudness of Sounds in Control (Dry) Conditions

In the fourth subexperiment, the subjects again matched the loudness of the dry sounds. An array of sixteen dry sounds was made out of the four dry prototypes. For each of the four efforts four sounds were produced, with the powers matching the powers of the reverberated sounds of parts two and three. The subjects were then supposed to match the loudness of each of the four, dry, equalized prototypes to the loudness of the four corresponding dry sounds from the matrix. This part of the test was meant as a control to remove possible bias which could arise from the comparison of the sounds produced at different sound levels, and to remove a possible effect of equalizing references varying in timbre.

4.2.3.5 Loudness of Sound Sources in Abnormal Reverberant Conditions

Finally, in the fifth part of the test, sixteen reverberated sounds were produced again, in a similar way as in the part two. This time, however, constant direct-to-reverberant ratios were assumed for each power level of the presentation. For each of the efforts, the ratios corresponding to the effort number 0 (first row in table 4.1) were applied. According to the acoustical model, sounds produced by sources varying as to the effort and having a common power as well as the same direct-to-reverberant energy ratio could not appear in one room. The sounds within each of the pairs could be thought of as being produced in the same room, however. The subjects were then informed that they would be presented with two players performing at two different positions–a distant and a close one, in a variety of rooms. As in part two, they were asked to think about the loudness of the sound *sources*, and they were supposed to match the loudness of the *players*. No leading sequence was reproduced this time to help them imagine the auditory perspective.

4.3 Results

The 'ecological' part of the test ("imagine two players in a room") was the most difficult one, therefore reliability of the performance was checked on its results. Pearson's correlation coefficient of 0.7 between the two loudness matches of identical pairs was assumed to be a reasonably good cut-off level to exclude some inconsistent subjects and limit the error. Sixteen of the twenty six subjects passed this criterion, and were selected for further study. For these subjects, the correlation coefficient for repeated results of the other parts of the test was even higher. Only ten of the sixteen people were able to come again and take the fourth part of the test, however, and their results are reported.

For the analysis, we were mostly interested in the relative *differences* in loudness between the reverberated sounds and the dry references. These differences were calculated as:

$$\Delta L_{kj} = 20 \log_{10} (g_{\mathrm{rev}_{kj}} / g_{\mathrm{dry}_k})$$

where:

g_{dry_k} – initial gain of the k-th effort dry reference, (k=0,1,2,3)

$g_{\mathrm{rev}_{kj}}$ – final gain of the k-th effort dry reference, after it was set as equally loud to the reverberated sound of the k-th effort and the j-th power level, (k,j = 0,1,2,3).

If power were the exclusive basis for the loudness match of the reverberated sound rev_{kj}, we would expect ΔL_{kj} to equal the power level j of the k-th sound, because the dry references were equal in power to the least powerful reverberated sound (0 dB). In contrast, systematic significant deviations from the power levels would show a possible reverberation influence on the sound loudness judgments. Since the reverberated sounds had the same prototypes as the dry reference sounds, reverberation was their only differentiating factor. Moreover, significant differences between the results of the second and third parts of the test should reveal that such deviations are caused by a conscious imaginative process. In other words, it would show that the subjects were affected differently by the instruction to imagine players in the room, and the auditory perspective rather than by the instruction to use the intensity in the ears alone.

4.3.1 Loudness of Stimuli in Dry Conditions

As we suspected that the estimated decibel differences of the loudness match could deviate from the sound levels of the stimuli even for the dry sounds, we included the control part in the test (subexperiment 4). If any extraneous factors such as timbre, absolute power level, or other interfered in the loudness match, the deviations of the estimated decibel differences from the four predefined levels would become apparent at this stage. In the analysis, we decided to work out the differences ΔL_{kj} obtained during the loudness match in reverberant conditions in relation to the corresponding differences ΔL_{kj} in the control (dry) conditions, so as to eliminate all the irrelevant agents, and attribute the effects to reverberation alone. Therefore, the results will be further discussed in the following sections along with comparison to the pertinent results of the loudness estimation in reverberant conditions.

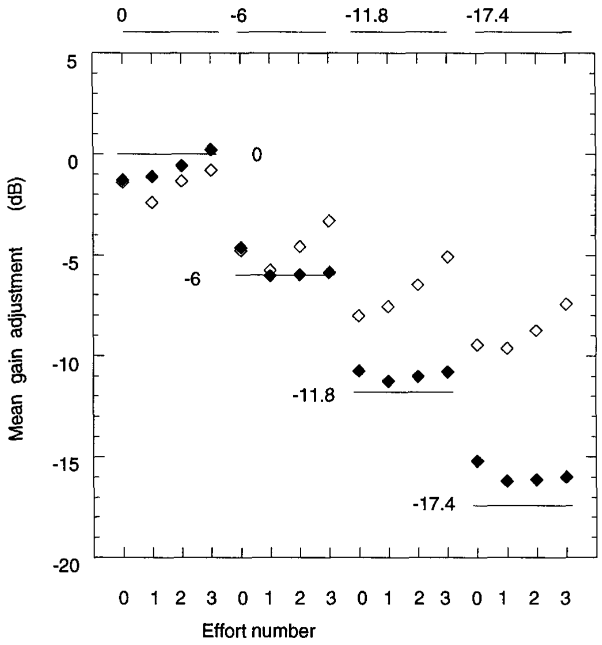
4.3.2 Loudness of Stimuli in Reverberant Conditions

The results of the "imagine two players in the room..." part of the test are shown in figure 4.5a. White diamonds denote mean gain adjustment differences ΔL in dB between the sixteen sound sources in the room (remote players), and the four loudness-matched dry references (close players), across the ten subjects. The adjustments are grouped by the four power levels, and arranged by the four efforts within each power level. Expected gain adjustments: 0 dB, -6 dB, -11.8 dB, and -17.4 dB are marked. According to the above discussion, if intensity were the only physical cue for loudness, the adjustments ΔL_{kj} would reflect relative loudness differences between the sounds presented at different values of distance and effort, and should correspond to the differences in power levels. The actual adjustments lie far above the levels of -11.8, and -17.4, and show a general decreasing tendency for increasing effort. This tendency reveals that the loudness differences expressed by the adjustments $\Delta L_{k\,j}$ were perceived as smaller for more reverberated sounds, and for larger effort sounds. Black diamonds in the figure mark mean gain adjustment differences ΔL in the control dry-to-dry loudness match (fourth part of the test). It can be seen that the adjustments follow the levels much more closely, although they are also smaller than expected for the levels of -11.8, and -17.4.

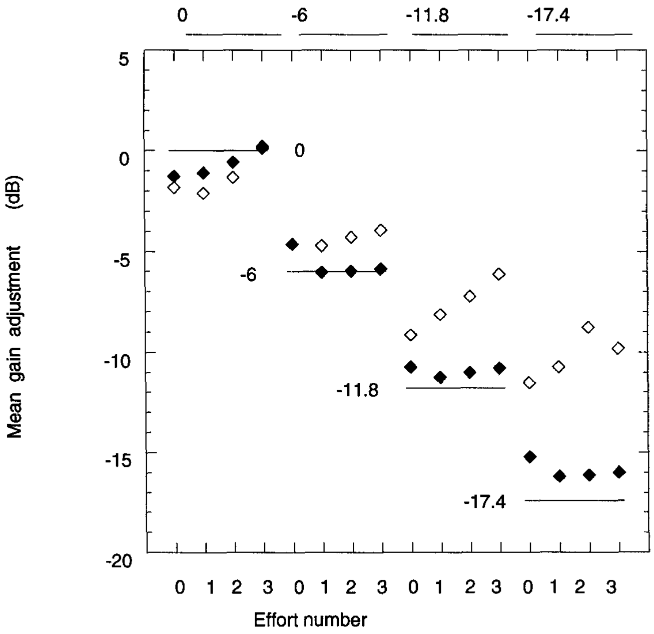
The results of the "do not think about any room..." part of the test (third subexperiment), are shown in figure 4.5b, in the same way as it was described above. The black diamonds (dry-to-dry loudness match) are the same on both plots. A similar tendency can be observed, i.e. that the loudness of reverberated sounds was perceived as greater than the loudness of the corresponding dry sounds, after a certain dry-to-reverberant energy ratio was exceeded.

The results of the loudness match in the anomalous reverberant conditions– "imagine two players in rooms..." (subexperiment 5), are shown in figure 4.5c. Constant direct-to-reverberant sound energy ratios used in this test caused an artificial situation, and according to acoustical principles the sounds could not be all located in one room. The ratios were equal across all efforts, but increased along the power levels; therefore, the effort dimension was no longer associated with the ratio. As in the other parts of the experiment, the tendency to overestimate the loudness of reverberated sounds is clearly visible. The importance of the increasing effort is apparent here as well, even though it is no longer associated with the increasing direct-to-reverberant energy ratio, because it was fixed at a given power level.

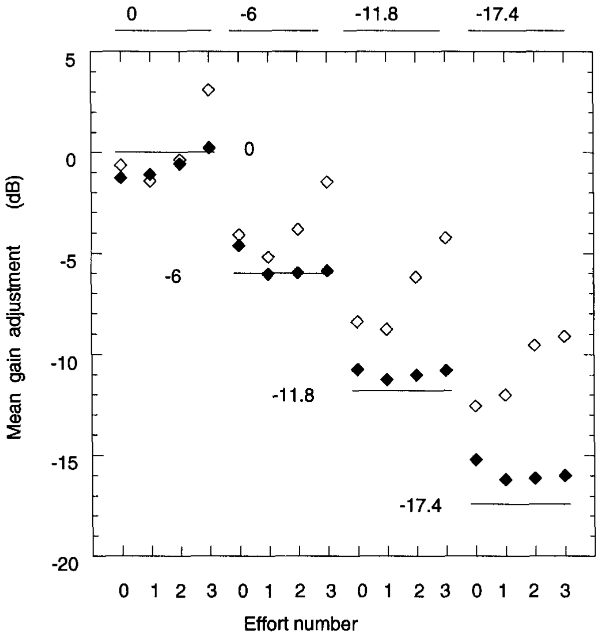
In the experiment, we were interested only in the relative loudness differences, i.e. the differences between the reverberated sounds in relation to the corresponding dry sounds. To correct for the deviations from the expected



(a) “imagine two players in a room...”



(b) “do not think about any room...”



(c) “imagine two players in rooms...”

Figure 4.5 (a)-(c) The comparison of mean gain adjustments of the reverberated sounds (white diamonds) and the dry sounds (black diamonds).

(a) the results of the test instructed as “imagine two players in a room...”

(b) the results of the test instructed as: “do not think about any room...”

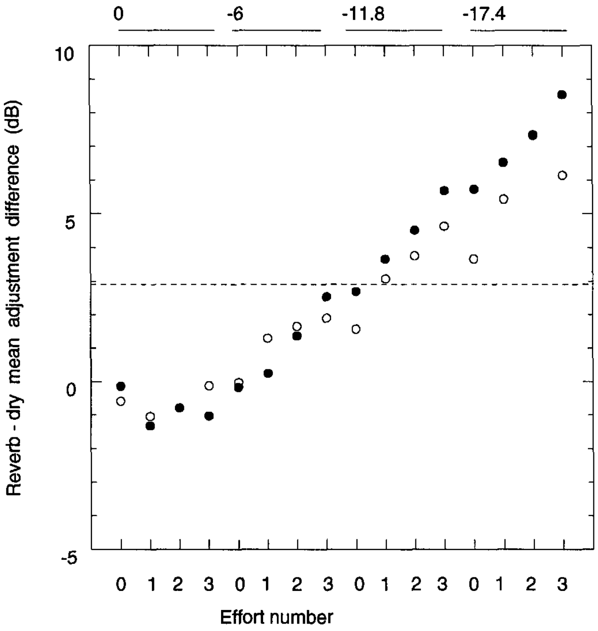
(c) the results of the test instructed as: “imagine two players in rooms...”

levels of the dry-to-dry sound loudness match, mean gain adjustments of the match were subtracted from the adjustments of the corresponding reverberated sounds. In this way the four predefined power levels were removed from the analysis, leaving only the relative deviations caused by the reverberation and instruction. The result is depicted in figure 4.6a. As in figure 4.5a, the ordinate shows the size of this difference in dB: the sixteen test conditions expressed by the power level, and effort of the sounds are marked on the abscissa. Black dots show the results of the "imagine two players in a room..." test (and correspond to those in figure 4.5a), whereas open dots show the results of the "do not think about any room..." test (and correspond to those in figure 4.5b). If reverberation did not affect loudness, the points should be distributed randomly around zero. The figure clearly shows, however, that subjects regarded the reverberated sounds as louder than the corresponding dry sounds. Loudness differences above the horizontal line are significantly different from zero with $p<0.021$ for the closest black diamonds, and $p<0.032$ for the closest white diamonds. However it should be clear, that even the points below the line were not obtained by chance, because for both tests the loudness (decibel) difference increases in a linear fashion.

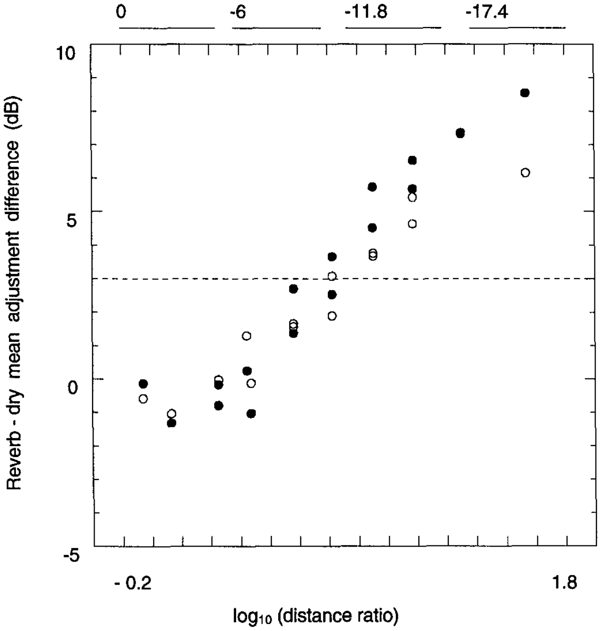
We should note, that the perceived distance ratio between the sounds in the experiment also grows monotonically along the effort within each power level, and it generally increases between the power levels as well. This suggests that the direct-to-reverberant sound energy ratio, associated with the distance ratio, is responsible for the change.

To examine this possibility, the loudness differences were plotted versus distance ratios, expressed in a logarithmic scale, in figure 4.6b. The conversion of each combination of the effort and the power level to a distance ratio was done according to the table 4.1. The relationship remains linear, with:

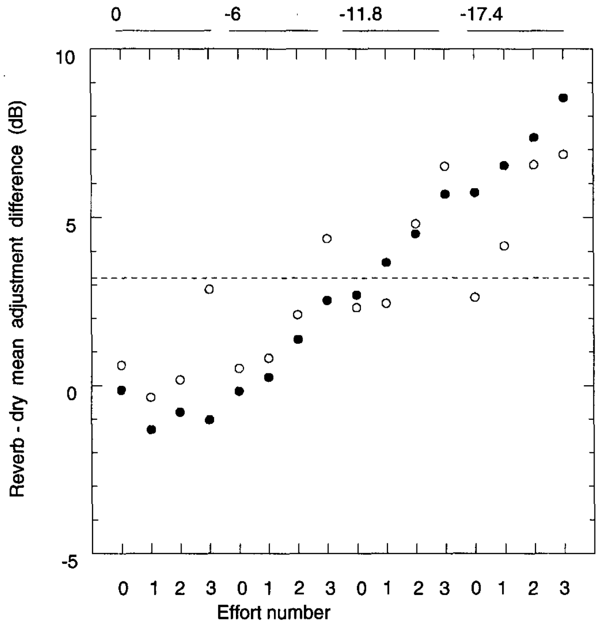
$$\Delta L_{kj} = 0.37 * \text{distance ratio} - 2.2$$



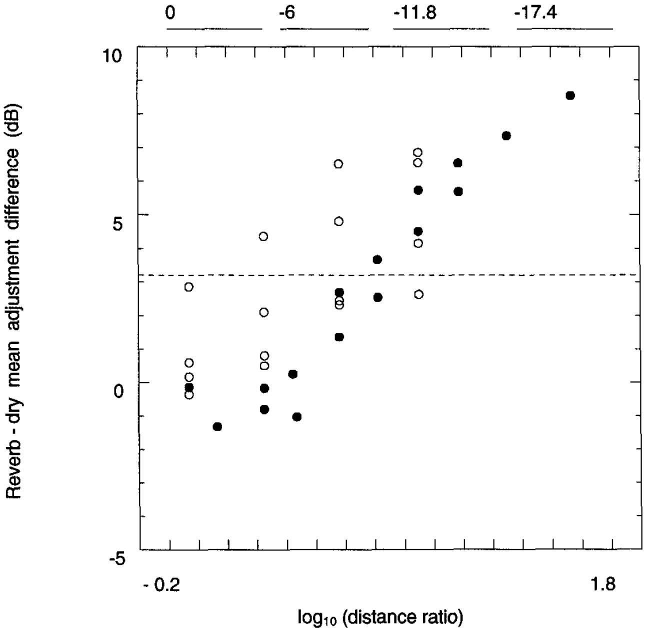
(a) "imagine two players in a room..." (black dots), and "do not think about any room..." (white dots)



(b) "imagine two players in a room..." (black dots), and "do not think about any room..." (white dots)



(c) "imagine two players in a room..." (black dots) , and "imagine two players in rooms..." (white dots)



(d) "imagine two players in a room..." (black dots), and "imagine two players in rooms..." (white dots)

Figure 4.6 (a)-(d). Mean gain adjustment differences between the loudness match of the reverberated sounds and dry sounds as a function of the physical distance, controlled in the experiment by the d/r ratio.

for the "imagine two players in a room..." test (adjusted multiple r=0.910), and

$$\Delta L_{k\,j} = 0.29 * \text{distance ratio} - 1.7$$

for the "do not think about any room..." test (adjusted multiple r=0.912)
Clearly, the two sets of results follow the same relationship along the distance.

The results of the loudness estimate in the atypical reverberant conditions ("imagine two players in rooms..."), corrected for the dry sound differences, are shown in figures 4.6 c, and d, along with the results of the 'ecological' sounds as a function of effort and power level (c), and distance differences (d). It can be seen that the sounds of the strongest effort (marked by arrows) were judged as much louder than the corresponding 'ecological' sounds. The general relationship is, however, very similar to the comparison between the "imagine two players in a room..." and "forget about any room..." parts of the experiment depicted in figures 4.6 a, and b.

4.3.3 Instruction as a Factor of Loudness Judgment

In the next analysis we wanted to discover whether the different instructions: "imagine two players in a room...", "do not think about any room..." had an effect on perception. A three-way repeated measure variance analysis of the loudness differences (instruction by effort by power level) was performed to this end. As expected, the F-ratio appeared highly significant for both effort and power level (F= 9.957, df=3/27 , $p < 0.001$; F=43.871, df=3/27, $p<0.001$), and we could conclude that they highly influenced the loudness judgement. As we pointed out in table 4.1, distance ratio, and consequently direct-to-reverberant energy ratio, are associated with both effort and level dimensions. Therefore, the ANOVA results formally confirmed the dependence of loudness on sound distance. The F-ratio for the instruction dimension, however, did not justify the importance of instruction for the loudness judgement (F=0.498 , df=1/9, p=0.498).

Reverberation was equally important for the loudness estimation, regardless of the instruction. Subjects seemed to be unable to get rid of the illusion of auditory depth and estimate the loudness at their ears. The only significant interaction effect, although not very strong, appeared to be the interaction between the instruction and the power level (F=3.294, df=3/27, p=0.036). To explain this result we need to go back to figures 4.5a and b. It can be seen that decibel differences deviate by more from the power levels of -11.8 dB and -17.4 dB, under the instruction "imagine two players in a room..." than under the instruction "do not think about any room...." Much smaller direct-to-reverberant energy ratios are associated with these levels than with the others, for which the results correspond quite closely to one another. It seems feasible that for the large distance ratios associated with the levels, the more elaborate auditory perspective provided in the second subexperiment worked in a more suggestive way; therefore, the remote sound sources appeared as louder in such conditions. Polynomial contrast analysis confirmed a very strong linear trend for both effort (F=23.059, df=1/9, p=0.001), and power (F=53.617, df=1/9, p<0.001). Direct-to-reverberant energy ratio monotonically associated with both effort and power levels was then the only decisive factor for the loudness match.

The results of the fifth subexperiment ("imagine two players in rooms...") were finally compared to the results of the second part of the test ("imagine two players in a room...") in a three-way repeated measure analysis of variance between the effort, power, and instruction. As expected from the inspection of the plots 4.5c and 4.6 c-d, the results show the importance of power (F=33.409, df=3/27, p<0.001). Effort is another highly significant main effect (F=15.812, df=3/27, p<0.001), whereas, as in the previous analysis, we cannot justify the importance of the instruction (F=0.144, df=1/9, p=0.713).

Inability to judge the loudness of sounds differently despite the varied instructions shows that subjects were unable to concentrate their attention on the required aspect of loudness. In particular, they seemed to be incapable of distinguishing between the loudness of the sound sources (players) from apparent loudness in the ears in reverberant conditions. Estimates of apparent loudness differed significantly from estimates of dry sounds, which shows that reverberation influences loudness regardless of instruction. It may mean that reverberation constitutes a separate category within acoustical events. The category would include 'sounds in a room' related to a specific room and distances of sound. Conversely it would be rather impossible to think about such sounds without an impression of the room, simply as some sounds of a modified complex timbre.

4.4 Discussion

The ability to estimate the loudness of sounds appears to depend on reverberation. The relation is linear on a decibel scale across direct-to-reverberant energy ratios expressed on a logarithmic axis, covering simulated distance ratios 1:34 in a live room. Reverberated sounds are generally perceived as louder than corresponding dry sounds. Differences in loudness larger than about 3 dB become statistically significant, which in our experiment corresponds to distance ratios greater than 6:1, or distance differences larger than sixteen meters, assuming the closest sound is at one meter from the listener. For the closer sounds, the difference in reverberation is too small to produce any perceptible effect, although the above relationship is linear in the whole range of distance differences.

This finding is in concordance with our knowledge about the acuity of distance perception in a reverberant environment. Sheeline (1983) has shown that it is rather low, decreasing for large distances. The results of our research on distance perception with the loudness cue removed (see experiment 1) also show that perception of distance differences in such conditions improves greatly when the physical distance between sounds is larger than two meters. The direct-to-reverberant energy ratio only becomes effective as a cue for the large distance ratios. In other words, a substantial ratio of reverberation is needed to provide the acoustical perspective of a room. From such a perspective, sounds located very far from the listener are perceived as louder than we would expect from a judgment based on their intensity alone.

The consistent relationship between the d/r ratio and gain adjustment has a form similar to the general formula for the loudness sone scale. Put differently, the d/r ratio seems to behave like loudness in the sense of Weber's law, and can possibly be traded for loudness. The results are also in accord with Warren (1973), who found that reverberation has an effect on the loudness judgments of speech stimuli. He attributed the anomalous loudness functions to both acoustical perspective and the non-natural way the direct-to-reverberant ratio of the speech stimuli was changing in his experimental procedure. He argued that conflicting distance cues arising because of the constant ratio at different power levels required a greater attenuation for the half-loudness estimate. The thesis, that we somehow use the reverberation features of the sound source to derive sound loudness, can be supported by our experiment as well. The results also conform to the first two postulates of Warren (1963) (see section 1.2.2), and possibly indicate that loudness is a derived percept secondary to sound distance. His third postulate, however, that loudness can be calculated from physical principles governing sound (including its distance) seems to be unrealistic, at least in light of our present understanding of this phenomenon.

The subjects' inability to distinguish between the instructions of parts 1, 2, and 5 of the test shows that reverberation is too strong a cue to be simply ignored in loudness judgment. It is extremely difficult to consciously and systematically follow an image of a room (rooms) given the auditory perspective provided by the direct-to-reverberant energy ratio. For this reason the loudness judgments of sound sources in a room (rooms) did not differ much from the judgments of 'just reverberated sounds' in the subjects' ears. The ability to consistently remain concentrated on the loudness of sound sources in a room is rare, even among such sophisticated listeners as composers and musicians. For a similar reason, the hypothesis that the constant direct-to-reverberant sound energy ratio is estimated, and acts as a misleading factor in this context cannot be sustained. Acuity of distance perception, which is based on the ratio, is not large enough to allow a systematic distinguishability between the loudness judgment of the sound source and the loudness of sound in the ears. Therefore, the thesis about a conscious use of distance in auditory perception in a similar way we estimate size in the visual world cannot be justified from this experiment.

Reverberation is a complex phenomenon whose profound effect on a dry sound can be attributed to many physical parameters, not only the d/r ratio. While it is convenient to control artificial reverberation by changing this ratio, and it is useful to introduce such a theoretical notion in a completely diffused field, it is difficult to imagine how the auditory system is able to come up with an estimation of this quantity. Does it judge some growing characteristic 'noisiness' of the sound added by reverberation? At such a point it may simply be driven by timbral changes. Other sound parameters are also correlated with distance and are proven to work perceptually, for example, the interaural cross-correlation coefficient, possibly showing the phase consistency between the direct and reverberant parts. Reverberation makes the dry sound longer. At the same time it changes the relative energy distribution of a dry sound. The energy, typically concentrated in the attack of a dry percussive sound, is shifted away from the onset of the sound, and becomes more evenly distributed over time. The steepness of the envelope of the attack part is smoothed by reverberation as well. Both the relative energy distribution and rapidness of the onset are also associated with playing effort, or articulation of sound. Unfortunately, their effect upon loudness is not known. It is normally assumed that, except for very short sounds, loudness is linearly proportional to the integrated energy or the SPL of sound. However, we intuitively feel that, for example, a percussion sound played backwards may not be as loud as the original one. Therefore one should not exclude the possibility that factors other than the direct-to-reverberant energy ratio, but specific to reverberation and correlated with physical distance, contributed to loudness in this experiment.

Chapter 5

Conclusions

In this research, interrelationships between distance perception and loudness perception were pursued. Such interrelationships were implied in the literature, but very little systematic experimental work was done in this direction. Although the importance of reverberation on distance perception was recognized long ago, loudness was regarded as a primary cue for distance in most of the previous research. However, loudness has also another obvious meaning: it is the subjective intensity of sound. In this meaning loudness is effective without any reference to the distance. In particular we can distinguish between a loud and a soft sound played from the same distance. For this reason loudness can be also a contradictory cue to distance, for example when a distant sound is played loudly, and a close one softly. A reciprocal relationship was also sporadically reported: sounds heard from a distance in reverberant environment seemed to be louder than equivalent sounds without the reverberation cue. There is a relative lack of experimental work to prove (or disprove) this hypothesis. This thesis was aimed to provide some evidence about these relationships.

In the first part of this work, the ability of distance judgment in absence of loudness was investigated. The major finding of this part was that *loudness is not necessary* as a cue to distance. It has to be stressed here that this does not mean that loudness is not *useful* for distance judgment. It has been shown, however, that relative distance recognition of two sounds at different distances was not significantly worse in absence of the loudness cue than in its presence, except when the sounds were spaced by less or equal to 2 meters. It has been also demonstrated that the distance judgments were independent of the loudness presence, which means that, at least for some subjects, loudness was not the primary factor of distance judgment. Secondary finding of this part was the unidimensionality of the distance perception in absence of loudness. A comprehensive analysis of the stimuli and the acoustical environment uncovered that the direct-to-reveberant energy ratio and the interaural cross-correlation coefficient (up to about the reverberation distance) were best correlated with the perceptual scale of distance, in the absence of loudness. Thus, distance perception can be best explained by a mutual effect of several parameters, rather than a single one.

The 'loudness constancy' hypothesis was the subject of the third experiment. This term was borrowed to provide an analogy to vision, where the 'size constancy' phenomenon can be observed. In this effect, given two objects of an equal size in the visual perspective on a plane, observers will normally see the more distant object as larger than the closer one. In the auditory world, loudness can be thought of as being an equivalent percept to size in vision. Similarly, reverberation has a potential to create the auditory perspective providing cues to distance of sound (in rooms). Therefore, according to the 'loudness constancy' postulates, sounds played from a large distance in a room should be louder than equivalent sounds played from a small distance in respect to the listener's position. The experiment has demonstrated that reverberation (associated with distance in a room) can indeed influence loudness of sound. Reverberated sounds were judged as louder than equivalent dry prototypes. Moreover, this relationship was systematic, positively correlated with the direct-to-reverberant energy ratio and therefore with simulated distance. 'Loudness constancy' also postulates that the auditory perspective is effective because the listener estimates the loudness at the source, rather than in the ears. This supposition cannot be definitely confirmed by the experimental results. The results have shown, that the ability to consistently estimate the loudness of a sound source is rare even among such sophisticated listeners as musicians. Furthermore, similar results were obtained, regardless whether subjects were asked to judge the loudness of sound sources (given additional cues to the auditory perspective), or just the loudness of the sounds in the ears, without any reference to the room. Yet the subjects were apparently influenced by the amount of reverberation in the sounds. One possible explanation is that they were unable to get rid of the auditory perspective illusion. This is to say, they were not able to use the auditory depth reliably and consistently because of the low acuity of our distance hearing, but at the same time they were unable to judge the loudness of the reverberated stimuli without the reference to distance, 'as they appeared in their ears'. This explanation is in favor of the 'loudness constancy' hypothesis, but also points to the fact that, unlike in vision, the perception of distance of sound is not very accurate. Hence, the result is not so strong. On the other hand, we cannot exclude the possibility that another physical attribute of the sound associated with the increasing distance, other than the d/r ratio used to control the distance acted to result in the increasing loudness. The change of energy distribution of the sound in time produced by the reverberation may also be a reasonable cause of the increased loudness. No research is known to explain the possible effect of different energy distributions in time whether in a reverberant environment or in a free field.

Musical or musical-like sounds were used in the first and third experiments, because the work was meant for musicians, but also because some previous experiments suggested a negative impact of familiarity of sound on distance perception. This hypothesis was checked in the second experiment. Contrary to the general opinion, the results have not confirmed the importance of familiarity. At least for musicians, previous knowledge of the sound, and familiarity with the acoustical features of the sound and acoustic environment have proved to be of no importance for relative distance recognition.

List of References

Distance Perception

Ashmead, D.H. DeFord LeRoy, & Odom, R.D., (1990). Perception of the relative distances of nearby sound sources. *Perception & Psychophysics* **47**, 326-331

Begault, D.R., (1987). Control of Auditory Distance. Ph.D. thesis, University of California San Diego

Begault, D.R., (1991). Preferred sound intensity increase for sensation of half distance. *Perceptual & Motor Skills* **72**, 1019-1029

Begault, D.R., (1992). Perceptual Effect of Synthetic Reverberation on Three-Dimensional Audio Systems, *J.Audio Eng.Soc.*, **40**(11), 895-904

Bekesy, G. von, (1960). Perception of the Distance of a Sound Source. In Experiments in Hearing, (McGraw-Hill Book Company, Inc.), 301-313.

Blauert, J., (1983). Spatial Hearing. Cambridge, MA: MIT Press

Butler, R.A., & Levy, E.T. & Neff, W.D., (1980), Apparent Distance of Sounds Recorded in Echoic and Anechoic Chambers. *Journal of Experimental Psychology: Human Perception and Performance* **6**(4) 745-750

Chomyszyn, J. (1992). Loudness as a Cue in Distance Perception, in *Proceedings of the International Computer Music Conference*, San Jose 1992

Coleman, P.D., (1962). Failure to localize the source distance of an unfamiliar sound. *J.Acoust.Soc.Am.* **34**, 345-346

Coleman, P.D., (1963). An Analisis of Cues to Auditory Depth Perception in Free Space. *Psychological Bulletin* **60**, 302-315

Coleman, P.D., (1968). Dual Role of Frequency Spectrum in Determination of Auditory Distance. *J.Acoust.Soc.Am.* **44**, 631-632

Gardner, M.B., (1969). Distance Estimation of 0 -oriented Speech Signals in Anechoic Space. *J.Acoust.Soc.Am.* **45**, 47-53

Gogel, W.C., (1968). The Measurment of Perceived Size and Distance, in W.D.Neff (Ed.), Contributions to Sensory Physiology vol.3 (Academic Press-NewYork), 125-148

Gogel, W.C., (1976). An Indirect Method of Measuring Perceived Distance from Familiar Size. *Perception and Psychophysics* **20**, 419-429

Gotoh T., Kimura Y., Kurahashi A., Yamada A., (1977). A Consideration of Distance Perception in Binaural Hearing. *The Journal of the Acoustical Society of Japan* **33** (12) pp. 667-671

Greene, D.C., (1968). Comments on Perception of the Range of a Sound Source of Unknown Strength, *J.Acoust.Soc.Am.* **44**, 634

Guski, R.(1990). Auditory localization: Effects of reflecting surfaces. *Perception* **19**(6), 819-830

Haustein,B.G. (1969). Hypothesen uber die einohrige Entfernungswahrnehmung des menschlichen Gehors [Hypotheses about the perception of distance in human hearing with one ear]. *Hochfrequentech und Elektroakustik* **78**, pp.46-57

Hirsch, H.R., (1968). Perception of the Range of a Sound Source of Unknown Strength. *J.Acoust.Soc.Am.* **43**, 373-374

Holt, R.E., & Thurlow, W.R., (1969). Subject Orientation and Judgement of Distance of a Sound of a Sound Source. *J.Acoust.Soc.Am.* **46**, 1584-1585

Julien,J.P., Lavandier',C., Warusfel',O. (1989).Compared analysis of the clarity index and the reverberation time, in *Proceedings of the13th International Congress on Acoustics*, Belgrade.

Kozhevnikova, E. V. (1990). Auditory estimate of the distance of a sound source. *Sensory Systems* **4**, 165-171

Kurozumi,K., and Ohgushi,K. (1983). The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality. *J.Acoust.Soc.Am.* **74**(6), 1726-1733.

Lambert, R.M., (1974). Dynamic Theory of Sound Source Localization, *J.Acoust.Soc.Am.* **56**, 165-171

Little, Alex D., (1986). Frequency Spectrum as a Cue to Perceived Auditory Egocentric Distance. MS thesis, North Carolina State University, Raleigh, NC, USA

Little, Alex D., (1992). Spectral content as a cue to perceived auditory distance. *Perception* **21**(3), 405-416.

McGregor, P.,Horn,A.G. and Todd,M.A. (1985). Are familiar sounds ranged more accurately? *Perceptual & Motor Skills* **61**(3), 1082

Mershon, D.H, Ballenger, W.L., Little, A.D., McMurtry, P.L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception* **18**, 403-416

Mershon, D.H. & Bowers, J.N., (1979). Absolute and Relative Cues for the Auditory Perception of Egocentric Distance. *Perception* **8**, 311-322

Mershon, D.H. & King, L.E., (1975). Intensity and Reverberation as Factors in the Auditory Perception of the Egocentric Distance. *Perception and Psychophysics* **18**, 409-415

Mershon, D.H. Hutson, W.E. (1991), Toward the indirect measurement of perceived auditory Distance. *Bulletin of the Psychonomic Society* **29**(2), 109-112.

Molino, J., (1973). Perceiving the Range of a Sound Source when the Direction is Known. *J.Acoust.Soc.Am.* **53**, 1301-1304

Sheeline C.W., (1983). An Investigation of Direct and Reverberant Signal Interactions on Auditory Distance Perception. Ph.D. thesis at the Center for Computer Research in Music and Acoustics, Stanford University (STAN-M-13)

Simpson, W.E., & Stanton, L.D., (1973). Head Movement does not Facilitate Perception of the Distance of a Source of Sound. *American Journal of Psychology,* **86**, 151-159

Strybel, T.Z., & Perrott, D.R. (1984). Discrimination of Relative Distance in the Auditory Modality: The Success and Failure of the Loudness Discrimination Hypothesis *J.Acoust.Soc.Am.,* **76**, 318-320

Warren, R.M. (1963). Are loudness estimates based on distance estimates? *J.Acoust.Soc.Am., **35**, 613-614*

Warren,R.M., Sersen,E.A, and Pores, E.B., (1958). A Basis for Loudness-Judgments. *J.Acoust.Soc.Am.,* **71**, 700-709

Loudness Perception - Selected References

Algom,D., Rubin,A., and Cohen-Raz,L (1989). Binaural and Temporal Integration of the Loudness of Tones and Noises. *Perception & Psychophysics*, **46**(2), 155-166

Chowning, J.M. (1990). Music from Machines: Perceptual Fusion & Auditory Perspective–for Ligeti–. Center For Computer Research in Music and Acoustics. Report No. STAN-M-64

Marks, L.E. (1979). A Theory of Loudness and Loudness Judgments. *Psychological Review* **86**, 256-285

Melara, R.D., Marks,L.E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics* **48** (2), 169-178.

Moore, B.C. (1989). An introduction to the Psychology of Hearing. New York: Academic Press

Scharf, B. (1978). Loudness. In Carterette,E.C., & Friedman,M.P. (Ed.), Handbook of Perception v **4**, chapter 6. New York: Academic Press

Schneider, B, and Parker,S. (1990). Does stimulus context affect loudness or only loudness judgments? *Perception and Psychophysics* **48**(5), 409-418

Schneider, B, Parker,S, Valenti, M, Farrell, G, and Kanow, G, (1978). Response Bias in Category and Magnitude Estimation od Difference and Similarity for Loudness and Pitch. *Journal of Experimental Psychology: Human Perception and Performance* **4**(3), 483-496

Stephens,S.D.G. (1970). Personality and the Slope of Loudness Function. *Quarterly Journal of Experimental Psychology*, **22**, 9-13

Stevens,S.S. and Guirao,M. (1962). Loudness, Reciprocality, and Partition Scales. *J.Acoust.Soc. Am.*, **34**(8) pt.2, 1466-1471

Warren, R.M. (1973). Anomalous loudness function for speech. *J.Acoust.Soc. Am.* **54(2)**, 390-396

Zwicker, E., Flottorp.G. and Stevens, S.S. (1957). Critical Bandwidth in Loudness Summation. *J.Acoust.Soc. Am.*, **29**, 548-557

Zwicker, E., & Scharf, B. (1965). A model of loudness summation. *Psychological Review*, **72**, 3-26

Related References

Allen,J.B. (1977). Short-term Spectral Analysis and Synthesis and Modification by Discreete Fourier Transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing.* ASSP-**25**(3), 235-238.

Allen,J.B., and Rabiner,L.R. (1977). A Unified Approach to Short-Time Fourier Analysis and Synthesis. *Proceedings of IEEE*, **65**(11), 1558-1564.

Ando, Y. (1985). Concert Hall Acoustics, Berlin, Springer-Verlag

Baird, J.C., and Noma, E. (1978). Fundamentals of scaling and psychophysics, New York, Wiley and Sons, Inc.

Beranek, L. (1954). Acoustics.New York, McGraw Hill Book Company

Beranek, L. (1992). Concert hall acoustics-1992, J.Acoust.Soc.Am. **92**(1), 1-39.

Borish, J., (1984). Electronic Simulation of Auditorium Acoustics. Ph.D. Dissertation, Department of Electrical Engineering, Stanford University

Bracewell,R.N. (1986). The Fourier Transform and its Applications. New York. McGraw-Hill Inc.

Cook, P. (1990). Scrubber, Software for Frequency Domain Noise Suppression. Available at ccrma-ftp.Stanford.edu.

David, H.A. (1969). The Method of Paired Comparisons. London: Charles Griffin & Co. Ltd.

Dixon, W.J., and Massey, F.J.Jr. (1983). Introduction to Statistical Analysis. New York: McGraw-Hill Book Company

Fletcher,H.F and Munson W.A.(1933). Loudness, its definition, measurement and calculation. *Journ. Acoust. Soc. Amer.,* **5**, 82-108

Florentine, M. and Buus,S. (1981). An Excitation-pattern model for Intensity Discrimination. *Journ. Acoust. Soc. Amer.*, **70**, 1646-1654.

Glasberg,B.R., and Moore,B.C.J. (1990), Derivation of auditory filter shapes from notched-noise data. *Hearing Research* **47**, 103-138.

Gower, J.C., (1985). Measures of similarity, dissimilarity, and distance. In Kotz,S. and Johnson, N.L. Encyclopedia of statistical sciences, vol.**5**. New York: Wiley and Sons, Inc.

Guilford, J.P. (1954). Psychometric Methods. New York: McGraw Hill Book Co.

Guttman, L.A. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, **33**, 469-506

Karjalainen, M (1984). Use of Computational Psychoacoustical Models in Speech Processing: Coding and Objective Performance Evaluation. *Proceedings of ICASSP 84*, 1.9.1-1.9.4.

Karjalainen, M (1985). A new Auditory Model for the Evaluation of Sound Quality of Audio Systems. *Proceedings of ICASSP 85*, vol.2, 608-611

Kruskal, J.B. (1964). Multidimensional Scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27

Kutruff, H. (1991). Room Acoustics. London: ElsevierScience Publishers Ltd.

Licklider,J.R.C.(1959). Three Auditory Theories. In Koch,S. Psychology: A Study of Science. v1. New York. McGraw Hill.

MacWiliiams,F.J., and Sloane,N.J.A. (1976). Pseudo-Random Sequences and Arrays. *Proceedings of the IEEE*, **64**(12), 1715-1728.

Mason, R.L., Gunst,R.F, and Hess, J.L. (1989). Statistical Design and Analysis of Experiments. New York: Willey & Sons

McAulay,R.J., Quatieri,T.F. (1986). Speech Analysis/Synthesis Based on a Sinusoidal Representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* ASSP-**34**(4), 744-754.

Moore,B.C.J., and Glasberg,B.R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calulating excitation patterns. *Hearing Research* **28**, 209-225.

Moorer J.A., (1979). About this reverberation Business. *Computer Music Journal*, **3**(2):13-28

Mosteller, F. (1951). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, **16**, 207-218

Oppenheim,A.V, Schafer,R.W. (1989). Discrete-Time Signal Processing. Englewood Cliffs, Prentice Hall Inc.

Patterson, R.D. (1976). Auditory Filter Shapes Derived with Noise Stimuli. *Journ. Acoust. Soc. Amer.*, **59**, 640-654

Patterson,R.D.,Nimmo-Smith,I.,Weber,D.L.and Milroy,R. (1982). The Deterioration of Hearing with Age: Frequency Selectivity, the Critical Ratio, the Audiogram, and Speech Threshold. *Journ. Acoust. Soc. Amer.*, **72**, 1788-1803

Portnoff,M.R.(1980). Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing.* ASSP-**28**(1), 55-68.

Rife,D.D, and Vanderkooy,J. (1989). Transfer-Function measurement with Maximum-Length Sequences. *J. Audio Eng. Soc.*, **37**(6), 419-443.

Schroeder,M.R.(1965). New Method of Measuring Reverberation Time. *Journ. Acoust. Soc. Amer.*, **37**, 409-412.

Schroeder,M.R.(1970). Digital Simulation of Sound Transmission in Reverberant Spaces. (Part 1). *Journ. Acoust. Soc. Amer.*, **47**(2):424-431

Serra, X. (1989). A System for Sound Analysis/ Transformation/ Synthesis Based on a Deterministic plus Stochastic Decomposition. Ph.D. dissertation, Stanford University

Serra, X, and Smith III, J. (1990). Spectral Modeling Synthesis: A Sound Analysis/ Synthesis System Based on a Deterministic plus Stochastic Decomposition. *Computer Music Journ.*, **14**(4): 12-24.

Shepard,R.(1962 a). The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function II. *Psychometrika.* **27**, 125-140

Shepard,R.(1962 b). The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function II. *Psychometrika.* **27**, 219-246

Shepard, R. (1974). Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika*, **39**(4), 373-421.

Shepard, R.(1981). Psychological Relations and Psychophysical Scales: On the Status of "Direct" Psychophysical Measurement. *Journal of Mathematical Psychology*, **24**(1), 21-57.

Simon, L.E., and Grubbs,F.E. (1952). Tables of the cumulative binomial probabilities. U.S. Army Ordnance Corps Pamphlet ORDP 20-1.

Smith,J.O., and Serra,X. (1987). PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation. *ICMC Proceedings 1987*, 290-297.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychol.Rev.*, **34**, 273-286

Torgerson, W.S. (1958). Theory and Methods of Scaling. New York: Wiley and Sons, Inc.

Young,F.W. Torsca-9, aFortran IV program for nonmetric multidimensional scaling. *Behavioral Science.* **13**, 343-344.

Van Duyne, S.A., Pierce, J.R., and Smith. J.O. (1994). Traveling Wave Implementation of a Lossless Mode Coupling Filter and the Wave Digital Hammer. *Proceedings of the 1994 International Computer Music Conference*, Aarhus, Denmark

Wilkinson, L., Hill,M., Welna,.J.P., Birkenbeuel,G.K. (1992). Systat for Windows, Version 5. Statistics. Evanston. Systat, Inc.

Zurek,P.M.(1980). The Precedence Effect and Its Possible Role in the Avoidance of Interaural Ambiguities. *Journ. Acoust. Soc. Amer.*, **67**(3), 952-964.