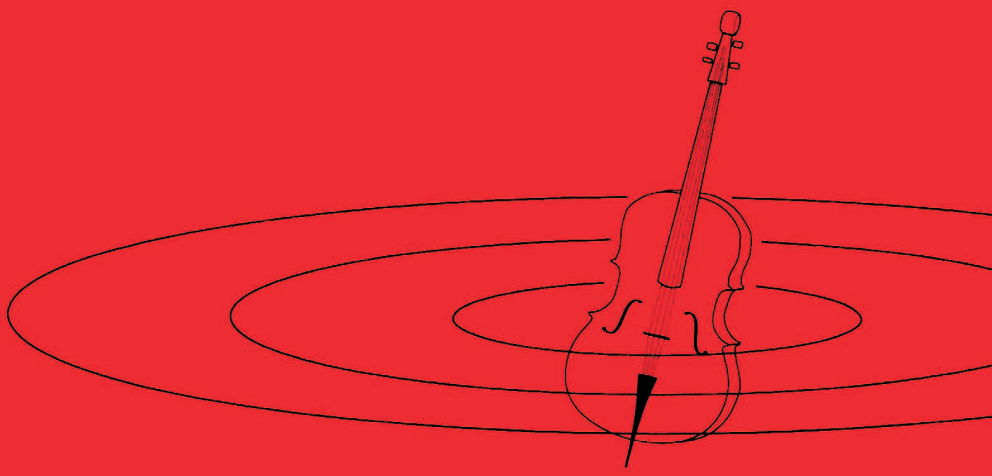


Sound Reproduction by Wave Field Synthesis

Edwin Verheijen

Sound Reproduction by Wave Field Synthesis

Edwin Verheijen

The first edition of this thesis was published in hard copy only. In 1997, a document format like PDF was hardly known. This second edition is only available as PDF, to provide full text search capabilities to the reader.

The original text is unchanged, except for two or three typing errors. As none of the original *computed* graphs have been preserved, I have included these as scans, taken from the hard copy. Unfortunately, Chapter 1 and two appendices were destroyed completely (on 3.5 inch diskette). For this reissue, I took the trouble to edit a recovered ASCII text version of Chapter 1 into a new Chapter 1. Besides the graphs, also the equations in this chapter will appear as scanned pictures. The appendices were not preserved in any form, except hard copy. These are included as scans, and will remain non-searchable.

Utrecht, 28th May 2010

SUPPORT
This work has been supported by the Dutch Technology Foundation (STW) and Philips Research Laboratories, The Netherlands.

Language: English, with additional summaries in German, French and Dutch.

De sorte qu'il faut qu'autour de chaque particule il se fasse une onde dont cette particule soit le centre.

Christiaan Huygens (1629-1695)
Traité de la Lumière, Leyden, Holland, 1690.

Contents

Chapter 1

Introduction

1.1 A century of sound reproduction

The history of sound reproduction starts with the invention of the phonograph by Thomas Alva Edison in 1877. The apparatus consisted of a diaphragm with a stylus that transferred sound waves into indentations in a piece of tinfoil wrapped around a rotating cylinder. In order to reproduce the `stored' sound, the cutting stylus had to be replaced by a playback stylus, sliding through the grooves with less pressure (Butterworth 1977). Edison presented his apparatus as a human voice recorder.

The acousto-mechanical phonograph integrated three essential components for sound reproduction: the recording transducer, the recording medium and the playback transducer. Through the application of electricity it was no longer necessary to keep the essential parts together in one single machine. In the 1920s the condenser microphone (Massa 1985) and the electrodynamic loudspeaker (Hunt 1982) were successfully introduced. These transducers would basically remain unaltered until the present, unlike the recording medium. Flat record discs were already common practice, but still had to compete with cylindrical records. Other media like metal particle tape and magnetic disc (`hard disc') were also considered, but proved premature (Camras 1985).

Until 1930, sound was reproduced in mono. The success of motion pictures in the 1930s ushered in an era of multi-channel sound. Great pioneering work was done by Blumlein (1931), who patented a system of sound pickup and reproduction that we would now call intensity stereophony. Steinberg and Snow (1934) used three spaced microphones to record sound, intended for playback through loudspeakers placed left, center and right. The work of De Boer

(1940) was of great importance for the understanding of stereophonic imaging. Multi-track recordings on magnetic tape became possible just after World War II, but the public had to wait until 1958 before two-channel stereophonic records became commercially available. The 16-inch long-playing vinyl disc, played at 33.3 rpm, would grow extremely popular in the next quarter of a century.

In the early 1970s, an attempt was made to extend the number of reproduction channels. A variety of multi-channel systems was proposed, collectively known as quadraphony (see e.g. Eargle 1971 and Cooper et al. 1972). Two additional loudspeakers were placed to the back or sides of the listener. The demand of compatibility forced researchers to develop so-called matrix systems, with four reproduction channels encoded in two transmission channels. Due to lack of standardization and moderate sound quality, quadraphony failed to conquer the consumer market.

The concept of matrix systems, however, was not abandoned by motion picture industry. Cinema already had a long tradition in multi-channel sound, and Dolby Stereo became a standard encoding matrix in the 1980s. Two-channel `surround sound' recordings on videotape (for use at home) could be decoded to four channels: left, center, right and rear. Surround sound entered the living-room.

Meanwhile, a great improvement in sound reproduction quality was offered by Compact Disc, introduced in 1982. This optical medium provides two-channel recordings with enlarged recording duration, bandwidth and dynamic range of the (digitally stored) audio signals. It is expected that digital techniques will give a new impulse to multi-channel audio and video systems in the near future.

1.2 Objectives

In this thesis a new method of sound reproduction is presented. This method is close to the physics of sound fields, because it attempts to re-create the original sound field. Therefore the original sound field is decomposed in its elementary temporal and spatial properties, transmitted, and finally synthesized to produce the original sound field again. This approach, based on wave field synthesis (Berkhout 1988 and Berkhout et al. 1993), offers a large listening area with equal and high reproduction quality for all listeners. Besides that, the sound engineer is given a powerful tool to re-arrange sources without degrading the sound image. Acoustic corrections to the original sound image can easily be applied, and if desired, the original acoustic impression can be replaced by the ambience of a different room.

This chapter will give a survey of room acoustics and psycho-acoustics. Afterwards the concepts of stereophonic reproduction systems will be discussed. Chapter 2 is a treatise on wave field synthesis. The driving functions for the synthesizing loudspeaker arrays will be derived from the starting point of the Kirchhoff-Helmholtz integral. Two types of loudspeaker arrays will be discussed in Chapter 3. In Chapter 4 the design of the complete reproduction system is described. Finally, Chapters 5 and 6 attempt to give an objective and subjective qualification of the reproduction system in comparison to other systems.

1.3 Room acoustics

Sound is a propagating pressure fluctuation in a fluid. If the medium is homogeneous, sound will extend radially, thus creating spherical waves (Figure 1.1). The magnitude of the pressure fluctuations is inversely dependent on the distance to the source of the sound. For an omnidirectional or monopole source, the sound pressure p can be written as:

$$p(r,t) = \frac{s(t-r/c)}{r} \tag{1.1}$$

where r is the distance from the source, t is the time, c is the propagation velocity (340 m/s in air at 1 atm., 20º C), and $s(t)$ is the source signal. This equation states that the source signal remains unaltered during propagation, except for its amplitude.

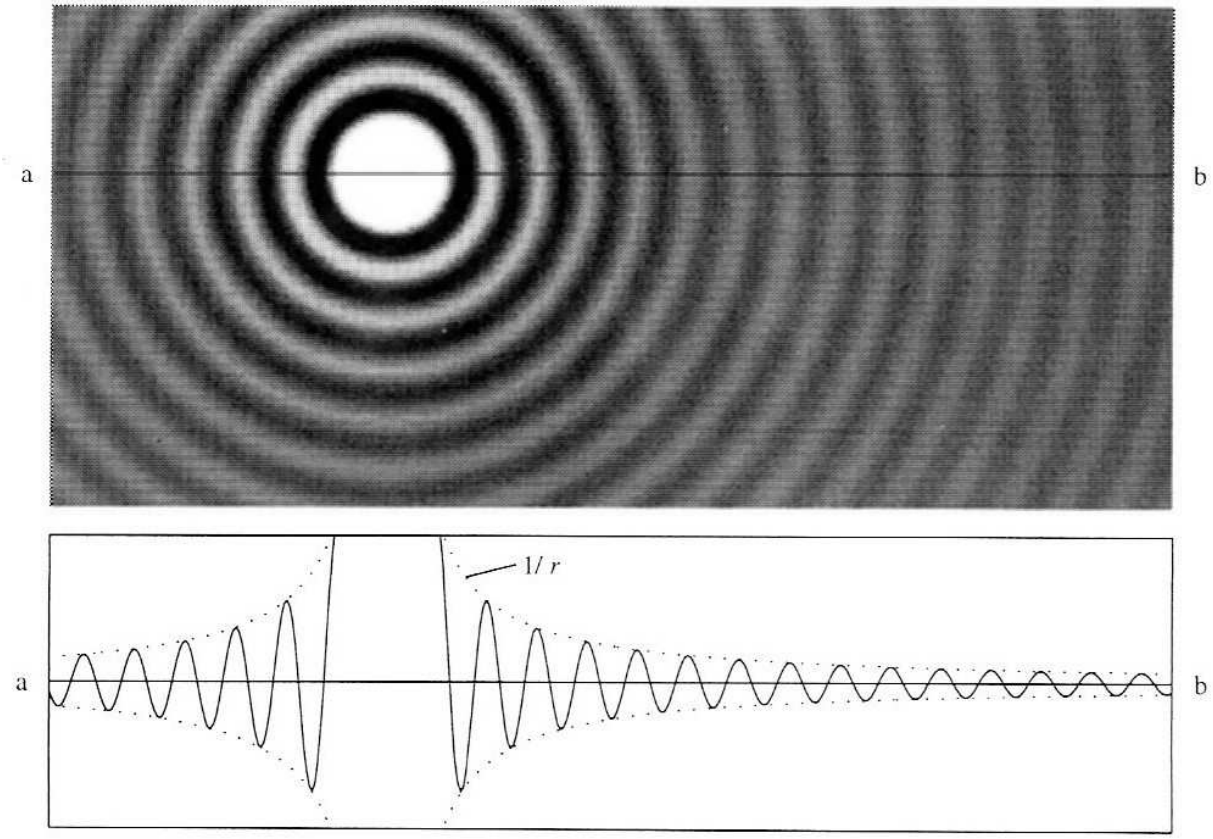


Figure 1.1 Snapshot of the sound pressure of a monochromatic sound source in free space. The magnitudes are displayed as grayscale levels. Cross-section a-b shows the momentary magnitude of the pressure. The decrease of the envelope is inversely proportional to the distance of the source.

A well-known measure related to sound pressure is the decibel (dB). The dB-scale is a relative scale: the root mean square of the sound pressure is compared with a reference pressure. The sound pressure level (SPL) is given by

$$L_p = 20\log\frac{p_{\text{rms}}}{p_0} \tag{1.2}$$

with p_0 = 20 μPa, the pressure of a just noticeable 1 kHz tone for people with normal hearing.

If sound is produced inside a room, the wave fronts will reach the walls within a few milliseconds. Most of the sound energy is bounced back into the room; only part of the energy is absorbed by the walls. The reflected sound wave will be bounced repeatedly by successive walls until it has vanished by absorption. In a small room, e.g. a living-room, one is hardly aware of these reflected waves during a conversation. A hand-clap, however, provides enough peaked energy to let the reflections decay audibly. The effect is much stronger in a large room such as a concert-hall or a church, while it is absent in free field. In acoustics, listening to a hand-clap has evolved to measuring an impulse response. Impulsive sounds, such as produced by alarm pistols, have been replaced by much more controllable measurement signals like sweep-tones (Berkhout et al. 1980) and bursts of maximum length sequences (Rife 1989).

The impulse response is widely regarded as the most meaningful characteristic in room acoustics. Many acoustic parameters related to concert-hall acoustics, such as the reverberation time T60, can be derived from it. For measurements of impulse responses, an omnidirectional sound source (a loudspeaker system producing spherically symmetric sound waves) is placed at a position of interest, e.g. on stage where musical instruments are found during concerts. The measurement signal, fed to the sound source, is received by a microphone in the audience area and processed afterwards to give a unique impulse response for that source-receiver combination.

Figure 1.2 shows an impulse response measured in a shoebox-shaped concert-hall. Within a few milliseconds after the direct sound the reflections arrive. The reverberation tail is said to begin where the reflection density is so high that no individual reflections can be discriminated anymore, which is usually after 80 to 120 ms. An important property of reverberation is that its sound pressure level for steady state signals is more or less independent of the source-receiver distance. If the steady state source signal is switched off, the reverberation level will decrease approximately linear in time due to absorption (i.e. the reverberation tail in an impulse response decays exponentially). The reverberation time T_{60} is defined as the time it takes before the sound pressure level has dropped by 60 dB. Typical values for T_{60} are about 1 second for lecture halls, 2 seconds for concert-halls and 5 seconds for cathedrals. Concert-hall acoustics is reviewed recently by Beranek (1996).

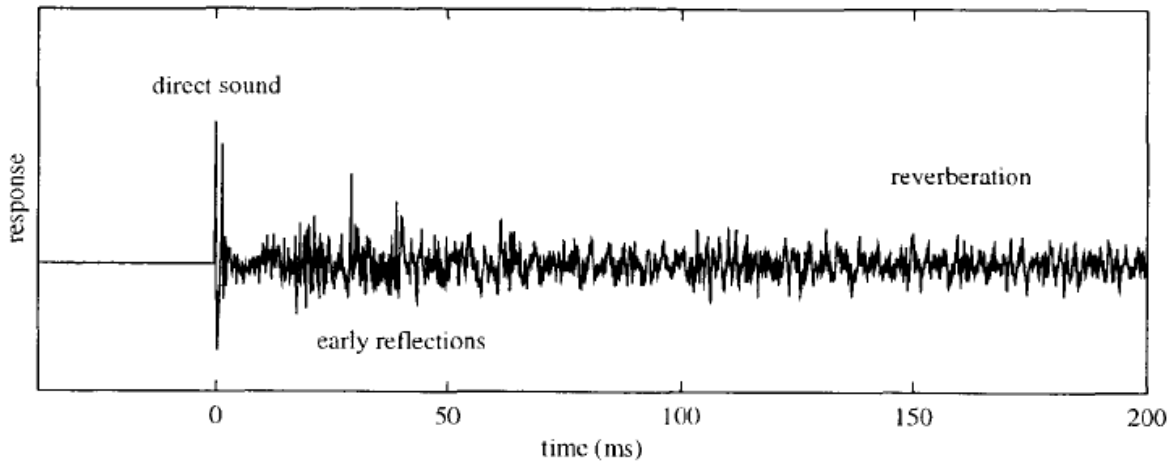


Figure 1.2 Impulse response of the *Stadsgehoorzaal* in Leyden, the Netherlands. The time-axis is aligned with the arrival of direct sound. Source and receiver are 13 m apart. Some strong early reflections can still be distinguished, but later reflections tend to vanish in reverberation.

Consider a trumpet being played on stage while a microphone has been placed in the audience area. Disregarding the directivity pattern of the trumpet, it is possible to calculate the pressure at the microphone $p(t)$, if the impulse response $h(t)$ for the corresponding source-receiver combination is known:

$$p(t) = \int_{-\infty}^{\infty} h(t')s_{tr}(t-t')dt' \quad , \tag{1.3a}$$

or, by introduction of the convolution symbol (asterisk),

$$p(t) = h(t)*s_{tr}(t) \quad . \tag{1.3b}$$

where $s_{tr}(t)$ is the trumpet's output signal as would be recorded by a microphone close to the horn. Note that for causal impulse responses the lower time boundary ($-\infty$) of the integral is replaced by 0. It may be more appropriate to rewrite this equation in the frequency domain. After substitution of the inverse Fourier transform of $h(t)$, given by

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)e^{j\omega t}d\omega, \tag{1.4}$$

and a similar expression for $s_{tr}(t)$ into integral (1.3a) we can evaluate the Fourier transform

$$P(\omega) = \int_{-\infty}^{\infty} p(t)e^{-j\omega t}dt \tag{1.5}$$

to

$$P(\omega) = H(\omega)S_{tr}(\omega) , \tag{1.6}$$

according to the rule `convolution in the time-domain equals multiplication in the frequency-domain'. In terms of signal processing it can be said that the trumpet's spectrum is filtered by the transfer function of the concert-hall. Any other system with the same transfer function can replace the influence of the concert-hall on the sound of the trumpet. But there is more to a concert-hall than only this one-dimensional transfer function. Most of the spatial information of the sound field is missed in the signal $p(t)$. Listening to the signal $p(t)$ would give a faint idea about the spatial environment, but it would fail to provide a *spatial* sound image[*]. Spatial perception of sound is the subject of the next paragraph.

1.4 Spatial perception

Some basic knowledge on the perception of sound is necessary to understand the requirements for a sound reproduction system. The human auditory system is able to retrieve a spatial sound image of an acoustic event. Table 1.1 gives a few examples of such sound images that would occur to a listener by careful listening (while disregarding information of other senses).

Table 1.1 Examples of sound images

acoustic event	temporal[a] information	spatial information
classic concert	the melody of a cello; the timbre of the tenor	concert hall; the cello is to the right; the singer is in front of the orchestra
jazz concert	rhythm of a snare drum; words of a singer	a small jazz-club; the stage is to the left the snare drum echoes from the backwall
street scene (film)	klaxon of a car; scream of a cat	narrow street; the car is nearby; the cat is far-away
nature scene (documentary)	shouting of a shepherd; rumbling of a herd	vast moor (`dry' sound); the shepherd is behind the herd

[a] Though spectral and temporal aspects of sound are quite distinct from a perceptual point of view, they are physically equivalent via Fourier theory.

By introspection of a sound image, the auditory system can distinguish to a certain extent between temporal information and spatial information. Temporal information is mainly related to the sound source, but can also be influenced by spatial conditions. Spatial information depends primarly on the geometry of the room, but is at its turn dependent on the kind of

[*] A spatial sound image is possible if the impulse responses are measured at both ears of a (dummy) head instead of with one free standing microphone, see Section 1.6.5.

source signal. A clear distinction between temporal and spatial perception is therefore not possible.

Three important phenomena of spatial hearing are discussed briefly here: localization, spaciousness and perception of reverberation. Finally, a remark is made on the interaction between the auditory and the visual system.

Localization

The localization of sound sources is due to a combination of directional and distance hearing (Blauert 1983). In anechoic circumstances, the localization blur* of sources in the horizontal plane is in the order of 1º (straight ahead). To the sides this is higher: 10º (Figure 1.3). In the median plane also a value in the order of 10º is found. Directional localization of impulsive sounds in the horizontal plane is independent of the room reverberation time (Hartmann 1983). It is therefore believed that the first arriving wave front (direct sound) establishes the direction of the auditory event.

The human auditory system uses interaural time differences (ITD) and interaural level differences (ILD) for localization in the horizontal plane. Sound arriving at the head from left will reach the left eardrum earlier than the right eardrum. For frequencies below 1500 Hz the ITD is the main cue, while for higher frequencies the ILD plays an important role, next to the ITD of the envelope of the ear signals. Directional hearing in the median plane is based on head and torso cues, i.e. on spectral effects of diffraction from head and torso. In case of con-

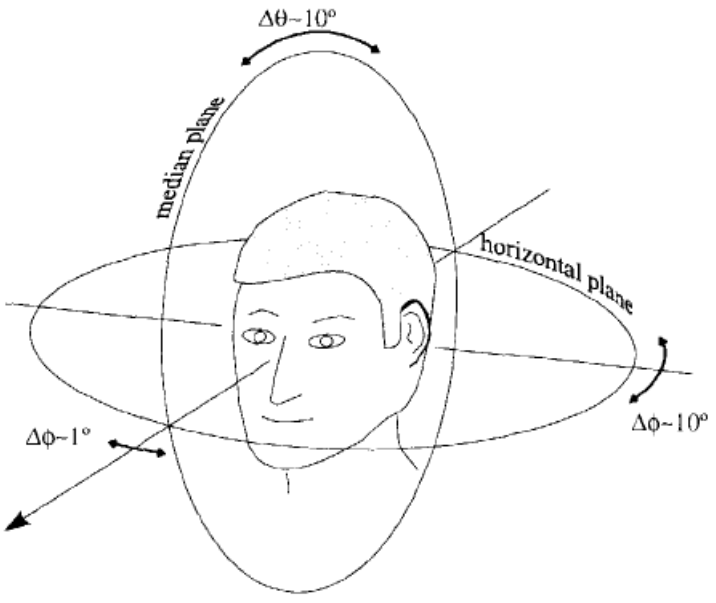


Figure 1.3 Accuracy of auditory directional localization.

* Localization blur is the smallest change of source position that can just be noticed by a subject.

flicting ITD's, ILD's and head and torso cues, the ITD cue is dominant as long as low frequencies are incorporated (Wightman et al. 1992).

For distance hearing in anechoic circumstances a reference sound is needed: if the listener is unfamiliar with the sound, the estimated distance depends only on the loudness, not on the real distance. This implies that the curvature of the wave front is not a cue for distance hearing, at least not for far-away (>1 m) sources. For nearby sources (<1 m), frequency-dependent diffraction of the waves around the head provides distance cues. In reverberant spaces, the main cue for distance determination is the level difference between direct sound and reverberation. Such a judgement is based on the fact that the reverberation level is more or less independent of distance, while the direct sound follows the $1/r$-law.

Spaciousness

In the foregoing, attention was paid to direct sound and reverberation. For a relatively long time the perceptual effects of reflections were scarcely understood. Since the late 1960s the notion has grown that reflections are responsible for a sense of spaciousness as found in good concert-halls. Barron (1971) recognized the importance of early lateral (from the sides) reflections, since they contribute to a `spatial impression'. Potter (1993) refers to this impression as `subjective broadening of the sound source'. He found that frequencies above 1500 Hz do not contribute significantly to spaciousness. Bradley et al. (1995) suggested that there may be two distinct notions of spaciousness: `apparent source width' and `listener envelopment', each connected to different temporal parts of the impulse response. In their opinion the apparent broadening of the source is related to the *early* lateral reflections, arriving no later than 80 ms after the direct sound. A sense of envelopment is felt due to *later* arriving lateral reflections.

Perception of reverberation

A gradual transition from late arriving reflections to reverberation is made after about 100 to 120 ms, where the fine structure of the reflections has more or less a stochastic nature. Reverberation is perceived as a gentle slurring of the sound. Reverberation is appreciated if no particular direction of incidence can be associated with it: if the direct sound stops abruptly, a diffuse and smoothly decaying sound is left.

Vision

The perception of a sound image may be influenced strongly by other sensations, especially those of the visual system. E.g. if visual cues are presented to the listener as well, these will dominate in localization. This phenomenon is known as the ventriloquist illusion. It states that sound will be localized at the apparent visual source if the discrepancies between visual and auditory cues are within certain boundaries.

It is the aim of this research to develop a reproduction system that can be used either with or without accompanying vision (e.g. film or television). The requirements for a reproduction

system must therefore be attuned to the more demanding case of sound reproduction in absence of vision.

1.5 Sound reproduction requirements

For many reasons it may be desired to record and reproduce the acoustic properties of a sonic event. For instance, a musical performance can only be attended by a limited number of people. By recording the event, and afterwards selling or broadcasting the recording, many more people are given the opportunity to have a notion of the original performance. If the suggestion of the reproduced event is so strong that a listener can hardly distinguish it from a real event, acoustic virtual reality becomes feasible. In this paragraph, requirements for realistic sound reproduction are discussed.

Since it is thought that the purpose of a reproduction system is to evoke the same sound image as would be done by the original sound field, an exact copy of the original sound field would be the simplest criterion. However, seen in the light of the perceptual limitations of the auditory system, a less demanding set of criteria will suffice. Such criteria must be based on psycho-acoustic knowledge. The question is therefore: which attributes of the (original) sound image are relevant? The answer can be divided in temporal and spatial attributes. In general, temporal requirements will be fulfilled if highly linear reproduction equipment with sufficient bandwidth and dynamic range is used. The spatial requirements follow the attributes of spatial hearing:

1. **Correct localization.** Direction and distance of reproduced sources should approach the position of the original sources as close as possible.
2. **Correct spaciousness.** The spatial impression of the original acoustic environment should be preserved.
3. **Correct degree of diffuseness of the reverberant field.** The reproduced reverberation must sound equally diffuse as the reverberation in the original environment.

The most important demand is however that the above requirements should be satisfied in a **large listening area**, large enough to accommodate several listeners. It is the motivation for the present research to come to a reproduction method that offers equal conditions for sound imaging in a large reproduction room. It is found that conventional reproduction methods have severe shortcomings at this point. An understanding of the underlying theory of present day reproduction systems is necessary to identify the causes of these problems. In the following paragraphs a review of current reproduction methods is given.

1.6 Two-channel stereophony

Two reproduction channels are a minimal condition for spatial sound reproduction. Two-channel reproduction basically falls apart in two groups of techniques, distinct by their objectives.

1. True stereophony aims at the reproduction of a natural sound image, such that a listener will get an accurate impression of the original sonic event. This reproduction method is mainly used for classical music. In its purest form, only two microphones pick up the sound, and no manipulation other than storage on a medium and power amplification is allowed to their signals, before feeding them to two loudspeakers.
2. The second group of reproduction techniques is used to create an artificial sound image, not necessarily related to one specific sonic event. Popular music is usually (re)produced this way. Microphones are placed as close as possible to the sound sources (`close miking'), so that each microphone signal contains the direct, non-reverberant, sound of only one source. The signals are recorded separately on a multi-track recorder and are often manipulated by various signal processing tools. Finally a two-channel mix is made for reproduction.

In most recordings (for television, radio and playback at home) a mixture of both techniques is used. Purely stereophonic recordings are rare. On the other hand, in popular music, `close miking' is not always appropriate: whenever a natural sound is required, such as for acoustic piano and choir, stereo microphone techniques cannot be avoided. Since it is the aim to investigate natural recording techniques, the former method, i.e. two-channel stereophony, will be discussed in more detail.

Since the introduction of two-channel stereophony much effort has been made to develop reproduction equipment and techniques that offered an accurate sound image to the listener. A combination of a specific loudspeaker configuration and stereo microphone techniques are the means to attain this goal.

1.6.1 Loudspeaker set-up

Two identical loudspeakers are placed in front of the listener, in the way that loudspeakers and listener constitute an equilateral triangle (Figure 1.4). The length of the sides of the triangle is preferably about 3 or 4 meters. The room in which the sound is reproduced should have a reverberation time that is short in comparison with the reverberation time of the rooms in which the recordings were made. Strong reflections should be avoided. Therefore, the listener as well as the loudspeakers should be at some distance (1 or 2 meters) from the walls and the ceiling. The ideal position of the listener is called the `sweet spot'.

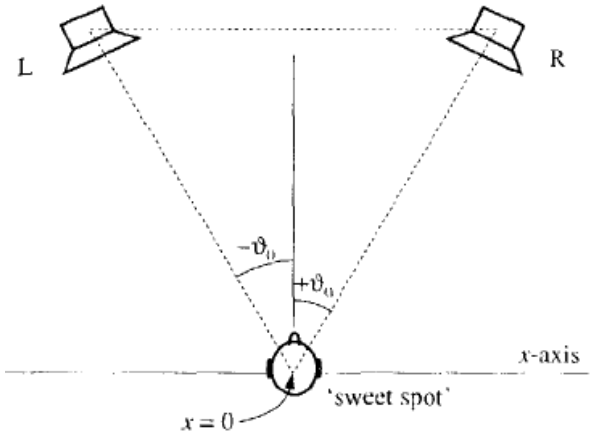


Figure 1.4 Stereophonic loudspeaker arrangement. In the standard configuration, the listener and the left and right loudspeaker constitute an approximate equilateral triangle ($\vartheta_0 = 30^o$).

1.6.2 Reproduction techniques

Stereophonic imaging is divided in two basic reproduction techniques that use loudspeaker feeds that differ either in intensity or in phase. The philosophy behind these imaging methods will be explained in two subsections called intensity stereophony and time-based stereophony.

A. Intensity stereophony

If there exists an intensity difference between the loudspeaker signals, a phantom source* will be heard somewhere between the loudspeakers. The position of the phantom source is a smoothly varying function of the applied intensity difference. Intensity stereophony provides a stable and well-defined image position for a listener at the sweet spot. Since there are no phase differences between the left and right channel, it allows for monophonic reproduction by just summing both signals, which is an important advantage when mono-compatibility is required.

Imaging with intensity stereophony can be understood by studying the local wave field near the sweet spot. If two loudspeakers, in stereophonic arrangement with angle $\pm\vartheta_0$, are sufficiently far from the x-axis (Figure 1.4), plane waves may be assumed from these loudspeaker at the sweet spot. The resulting sound pressure at the x-axis is given in the frequency domain by

$$P(x, \omega) = R\exp(jk_x x) + L\exp(-jk_x x) \tag{1.7}$$

where R and L are the sound pressures of the right and left loudspeaker signals at the sweet spot, and $k_x = k \sin\vartheta_0$ the wavenumber in the x-direction, see Figure 1.5a. This equation can be re-

* Stereophonic image sources are called phantom sources.

written as

$$P(x, \omega) = (R + L)\cos(kx\sin\vartheta_0) + j(R - L)\sin(kx\sin\vartheta_0). \tag{1.8}$$

The resulting phase function along the *x*-axis, defined as

$$\varphi(x, \omega) = \operatorname{atan}\left(\frac{\Im\{P(x, \omega)\}}{\Re\{P(x, \omega)\}}\right), \tag{1.9}$$

can be approximated for low frequencies ($\tan(kx) \approx kx$) yielding

$$\varphi(x) \approx \frac{R - L}{R + L} kx\sin\vartheta_0 . \tag{1.10}$$

Since the phase function is linear in x (at least near $x = 0$), locally a plane wave is reconstructed. The angle of incidence ψ of that plane wave is found from its phase function as

$$kx\sin\psi = \varphi(x), \tag{1.11}$$

leading to the well-known stereophonic law of sines (Clark et al. 1958 or Bennett et al. 1985)

$$\sin\psi = \frac{R - L}{R + L}\sin\vartheta_0 . \tag{1.12}$$

The imaging properties of this law are displayed in Figure 1.5b. The predicted image direction is in good agreement with perceptual results (Blauert 1983). If the level difference exceeds 30

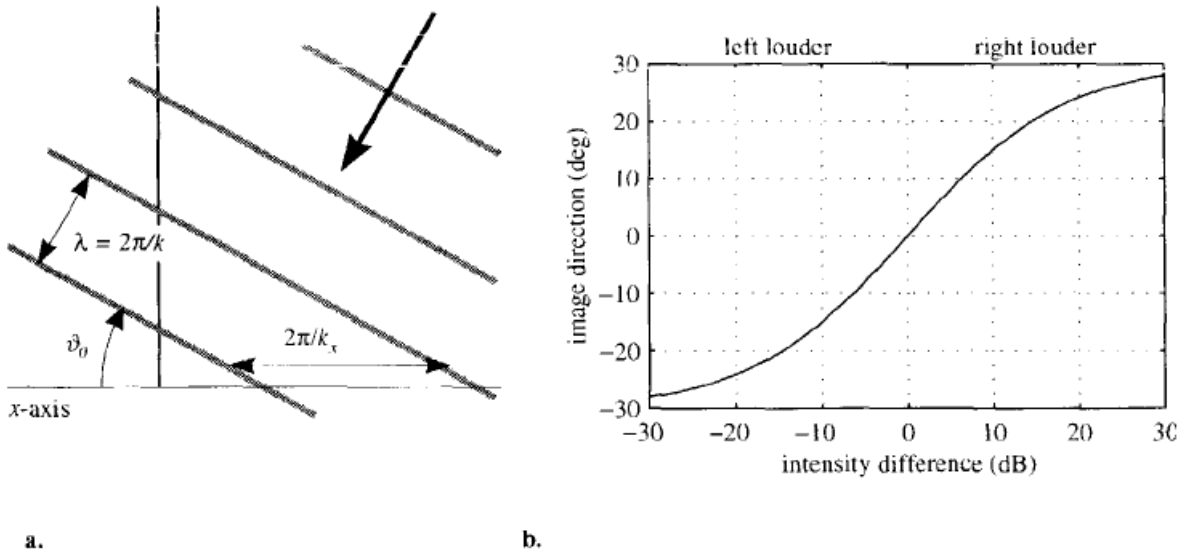


Figure 1.5 **a.** A plane wave at the x-axis as produced by a far-away source at an angle ϑ_0. **b.** Image direction versus interchannel intensity difference for $\vartheta_0 = 30°$, according to the law of sines (1.12).

dB the image appears in one loudspeaker. A smoothly varying function is found in between the extremes. Because locally a wave front is constructed, directional hearing with intensity stereophony is close to natural directional hearing.

The behavior of the wave fronts in a larger listening area is studied by using the exact expression for the sound pressure field. The superposition of the sound pressure of the left and the right (monopole) loudspeaker in the frequency domain yields

$$P(\boldsymbol{r}, \omega) = \frac{S_L(\omega)\exp(-j\omega r_L/c)}{r_L} + \frac{S_R(\omega)\exp(-j\omega r_R/c)}{r_R}, \qquad (1.13)$$

with r_L and r_R the distances to left and right loudspeaker. For sinusoidal signals S_L and S_R, an interference pattern is produced as shown in Figure 1.6. The wave fronts arriving at the listener's position are tilted towards the loudspeaker with the higher magnitude. Along the central axis (ahead of the listener), the interference pattern shows regions where the local form of the wave front is more or less the same as that of a single source somewhere between the - loudspeakers. There are, however, large phase transitions between neighboring regions, visible as light and dark shading. The width of the local wave fronts depends on the wavelength and the angle between the loudspeakers as seen from the listener's view-point: the further he moves away, the smaller the angle (at infinity this angle is zero and plane wave fronts will be received).

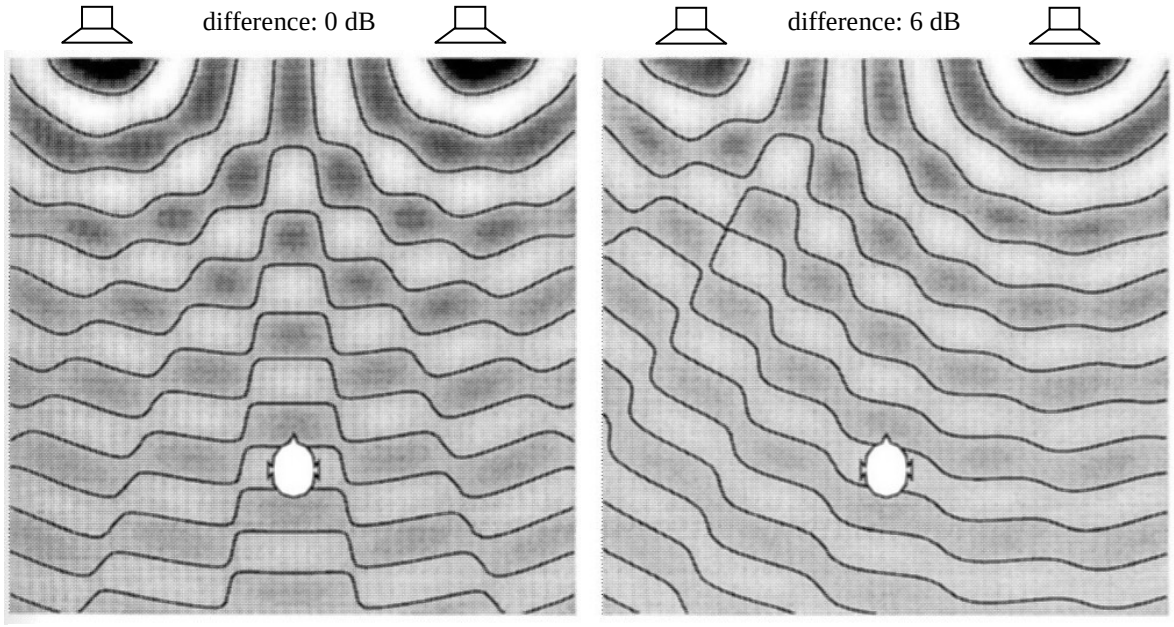


Figure 1.6 Momentary pressure distribution in a 5×5 m^2 area as a result of level differences between two loudspeakers (f = 400 Hz). The (momentary) phase is positive in the light regions, while it is negative in the dark regions. Iso-phase lines are drawn at phase 0 and π (alternating) to indicate the orientation of the local wave fronts.

a. No level difference. **b**. The right loudspeaker has a 6 dB higher intensity level.

The frequency dependence and the angle of incidence (or direction of propagation) of the local wave front can be investigated for the listener position. Therefore, the exact expression for the phase function, calculated from (1.13), is taken. The angle of incidence ψ is now obtained from the phase difference between two receivers at $x = x_1$ and $x = x_2$ (Figure 1.7)

$$\psi(\omega) = \operatorname{asin}\left(\frac{c}{\omega \Delta x} \operatorname{atan} \frac{\Im\{P(x_2, \omega)/P(x_1, \omega)\}}{\Re\{P(x_2, \omega)/P(x_1, \omega)\}}\right) \qquad \omega < \frac{2\pi c}{\Delta x}, \tag{1.14}$$

with $\Delta x = x_2 - x_1$ the spacing between the receivers. Note that this equation is exact for plane waves with angular frequency ω, and still very accurate for spherical waves if the source distance is larger than Δx. Equation (1.14) is ambiguous for wavelengths smaller than the spacing Δx; these are not considered by restricting the angular frequency to $\omega < 2\pi c/\Delta x$.

If Δx = 17 cm is taken, the receivers may be thought of as ears. It should be noted, however, that (1.14) cannot be used to predict the actual image direction (as perceived by a listener) because it does not take into account wave diffraction around the head, and furthermore, no binaural localization model is incorporated. Nevertheless, it is a useful equation to gain insight in the (local) propagation direction of wave fronts.

The angle of incidence ψ is shown for four values of the interchannel intensity level difference in Figure 1.8. For low frequencies the image source direction is in agreement with (1.12). For higher frequencies the image direction collapses. This can be understood as follows. The width of the local wave fronts decreases with increasing frequency. For frequencies approaching $f = c/\Delta x$ the receivers (`ears') will reach the borders of the local wave fronts.

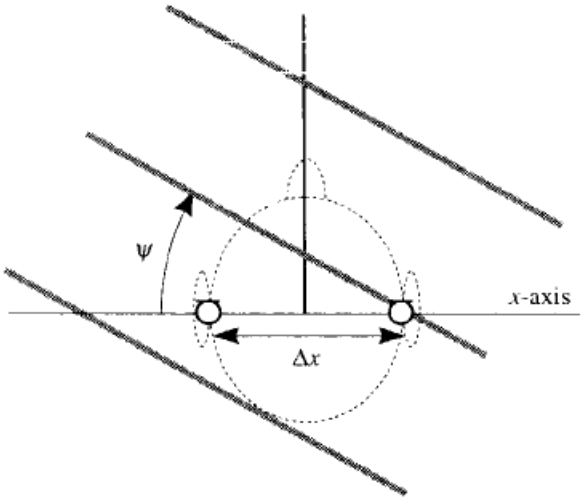


Figure 1.7 The frequency-dependent angle of incidence is calculated from the phase difference between two receivers.

It is well known that the listening area for stereophony is restricted to the central axis. For listeners at off-axis locations the phantom source tends to shift towards the closest loudspeaker. It can be shown that there is a strong frequency dependence for the image direction as calculated from the phase differences between two receivers (centered off-axis). In Figure 1.9, Equation (1.14) has been applied for a position 0.5 m to the right of the sweet spot. For very low frequencies (f < 200 Hz) the receivers are still in the same local wave front region. For higher frequencies the phase relation between the receivers is unstable, as can be seen from the strong fluctuations of the image direction.

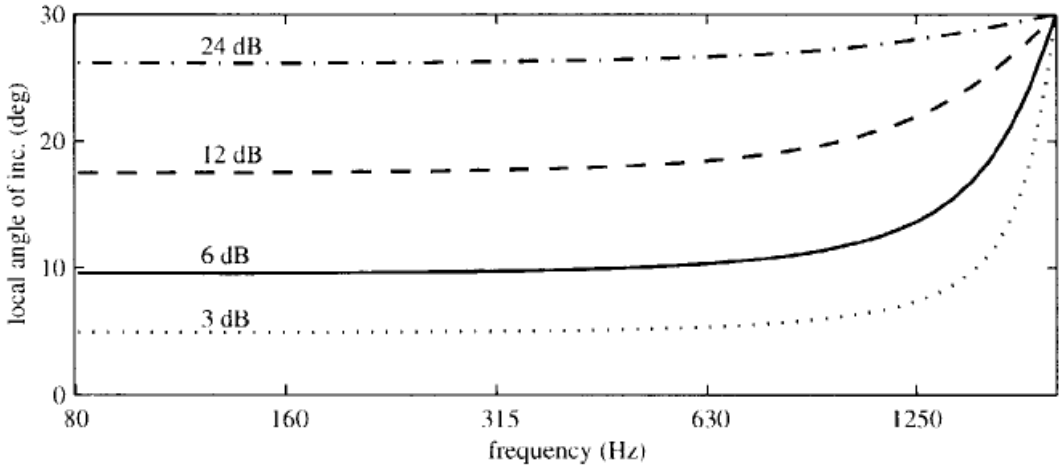


Figure 1.8 On-axis imaging with intensity stereophony. Position of receivers: 8.5 cm to the left and the right of the sweet spot.

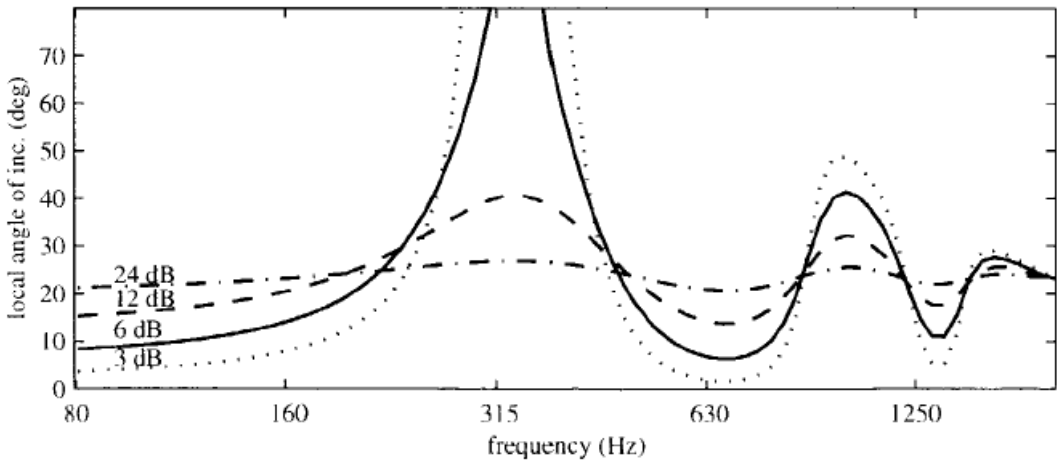


Figure 1.9 Off-axis imaging with intensity stereophony. Position of receivers: 41.5 cm and 58.5 cm to the right of the sweet spot.

B. Time-based stereophony

If a small time delay is applied to the loudspeaker signals, a phantom source is heard towards the loudspeaker with the earliest signal. The imaging mechanism is based on the theory that the interaural time difference (ITD) is the main cue for directional hearing. For a real source, the maximal ITD (for a source 90º off-center) is approximately 0.7 ms. In other words, 0.7 ms is the largest natural signal delay between both ears. If two loudspeakers are used, with an increasing interchannel delay from 0 to 1.5 ms, the image will gradually move towards the earliest channel (Bartlett 1991). This effect is referred to as `summing localization' (Blauert 1983). If crosstalk from right loudspeaker to left ear and from left loudspeaker to right ear could be neglected, a delay of 0.7 ms would be enough to place the phantom image entirely to one side. Apparently, the resultant signals at the two ears suffer from crosstalk, and an exaggerated delay is necessary to shift the image to one loudspeaker.

For time-based stereophony, the perceived direction of the phantom image is strongly dependent on the source signal (Blauert 1983). If narrow band signals (1/3 octave band pulses and sinusoidal signals) are used, the direction of the image between the loudspeakers is not uniformly increasing with increasing delay time. This effect can also be heard with single tones in music reproduction: successive tones played at one instrument sometimes appear to come from different directions.

The effect is studied by looking at the reproduced wave field. Figure 1.10 displays the pressure distribution for an interchannel time difference of 0.5 ms and 1.0 ms. In comparison with Figure 1.6 it is clear that a time difference between the microphone signals leads to a shifted interference pattern. By further delaying the left channel, the pattern will shift until it is eventually returned to the starting-point.

Note that even for low frequencies the phase difference between the two receivers (`ears') on either side of the central axis can be high. This is illustrated in Figure 1.11 by applying Equation (1.14) for two receivers placed at either side of the sweet spot. The image direction suddenly blows up when the phase difference between the receivers is high. This behavior seems to be in agreement with the previously mentioned observations that localization of sinusoids suffers from divergence.

In general, the sound image produced by time-based stereophony does not correspond to a sound image heard in natural situations. The image direction varies from listener to listener, is affected by small head movements, and is quite signal dependent (Lipshitz 1986).

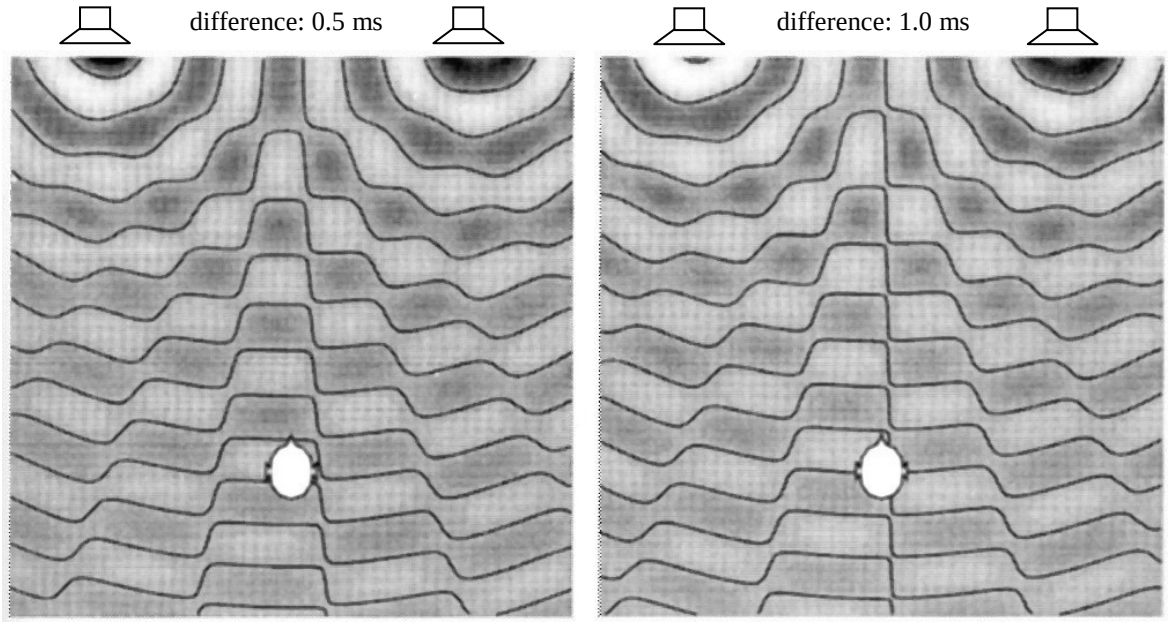


Figure 1.10 Momentary pressure distribution in a 5×5 m^2 area as a result of time differences between two loudspeakers (f = 400 Hz). Contour lines are drawn at zero pressure for emphasis of wave fronts.
a. 0.5 ms time difference. **b**. 1.0 ms time difference.

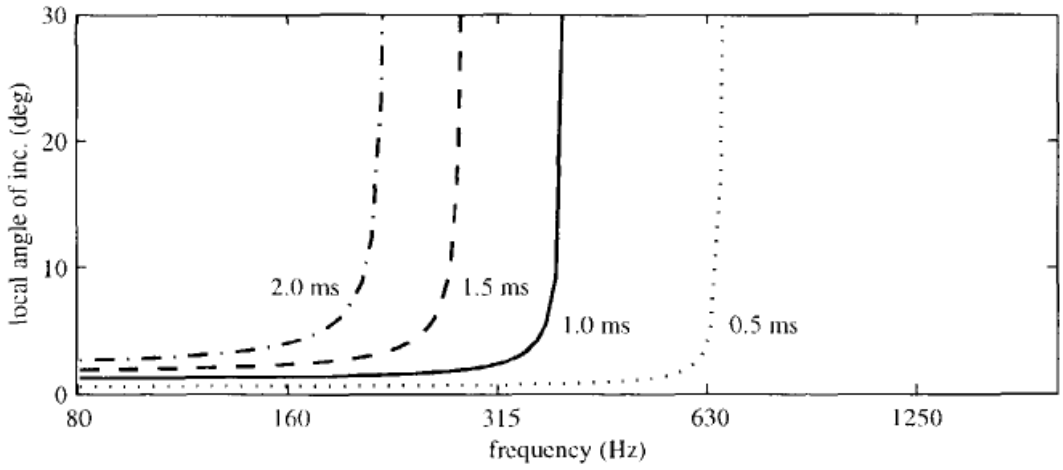


Figure 1.11 On-axis imaging with time-based stereophony. Position of receivers: 8.5 cm to the left and the right of the sweet spot.

1.6.3 Microphone and mixing techniques

Microphones can be classified on the basis of their directivity characteristics. There are two main types of directivity patterns: omnidirectional and bidirectional (Figure 1.12). From these two, most other patterns can be constructed, e.g. the cardioid pattern is a sum pattern of the omnidirectional and bidirectional pattern. For stereophonic sound reproduction a set of two microphones is used to pick up the sound at some distance from the source(s). The configuration of these microphones and the appropriate distance to the sources is largely a matter of taste of the recording engineer, though a classification of microphone techniques, each with their own drawbacks and advantages, can be made. A thorough overview has been given by -Streicher et al. (1985) and Bartlett (1991). Only the main classes are described here: coincident pair, spaced pair and combined techniques.

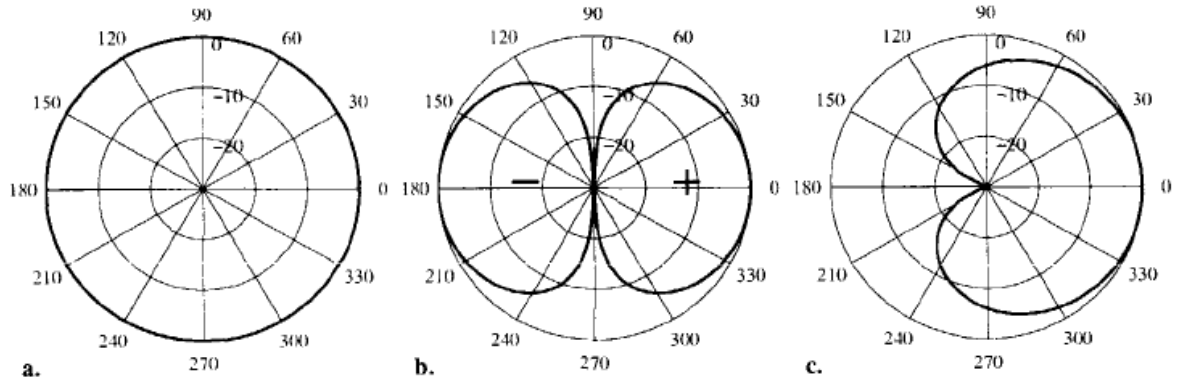


Figure 1.12 Directivity patterns of commonly used microphones. **a**. Omnidirectional or pressure microphone. **b**. Bidirectional or velocity microphone (with lobes in anti-phase). The pattern follows cos φ. **c**. Cardioid microphone. The directivity pattern is given by ½ + ½ cos φ.

A. Coincident pair techniques

Two directional microphones are mounted closely to each other (nearly touching) while aiming at the left and right sides of the source area (Figure 1.13a). Intensity differences will arise, but phase differences between the signals are excluded. Therefore the method perfectly fits the requirements of intensity stereophony. The closer the source is to the axis of one of both microphones, the larger the intensity difference between the channels. A special form of coincident pair techniques is the Blumlein pair: two velocity microphones crossed at 90º, see Figure 1.14a. This technique was originally developed by Blumlein (1931) and is famous for its excellent localization properties.

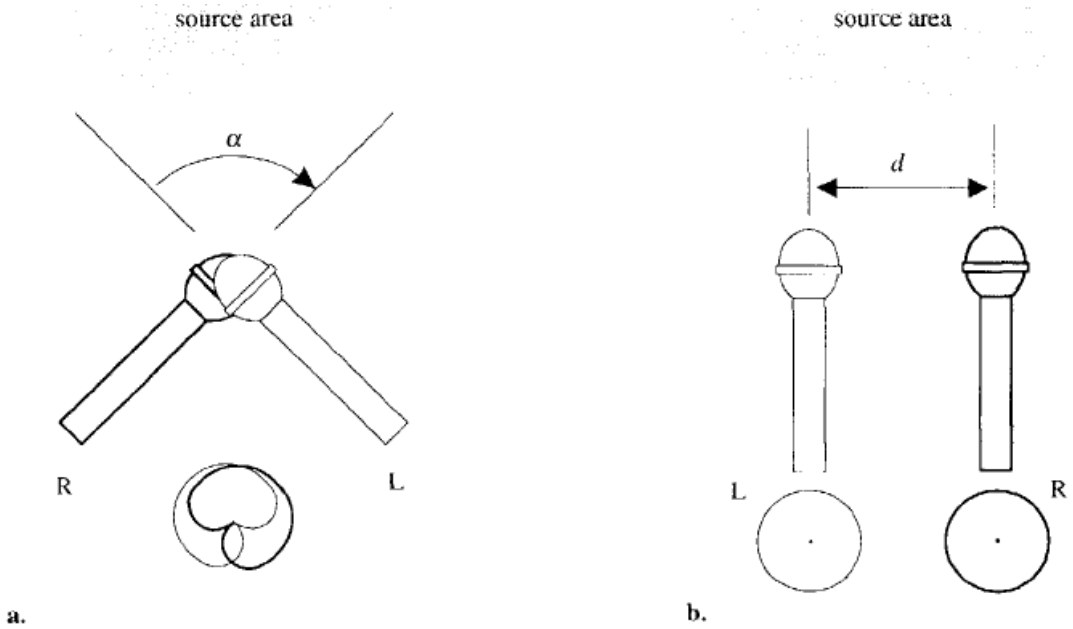


Figure 1.13 Stereophonic microphone techniques. **a**. Coincident pair or XY-technique. **b**. Spaced pair or AB-technique.

The microphones signals of a Blumlein pair are given by

$$L(\omega) = S(\omega)\cos(\beta + 45^\circ), \tag{1.15a}$$

and

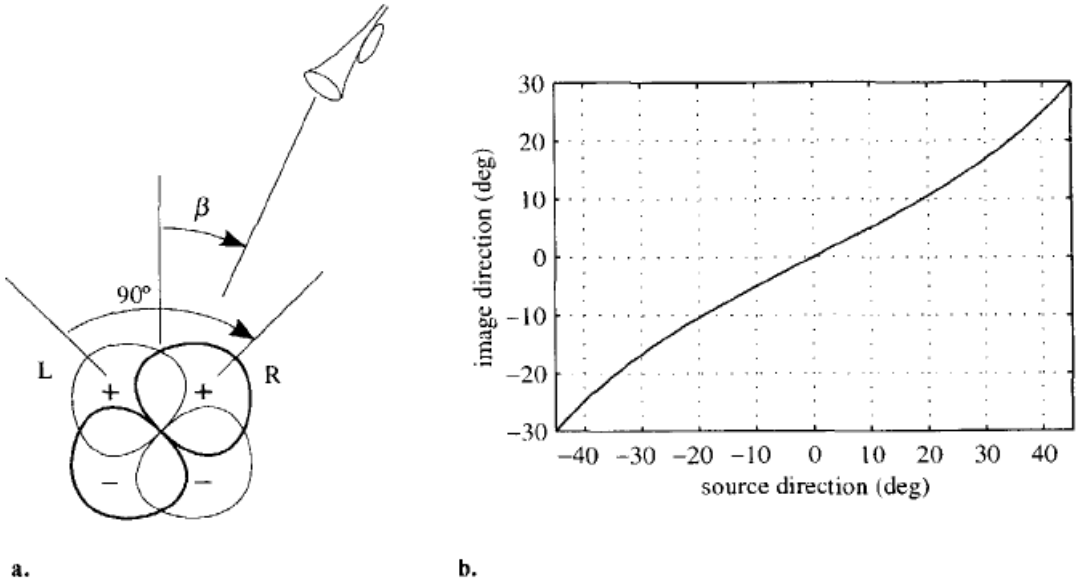
$$R(\omega) = S(\omega)\cos(\beta - 45^\circ), \tag{1.15b}$$

where $S(\omega)$ is the source signal and β the source angle. Note that the cosine factors represent the directivity of the velocity microphones. Clark et al. (1958) combined these signals to a formula for the intensity differen-ces at the left and right channel for the Blumlein pair:

$$\frac{R-L}{R+L} = \tan\beta\,. \tag{1.16}$$

Finally, from (1.16) and the law of sines (1.12) it follows that stereophonic imaging with a Blumlein pair is described by

$$\sin\psi = \tan\beta\ \sin\vartheta_0\,. \tag{1.17}$$



a. b.

Figure 1.14 **a**. A source at angle β is seen by a pair of crossed velocity microphones. **b**. Stereophonic imaging with a Blumlein pair according to the `law of sines' ($\vartheta_0 = 30^o$).

Figure 1.14b plots the true source direction b against the direction of the sound image ψ. Apparently, the Blumlein pair has fair imaging capabilities, though the images are slightly closer to the center than the true sources. In practice, a disadvantage of the used velocity microphones may be that they are equally senstive to sound from the rear as from the front. If the direct/reverberation ratio of the signals is to low, other directional microphones, such as cardioids can be used for intensity stereophony.

In most contemporary stereophonic recordings also spot microphones are used. In addition to a main microphone pair, the signals of microphones close to the sources are recorded as well. For that purpose directional microphones are often used, but omnidirectional microphones may be preferred sometimes for sound quality reasons. The signals, that primarily contain the direct sound of one or a few sources, are added to the main microphone signals via a panoramic potentiometer circuit (pan pot) on the mixing console. This circuit distributes the signal energy between left and right channel. Therefore, imaging properties as described for coincident pair signals also apply for panned signals.

An advantage of the use of spot microphone signals is that the direct/reverberation ratio and the balance between various sources within the stereophonic image can be controlled better. The balancing engineer is able to add more direct sound if the image is too reverberant, or amplify the sound of one source in favor of others. Also, artificial reverberation or other signal manipulation can be used on spot microphone signals.

B. Spaced pair techniques

The second class of microphone techniques is the spaced pair or AB method. This method is appreciated for its ability to reproduce a spacious sound image. Two microphones, mostly with omnidirectional characteristics, are placed some distance *d* apart (Figure 1.13b). The sound from a source that is off-center has a different travel-time to each of the microphones. If a source is at a large distance compared to the distance *d* between the two pressure microphones, the pressure difference can be neglected, and the time delay to the farther microphone is the only signal difference. Imaging with a spaced pair technique is therefore described by the theory of time-based stereophony.

It has been argued yet, that an exaggeration of the applied time delays is necessary because of the situation during reproduction, where `crosstalk' from the loudspeakers to the opposite ears must be compensated. A second compensation is necessary due to the absence of a shielding head between the microphones during recording. Interaural level differences, that accompany the interaural time differences under natural circumstances, are missing. Except extra time delays, sometimes a shielding disc (Jecklin) or sphere is used in between the spaced microphones. This method is closely related to binaural recording and reproduction (Section 1.6.5).

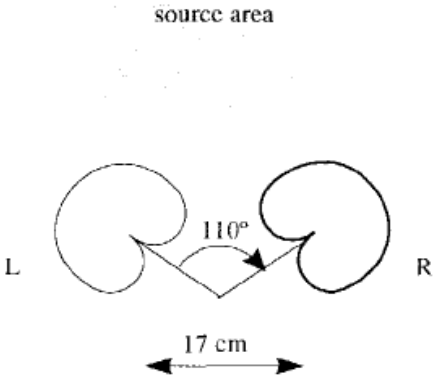


Figure 1.15 The ORTF microphone technique uses two cardioids, angled 110º, and spaced by 17 cm.

C. Combined techniques

Various other stereophonic microphone techniques are employed. Most of them can be regarded as intermediates between coincident and spaced techniques. They attempt to combine the benefits of both classes while avoiding their disadvantages. An example is the ORTF system, originally designed by the French Broadcasting Organization (Figure 1.15). This system has been tuned to yield stable images with reasonable spaciousness. Subjective comparisons of stereophonic microphone techniques are given by Ceoen (1972) and Bartlett (1991).

1.6.4 Spectral properties of stereophony

Because two loudspeakers with correlated signals are used, stereophonic reproduction suffers from comb filter effects. In an anechoic environment, which, of course, is not a natural location for reproduction, the spectral distortion is easy to perceive. The thick line in Figure 1.16 shows the response at a receiver placed 8.5 cm to the right of the sweet spot for two identical signals fed to the loudspeakers. At 2.0 + 4.0*n* kHz (*n* is a positive integer) the left and right loudspeaker contributions are in anti-phase. If the receiver is moved 0.5 m to the right, much more notches occur, but they are less deep. Due to room reflections in a normal listening environment, the comb filter effects are partly masked by reflective sound.

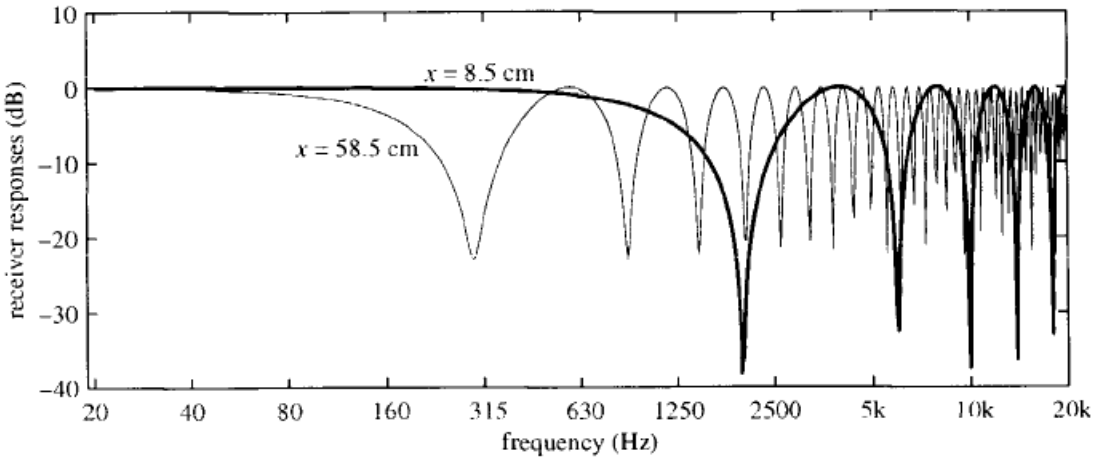


Figure 1.16 Simulated stereophonic responses. Left and right loudspeaker are driven by identical signals. The receiver positions are 8.5 cm (thick line) and 58.5 cm (thin line) to the right of the sweet spot.

1.6.5 Binaural recording and reproduction

A special form of two-channel stereophony is binaural reproduction. During recording, two microphones are placed at the entrance of the ear canals of an artificial head. A headphone is used to reproduce the recorded sound. Binaural imaging is based on the assumption that a listener would perceive the same sound image whether he were present at the recording session or at the reproduction site, if the sound pressure at his eardrums in the latter situation is an exact copy of that in the former. It is not surprising that binaural reproduction can offer an extremely realistic 3D sound image.

A few disadvantages of binaural reproduction are mentioned here. When the head is moved (during playback), the recorded ear signals do not vary accordingly, which may violate

the realism of the sound image. A second objection is due to shape of the dummy torso and pinnae (shell of the ear). Because these external reflectors and diffractors may differ considerably from the listener's own torso and pinnae, the head related transfer functions (HRTFs) from source to dummy ears will differ from the listener's own HRTFs. Therefore, the perceived image may sound not as natural to all listeners. And finally, coloration of the binaural signals due to HRTFs will disqualify them for use with other reproduction techniques.

1.7 Surround sound systems

For a better understanding of current surround sound systems and the effort taken to remain compatible with two-channel stereophony, it is worthwhile to study the problems of quadraphony. Though commercially considered a fiasco, quadraphony used premises and principles that are still of interest.

1.7.1 Quadraphony

It was recognized that two loudspeakers were not able to re-create a true 3D image of the original sound field: sound coming from the sides, the back and above were lacking. If only the horizontal plane was considered, at least two loudspeakers were required to fill the gap (Figure 1.17a). This suggested that at least two more audio channels were necessary. However, not any extra playback channel would contribute as much to the total stereo image as the front loudspeakers: with each extra channel more redundancy was added. Moreover, compatibility with two-channel records was an important demand. A solution was therefore sought in matrix systems (Ten Kate 1993).

A matrix equation describes how the loudspeaker signals are related to the original signals. Most quadraphonic proposals use a 4-2-4 scheme: four initial (studio) channels, two

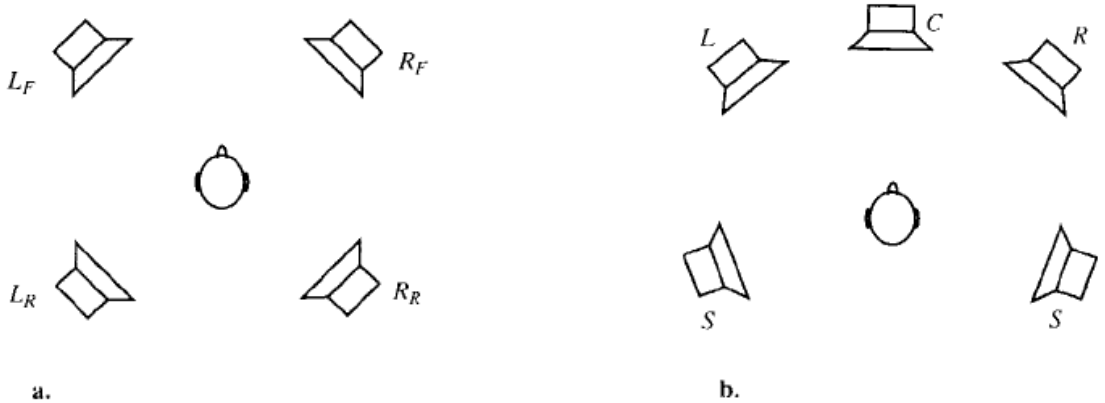


Figure 1.17 a. Quadraphonic loudspeaker arrangement. **b**. Surround sound loudspeaker arrangement, such as used by Dolby Surround and some other systems.The surround signal *S* can be delayed by a few tens of milliseconds to ensure that sound is not localized at the surround loudspeakers.

transmission channels and four loudspeaker channels. Such systems consist of a 2×4 encoding matrix and a 4×2 decoding matrix. The transmission channels are encoded as

$$\begin{bmatrix} L_t \\ R_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \begin{bmatrix} L_F \\ R_F \\ L_R \\ R_R \end{bmatrix} \tag{1.18}$$

where L_F , R_F , L_R and R_R represent the signals of initial channels for left front, right front, left rear and right rear loudspeakers. The transmission channels must serve as reasonable two-channel reproduction channels for compatibility requirements. The quadraphonic loudspeaker channels, denoted with primes, are retrieved by the decoding scheme

$$\begin{bmatrix} L'_F \\ R'_F \\ L'_R \\ R'_R \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix} \begin{bmatrix} L_t \\ R_t \end{bmatrix}. \tag{1.19}$$

Power conservation yields $a_{1i}^2 + a_{2i}^2 = 1$ for $i \in \{1, 2, 3, 4\}$ and $b_{1j}^2 + b_{2j}^2 + b_{3j}^2 + b_{4j}^2 = 1$ for $j \in \{1, 2\}$. A 4×4 matrix results after substitution of (1.18) in (1.19). This matrix can be optimized according to certain premises (Eargle 1971).

Besides disagreement on the number of channels, also obscurity about the recording strategy and the aims to be achieved, caused a stagnation in the development of multi-channel surround (Gerzon 1977). At last, Dolby Laboratories introduced the Dolby Surround system. Being derived from their cinema surround sound standard Dolby Stereo, this domestic system became successful as well.

1.7.2 Dolby Surround

The Dolby Surround system uses a left, center, right and surround channel recovered from two transmission channels (Figure 1.17b). Such a configuration is appropriate if a television set is used in combination with audio channels. The central phantom image, as heard between left and right loudspeakers, is now anchored in the central loudspeaker. The transmission scheme is given by (Dressler 1993)

$$\begin{bmatrix} L_t \\ R_t \end{bmatrix} = \begin{bmatrix} 1 & q & 0 & -jq \\ 0 & q & 1 & jq \end{bmatrix} \begin{bmatrix} L \\ C \\ R \\ S \end{bmatrix}, \tag{1.20}$$

with $q = \frac{1}{2}\sqrt{2}$ (or –3 dB) and j is a 90° phase shift. The loudspeaker signals are recovered as

$$\begin{bmatrix} L' \\ C' \\ R' \\ S' \end{bmatrix} = \begin{bmatrix} 1 & q & 0 & -jq \\ q & 1 & q & 0 \\ 0 & q & 1 & jq \\ jq & 0 & -jq & 1 \end{bmatrix} \begin{bmatrix} L \\ C \\ R \\ S \end{bmatrix}. \tag{1.21}$$

Crosstalk between adjacent channels is –3 dB, while it is –∞ dB between opposite channels. The *S*-signal is bandlimited from 100 Hz to 7 kHz, encoded with a modified Dolby B-type noise reduction and phase shifted over ± 90º. Since the crosstalk would lead to intolerable blurring of the image resolution (Bamford et al. 1995), additional active processing is done by a Pro Logic Decoder. The decoder detects the dominant direction in the playback channels and then applies enhancement in the same direction and in proportion to that dominance.

Compatibility with two-channel stereophony is supported to such an extent that a fairly satisfactory downmix is given. However, the harm done to stereophonic imaging by the encoding procedure is rather contrasting with the care taken in setting up an accurate stereophonic sound image (see Section 1.6).

1.7.3 Ambisonics

Ambisonics has its roots in quadraphonic matrixing (Cooper et al. 1972) and has been proposed as a successor of stereophony (e.g. Gerzon 1985 and Lipshitz 1986). Ambisonics uses a different microphone and imaging technique than stereophony. It pursues local wave front reconstruction: a listener at the sweet spot receives wave fronts that are copies of the original wave fronts. In this section a simple example is given with the following assumptions:

1. Analysis and reconstruction are restricted to a small area of the order of a wavelength. Therefore all incident waves in the original and the reproduced field may be considered to be plane waves.
2. Analysis and reconstruction are restricted to the horizontal plane.
3. Four loudspeakers are used for reconstruction.
4. Only the zeroth and first order spherical harmonics are taken into account. These can be measured with pressure and velocity microphones, respectively.

Only the first restriction is essential for ambisonic theory. The second and third restriction are made here for simplicity in derivation, while the fourth restriction originates from the practical circumstance that second and higher order microphones have not been constructed yet.

Consider a plane wave with an angle of incidence ψ with respect to the *x*-axis (Figure 1.18a). Pressure and velocity at the origin are given by

$$P_0 = P, \tag{1.22}$$

$$V_x(\psi) = V\cos\psi\,, \tag{1.23a}$$

$$V_y(\psi) = V\sin\psi\,, \tag{1.23b}$$

with P the pressure and V the velocity of the plane wave. Ambisonics uses a special microphone with nearly coincident diaphragms to acquire P, V_x and V_y. These signals can be encoded for transmission in so-called UHJ format (Gerzon 1985), and recovered at the reproduction site, where the loudspeaker inputs are calculated according to (Bamford et al. 1995)

$$L_n = \frac{1}{N}(P_0 - 2\rho_0 c V_x \cos\varphi_n - 2\rho_0 c V_y \sin\varphi_n) \tag{1.24}$$

with L_n the feed, and φ_n the direction of the n^{th} loudspeaker, with N the total number of loudspeakers, and $\rho_0 c$ the acoustic impedance of a plane wave. For the arrangement shown in Figure 1.18b, loudspeaker signals (1.24) are evaluated as

$$L_1 = \tfrac{1}{4}(P_0 + 2\rho_0 c V_x)\,, \tag{1.25a}$$

$$L_2 = \tfrac{1}{4}(P_0 + 2\rho_0 c V_y)\,, \tag{1.25b}$$

$$L_3 = \tfrac{1}{4}(P_0 - 2\rho_0 c V_x)\,, \tag{1.25c}$$

$$L_4 = \tfrac{1}{4}(P_0 - 2\rho_0 c V_y)\,. \tag{1.25d}$$

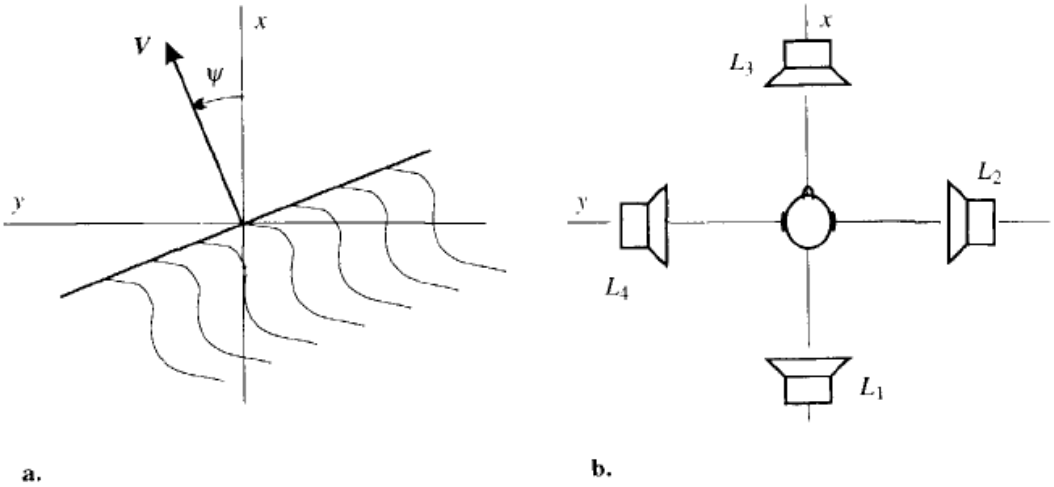


Figure 1.18 Ambisonic reproduction example. **a**. Plane wave incidence at origin. **b**. Reconstruction of the plane wave by composition with loudspeakers.

The reconstructed wave at the origin is found by superposition of the loudspeaker outputs. The pressure is given by summation of all contributions

$$P'_0 = \sum L_n = P_0 \,, \tag{1.26}$$

and the velocity by summing *x* and *y* components (note that opposite loudspeakers have opposite signs for their velocity components)

$$V'_x = \frac{1}{\rho_0 c} \sum_x L_n = \frac{1}{\rho_0 c} [L_1 + (-L_3)] = V_x \,, \tag{1.27a}$$

$$V'_y = \frac{1}{\rho_0 c} \sum_y L_n = \frac{1}{\rho_0 c} [L_2 + (-L_4)] = V_y \,. \tag{1.27b}$$

Though in theory the original pressure and velocity are fully recovered, problems may arise due to the prescription that opposite loudspeakers have anti-phased velocity information. Certainly such cancelation is very critical, and will probably require careful adjustment of the decoding circuitry and accurate positioning of loudspeakers. At high frequency, moreover, cancelation of oppositely propagating waves will not occur at all, because firstly, there is a delay between left and right waves at the listener's ears, and secondly, the head shields the wave field.

1.7.4 Surround sound in the cinema

The first surround sound project in motion picture industry has been Disney's discrete six-channel *Fantasia* in 1941. Since then, various systems have been designed, of which Dolby Stereo has been most successful. This surround sound system decodes two channels from 35 mm optical prints into left (L), center (C), right (R) and surround (S) loudspeaker channels. It is the professional version of the Dolby Surround system, as described in section 1.7.2.

In the 1990s, the digital era in cinema was entered with three non-compatible digital surround sound systems: Dolby Stereo Digital, Digital Theater Systems and Sony Dynamic Digital Sound (Blake 1995). All systems employ the so-called 5.1 format: a discrete six-channel configuration with L-C-R-S1-S2 and a subwoofer channel (Figure 1.19). For cinemas with less than five loudspeaker clusters, also a three-channel or four-channel sub-set can be decoded from the 5.1 format. On the other hand, extensions up to the 7.1 format are supported as well (see Figure 1.19).

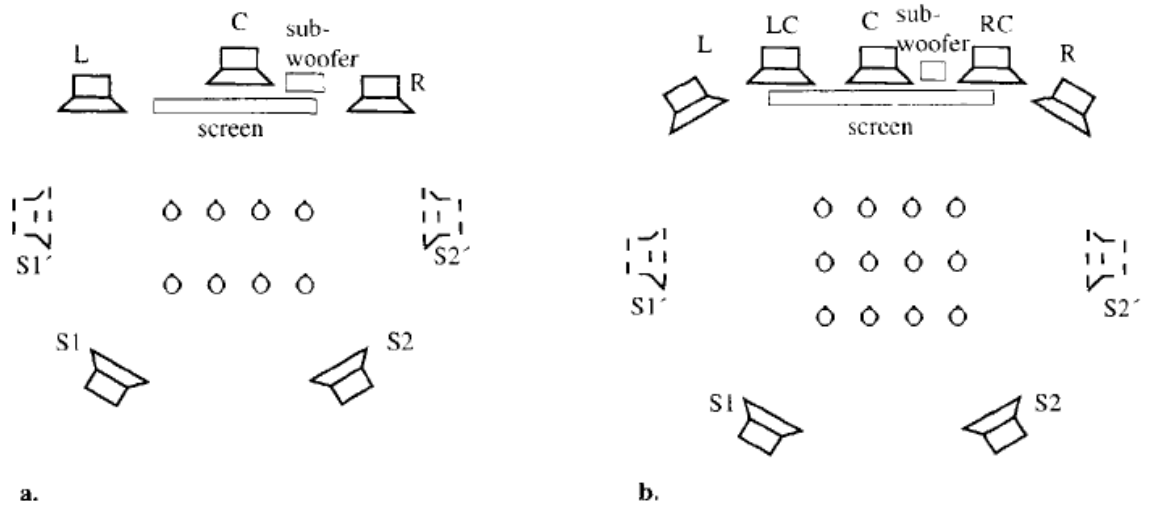


Figure 1.19 Surround sound configurations. The S1´ and S2´ surround loudspeakers are optional. **a**. The 5.1 surround sound format. **b**. The 7.1 surround sound format.

1.7.5 Discrete surround sound at home

With the arrival of advanced widescreen television, new opportunities for the audio industry were provided. A recommendation of the ITU-R (1991) committee has standardized multi-channel transmission formats and loudspeaker arrangements. The transmission format is 5.1 (five wide band channels and one narrow band subwoofer channel) with three front channels and two surround channels (denoted as 3/2) as to be compatible with motion picture formats and arrangements. Also, sub-sets of this configuration are possible, such as 3/1, as used by Dolby Surround, and 2/0, which is the conventional two-channel stereophony. Others propose the use of side surround loudspeakers, S1´ and S2´ in Figure 1.19, because the imaging of lateral sources and reflections is problematic with only front and rear loudspeakers (Griesinger 1996 and Zieglmeier et al. 1996).

Preparation of audio signals for the 3/2 system is largely an extension of stereophonic imaging. Reports on recordings for the new standard mention conventional microphone techniques. Schneider (1996) reports that a conventional AB pair could not provide a satisfactory pan pot mix for the frontal channels L, C and R. Finally, three discrete microphone signals, with 3 dB attenuation for the central channel, were used for these channels to image a symphony orchestra. Spikofski et al. (1992) propose a reversed ORTF-pair (Figure 1.20a) for the surround channels. Theile (1996) uses a similar microphone configuration for overall sound pick-up (Figure 1.20b). He argues that the center channel should be ignored as long as no adequate three-channel main microphone exists, leaving a 2/2 configuration.

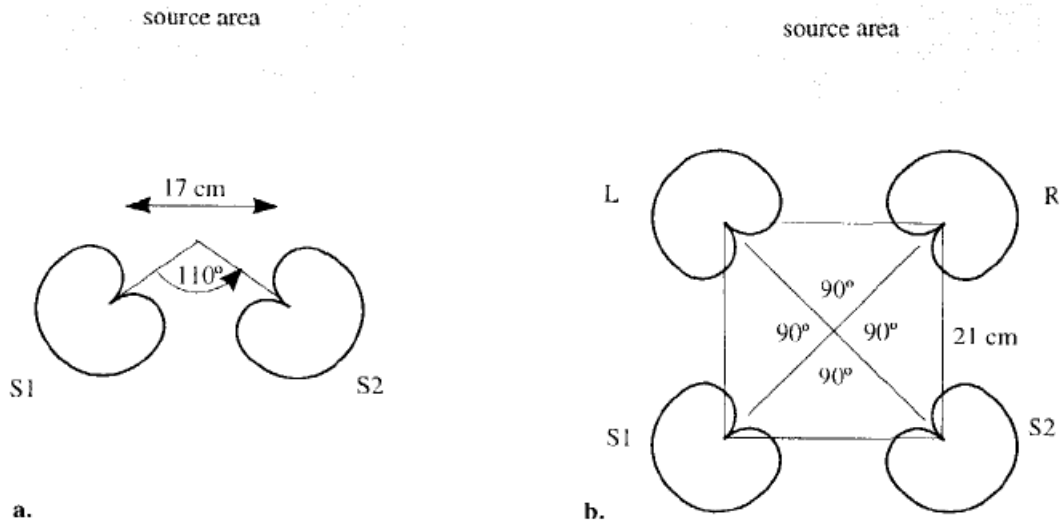


Figure 1.20 **a**. ORTF-pair used for surround sound pick-up. **b**. Modified ORTF-pairs for a 2/2 system.

1.8 Conclusions

Since the early 1970s, the audio industry is vainly trying to enrapture the public for multi-channel reproduction. The motivation for these efforts was the general idea that the listener should be enveloped with sound, as in natural sound fields. However, sound reproduction and two-channel stereophony are still synonymous to the consumer. The commercial failure of multi-channel sound so far, is regarded to be caused by moderate enhancement of sound quality and lack of standardization.

It is believed that these problems have been overcome by digital technology and a standardized surround format, the so-called 3/2 format. Though presented as an innovative solution, the new surround sound proposals still lean heavily on stereophonic imaging. Therefore, important drawbacks of stereophony will unavoidably be adopted as well.

In this chapter, a reformulation of the aims and requirements for sound reproduction has been given. It is shown that stereophony cannot fulfill such demands: imaging is restricted to an extremely narrow listening area. Outside this area, spatial and temporal distortion of the sound image occurs. The main goal of the present research is to develop a method of sound reproduction that offers a natural high quality sound image in a large listening area.

Chapter 2

Wave Field Synthesis

2.1 Introduction

This chapter gives the theoretical basis for spatially correct sound reproduction by applying the concept of wave field synthesis (Berkhout 1988 and Berkhout et al. 1993). It will be shown that – starting from the Kirchhoff-Helmholtz wave field representation – a synthesis operator can be derived that allows for a spatial and temporal reconstruction of the original wave field. The general 3D-solution (Section 2.2) can be transformed into a so-called 2½D-solution (Section 2.3), which is sufficient for reconstructing the original sound field in the (horizontal) plane of listening. For that purpose a linear array of loudspeakers is used to generate the sound field of virtual sources lying behind the array. A special manifestation of the operator appears in the case that sources in front of the array need to be synthesized. In that case, the array radiates convergent waves towards a focus point, from which divergent waves propagate into the listening area (Section 2.3.2).

For finite or corner-shaped arrays diffraction occurs. In most situations diffraction waves cause no serious distortion of the perceived sound image. However, if the main contribution of the wave at a certain listening position is generated near a discontinuity of the array, diffraction can have audible effects. Solutions for diffraction are offered in Section 2.4.

Discretization of continuous source distributions has influence on the frequency response and the spatial behavior of the synthesized wave field. Section 2.5 discusses artefacts and solutions. Finally, related research in the field of wave field synthesis is summarized in Section 2.6.

2.2 Kirchhoff-Helmholtz Integral

From Green's theorem and the wave equation, the Kirchhoff-Helmholtz integral can be derived for the pressure P_A at a point A within a closed surface S (e.g. Skudrzyk 1954 or Berkhout 1982)

$$P_A = \frac{1}{4\pi}\int_S (P\nabla G - G\nabla P)\boldsymbol{n}dS\,, \tag{2.1}$$

where G is called the Green's function, P is the pressure at the surface caused by an arbitrary source distribution outside the enclosure, and $\boldsymbol{n}$ is the inward pointing normal unit vector to the surface, see Figure 2.1. G should obey the inhomogeneous wave equation for a monopole source at position A

$$\nabla^2 G + k^2 G = -4\pi\delta(r - r_A)\,. \tag{2.2}$$

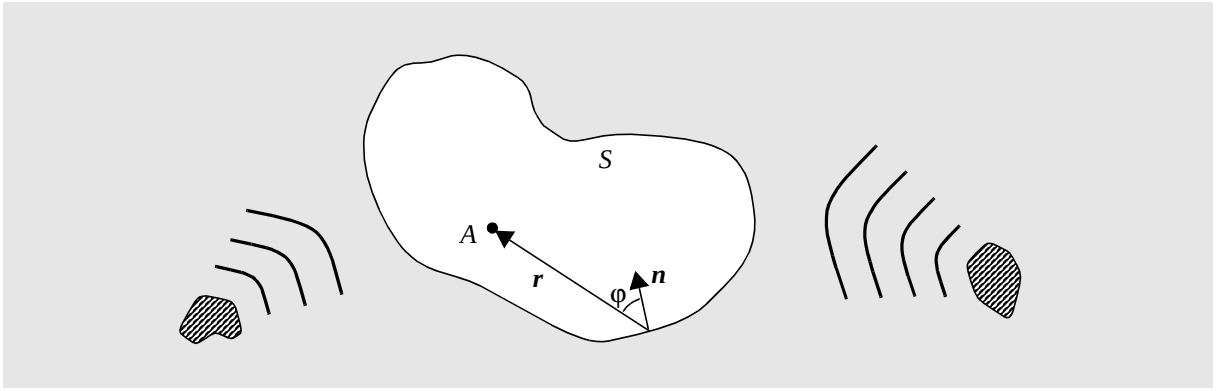


Figure 2.1 The pressure in a point A can be calculated if the wave field of an external source distribution is known at the surface of a source-free volume containing A.

The general form of the Green's function is given by

$$G = \frac{\exp(-jkr)}{r} + F\,, \tag{2.3}$$

where F may be any function satisfying the wave equation (2.2) with the right-hand term set to zero. For the derivation of the Kirchhoff integral $F = 0$ is chosen, and the space variable r is chosen with respect to A

$$r = \sqrt{(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2}\,. \tag{2.4}$$

Thus G represents the wave field of a point source in A. The physical interpretation of these choices is that a fictitious point source must be placed in A to determine the acoustical

transmission paths from the surface towards *A*. Substitution of the solution for *G* and the equation of motion

$$\frac{\partial P}{\partial n} = -j\omega\rho_0 V_n \tag{2.5}$$

into integral (2.1) finally leads to the Kirchhoff integral for homogeneous media

$$P_A = \frac{1}{4\pi}\int_S \left[\left(P\frac{1+jkr}{r^2}\cos\varphi\exp(-jkr)\right) + \left(j\omega\rho_0 V_n \frac{\exp(-jkr)}{r}\right)\right] dS\,. \tag{2.6}$$

The first term of the integrand of (2.6) represents a dipole source distribution driven with the strength of the pressure *P* at the surface, while the second term represents a monopole source distribution driven with the normal component of the particle velocity V_n at the surface (Figure 2.2). The original source distribution may be called the *primary* source distribution; the monopole and dipole sources may be called the *secondary* source distribution.

Since *A* can be anywhere within the volume enclosed by *S*, the wave field within that

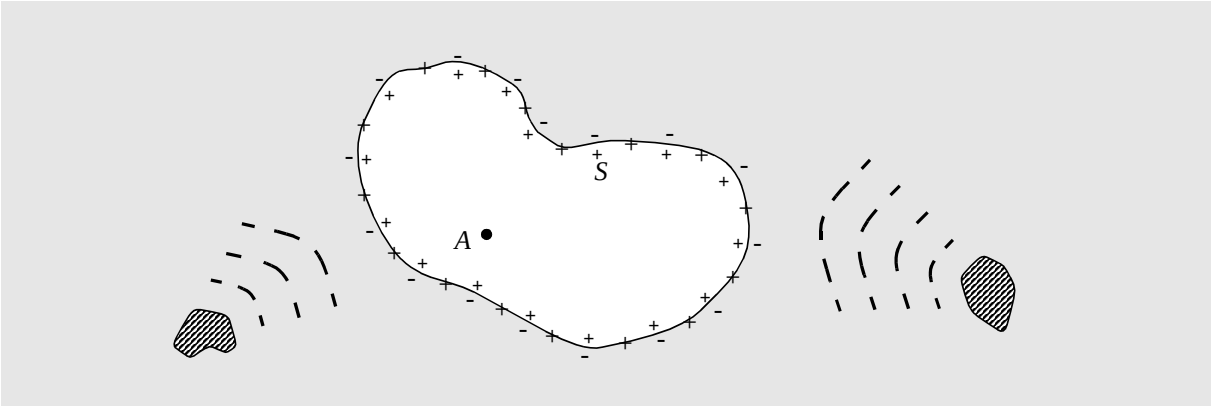


Figure 2.2 The pressure in a point *A* can be synthesized by a monopole and dipole source distribution at a surface *S*. The strength of the distributions depends on the velocity and pressure of external sources measured at the surface.

volume is completely determined by (2.6), whereas the integral is identically zero outside the enclosure. This can be understood in an intuitive manner by considering that the positive lobes of the dipoles interfere constructively with the monopoles *inside* the surface, while the negative lobes of the dipoles exactly cancel the single positive lobe of the monopoles (see also Figure 1.12) *outside* the surface. Apparently, part of the complexity of the Kirchhoff expression is involved in canceling the wave field of the secondary sources outside the closed surface. For the application of the Kirchhoff integral in wave field synthesis, this cancelation property is of less importance. Therefore, two special solutions of the Green's function are chosen that will considerably simplify the Kirchhoff integral, be it at the cost of a fixed surface geometry and a

non-zero wave field outside the closed surface. Of course, the wave field *inside* the closed surface is correctly described by these solutions, known as the Rayleigh I and II integrals.

2.2.1 Rayleigh I integral

The Rayleigh I integral can be found by choosing a particular surface of integration for integral (2.1) and a suitable function F in (2.3). The surface consists of a plane at $z = 0$ and a hemisphere in the half space $z > 0$, as drawn in Figure 2.3. All sources are located in the half space $z < 0$, so for any value of the radius R the volume enclosed by S_1 and S_2 is source-free. If $G(r)$ falls off faster than or proportional to $1/r$, the Sommerfeld condition is satisfied and the integral over this part of the surface vanishes for $R \to \infty$. The pressure in A is now found by substitution of (2.3) in integral (2.1):

$$P_A = \frac{1}{4\pi}\int_{S_1}\left(P\frac{\partial}{\partial n}\left[\frac{\exp(-jkr)}{r}+F\right]-\left[\frac{\exp(-jkr)}{r}+F\right]\frac{\partial P}{\partial n}\right)dS\,, \tag{2.7}$$

where S_1 is the infinite plane surface at $z = 0$.

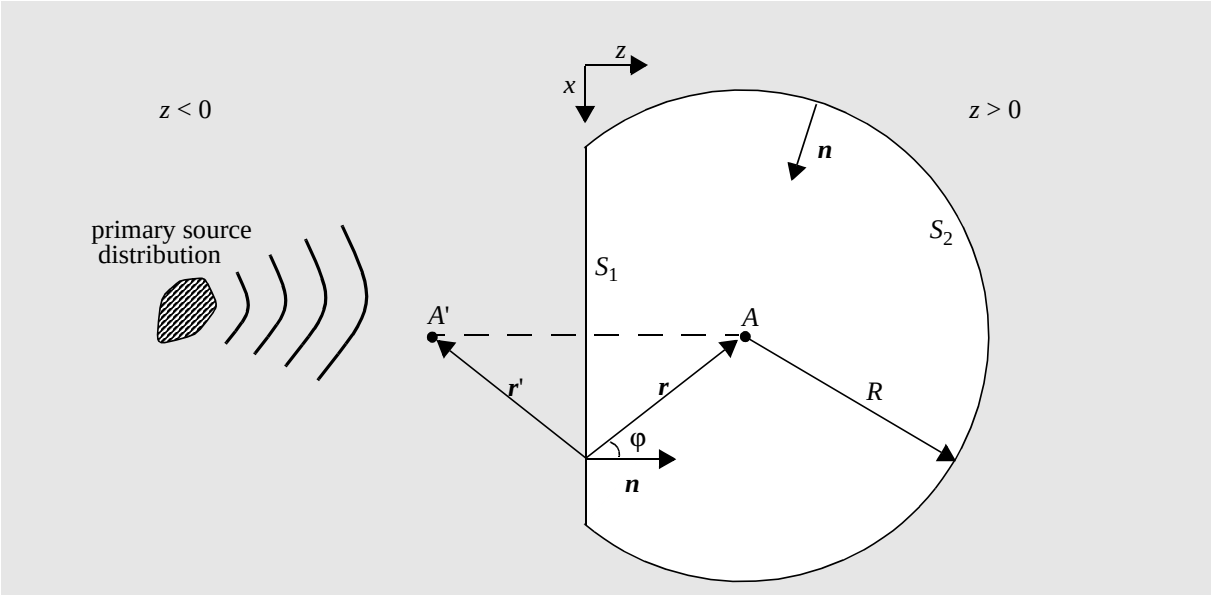


Figure 2.3 The integral for the Rayleigh representations runs over a hemisphere surface S_2 and a plane surface S_1.

The aim is to choose a function F that causes the first term of the integrand to vanish. For that to be true, the normal component of the gradient of F should have a sign opposite to that of the normal component of the gradient of $\exp(-jkr)/r$. This will be the case for

$$F = \frac{\exp(-jkr')}{r'} \tag{2.8}$$

which is the pressure of a monopole at A', the image of A mirrored in the plane S_1. On S_1, $r' = r$

applies and therefore

$$\frac{\partial F}{\partial n} = \frac{\partial}{\partial r'}\left[\frac{\exp(-jkr')}{r'}\right]\frac{\partial r'}{\partial n}$$

$$= \frac{\partial}{\partial r}\left[\frac{\exp(-jkr)}{r}\right]\frac{-\partial r}{\partial n}$$

$$= -\frac{\partial}{\partial n}\left[\frac{\exp(-jkr)}{r}\right], \tag{2.9}$$

which causes the first term of the integrand of (2.7) to vanish indeed, yielding the Rayleigh I integral after substitution of (2.5)

$$P_A = \frac{1}{2\pi}\int_{S_1} j\omega\rho_0 V_n\left[\frac{\exp(-jkr)}{r}\right]dS\,. \tag{2.10}$$

This equation states that a monopole distribution in the plane $z = 0$, driven by two times the strength of the particle velocity components perpendicular to the surface, can synthesize the wave field in the half space $z > 0$ of a primary source distribution located somewhere in the other half space $z < 0$ (see Figure 2.4).

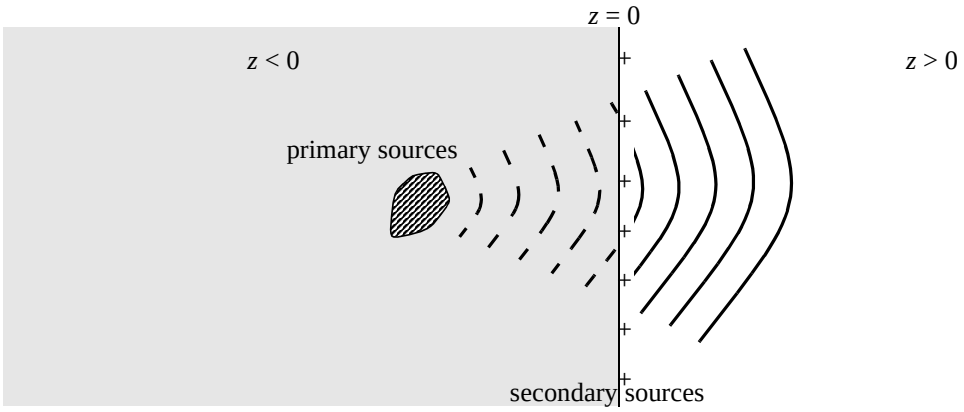


Figure 2.4 With the aid of the Rayleigh I integral the wave field of a primary source distribution can be synthesized by a monopole distribution at $z = 0$ if the velocity at that plane is known.

Note that such a distribution of monopoles will radiate a ‘mirror’ wave field into the half space $z < 0$, since there are no dipoles to cancel this wave field. Therefore the Rayleigh I solution can be applied for synthesis if the wave field in the half space $z < 0$ is of no interest.

2.2.2 Rayleigh II integral

In order to find the Rayleigh II solution, a function F is required that will eliminate the second term of (2.7). The situation as depicted in Figure 2.3 is considered again. A choice

$$F = -\frac{\exp(-jkr')}{r'} \tag{2.11}$$

will obviously have the desired effect, because at the plane of integration $r' = r$ applies, causing the second term of (2.7) to disappear. This time, G represents the wave field of a monopole at A, together with an anti-phase monopole at A'. The Rayleigh II integral can now be derived as

$$P_A = \frac{1}{2\pi}\int_{S_1} P\frac{1+jkr}{r^2}\cos\varphi\exp(-jkr)dS\,. \tag{2.12}$$

This expression states that the wave field in the half space $z > 0$ can be synthesized by a distribution of dipoles, driven by two times the strength of the pressure (measured at $z = 0$) of a source distribution in the half space $z < 0$ (Figure 2.5). Note that such a distribution of dipoles will radiate a 'mirror' wave field into the half space $z < 0$, the pressure of which is in anti-phase with the pressure at $z > 0$.

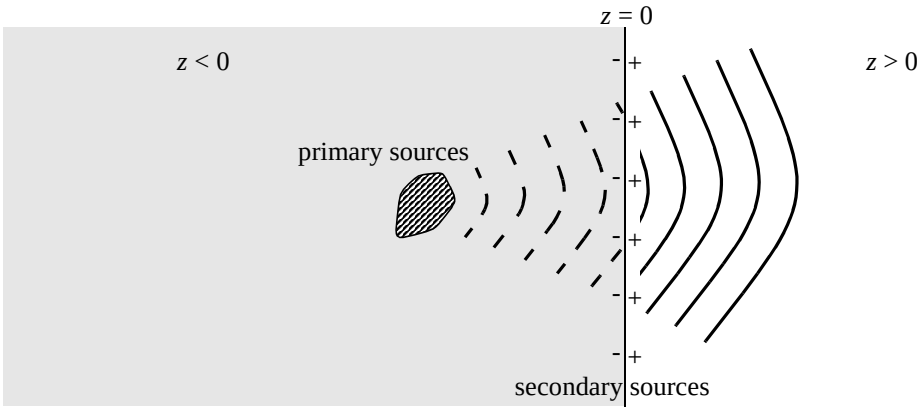


Figure 2.5 With the aid of the Rayleigh II integral the wave field of a primary source distribution can be synthesized by a dipole distribution at $z = 0$ if the pressure at that plane is known.

2.3 Synthesis operator for line distributions

In principle, the Rayleigh I and II integrals can be used to synthesize the wave field of any primary source distribution by a planar array of loudspeakers. The discretization of the operators, from a continuous *distribution* of secondary sources to a discrete *array* of loudspeakers, will be treated in Section 2.5. In this section an intermediate step is made, from pla-

nar to linear distributions. Such a reduction of dimensions is allowed as long as the requirements for spatial sound reproduction are being respected (Section 1.5).

A derivation of the Rayleigh I and II integrals in two dimensions (abbreviated: 2D) is given by Berkhout (1982). Analogously to the 3D situation, the wave field of a primary source distribution can be represented by a distribution of secondary sources at a straight line. However, the 2D analogy of a secondary point source behaves as a line source in 3D, i.e. instead of the $1/r$-law, a $1/\sqrt{r}$-law is followed. Such solutions are of little practical value for wave field synthesis. Therefore, Vogel (1993) has proposed a different approach, resulting in an expression for a so-called 2½D-operator. This operator can be regarded as an intermediate form between the 3D and 2D Rayleigh operators, because the secondary sources have 1/*r*-attenuation while the derivation is given in a plane.

De Vries (1996) assumed monopole characteristics for the primary source and incorporated arbitrary directivity characteristics of the secondary sources (loudspeakers) in the synthesis operator. He showed that the characteristics of the secondary sources and the primary source are inversely interchangeable. Furthermore, his simulations show that if the directivity patterns of one or more array loudspeakers deviate from the others, serious synthesis errors may occur. Therefore all secondary sources should have identical characteristics. The derivation given here starts off with arbitrary characteristics for the primary source, and monopole characteristics for the secondary sources. It follows the treatment given by Start (1996), who derived an expression for the 2½D-operator by reducing the (exact) 3D Rayleigh I surface integral to a line integral.

The position of the primary source is not restricted to the half plane behind the secondary line source, as will be shown in Section 2.3.2. The presence of sources in front of the loudspeaker arrays may contribute strongly to the imagination of the listener.

A summary and generalization of 2½D-operators is given in Section 2.3.3.

2.3.1 2½D Synthesis operator

Consider a primary source *S* in the *xz*-plane, as drawn in Figure 2.6. The pressure field of the primary source will be given by

$$P(\boldsymbol{r}, \omega) = S(\omega) G(\varphi, \vartheta, \omega) \frac{\exp(-jkr)}{r}, \tag{2.13}$$

where $S(\omega)$ is the source function and $G(\varphi, \vartheta, \omega)$ is the directivity characteristic of the source in the far-field approximation ($kr >> 1$). Across the *xy*-plane, a secondary source distribution with monopole characteristics is situated, which will synthesize the wave field of the primary source at the receiver position *R* according to the Rayleigh I integral (2.10)

$$P_{\text{synth}} = \frac{1}{2\pi} \int_{xy\text{-plane}} j\omega\rho_0 V_n(\boldsymbol{r}, \omega) \left[\frac{\exp(-jk\Delta r)}{\Delta r} \right] dx\,dy \tag{2.14}$$

with $V_n(\boldsymbol{r}, \omega)$ the velocity component of the primary source perpendicular to the *xy*-plane. The

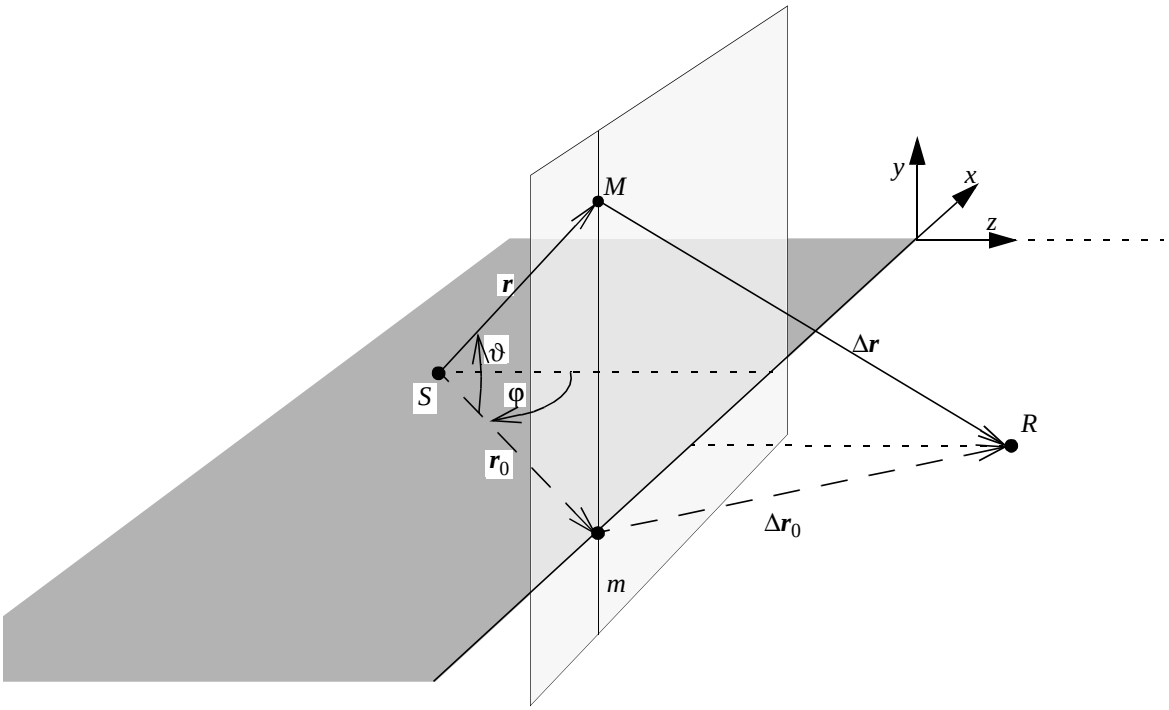


Figure 2.6 The Rayleigh I integral is evaluated for a primary source *S* and receiver *R*, both located in the *xz*-plane. Vector ***r*** points from the primary source to a secondary source *M* at the line *m*. Vector Δ***r*** points from secondary source to receiver. The projections on the *xz*-plane are denoted by subscript 0.

surface integral will be reduced to a line integral along the *x*-axis by evaluating the integral in the *y* direction along the line *m*. By applying the equation of motion (2.5) to Equation (2.13) it follows that

$$j\omega\rho_0 V_n(\boldsymbol{r}, \omega) = -\frac{\partial}{\partial n} P = -S(\omega)\frac{\partial}{\partial z}\left[G(\varphi, \vartheta, \omega)\frac{\exp(-jkr)}{r}\right]. \tag{2.15}$$

In the chosen geometry of Figure 2.6, with $\varphi = \operatorname{atan}(x/z)$ and $\vartheta = \operatorname{atan}(y/\sqrt{x^2 + z^2})$, this expression is written as

$$j\omega\rho_0 V_n(\boldsymbol{r}, \omega) = S(\omega)\frac{\exp(-jkr)}{r}\left\{\frac{\sin\varphi}{r\cos\vartheta}\frac{\partial G}{\partial\varphi} + \frac{\cos\varphi\sin\vartheta}{r}\frac{\partial G}{\partial\vartheta} + \frac{1 + jkr}{r}G\cos\varphi\cos\vartheta\right\}. \tag{2.16}$$

The integration along the line *m* can be carried out by applying the stationary phase method (Bleistein 1984). This method is concerned with integrals of the form

$$I = \int_{-\infty}^{\infty} f(y)\exp(j\phi(y))dy\,. \tag{2.17}$$

Equation (2.14), with substitution of (2.16), indeed takes this form. The amplitude part is

$$f(y) = \frac{1}{2\pi}\frac{S(\omega)}{r\Delta r}\left\{\frac{\sin\varphi}{r\cos\vartheta}\frac{\partial G}{\partial\varphi} + \frac{\cos\varphi\sin\vartheta}{r}\frac{\partial G}{\partial\vartheta} + \frac{1+jkr}{r}G\cos\varphi\cos\vartheta\right\}, \tag{2.18a}$$

in which the variables ϑ, r and Δr implicitly depend on y, while the phase is given by

$$\phi(y) = -k(r+\Delta r)\,. \tag{2.18b}$$

For $\phi >> 1$, the approximate solution to (2.17) is

$$I \approx f(y_0)\exp(j\phi(y_0))\sqrt{\frac{2\pi j}{\phi''(y_0)}}\,, \tag{2.19}$$

where y_0 is the stationary phase point and $\phi''(y_0)$ the second derivative of ϕ evaluated at y_0. The stationary phase point is defined as the point where $\phi(y)$ reaches its stationary value, i.e. where $\phi'(y) = 0$. From inspection of Figure 2.6 it follows that

$$y_0 = 0\,, \tag{2.20a}$$

(and therefore $\vartheta_0 = 0$),

$$\phi(y_0) = -k(r_0+\Delta r_0)\,, \tag{2.20b}$$

$$\phi''(y_0) = -k\frac{r_0+\Delta r_0}{r_0\Delta r_0}\,, \tag{2.20c}$$

$$f(y_0) = \frac{1}{2\pi}\frac{S(\omega)}{r_0\Delta r_0}\left\{\frac{\sin\varphi}{r_0}\frac{\partial G(\varphi,0,\omega)}{\partial\varphi} + \frac{1+jkr_0}{r_0}G(\varphi,0,\omega)\cos\varphi\right\}. \tag{2.20d}$$

For $kr_0 >> 1$, the second term of $f(y_0)$ is dominating, and therefore (2.14) is finally reduced to

$$P_{\text{synth}} = S(\omega)\sqrt{\frac{jk}{2\pi}}\int_{-\infty}^{\infty}\sqrt{\frac{\Delta r_0}{r_0+\Delta r_0}}G(\varphi,0,\omega)\cos\varphi\frac{\exp(-jkr_0)}{\sqrt{r_0}}\frac{\exp(-jk\Delta r_0)}{\Delta r_0}dx\,. \tag{2.21}$$

Integral (2.21) states that the contributions of all secondary sources M along the line m can be approximated by the contribution of a secondary point source at the intersection of m with the x-axis. The driving function $Q_{\text{m}}(x,\omega)$ of that secondary monopole point source at x is

$$Q_{\text{m}}(x,\omega) = S(\omega)\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{\Delta r}{r+\Delta r}}G(\varphi,0,\omega)\cos\varphi\frac{\exp(-jkr)}{\sqrt{r}}\,, \tag{2.22a}$$

yielding for the integral (2.21)

$$P_{\text{synth}} = \int_{-\infty}^{\infty} Q_{\text{m}}(x, \omega) \frac{\exp(-jk\Delta r)}{\Delta r} dx\,, \tag{2.22b}$$

in which the subscripts of the variables r_0 and Δr_0 have been dropped for convenience.

It will be verified here that the obtained driving function is valid within the approximations of the stationary phase method. Therefore this method is applied to integral (2.22b). This time, the amplitude $f(x)$ and phase $\phi(x)$ function are

$$f(x) = \frac{S(\omega)}{\Delta r \sqrt{r}} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta r}{r + \Delta r}} G(\varphi, 0, \omega) \cos\varphi \tag{2.23a}$$

and

$$\phi(x) = -k(r + \Delta r) \tag{2.23b}$$

respectively, in which φ, r and Δr are implicitly dependent on x. Again a stationary phase point has to be found. From Figure 2.7 it is obvious that it lies at the intersection of the line RS with the x-axis. Introducing ρ as the vector pointing from primary source to stationary phase point, the second derivative of the phase function of the integrand of (2.21) is found as

$$\phi''(x_0) = -k \frac{z_0^{\,2}}{\rho^3} \left(\frac{z_0 + \Delta z_0}{\Delta z_0} \right), \tag{2.24}$$

where z_0 and Δz_0 are the respective *distances* (positive) of the primary source and the receiver to the x-axis. The amplitude part of the integrand, evaluated at the stationary phase point, yields

$$f(x_0) = \frac{S(\omega)}{\Delta\rho \sqrt{\rho}} \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta z_0}{z_0 + \Delta z_0}} G(\varphi_0, 0, \omega) \cos\varphi_0\,. \tag{2.25}$$

Realizing that $\cos\varphi_0 = (z_0 + \Delta z_0)/(\rho + \Delta\rho)$, substitution of (2.24) and (2.25) into (2.19) gives

$$P_{\text{synth}} = S(\omega) G(\varphi_0, 0, \omega) \frac{\exp(-jk(\rho + \Delta\rho))}{(\rho + \Delta\rho)}\,. \tag{2.26}$$

Comparison of this expression with the true wave field (2.13) shows that the synthesized wave field is equal to the true wave field of the primary source within the plane containing this source and the distribution of secondary sources. Now consider the situation of different receiver points

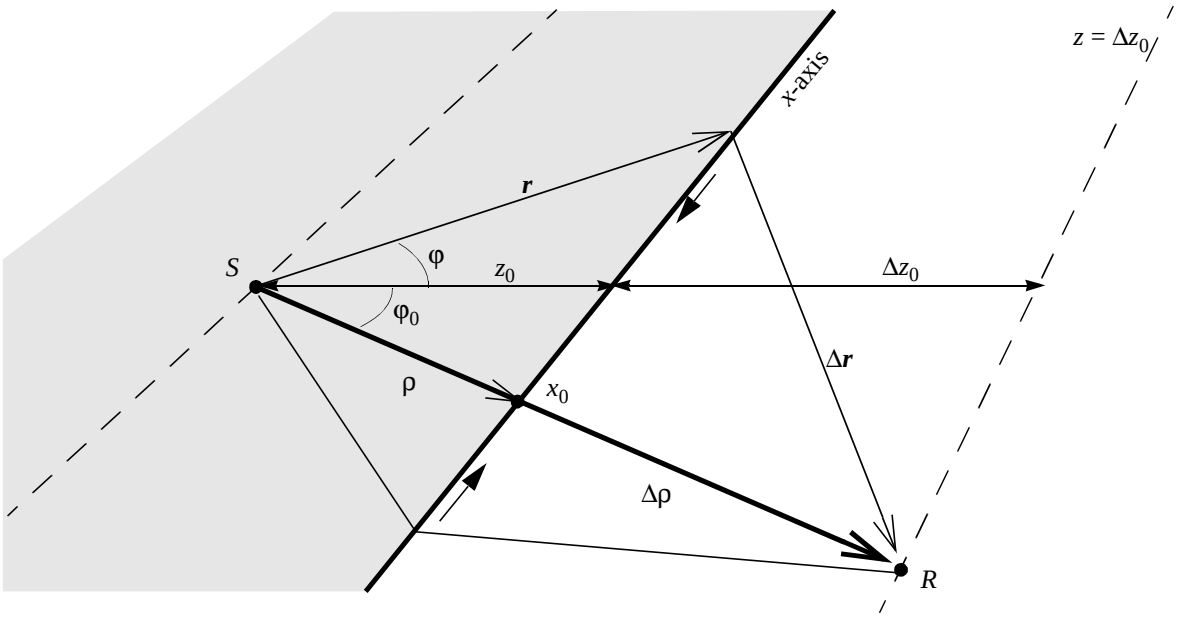


Figure 2.7 The phase function $-k(r + \Delta r)$ approaches its stationary value if $x \dashrightarrow x_0$. The value of x_0 is found at the intersection of the line *RS* with the *x*-axis.

which all are on the line $z = \Delta z_0$ as shown in Figure 2.7. By choosing the driving function

$$Q_{\mathrm{m}}(x, \omega) = S(\omega)\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{\Delta z_0}{z_0 + \Delta z_0}}G(\varphi, 0, \omega)\cos\varphi\frac{\exp(-jkr)}{\sqrt{r}}, \tag{2.27}$$

where the factor $\sqrt{\Delta r/(r + \Delta r)}$ in Equation (2.22a) has been fixed to $\sqrt{\Delta z_0/(z_0 + \Delta z_0)}$, the synthesized wave field is still correct for all the receivers at the line $z = \Delta z_0$. This adjustment yields a driving function that is independent of the receiver position at the line $z = \Delta z_0$, and is therefore practically realizable in a wave field synthesis system. Notice that the amplitude of the synthesized wave field is not correct at other receiver distances.

This effect is investigated in a simulation. Figure 2.8 shows the wave field of a monopole source, i.e. $G(\varphi) = 1$. The source signal is a repetitive broadband Gaussian pulse (shown in the upper left corner). Figure 2.9 shows the wave field synthesized by a distribution of secondary sources with monopole characteristics, the driving function being given by (2.27). Vertical cross-sections of the synthesized (dotted) and original wave field are shown in Figure 2.10. At the line $z = \Delta z_0$ (at $z = 2.5$ m in the cross-section) the synthesized wave field is correct in amplitude and phase, while elsewhere it is correct in phase and slightly deviating in amplitude in comparison to the true wave field of the primary source. This deviation is a consequence of reducing the Rayleigh-operator from planar to linear, yielding a gain factor $\sqrt{\Delta z_0/(z_0 + \Delta z_0)}$ that gives the correct amplitude at one receiver line: the pre-set *reference line*. By choosing an appropriate distance for the reference line, the amplitude errors can be kept small in a large listening area.

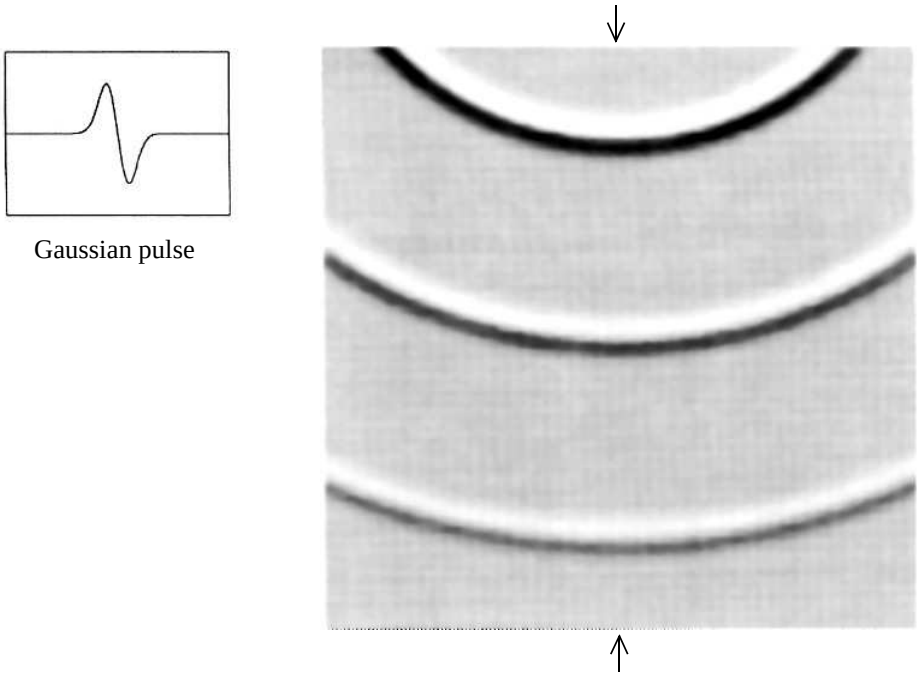


Figure 2.8 Wave field of a pulsating monopole source. The arrows indicate the position of the cross-section shown in Figure 2.10.

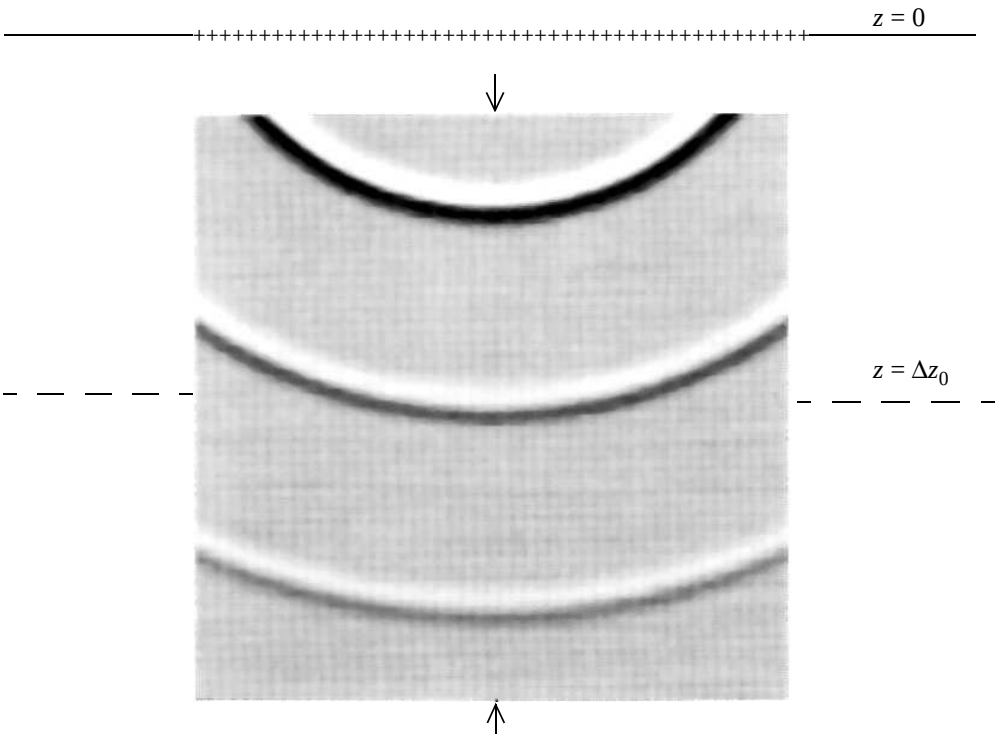


Figure 2.9 A line distribution of secondary monopole sources synthesizes the wave field of a pulsating monopole source.The arrows indicate the position of the cross-section shown in Figure 2.10.

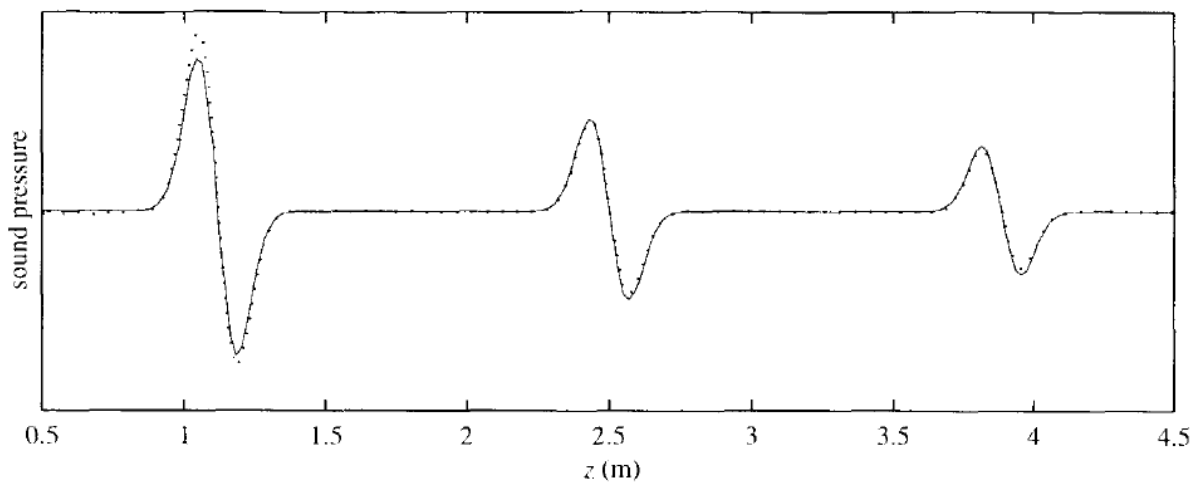


Figure 2.10 The sound pressure along the central cross-sections of Figure 2.8 and 2.9. The solid line represents the pressure of the original primary source (position: $z = -1$ m). The dotted line shows the synthesized pressure of the secondary sources. The driving function is tuned at $\Delta z_0 = 2.5$ m, so that the middle synthesized pulse matches the original one closely. The synthesized pulse near $z = 1.1$ m is about 1.5 dB stronger than the original one, while the synthesized pulse at $z = 3.9$ m is 0.5 dB weaker than the pulse of the primary source.

A second example, shown in Figure 2.11, demonstrates the synthesized wave field for a primary source with non-trivial directivity function $G(\varphi) = \cos(\varphi - 90^\circ)$, which is a dipole source with its zero pointing downwards (see also Figure 1.12b). The source signal is a sinusoid. Clearly, since the left and right lobes of the dipole have opposite phase, the alternation of dark and light zones is reversed between left and right half of the picture. A distribution of secondary sources in Figure 2.12 is used to synthesize the wave field of this primary source showing good comparison with Figure 2.11.

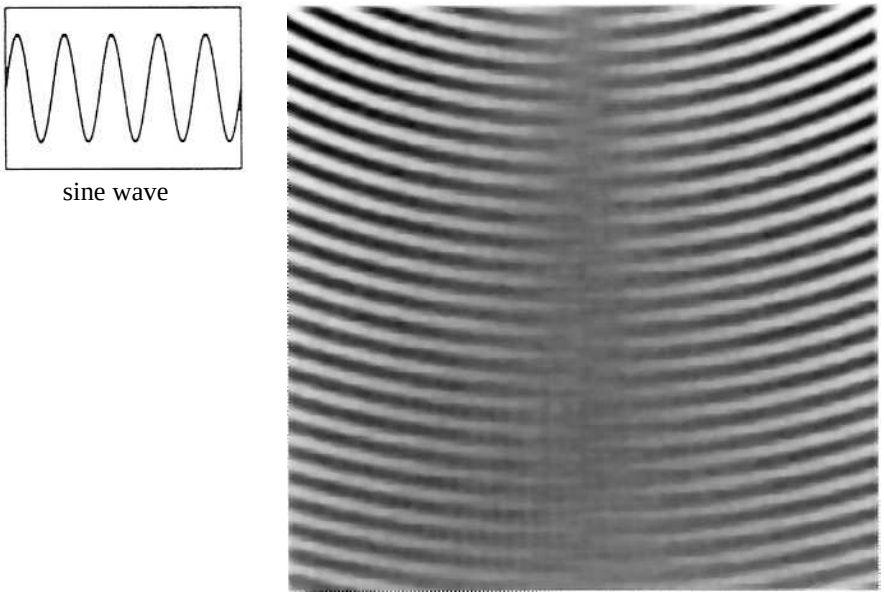


Figure 2.11 Wave field of a monochromatic (single frequency) dipole source with directivity characteristics $G(\varphi) = \cos(\varphi - 90^{o})$.

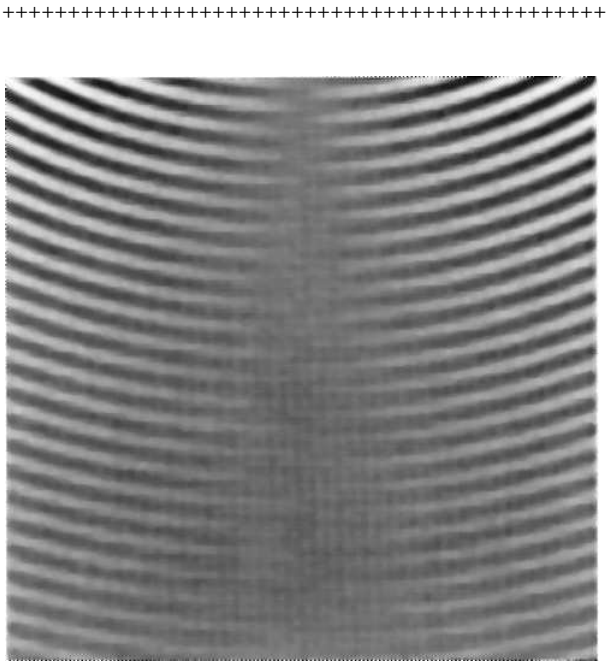


Figure 2.12 A line distribution of secondary monopole (+++) sources synthesizes the wave field of a monochromatic dipole source. The primary source has directivity characteristics $G(\varphi) = \cos(\varphi - 90^{o})$.

Instead of a distribution of *monopole* sources, also *dipole* sources can be used to synthesized a wave field. Starting from the Rayleigh II formulation (2.12), it can be shown analogously to the derivation of the monopole driving function that in this case the synthesized wave field is equal to

$$P_{\text{synth}} = \int_{-\infty}^{\infty} Q_{\text{d}}(x, \omega) jk \cos\psi \frac{\exp(-jk\Delta r)}{\Delta r} dx \tag{2.28a}$$

where ψ is the angle between the normal $\boldsymbol{n}$ (to the secondary source line) and the direction of the receiver (see Figure 2.13). Q_{d} is the dipole driving function given by

$$Q_{\text{d}}(x, \omega) = S(\omega) \sqrt{\frac{1}{2\pi jk}} \sqrt{\frac{\Delta z_0}{z_0 + \Delta z_0}} G(\varphi, 0, \omega) \frac{\exp(-jkr)}{\sqrt{r}}, \tag{2.28b}$$

where z_0 and Δz_0 are the respective distances of S and R to the x-axis.

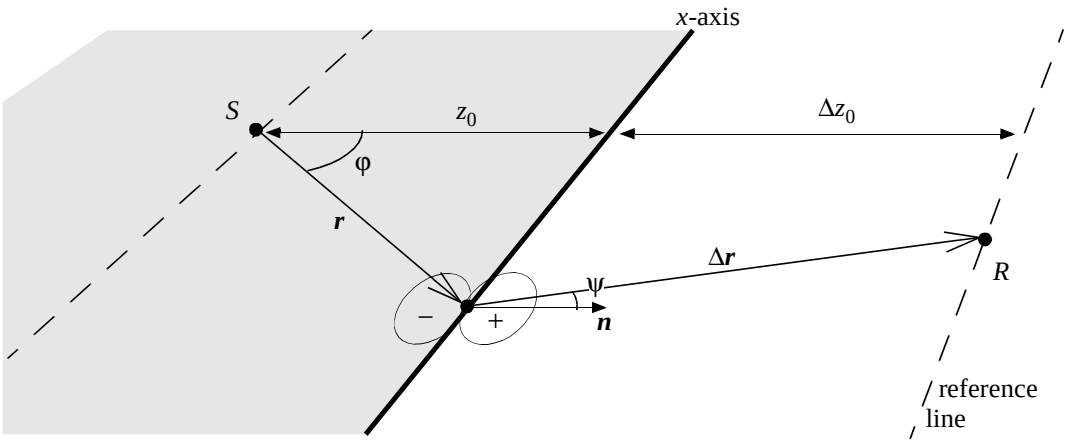


Figure 2.13 The wave field of a primary source can be synthesized by a secondary dipole source distribution along the x-axis.

2.3.2 Focusing operator

From the principle of reciprocity between source and receiver, which states that the acoustic transmission paths are reversible, it is obvious that it must be possible to focus sound waves from a distribution of secondary sources towards a focus point. Focusing of sound is applied in seismic data processing (Berkhout 1997b and Thorbecke 1997) and in ultrasonic scanning of human tissues. Focusing of sound also appears in nature: in outdoor sound propagation, so-called caustic surfaces may occur if sound waves are affected by wind gradients (Pierce 1991).

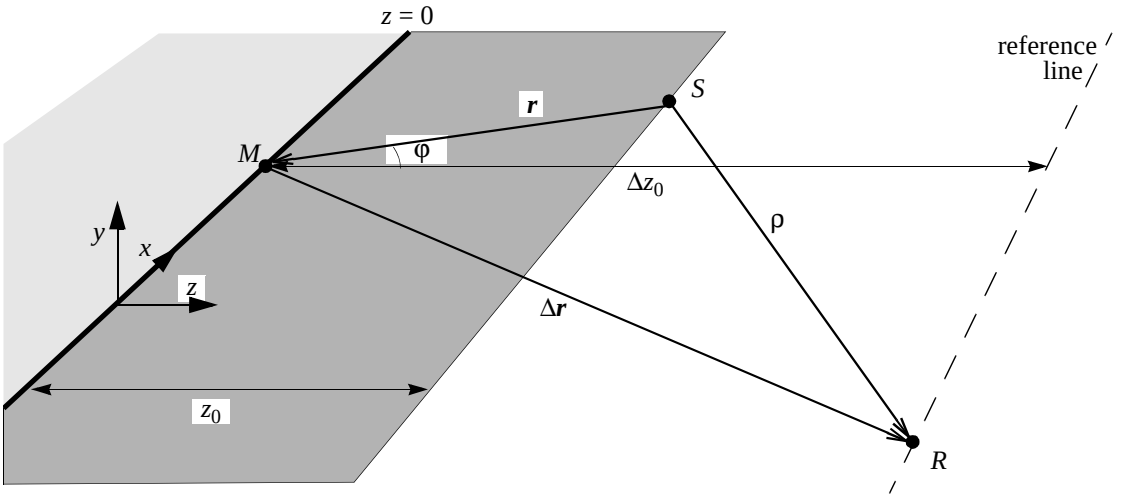


Figure 2.14 The wave field of a primary source *S* is synthesized at the reference line by secondary sources at $z = 0$.

A 2½D focusing operator can be derived for a secondary source distribution along a straight line. From inspection of the geometry of Figure 2.14, with the secondary sources at $z = 0$ and both the primary source and the receiver line at $z > 0$, it is clear that the waves of the primary source are propagating in the direction opposite to the waves of the secondary sources in the dark area where $0 < z < z_0$. In the space $z > z_0$, the waves of the primary source *and* the waves of the secondary sources are traveling in positive z-direction. In that area, a reconstruction of the primary wave field by a distribution of secondary monopole or dipole sources is possible, as shown in Appendix A. The synthesized wave field for a distribution of secondary monopole sources is given by (A.3)

$$P_{\text{synth}} = \int_{-\infty}^{\infty} Q_{\text{m}}^{\text{foc}}(x, \omega) \frac{\exp(-jk\Delta r)}{\Delta r} dx \tag{2.29a}$$

with driving function (A.15)

$$Q_{\text{m}}^{\text{foc}}(x, \omega) = S(\omega) \sqrt{\frac{k}{2\pi j}} \sqrt{\frac{\Delta z_0}{\Delta z_0 - z_0}} \cos\varphi \frac{\exp(+jkr)}{\sqrt{r}} \quad . \tag{2.29b}$$

Figure 2.15 displays the wave field of a focusing secondary source distribution. In the area $z < z_0$, the waves are convergent towards the focus (concave wave fronts). Figure 2.16 shows a cross-section through the focus. Note the change of the waveform at the focus, which is quantified by Equation (A.22): the phase of the waves is shifted by 90° when passing the focus.

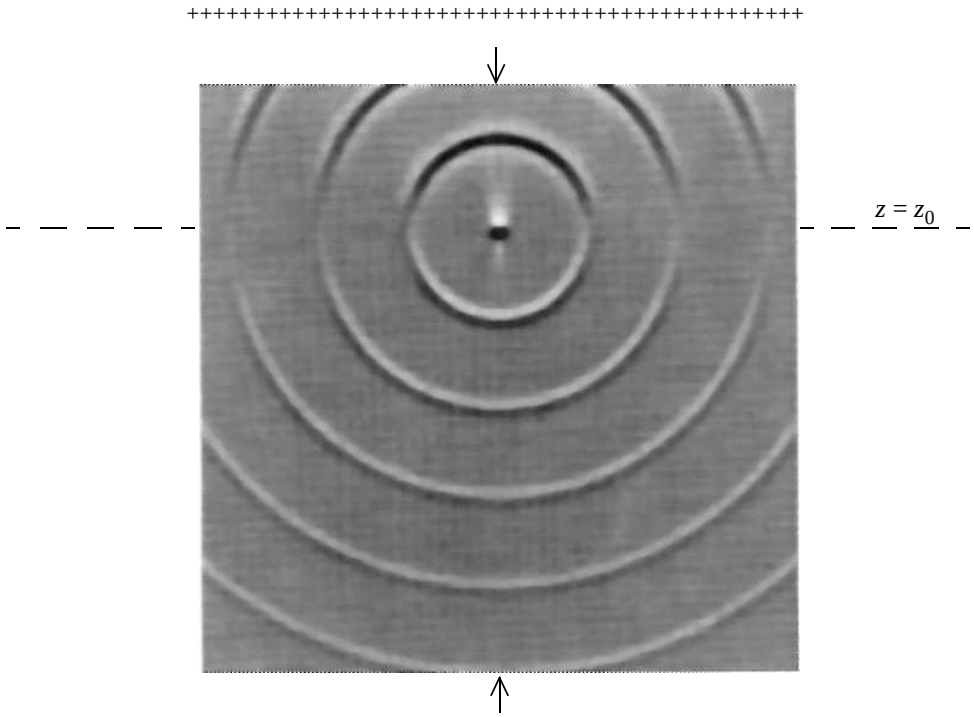


Figure 2.15 Wave field of a focusing monopole distribution. The source signal is a repetitive pulse. The arrows indicate the position of the cross-section shown in Figure 2.16.

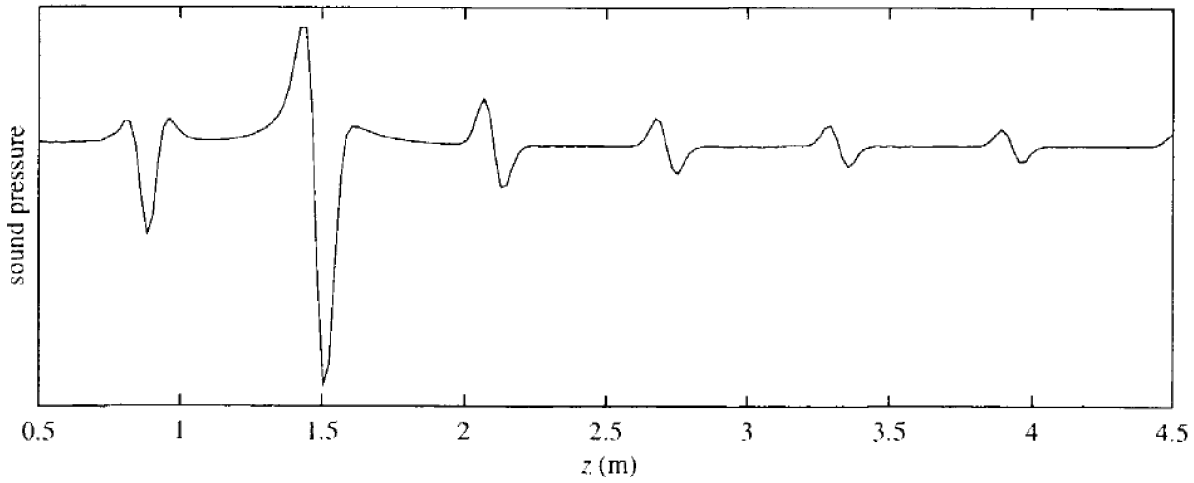


Figure 2.16 Cross-section of the wave field of Figure 2.15. The position of the primary source (focus) is at z_0 = 1.5 m.

2.3.3 Generalization of 2½D-operators

A. Focusing and non-focusing generalization

The non-focusing operator (2.27) with $G(\varphi) = 1$ bears strong resemblance to the focusing operator (A.15). They can be combined in one equation by use of the sign function:

$$Q_{\mathrm{m}}^{\mathrm{gen}}(x, \omega) = S(\omega)\sqrt{\frac{\mathrm{sign}(\zeta)k}{2\pi j}}\sqrt{\frac{\zeta}{\zeta - 1}}\cos\varphi\frac{\exp(\mathrm{sign}(\zeta)jkr)}{\sqrt{r}}, \tag{2.30}$$

where $\zeta = z_R / z_S$, the ratio between the respective (signed) z-coordinates of the reference line and the primary source*, for a line of secondary *monopole* sources situated at $z = 0$. Note that ζ is positive for the focusing operator and negative for the non-focusing operator. Also, ζ is bounded, i.e. $0 \le \zeta \le 1$ is inhibited, because for the focusing operator the primary source must lie between the secondary sources and the receiver line .

Similarly, a driving function for a secondary *dipole* source line can be found by combination of (2.28b), with $G(\varphi) = 1$, and (A.17) that holds for a primary monopole source on the same or other side of the secondary source line at $z = 0$:

$$Q_{\mathrm{d}}^{\mathrm{gen}}(x, \omega) = \frac{S(\omega)}{j}\sqrt{\frac{\mathrm{sign}(\zeta)}{2\pi jk}}\sqrt{\frac{\zeta}{\zeta - 1}}\frac{\exp(\mathrm{sign}(\zeta)jkr)}{\sqrt{r}} \tag{2.31}$$

with the same considerations for $\zeta = z_R / z_S$ as for the secondary monopole sources.

B. Arbitrary shape of secondary source line and receiver line

Start (1996, 1997) generalized the non-focusing 2½D-operator for an arbitrary shape of the secondary source line and receiver line (Figure 2.17). The pressure at the receiver line is given by

$$P_R = \int_{-\infty}^{\infty} Q_{\mathrm{m}}(\boldsymbol{r}, \omega)\frac{\exp(-jk\Delta r)}{\Delta r}dl, \tag{2.32a}$$

where the integration path is along the secondary source line. Q_{m} is the driving function

$$Q_{\mathrm{m}}(\boldsymbol{r}, \omega) = S(\omega)\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{\Delta r}{r + \Delta r}}\cos\varphi\frac{\exp(-jkr)}{\sqrt{r}}, \tag{2.32b}$$

where r is the length of the vector $\boldsymbol{r}$ from the primary source to the secondary sources, φ is the angle of incidence of $\boldsymbol{r}$ at the secondary source line, and Δr is the length of the vector $\Delta\boldsymbol{r}$ from secondary sources to reference line, pointing in the same direction as $\boldsymbol{r}$. This driving function allows a flexible configuration of synthesis arrays.

* Hence $z_R = +\Delta z_0$ and $z_S = +z_0$ in Figure 2.14, while $z_R = +\Delta z_0$ and $z_S = -z_0$ in Figure 2.7.

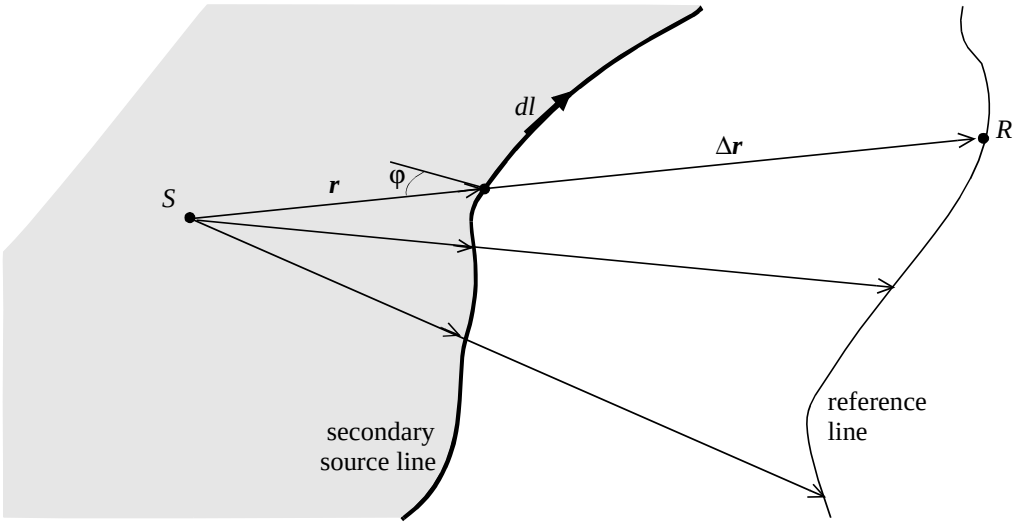


Figure 2.17 Synthesis with a curved secondary source line and reference line is allowed as long as each point R on the reference line 'sees' only one stationary phase point on the secondary source line.

In case of an arbitrarily shaped line of secondary *dipoles*, the synthesis integral reads

$$P_R = \int_{-\infty}^{\infty} Q_{\mathrm{d}}(\boldsymbol{r}, \omega) jk \cos\psi \frac{\exp(-jk\Delta r)}{\Delta r} dl \tag{2.33a}$$

with

$$Q_{\mathrm{d}}(\boldsymbol{r}, \omega) = S(\omega) \sqrt{\frac{1}{2\pi jk}} \sqrt{\frac{\Delta r}{r + \Delta r}} \frac{\exp(-jkr)}{\sqrt{r}} . \tag{2.33b}$$

Angle ψ is defined as in (2.28a) and depicted in Figure 2.13. For the stationary phase point the relation $\psi = \varphi$ applies.

C. Remark

It seems plausible to generalize driving function (2.30) for a smooth curvature of the receiver line and the secondary source line by setting $\zeta = \Delta r / r$ as in (2.32b), thus extending the validity of (2.32b) to *focused* primary sources. Because Δr and r are lengths of vectors rather than signed coordinates, a sign has to be implemented in ζ artificially, if the secondary source line is not straight as for driving function (2.30). In this way the synthesis operator is able to discriminate between focusing and non-focusing situations. From symmetry and reciprocity (and simulation results) such a generalization appears reasonable.

2.4 Diffraction waves

Diffraction waves in the synthesized wave field are caused by irregularities in the secondary source distribution. Two such causes are discussed here: truncations and corners.

2.4.1 Truncation effects

Effects of truncation or finite aperture will appear in the synthesized wave field if the secondary source distribution is not infinitely long. The area in which the wave field is reconstructed is limited to those positions for which a stationary phase point can be located at the secondary source line. This principle is visualized in Figure 2.18. The approximate reconstruction area can be found by drawing lines from the primary source towards the edges of the secondary source line. Inside this area diffraction waves will interfere with the synthesized wave field. Outside this area the wave field bends around the edges, i.e. the apparent center of the diffraction waves lies at the edges of the secondary source line.

This effect is studied in Figure 2.19. A plane wave is partly reconstructed by a truncated secondary source line. To the left a monochromatic source signal is used, to the right a repetitive pulse is synthesized. The strength of the driving function of the secondary sources is indicated by the (vertical) width of the bar on top of the picture. Spherical diffraction waves, much weaker than the synthesized plane waves (about –20 dB), are originating from the edge. The phase of these waves depends on the direction (away from the edge). This can be seen from the black-white order of the wave fronts in Figure 2.19b: inside the reconstruction area the diffraction waves appear to be in anti-phase with the phase of the plane waves. This is in agreement with diffraction theory (see e.g. Pierce 1991). Outside the reconstruction area the plane waves bend around the edge smoothly, apparently without phase disturbance.

Two methods will be given here to reduce these truncation artefacts.

De Vries et al. (1994) showed how to suppress these waves by compensating them with an additional secondary point source at the edge, driven by

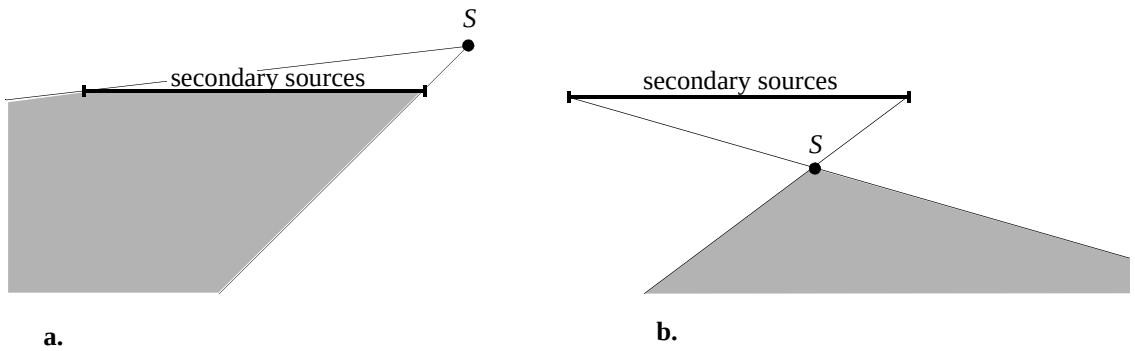
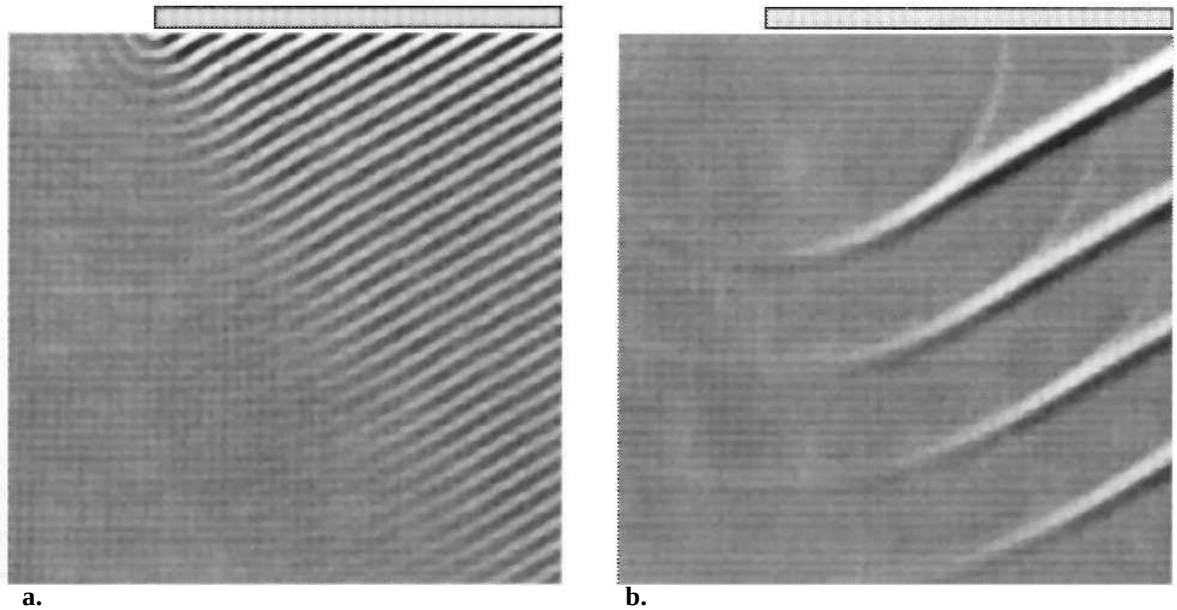


Figure 2.18 Truncation of the secondary source line causes a limitation of the reconstruction area. The approximate reconstruction area is marked gray. **a.** Source behind secondary sources; **b.** Source in front of secondary sources (focused).



a. **b.**

Figure 2.19 Diffraction waves for a synthesized far-away source, producing plane waves.
a. For a monochromatic source signal.
b. For a repetitive pulse.
The secondary source line is truncated only at its left side.

$$Q_m(\boldsymbol{r}_{edge}, \omega) = S(\omega)\sqrt{\frac{1}{2\pi jk}}\sqrt{\frac{\Delta z_0}{z_0 + \Delta z_0}}\frac{\cos\varphi_{edge}}{\sin\varphi_{edge} - \sin\beta_0}\frac{\exp(-jkr_{edge})}{\sqrt{r_{edge}}}, \qquad (2.34)$$

with r_{edge} and φ_{edge} the polar coordinates of the edge with respect to the primary source, and β_0 a fixed angle for which the diffraction waves are compensated optimally, as drawn in Figure 2.20a. The driving signal of the edge sources has a different frequency dependence ($1/\sqrt{jk}$) than the driving function of the secondary monopole source distribution ($\sqrt{jk}$).

This method of reduction of diffraction waves is demonstrated in Figure 2.21a. Again, the wave field of the far-away source of Figure 2.19b is synthesized. At the position of the circle (next to the bar above the picture), a monopole source is driven with signal (2.34) in which β_0 has been set to some angle inside the reconstruction area. In this area a reduction of diffraction waves of about 15 dB is accomplished. Note that the sign of (2.34) depends on whether $\beta_0 > \varphi_{edge}$ or $\beta_0 < \varphi_{edge}$ is chosen. Since in practice β_0 is fixed, this implies that the compensation worsens the situation *outside* the reconstruction area, where the sign of the diffraction waves is approximately equal to that of the compensating signal.

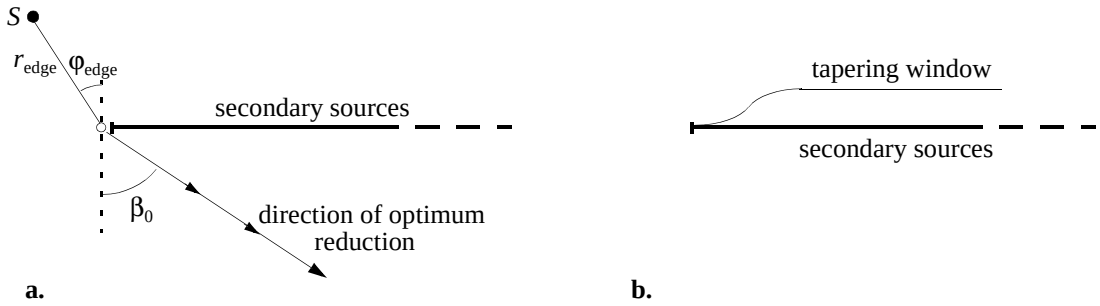


Figure 2.20 Reduction of diffraction effects. **a.** A single monopole source at the edge of the secondary source line driven by signal (2.34). After De Vries et al. (1994); **b.** A cosine taper applied to the amplitude of the driving function of the secondary sources. After Vogel (1993).

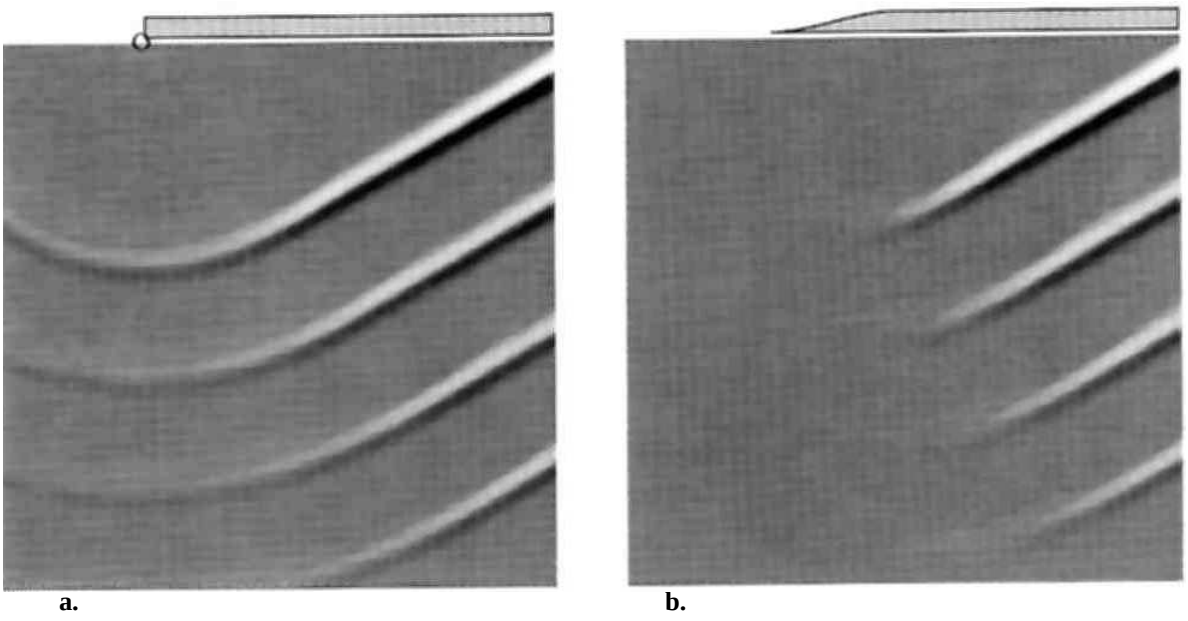


Figure 2.21 Reduction of diffraction waves by using:
a. a point source (position O) at the edge of the secondary source line.
b. a cosine or half-Hanning window as taper.
The bar at the top of the picture shows the amplitude of the driving function. The secondary source line is truncated only at its left side.

A different method for reduction of diffraction effects, known as *tapering*, may therefore be more efficient (Vogel 1993). The amplitude of the driving function is tapered, i.e. gradually decreased, towards the edge(s) of the secondary source distribution. This has the effect of smearing out the diffraction energy over a larger area. This method is the spatial equivalent of the application of a time-window in signal theory. Best results are obtained by applying a cosine taper, such as a half-Hanning window (Figure 2.20b). A taper-width of about 1 m

reduces the amplitude of the diffraction waves by 6 to 10 dB. Figure 2.21b illustrates the effect of a cosine taper applied to the secondary source distribution of Figure 2.19b. Though the reduction is smaller than with the method of De Vries et al., the diffraction waves are now attenuated inside as well as outside the reconstruction area.

In contrast with the compensation method, tapering is easily applicable in wave field processors without extra filtering per primary source. Therefore tapering is favorable in real-time wave field processing.

2.4.2 Corner effects

Diffraction also occurs if two straight line distributions with different orientation are attached to each other, thus constituting a corner*. However, in that case the diffraction waves are weaker than at truncations. For the 90º corner of Figure 2.22, the amplitude of the diffraction waves is approximately 27 dB less than that of the plane wave fronts. The strength of the corner diffraction waves is therefore comparable to the strength of residual effects after tapering with a single linear source distribution.

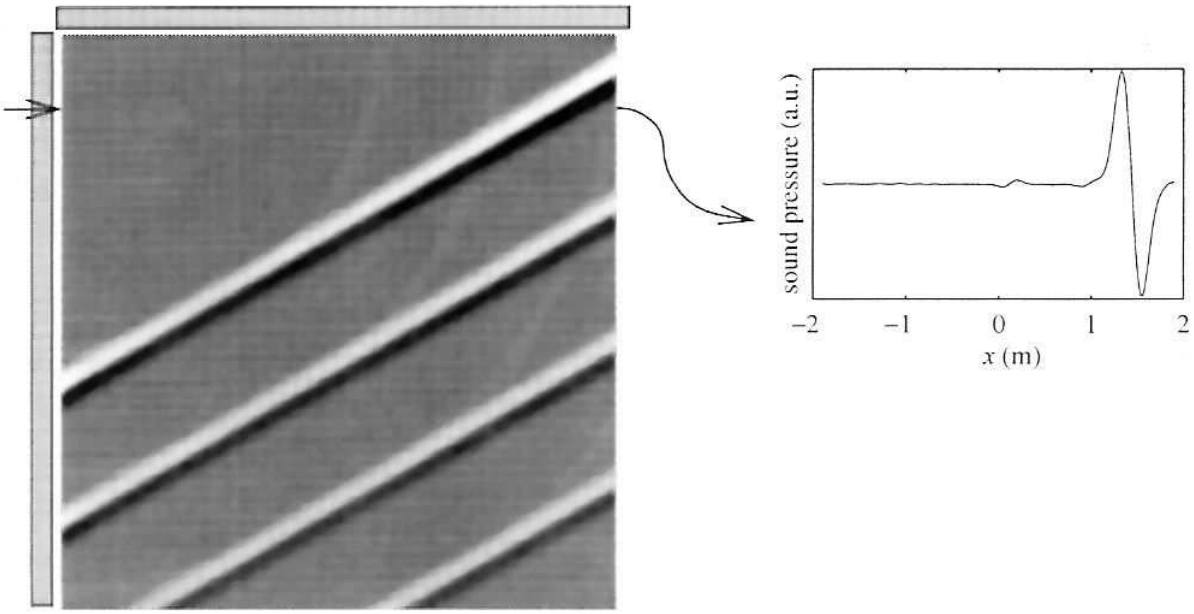


Figure 2.22 Planar waves synthesized by a secondary source distribution around the corner (shown by the bars). A cross-section of the wave field (at the indicated position) is displayed to the right. At $x = 0.1$ and $x = 1.0$ some residual diffraction effects are visible.

* The application of corner arrays is described in Chapter 4.

2.5 Discretization

The practical application of the theory of wave field synthesis requires continuous distributions of secondary sources to be replaced by discrete arrays of loudspeakers. In that case the primary sources are referred to as *virtual sources*. Within some restrictions, to be explained in this section, the integral equations in the previous sections can be read as summations

$$\int F(x)\,dx \rightarrow \sum_n F_n \Delta x$$

where $F_n = F(x = x_n)$ and Δx is the sampling distance between the loudspeakers. The sampling distance determines the maximum frequency component of the primary wave field that is synthesized correctly by the loudspeaker array:

$$f_{\max} = \frac{c}{2\Delta x} \tag{2.35}$$

where c is the speed of sound. The above relation holds for a worst case situation — in fact there is an angle dependence for the effects of undersampling. Before investigating these effects mathematically, a few examples will graphically show the frequency dependence of undersampling artefacts.

Figure 2.23 illustrates how loudspeakers are used to synthesize a wave front. The three figures to the left show the construction of a wave front of a virtual source behind the array for decreasing sampling width Δx. In Figure 2.23a only three loudspeakers (2 m spacing) are used. For this value of Δx only frequencies below 85 Hz are handled correctly. Since this frequency component is not present in the used Gaussian wavelet, no synthesis results. In Figure 2.23b already a synthesized wave front is recognizable. The wave front contains frequencies components up to 425 Hz, while higher frequency components do not interfere constructively yet. A clear wave front is visible in Figure 2.23c.

The same maximum frequency (2.35) holds for a virtual source in front of the array, but the spatial situation is different. In this case (Figure 2.23d-e-f), the residual wave fronts of the individual loudspeakers are produced ahead of the synthesized wave front.

In the remainder of this chapter, the frequency and angle dependence of the effects of undersampling are investigated. Therefore, the synthesized wave field has to be studied in the so-called xt-domain and $k_x k$-domain (Section 2.5.1). With the aid of plane wave decomposition, a comparison can be made between temporal sampling and spatial sampling (Section 2.5.2). It is shown that the effects of undersampling can be reduced by applying a k_x-filter to the sampling arrays. Practical implementations of such a filter are designed in the next chapter.

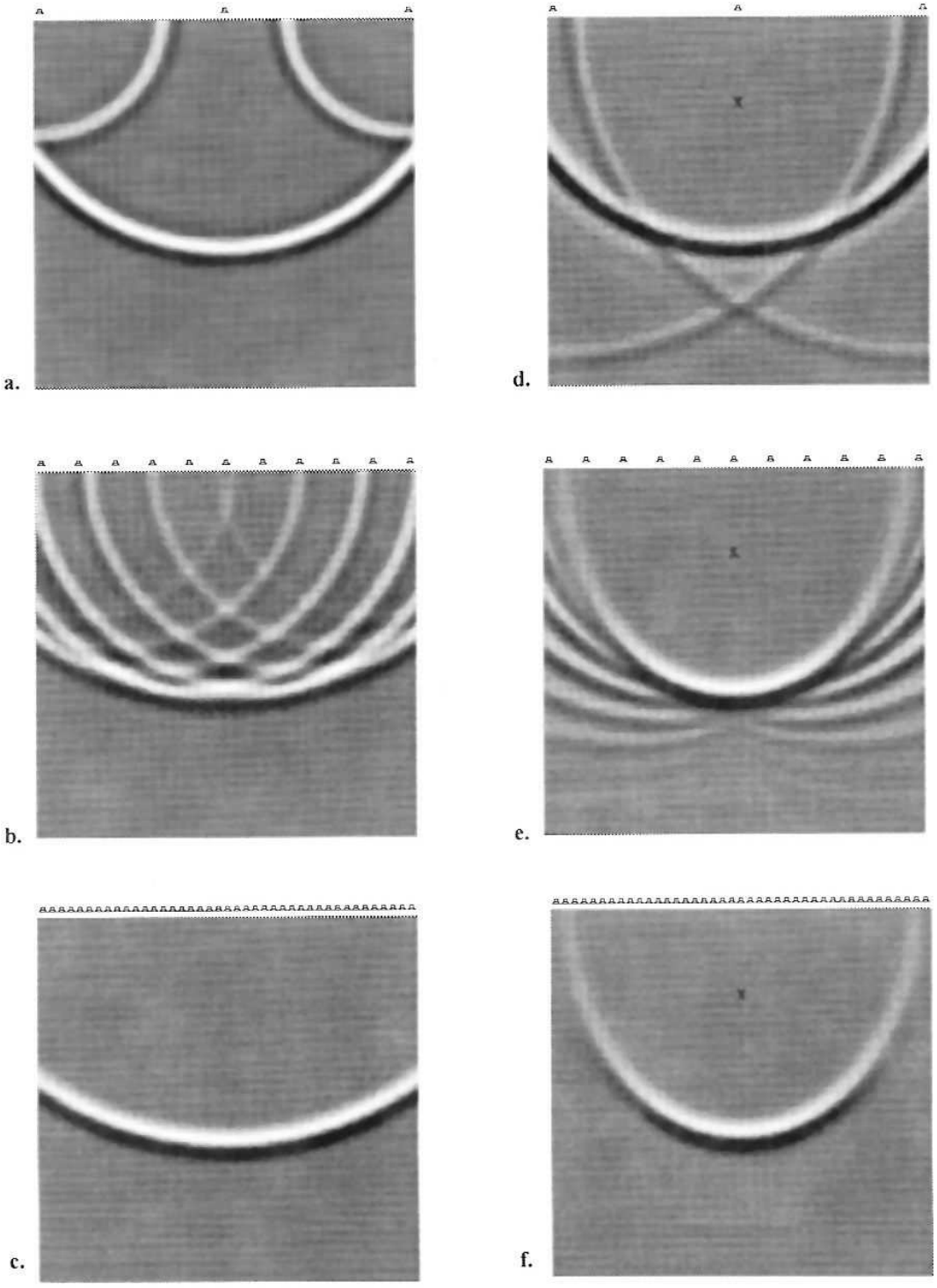


Figure 2.23 Synthesis with discrete loudspeaker array for a Gaussian wavelet with a -15 dB point at 1700 Hz.

Virtual source 1 m behind the array

a. Δx = 2.0 m.

b. Δx = 0.4 m.

c. Δx = 0.1 m.

Virtual source 1 m before the array (at position marked 'x')

d. Δx = 2.0 m.

e. Δx = 0.4 m.

f. Δx = 0.1 m.

2.5.1 Plane wave decomposition

Consider a plane wave entering the x-axis with an angle α_0, as displayed in Figure 2.24a. The direction of propagation is chosen in the xz-plane, perpendicular to the y-axis. If this wave is recorded by a microphone array stretching from $x = -4$ to 4 m along the x-axis, the xt-diagram shown in Figure 2.24b can be drawn. Each vertical line represents an impulse response taken at the position shown along the horizontal axis. Negative pressures at the microphones are blackened for clarity. If a double Fourier transform, given by

$$\tilde{P}(k_x, \omega) = \iint p(x, t) e^{-j\omega t} e^{jk_x x} dx dt , \tag{2.36}$$

is applied to these $p(x, t)$ data, a $k_x k$-diagram (Figure 2.24c) can be constructed. In this diagram, the original frequency parameter ω (along the vertical axis) has been replaced by $k = \omega/c$. Each point in a $k_x k$-diagram represents a monochromatic plane wave, the frequency of which can be read from the vertical axis. Since the Gaussian wavelet in this example has a broadband spectrum, a line instead of a point is found in the $k_x k$-diagram. Angle β_0, given by

$$\beta_0 = \operatorname{atan}(k_x / k) , \tag{2.37}$$

is related to the angle of propagation α_0 of the plane wave via the equation

$$k_x = k \sin\alpha_0 . \tag{2.38}$$

Combining (2.37) and (2.38) yields

$$\sin\alpha_0 = \tan\beta_0 . \tag{2.39}$$

Figure 2.24d-e-f shows the same analysis for a nearby source producing spherical waves. Here, the set of impulse responses is measured at $z = 1$ m. After about 3 ms the wave front arrives at the closest microphone. The shape of the wave front in the xt-diagram is a hyperbola. The corresponding $k_x k$-diagram is a plane wave decomposition diagram. Again, each point in the diagram represents a monochromatic plane wave traveling in the direction given by relation (2.39). Since the wave is traveling mainly leftward along the microphone array, i.e. towards the negative x, the energy in the $k_x k$-diagram is concentrated at the side of negative k_x. From this diagram it can also be seen that there is a sharp boundary at $\beta = -45^\circ$. Clearly such a limit must exist, because then $k_x = -k$ applies with $\alpha = -90^\circ$. A similar boundary at $k_x = k$ exists.

Plane wave decomposition of wave fields is of great importance as an analysis tool because it reveals the orientation of different components in the wave field. In the next paragraph it is shown that the $k_x k$-diagram is a spatial analogy of the frequency spectrum for time series.

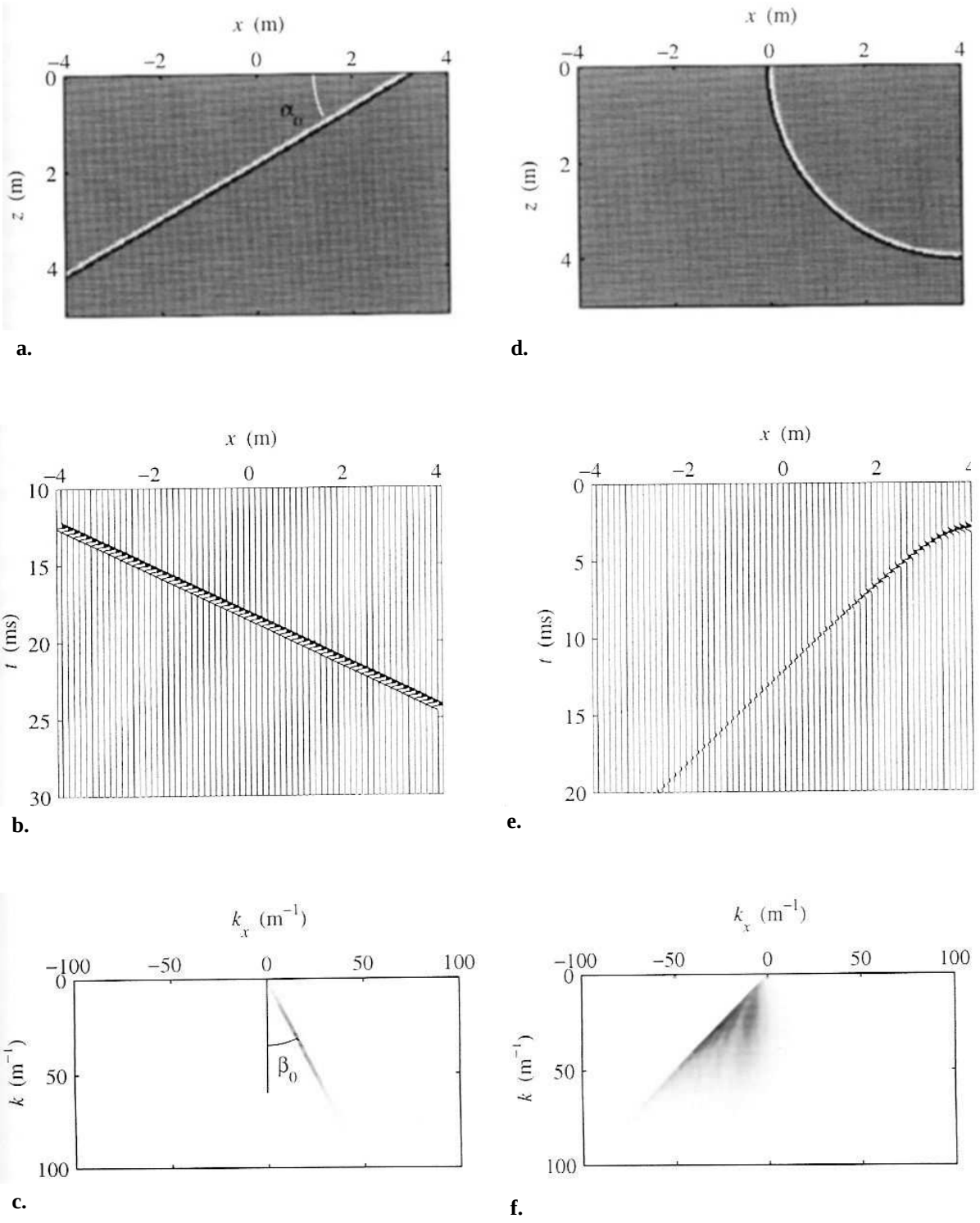


Figure 2.24 Two example wave fields in three different representations.
Plane wave
a. *xz*-diagram, space-space domain ($\alpha_0 = 30°$).
b. *xt*-diagram, space-time domain ($z = 1$; t offset).
c. $k_x k$-diagram, wavenumber-wavenumber domain ($\beta_0 = 27°$).
Source at $(x, z) = (4, 0)$
d. *xz*-diagram, space-space domain.
e. *xt*-diagram, space-time domain ($z = 1$).
f. $k_x k$-diagram, wavenumber-wavenumber domain.

2.5.2 Spatial sampling and reconstruction

Spatial sampling is easiest understood in comparison with temporal sampling, which is more common practice. In temporal sampling, a continuous-time signal is cut in regularly spaced time-slices. The reverse process attempts to reconstruct an analog signal from discrete-time data (Watkinson 1989). Figures 2.25a and 2.26 illustrate temporal discretization for a certain signal, drawn in Figure 2.26a. Since this continuous-time signal will in general contain more information than its discrete version, a data-reduction step is involved in the discretization process: it is passed through a low-pass filter. The purpose of this filter is to avoid aliasing during sampling: otherwise high frequency information would fold back into the frequency band from 0 Hz to the Nyquist frequency. The Nyquist frequency f_{Nyq} equals half the sample rate f_s. In the example signal (Figure 2.26a), the high frequency contents is represented by a transient peak, that is flattened by the (low-pass) anti-aliasing filter (Figure 2.26b). This information is unrecoverably lost at this stage, due to the applied low-pass filter. The quantizer in an actual analog-to-digital converter (ADC), which rounds off the height of the pulses to the nearest quantizing (bit) level, is not essential for the analogy to spatial sampling and is therefore not considered here. The discrete-time signal is shown in Figure 2.26c.

In the reverse process a continuous-time signal is constructed from the discrete-time series (Figure 2.26d). The choice of the procedure for interpolation between the samples is arbitrary*, since a reconstruction filter with the same cut-off frequency as the anti-aliasing filter will suppress any high frequent interpolation errors. The reconstruction filter output (2.26e) is identical to the anti-aliasing filter output (2.26b).

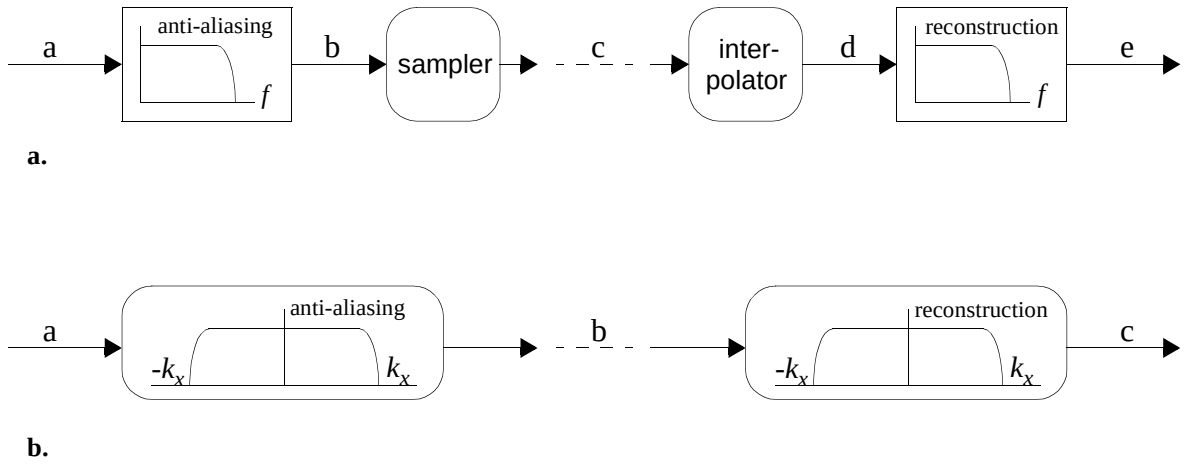


Figure 2.25 Conversion schemes: analogies between the temporal and spatial discretization.
a. Temporal sampling and reconstruction. Stages a, b, c, d and f refer to sub-plots of Figure 2.26.
b. Spatial sampling and reconstruction. Stages a, b and c refer to sub-plots of Figure 2.27.

* In this example stepwise interpolation is used, which is common practice: a discrete pulse is hold at its DC level until the next pulse occurs. This produces a repetitive spectrum in which consecutive copies fall off proportional to $\sin(\pi[f/f_s])/\pi[f/f_s]$. Since the high frequency components *within* the pass-band are also affected by this sinc-function, a compensating reconstruction filter is used in practice.

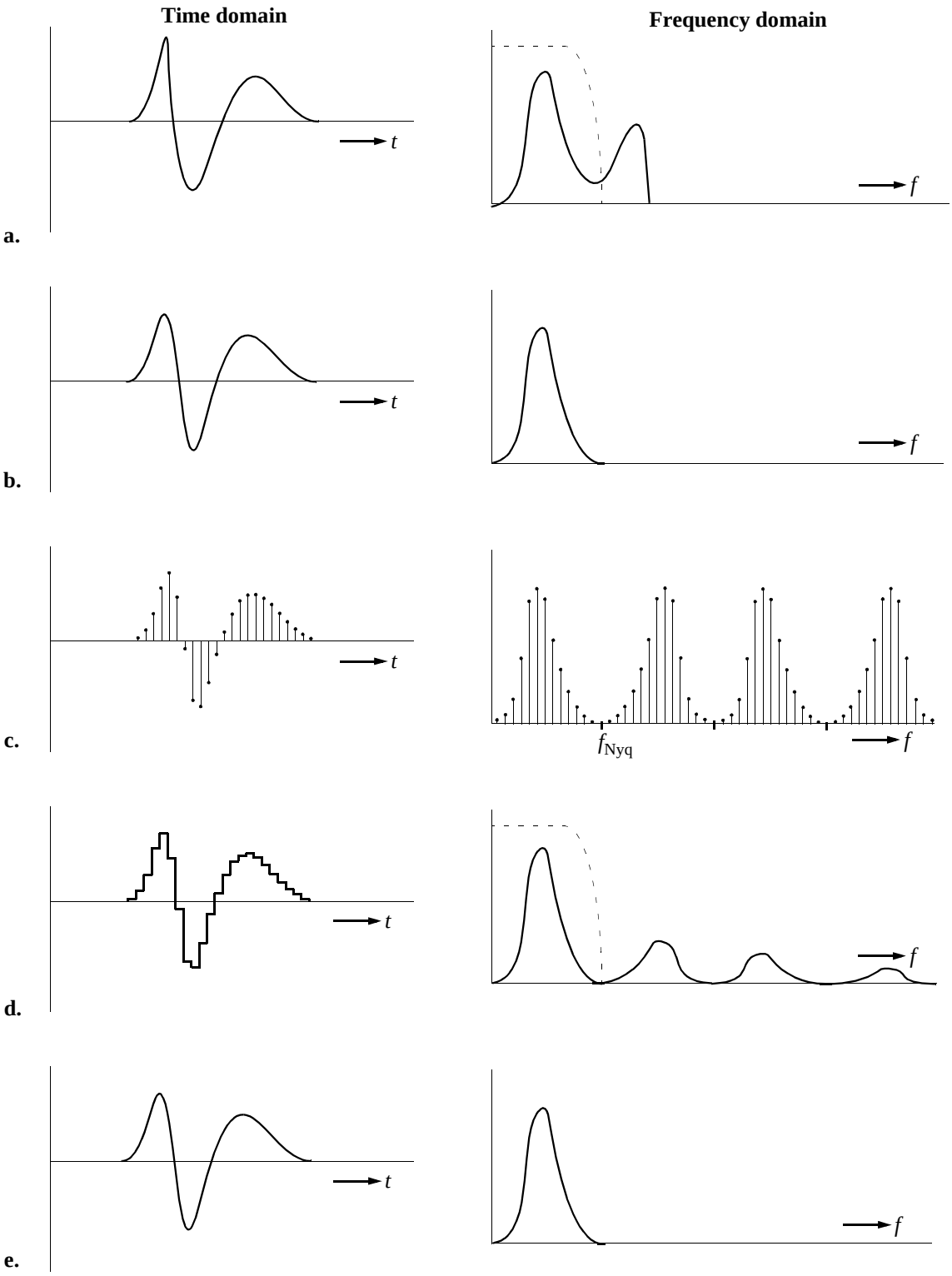


Figure 2.26 Schematic representation of temporal sampling.
Time and frequency domain signals at the five stages of Figure 2.25a.
a. Original analog signal.
b. Analog signal, passed through anti-aliasing filter.
c. Sampled signal.
d. Stepwise interpolated analog signal.
e. Analog signal, passed through reconstruction filter. The signal at stage **e** is identical to that at stage **b**.

For sampling in the spatial domain, the flow chart of Figure 2.25b applies. The different stages of spatial sampling are visualized in Figure 2.27. First, the *xz*-diagram of the original wave field is considered (Figure 2.27a). To remain within the analogy to temporal sampling, a continuous *xt*-diagram or $k_x k$-diagram would be required. Such a continuous diagram can be simulated by an array with an extremely small spacing between the microphones. In that case, it follows from criterion (2.35) that also the highest frequencies are treated correctly. The virtual microphone array is situated at $z = 0.75$ m parallel to the *x*-axis. The $k_x k$-diagram* is shown to the right of the *xz*-diagram in Figure 2.27a. In comparison with Figure 2.24f it is obvious that since the source is in the middle of the microphone array, a symmetrical diagram is expected.

The spatial sampling step and the application of a spatial anti-aliasing filter can be combined by using highly directive microphones. Such microphones act as low-pass k_x-filters†, i.e. they have a limited response to high frequency plane wave components at large angles of incidence. Figure 2.27b shows the wave field sampled by the microphone array (at $z = 0.75$ m). The middle impulse responses in the *xt*-diagram are left unaffected by the microphone's directivity, but the outer signals are low-pass filtered.

The corresponding $k_x k$-diagram reveals a periodical spectrum. The period equals two times the spatial Nyquist frequency

$$k_{x,\,\mathrm{Nyq}} = \frac{\pi}{\Delta x}, \tag{2.40}$$

where Δx is the spacing between the microphones. The influence of the microphone directivity is seen as a symmetrical low-pass filter along the k_x-axis. In absence of that filter, the wings of adjacent spectral copies would fold into the middle band, causing unrecoverable distortion in the overlapping zones. Since $k_x = k \sin\alpha$, the maximum frequency component that is correctly sampled by the microphone array has angle dependence

$$f_{\mathrm{Nyq}} = \frac{c}{2\Delta x \sin\alpha}. \tag{2.41}$$

This expression is a refinement to (2.35) in which $\alpha = 90^\circ$ was taken to hold for worst case incident angles. The signals of the microphones are fed to an array of highly directive loudspeakers. The loudspeaker characteristics have the same k_x-dependence as the microphones, but according to the Rayleigh II operator, an additional $\cos\alpha$ directivity is necessary. This filter now acts as a reconstruction filter, avoiding aliased‡ waves in the synthesized wave field (Figure 2.27c). A comparison of the *xz*-diagrams before (Figure 2.27a) and after (Figure 2.27c) sampling shows that the synthesized wave field is identical to the original wave field only in the

* The corresponding *xt*-diagram is left out of consideration here.

† The relation between the directivity of an array element and its k_x-response is proved in Section 3.3.1.

‡ Strictly speaking, *aliasing* is an effect caused by improper sampling, while *grating lobes* occur with erroneous reconstruction. In this thesis, the term *spatial aliasing* refers to either effect.

middle area. To the sides, the effects of low-pass filtering are visible as a widening of the wavelet. If a proper reconstruction in all directions is desired, a smaller spacing between the array elements is necessary.

Instead of a microphone array, also a single microphone close to the source can be used to record the original wave field. In that case, the loudspeaker feeds (Figure 2.27b) are given by the driving functions of Section 2.3.3, to which the previously mentioned low-pass k_x-filter

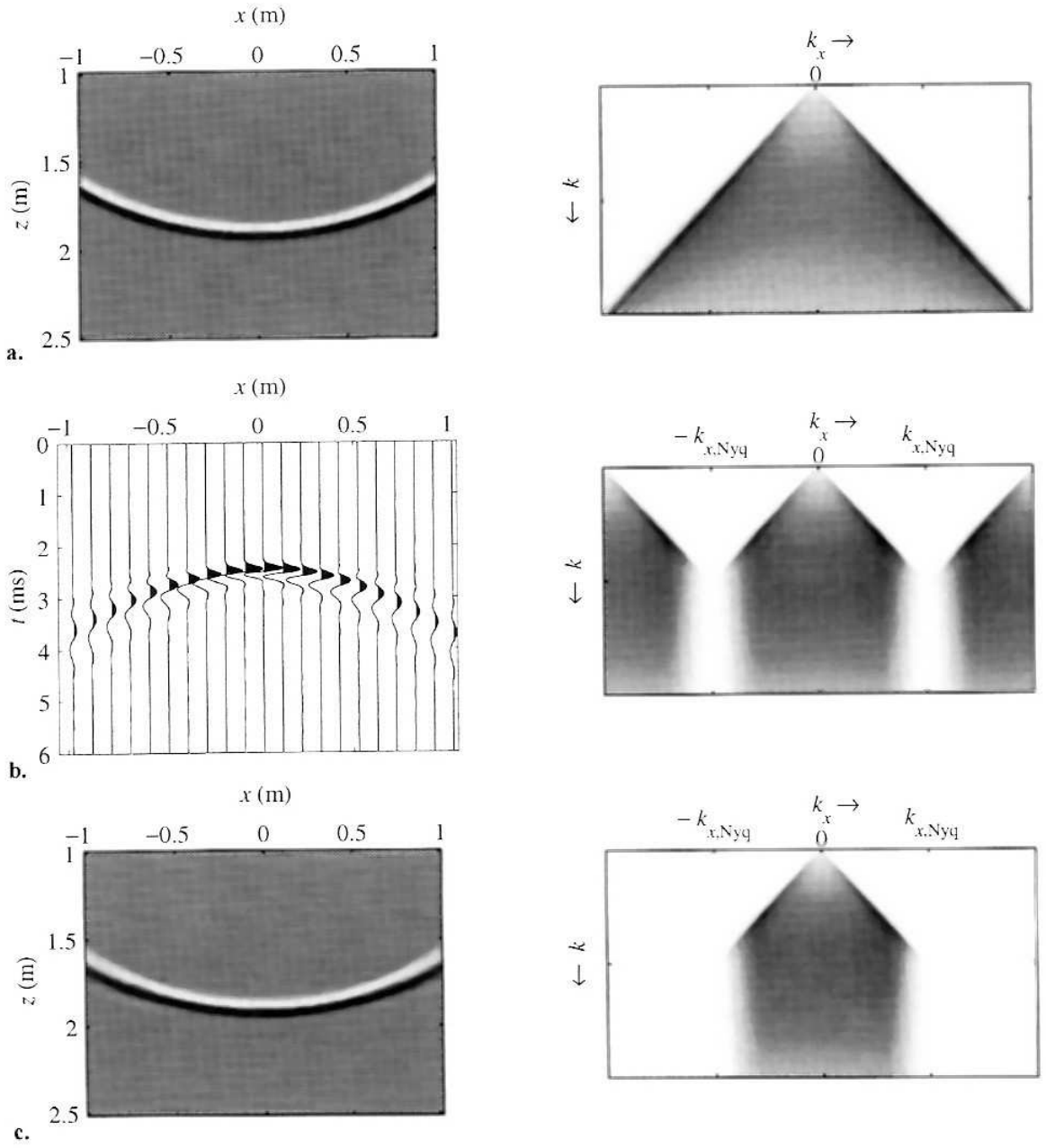


Figure 2.27 Schematic representation of spatial sampling. The source is at $(x, z) = (0, 0)$.
a. xz-diagram and k_xk-diagram of the original wave front.
b. xt-diagram and k_xk-diagram of the sampled wave front.
c. xz-diagram and k_xk-diagram of the synthesized wave front.

is applied via spatial convolution (signal processing). The reconstruction step is equal to the previously described reconstruction step.

Start (1997) gives a thorough treatise on spatial sampling, convolution and reconstruction. He also proposes a different method of anti-aliasing and reconstruction filtering, with constant bandwidth in finite aperture.

2.6 Applications of wave field synthesis

The theory of wave field synthesis has been investigated previously by Vogel (1993) and Start (1997) for use as a spatially correct sound reinforcement system. Here, the sound from a performer on stage, received through close-miking techniques or a microphone array, is amplified by a loudspeaker array hanging over the stage area. A source-tracking system acquires the source coordinates. If the virtual source position is chosen to coincide with the true source position, the audience localizes the reinforced sound at any time in the same direction as the (visual) source. Also the early reflections can be controlled in this way.

A variable acoustics system based on wave field synthesis can be applied in multi-purpose halls, where at one time speech intelligibility and at another time concert-hall acoustics are required (Reijnen et al. 1995). The above method of sound reinforcement is used for direct sound and early reflections, while the reverberant field is adaptively controlled by loudspeaker arrays towards a certain desired hall response.

2.7 Conclusions

Wave field synthesis states that the spatial and temporal reconstruction of a wave field is possible by applying the Rayleigh operators to the primary source signals, yielding the driving signals of a planar distribution of secondary sources. The wave field of a primary source at one side of that plane is correctly synthesized at the other side. Therefore, in principle wave field synthesis offers the solution for the spatial requirements for sound reproduction.

If the *planar* distribution of secondary sources is replaced by a *linear* distribution of secondary sources, reconstruction of the wave field is possible in the (horizontal) plane containing that linear distribution and the primary source. Finite linear source distributions exhibit truncation effects that are efficiently reduced by tapering the driving functions.

Spatial discretization theory allows for a practical implementation of the linear source distribution as loudspeaker arrays. Undersampling artefacts (spatial aliasing) can be minimized by applying directional loudspeakers.

Chapter 3

Loudspeaker Arrays

3.1 Introduction

The use of arrays is an essential part of wave field synthesis. Once it has been established that sound is a wave phenomenon with spatial characteristics, arrays appear to be indispensable for proper control of sound fields. In the previous chapter, the foundation of wave field synthesis was built through the constitution of the synthesis operators, yielding the array driving functions. In this chapter special attention will be paid to loudspeaker arrays.

The principles of the operation of electrodynamic and electrostatic loudspeakers are described in Section 3.2. At low frequencies, the radiation behavior of a closed box electrodynamic loudspeaker approaches that of a monopole point source, whereas the electrostatic loudspeaker has more or less dipole characteristics (Section 3.3). The deviation from ideal monopole behavior is usually described by directivity characteristics. It is shown that loudspeaker directivity can be expressed in terms of spatial convolution filters. Reversely, the directivity of a single array element can be controlled by using the spatial convolution property of loudspeaker subarrays.

Section 3.4 deals with the design of an electrodynamic and an electrostatic array, that have been tested by comparative measurements. The electrodynamic array is constructed of light-weight drivers that are active in a large frequency range. The electrostatic array consists of elements with controlled directivity in order to minimize effects of spatial aliasing. Finally a decision is made which array is most appropriate for use in the laboratory wave field synthesis system for sound reproduction.

3.2 Loudspeaker types

Two types of loudspeakers and their applicability for wave field synthesis are investigated: the electrodynamic loudspeaker, which is well-known for its low cost and high (power) performance, and the electrostatic loudspeaker, of which the directivity characteristics can easily be manipulated. The purpose of this section is to find the (approximate) frequency response of both loudspeaker systems, given a voltage E across the terminals, and to find some relevant parameters for the design of an electrodynamic and electrostatic array element.

3.2.1 Canonical equations

The action of an electromechanical transducer, shown schematically in Figure 3.1, is characterized by the two canonical equations that relate the basic electrical properties to the mechanical properties of the transducer (Hunt 1982):

$$E = Z_e I + T_{em} V, \tag{3.1a}$$

$$F = T_{me} I + Z_m V, \tag{3.1b}$$

with E the voltage, F the force, I the current, V the velocity, Z the electrical or mechanical impedance, and T the transduction coefficients. The transduction coefficients of a physically realizable transducer are either reciprocal, with $T_{em} = T_{me}$ (e.g. an electrostatic transducer), or anti-reciprocal, with $T_{em} = -T_{me}$ (e.g. an electrodynamic transducer).

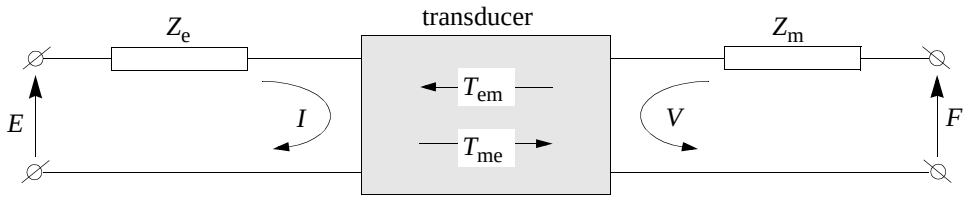


Figure 3.1 Schematic representation of an electromechanical transducer (after: Hunt 1982).

3.2.2 Electrodynamic loudspeaker

In an electrodynamic loudspeaker, the conversion of electrical energy to mechanical energy is realized by a voice coil moving in a permanent magnetic field, see Figure 3.2a. The transduction coefficients are found from the Lorentz force

$$F = -BlI \tag{3.2a}$$

and the induction law

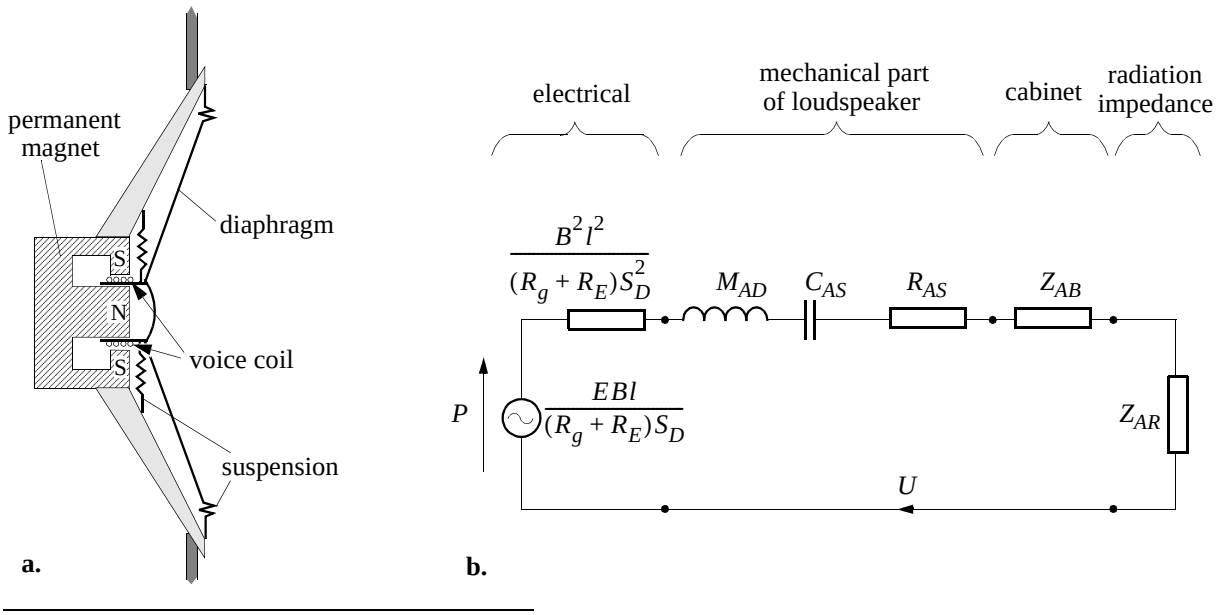


Figure 3.2 a. Cross-section of the electrodynamic loudspeaker.
b. Analogous acoustical impedance circuit of a loudspeaker in a cabinet (after: Beranek 1954). Parameters:

E	amplifier voltage	S_D	effective area of diaphragm	Z_{AB}	loading impedance of the cabinet
R_g	amplifier resistance	R_{AS}	acoustical resistance of suspensions	U	$= S_D V$ volume velocity
R_E	voice coil resistance	C_{AS}	acoustical compliance of suspensions	P	$= F/S_D$ pressure
l	voice coil wire-length	Z_{AR}	radiation impedance at the front		
B	air-gap flux density	M_{AD}	acoustical mass of diaphragm and voice coil		

$$E = BlV, \tag{3.2b}$$

with B the magnetic field and I the current through the voice coil with wire-length l, giving $T_{\text{em}} = Bl$ while $T_{\text{me}} = -Bl$. The electrical and mechanical impedances of the canonical equations (3.1a) and (3.1b) are complicated functions of various system parameters. If the loudspeaker is mounted in a cabinet (closed box), it is favorable to use the acoustical impedance analogy circuit as drawn in Figure 3.2b.

At low frequencies, the radiation impedance Z_{AR} is dominated by the mass of the air in front of the diaphragm, yielding $Z_{AR} \approx j\omega M_{AR}$. The loading impedance Z_{AB} at the back is written as a series resistance, mass and compliance

$$Z_{AB} = R_{AB} + j\omega M_{AB} + \frac{1}{j\omega C_{AB}}. \tag{3.3}$$

With these considerations, the equivalent scheme of Figure 3.2b reduces to a second order network with a total acoustical resistance

$$R_A = R_{AS} + R_{AB} + \frac{B^2 l^2}{(R_g + R_E)S_D^2}, \tag{3.4a}$$

a total acoustical mass

$$M_A = M_{AD} + M_{AB} + M_{AR}, \tag{3.4b}$$

and a total acoustical compliance

$$C_A = \frac{C_{AS} C_{AB}}{C_{AS} + C_{AB}}, \tag{3.4c}$$

all connected in series. The resonance frequency of this system is given by

$$\omega_0 = \frac{1}{\sqrt{M_A C_A}}, \tag{3.5}$$

with a quality factor

$$Q_0 = \frac{1}{R_A} \sqrt{\frac{M_A}{C_A}}. \tag{3.6}$$

Below resonance, the system is stiffness-controlled, i.e. the volume velocity U is proportional to $j\omega C_A$, while the system is mass-controlled above resonance (Beranek 1954):

$$U(\omega) = \frac{E(\omega) Bl}{j\omega M_A S_D (R_g + R_E)}. \tag{3.7}$$

If the cabinet is small compared to the wavelength, the loudspeaker has monopole characteristics. Then, the free field sound pressure can be approximated by that of a volume velocity source

$$P(\boldsymbol{r}, \omega) = \frac{j\omega\rho}{4\pi} U(\omega) \frac{\exp(-jkr)}{r}, \tag{3.8}$$

or, with Equation (3.7),

$$P(\boldsymbol{r}, \omega) = \frac{\rho E(\omega) Bl}{4\pi M_A S_D (R_g + R_E)} \frac{\exp(-jkr)}{r}. \tag{3.9}$$

This expression shows that the sound pressure is proportional to the steering voltage for frequencies above resonance (which is the operational frequency range for normal loudspeakers). The frequency factor $1/j\omega$ of the volume velocity is fully compensated by the factor $j\omega$ of the radiation impedance of the monopole.

If multiple electrodynamic loudspeakers are used to form a linear array for wave field synthesis, the only frequency dependence that is left originates from driving function (2.30). It states that the input signals of the loudspeakers should be filtered by +3 dB/octave ($\sqrt{k}$). In addition, for a primary source behind the array a frequency-independent phase shift of +45° is necessary, while it is – 45° for a source in front of the array.

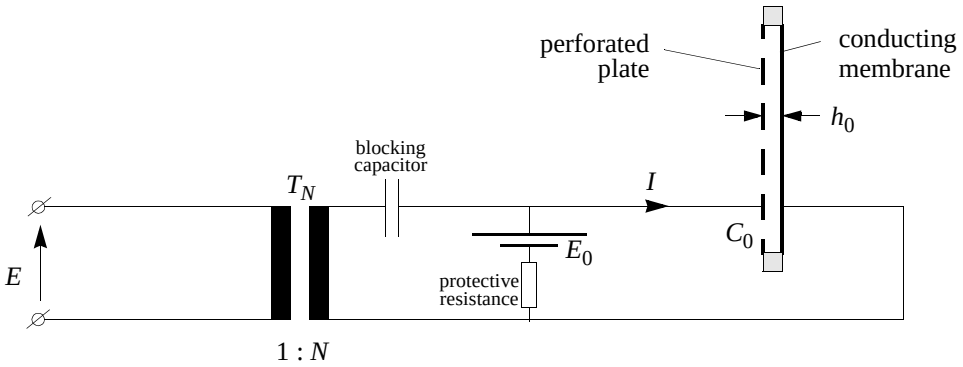
3.2.3 Electrostatic loudspeaker

The electrostatic loudspeaker consists of a conducting membrane and a perforated rigid plate, together forming a capacitor (Figure 3.3a). By applying an alternating voltage between the plates, the membrane is forced to move to and fro. The steering voltage E is increased by a transformer and added to the polarizing voltage E_0. Instead of one rigid plate, often two plates are used, one at either side of the membrane. The following relations also hold for this balanced ('push-pull') version, if the capacitance C_0 is read as $C_0/2$ (two capacitors C_0 in series). The canonical equations for the electrostatic loudspeaker are given by (Hunt 1982)

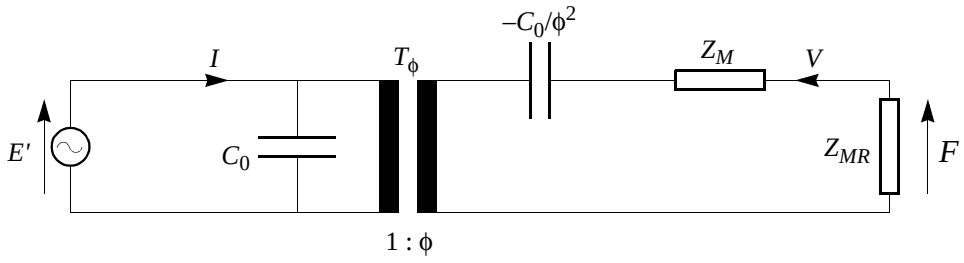
$$E' = \frac{I}{j\omega C_0} + \frac{E_0 V}{j\omega h_0} \tag{3.10a}$$

with $E' = EN$ (N the ratio of turns of the transformer) and V the velocity of the membrane, and

$$F = \frac{E_0 I}{j\omega h_0} + Z_M V, \tag{3.10b}$$



a.



b.

Figure 3.3 **a.** Electric circuitry for a single-sided electrostatic loudspeaker in its equilibrium state (with $E = 0$). The transformer T_N brings the input signal voltages (~1 volt) in the range of 100 volt. A voltage source provides a constant polarization voltage E_0.
b. Analogous electromechanical circuit of the electrostatic loudspeaker. Transformer T_N has been removed from the circuit by incorporating the turns-ratio in the signal source voltage $E' = NE$. Transformer T_ϕ represents the transduction coupling between the electric circuit to the left and the mechanical analogy circuit to the right. The ratio of turns, $\phi = E_0 C_0/h_0$, is calculated as the transduction coefficient T_{em} divided by the impedance in shunt position $1/j\omega C_0$.The impedance Z_M is due to the mechanical construction of the loudspeaker, while Z_{MR} is the mechanical radiation impedance. The negative capacitance $-C_0/\phi^2$ is caused by electrostatic attraction between the membrane and the perforated plate (after: Hunt 1982).

where Z_M is the mechanical impedance of the loudspeaker construction. These equations show that the transduction coefficient is symmetrical: $T_{\mathrm{em}} = T_{\mathrm{me}} = E_0/j\omega h_0$. The corresponding electromechanical analogy circuit is drawn in Figure 3.3b. With the aid of this scheme, it is found that

$$F = \frac{\phi}{1 - \phi^2/(j\omega C_0 Z_{MR}) + Z_M/Z_{MR}} E', \tag{3.11}$$

in which $\phi = E_0 C_0/h_0$ and the mechanical radiation impedance is defined by $Z_{MR} = F/V$. If the denominator of (3.11) is close to unity, a more or less frequency-independent relation results between the voltage E' and the force F. In that case,

$$F(\omega) = \frac{E_0 C_0 N}{h_0} E(\omega). \tag{3.12}$$

At low frequencies, the electrostatic loudspeaker acts more or less as a dipole. The far-field sound pressure for such a source can be calculated from the sound pressure of two monopole sources with volume velocities $+U$ and $-U$ at a small distance Δr from each other, giving

$$P(\boldsymbol{r}, \omega) = \frac{j\omega\rho U(\omega)\Delta r}{4\pi} jk\cos\varphi \frac{\exp(-jkr)}{r}, \tag{3.13}$$

with φ the angle between the point of observation and the line connecting the imaginary monopoles. Since, according to Newton's second law, the dipole's force acting on the surrounding medium, is given by

$$F(\omega) = j\omega\rho U(\omega)\Delta r, \tag{3.14}$$

the sound pressure can be evaluated by combining (3.12), (3.13) and (3.14):

$$P(\boldsymbol{r}, \omega) = \frac{E_0 C_0 N}{4\pi h_0} E(\omega) jk\cos\varphi \frac{\exp(-jkr)}{r}. \tag{3.15}$$

The frequency response in the far-field increases with 6 dB per octave, because of the factor jk. Since in normal use, e.g. for stereophonic reproduction, a flat frequency response is desired, the electrostatic loudspeaker is usually furnished with an internal filter of –6 dB per octave (and a frequency indepedent phase shift of –90º). Instead, if the electrostatic loudspeaker is used for wave field synthesis, an internal filter of –3 dB per octave ($1/\sqrt{k}$) is needed, as can be seen from driving function (2.31). The filter should have a phase shift of $-45°$ for a source behind the array and $-135°$ for a source in front of the array.

3.3 Directivity

In Section 2.5.2 it has been shown that controlling directivity is of great importance in wave field synthesis, because it allows proper control of spatial aliasing. The directivity of transducers can be either frequency-independent or frequency-dependent. For instance, the (ideal) dipole has frequency-independent directivity, because of the factor cos(φ) being independent of frequency. Since the effect of spatial aliasing depends on frequency *and* direction, it is obvious to use array elements with frequency-dependent directivity in order to control spatial aliasing.

Frequency-dependent directivity is caused by different parts of the radiating (or receiving) diaphragm having different distances to the point of observation (or emission), see Figure 3.4. At low frequencies, where the dimensions of the diaphragm are small compared to the wavelength, the transducer essentially behaves either as an ideal monopole or as an ideal dipole, according to its mechanical construction. At higher frequencies the shape and excursion of the diaphragm must be taken into account.

Apart from the directivity of a single transducer, an array of transducers can also generate a directivity pattern, since an array may be considered as one large, but discretized diaphragm. In this section it will be shown that, through relating directivity to spatial convolution, a thorough understanding and control of directivity is possible.

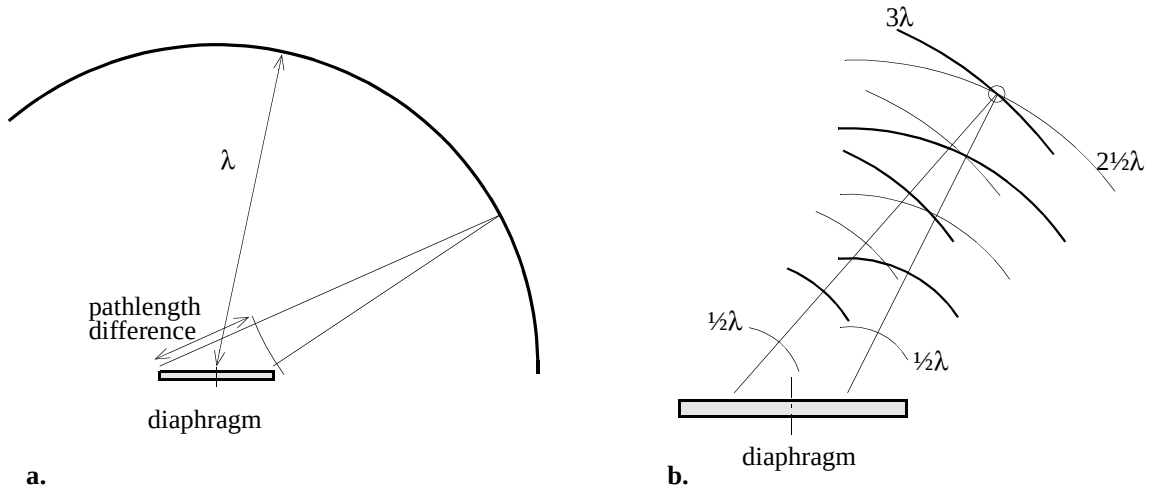


Figure 3.4 Frequency-dependent directivity. **a.** If the width of the diaphragm is small compared to the wavelength, the maximum pathlength difference will be likewise and directivity is negligible. **b.** If the width of the diaphragm is large compared to the wavelength, destructive interference occurs in certain directions if the waves originating from different parts of the diaphragm have opposite phases.

3.3.1 Spatial Fourier transform and directivity

The directional behavior of loudspeakers* is usually studied by considering directivity patterns for several frequencies. If the diaphragm is moving coherently, the directivity patterns for all frequencies can be calculated from the typical directivity function for that shape of diaphragm. For example, the far-field sound pressure of a rigid circular piston, mounted in an infinite baffle, is given by (Beranek 1954)

$$P = \frac{j\omega\rho U}{2\pi} L(ka\,\sin\varphi)\frac{\exp(-jkr)}{r}, \tag{3.16a}$$

where φ is the radiation angle and L the directivity function

$$L(ka\,\sin\varphi) = \frac{2J_1(ka\,\sin\varphi)}{ka\,\sin\varphi}, \tag{3.16b}$$

with J_1 the first order Bessel function of the first kind. Apart from the factor L, the sound pressure (3.16a) is twice that of a monopole volume velocity source (3.8) in a free field, due to the infinite baffle. For this source the directivity function and a few directivity patterns are drawn in Figure 3.5.

Electrodynamic loudspeakers behave as rigid pistons in the low and mid frequency range. Above some frequency, usually a few kHz, they exhibit cone break-up effects, i.e. frequency-dependent modes of bending waves occur over the diaphragm. This is a result of the construction: the driving force from the voice coil is not uniformly spread across the radiating surface. Because of this property, the directivity characteristics of electrodynamic loudspeakers are difficult to model, and hence to control. Electrostatic diaphragms generally do not suffer from break-up effects because the electrostatic force is almost homogeneous across the surface of the membrane. Therefore, electrostatic loudspeakers are well-suited for designing directivity characteristics. In the following, the directivity function of an arbitrarily shaped (electrostatic) dipole diaphragm will be derived.

For a dipole source distribution across a surface A within the xy-plane, the Rayleigh II integral can be used to calculate the sound pressure in a point R far away from that surface

$$P_R = \frac{1}{2\pi}\int_A P(\omega)\frac{1+jkr}{r^2}\cos\varphi\exp(-jkr)\,dx\,dy, \tag{3.17}$$

where $P(\omega)$ is the sound pressure at the surface A, and φ is the angle with the z-axis. In general, the surface of integration of Rayleigh II consists of the whole xy-plane, but since $P(\omega)$ is zero outside A the integration is confined to A only.

* For convenience the theory is illustrated for loudspeakers only, but because of reciprocity the results also apply to microphones.

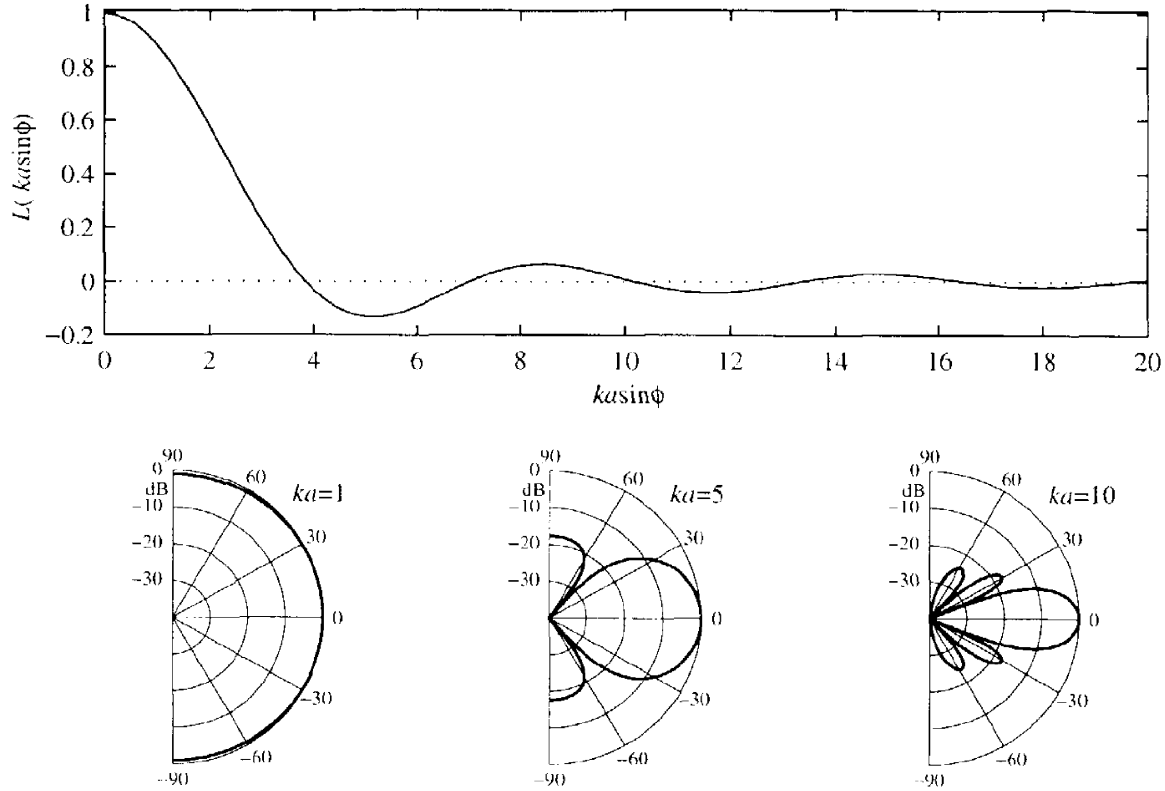


Figure 3.5 Directivity function and patterns for a rigid circular piston.

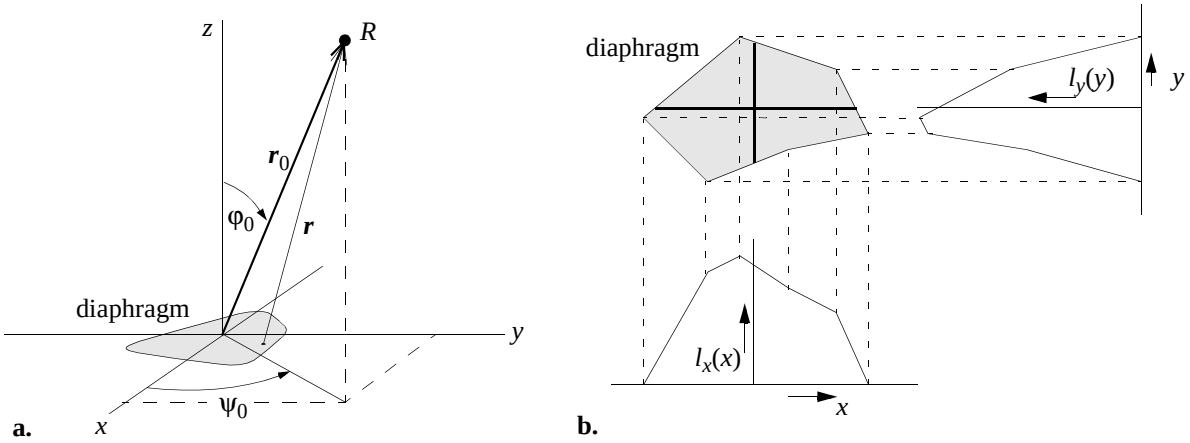


Figure 3.6 **a.** Fraunhofer approximation for radiating diaphragm. The vector $\boldsymbol{r}$ from a point at the diaphragm to R can be expressed in terms of $\boldsymbol{r}_0$ (pointing from the center of the diaphragm to R), the angles φ_0 and ψ_0, and coordinates x and y, as given by (3.18). **b.** Projections on the x- and y-axis yields the strength distributions l_x and l_y.

In the Fraunhofer area, where $r \gg d^2/\lambda_{min}$ with d the diameter of the diaphragm, the following approximation is valid (Berkhout 1987)

$$r \approx r_0 - x\sin\varphi_0\cos\psi_0 - y\sin\varphi_0\sin\psi_0\,, \tag{3.18}$$

where r_0 is the distance from the center of the diaphragm to receiver position R, see Figure 3.6a. Since $k_x = k\sin\varphi_0\cos\psi_0$ and $k_y = k\sin\varphi_0\sin\psi_0$, the sound pressure at R can be written as

$$P_R = jk\cos\varphi_0\frac{\exp(-jkr_0)}{2\pi r_0}\int_A P(\omega)e^{jk_x x}e^{jk_y y}\,dxdy\,. \tag{3.19}$$

The surface integral can be factorized by realizing that the driving force of a small (vertical) column of the surface of the diaphragm, for which k_x is constant, can be assigned to a point source at the x-axis, yielding a strength function $l_x(x)$ for a line distribution along the x-axis, see Figure 3.6b. An analogous procedure can be used for the y-direction. Integral (3.19) then changes into

$$P_R = jk\cos\varphi_0\frac{\exp(-jkr_0)}{4\pi r_0}F(\omega)\int l_x(x)e^{jk_x x}dx\int l_y(y)e^{jk_y y}dy\,, \tag{3.20}$$

in which $F(\omega) = 2AP(\omega)$ is the force of the diaphragm acting on the air. The strength functions $l_x(x)$ and $l_y(y)$ are normalized: $\int l_x(x)dx = 1$ and $\int l_y(y)dy = 1$. With spatial Fourier transforms $L_x(k_x)$ and $L_y(k_y)$ defined as

$$L_x(k_x) = \int_{-\infty}^{\infty} l_x(x)e^{jk_x x}dx \tag{3.21a}$$

and

$$L_y(k_y) = \int_{-\infty}^{\infty} l_y(y)e^{jk_y y}dy\,, \tag{3.21b}$$

the sound pressure at R evaluates to

$$P_R = L_x(k_x)L_y(k_y)\frac{F(\omega)}{4\pi}jk\cos\varphi\frac{\exp(-jkr)}{r}\,, \tag{3.22}$$

in which the subscript '0' has been omitted. Notice that this expression reduces to (3.13), with substitution of (3.14), if k approaches zero, i.e. for low frequencies where $L_x(k_x) \approx 1$ and $L_y(k_y) \approx 1$. This equation states that the sound pressure in the far-field is proportional to the spatial Fourier transforms of the diaphragm's strength functions in x- and y-direction. It thereby provides a clear method to design directivity characteristics for well-behaving electrostatic loudspeakers. In the next paragraph, an ideally shaped array loudspeaker is designed.

3.3.2 Spatial low-pass filter by diaphragm shaping

The requirements for a spatial reconstruction filter for loudspeaker arrays have been discussed in Section 2.5.2. It has been shown there that a low-pass k_x-filter is needed as a reconstruction filter. The spatial cut-off frequency is determined by the spatial Nyquist frequency $k_{x,\mathrm{Nyq}} = \pi/\Delta x$, where Δx is the spacing between the array elements. Theoretically a block-filter in the k_x-domain would be ideal, but because of the Fourier coupling (3.22) this would require a sinc-filter ($\sin(x)/x$) in the x-domain. Such a filter consists of many side lobes, i.e. side diaphragms alternating in phase and anti-phase. The feasibility of such an ideal filter is studied in the next paragraph. In this paragraph a simple and practical design is given, based on a single diaphragm with an optimized shape.

As a first trial, a rectangular diaphragm is taken, see Figure 3.7. The discretization of the driving functions for wave field synthesis in the xz-plane (Section 2.5) did not impose any requirements for the vertical directivity (k_y-filter). Therefore the designer is free to fulfill other demands. If it is desirable that ceiling and floor reflections are avoided, the width of the diaphragm in the y-direction should be large. If, however, part of the audience is sitting close to the arrays (viewing the array under a large vertical angle), vertical directionality should be small to obtain a frequency-independent response for the whole audience area. In that case, the width in the y-direction should be small. If only the x-direction is considered, the frequency-dependent part of the directivity function is given by the Fourier transform of a block-function with width $2a$:

$$L_{\mathrm{block}}(k_x) = \frac{\sin(k_x a)}{k_x a}. \tag{3.23}$$

If the loudspeakers are mounted tightly, the spacing between the centers of two adjacent diaphragms equals the width $\Delta x = 2a$. The attenuation of this k_x-filter at the Nyquist frequency is therefore $L_{\mathrm{block}}(k_{x,\mathrm{Nyq}}) = 2/\pi$, giving a level of only –3.9 dB with respect to the main lobe. The filter has side lobes of –13 dB at $k_x = \pm\, 1.43\pi/a$.

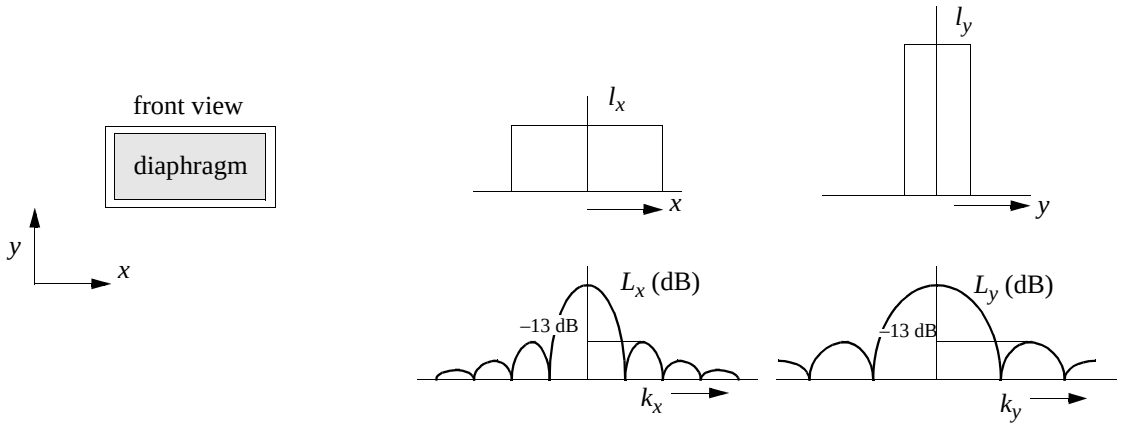
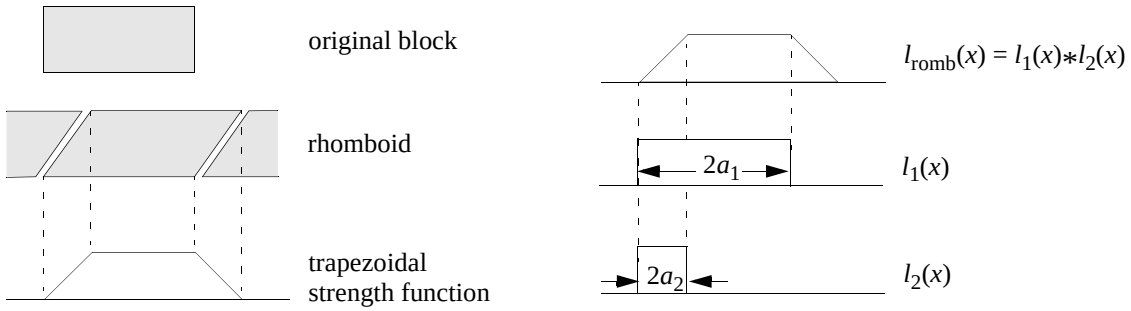
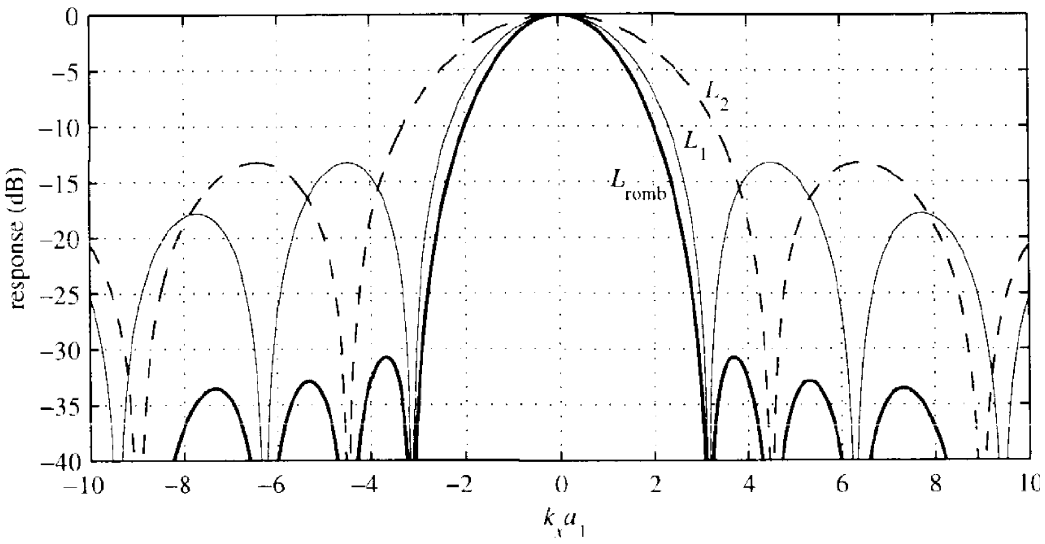


Figure 3.7 Directivity functions for a rectangularly shaped diaphragm. The width of the block-function is inversely proportional to the width of its Fourier transform, a sinc-function.



a.



b.

Figure 3.8 a. Construction of a rhombic diaphragm. **b.** k_x-filter for ratio of widths $a_1/a_2 = 1.43$.

The performance of the filter can be improved by smoothening the edges, but without increasing the spacing between adjacent array elements. This is realized by a rhombic diaphragm as shown in Figure 3.8a. Its strength function $l_{romb}(x)$ is trapezoidal: it is a convolution of two block functions with widths $2a_1$ and $2a_2$ respectively, $a_1 > a_2$. Therefore the directivity function is the product of two sinc functions:

$$L_{romb}(k_x) = L_1(k_x)L_2(k_x) \tag{3.24a}$$

with

$$L_1(k_x) = \sin(k_x a_1)/(k_x a_1) \tag{3.24b}$$

and

$$L_2(k_x) = \sin(k_x a_2)/(k_x a_2)\,. \tag{3.24c}$$

The performance of this filter can be optimized with certain constraints. The optimization task will be to find the ratio a_1/a_2 for which side lobes are reduced maximally while maintaining a reasonable response within the pass-band $|k_x| < k_{x,Nyq}$. It will be advantageous to choose a ratio

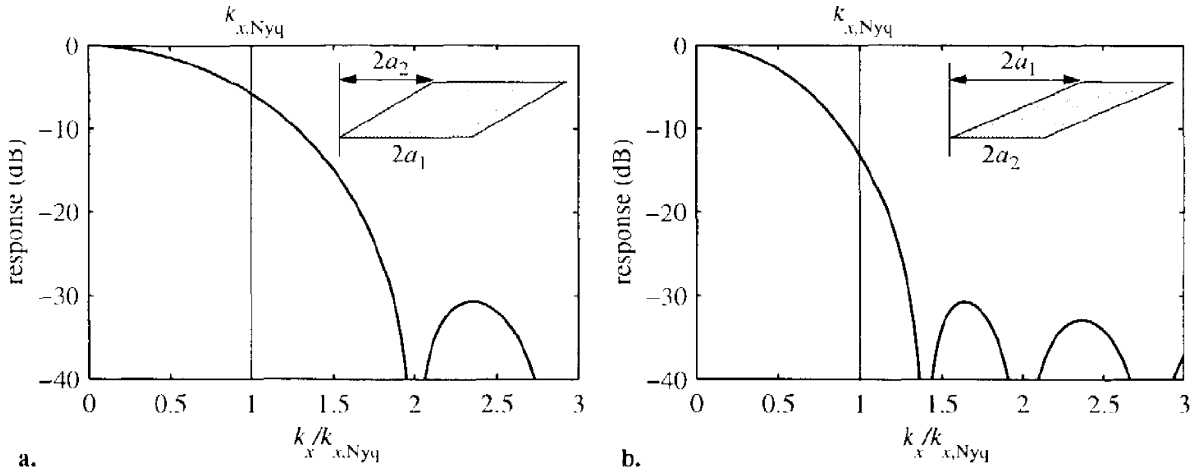


Figure 3.9 Spatial Nyquist frequency for a k_x-filter with $a_1/a_2 = 1.43$.

a_1/a_2 in such a way that the first side lobe of L_1 is canceled by the first zero of L_2, see Figure 3.8b. Because L_1 has its first side lobe at $k_x = \pm 1.43\pi/a_1$, while L_2 has its first zero at $k_x = \pm\pi/a_2$, it is clear that $a_1/a_2 = 1.43$ meets the requirement.

For any ratio, two possible rhombic diaphragms can be chosen: $\Delta x = 2a_1$ or $\Delta x = 2a_2$. The choice does not affect the level of the side lobes, but defines the position of the spatial Nyquist frequency with respect to the filter shape. In Figure 3.9 both settings are compared. The narrow filter of Figure 3.9b gives the best rejection in the stop-band (where $|k_x| > k_{x,\mathrm{Nyq}}$), but it also affects the pass-band considerably (–13 dB at $k_{x,\mathrm{Nyq}}$). The following paragraph describes how a better fit can be achieved by applying a discrete k_x-filter to the driving signals.

3.3.3 Spatial low-pass filter by discrete spatial convolution

Consider an array of identical elements, each with strength function $l(x)$. Now, instead of driving only one element with some driving function $Q(x, \omega)$, a series of $2N+1$ elements is driven by the same signal but at different gain. This results in a total strength function

$$l_N(x) = c_{-N} l(x + N\Delta x) + c_{-(N-1)} l(x + (N-1)\Delta x) + \ldots$$

$$c_{-1} l(x + \Delta x) + c_0 l(x) + c_1 l(x - \Delta x) + \ldots + c_{N-1} l(x - (N-1)\Delta x) + c_N l(x - N\Delta x)\,, \qquad (3.25)$$

where c_n is the weight of the nth array element, and Δx the spacing between the elements. The total strength function can be written as a sum

$$l_N(x) = \sum_{n=-N}^{N} c_n l(x - n\Delta x)\ . \qquad (3.26)$$

The Fourier transform of this sum,

$$L_N(k_x) = \int_{-\infty}^{\infty} \sum_{n=-N}^{N} c_n l(x-n\Delta x) e^{jk_x x} dx \ , \tag{3.27}$$

can be evaluated by interchanging summation and integration and calculating the Fourier transform of the elements

$$\int_{-\infty}^{\infty} c_n l(x-n\Delta x) e^{jk_x x} dx = \int_{-\infty}^{\infty} c_n l(x') e^{jk_x (x' + n\Delta x)} dx'$$

$$= c_n e^{jk_x n\Delta x} L(k_x) \, . \tag{3.28}$$

where $L(k_x)$ is the Fourier transform of the strength function $l(x)$ of a single element. Therefore, (3.27) reads

$$L_N(k_x) = L(k_x) \sum_{n=-N}^{N} c_n e^{jk_x n\Delta x} \, , \tag{3.29}$$

or, for symmetrical weights $c_n = c_{-n}$,

$$L_N(k_x) = L(k_x) C_N(k_x) \, , \tag{3.30a}$$

with

$$C_N(k_x) = c_0 + 2 \sum_{n=1}^{N} c_n \cos(k_x n\Delta x) \, . \tag{3.30b}$$

This directivity function is the product of a continuous part $L(k_x)$ and a periodic part $C_N(k_x)$ with period $2\pi/\Delta x$. In the space domain, the resultant strength function $l_N(x)$ can be written as a convolution of a continuous strength function $l(x)$ with a pulse sequence, see Figure 3.10.

Therefore, it is possible to reshape the directivity function of a single element by optimizing the weights of neighboring array elements. Of course, there is a restriction because of the periodicity of the Fourier transform of the discrete pulse sequence. Furthermore, the far-field conditions should be satisfied, which means that only a small number of array elements may be involved in the spatial convolution.

Two examples demonstrate the influence of a discrete spatial convolution filter with $N = 2$ (five array elements). Figure 3.11 shows the optimization result for a flat pass-band, i.e. for $|k_x| < k_{x,\mathrm{Nyq}}$. The spacing between the rhombic elements is $\Delta x = 2a_1$ (with $a_1/a_2 = 1.43$). The resultant k_x-filter (thick line) is a composition of the rhomboid's k_x-filter (Figure 3.9a) and the spatial convolution filter applied to the array elements. Only a slight change (2 to 3 dB) in the response is necessary to obtain a nearly flat pass-band. Note that, due to the periodicity of C_N, it is impossible to reduce the response in the stop-band without attenuating the response in the pass-band. The reason for this is that the spatial convolution filter is sampled at the same interval Δx as the synthesis operator. Therefore the same Nyquist condition is valid for both discretized functions of x.

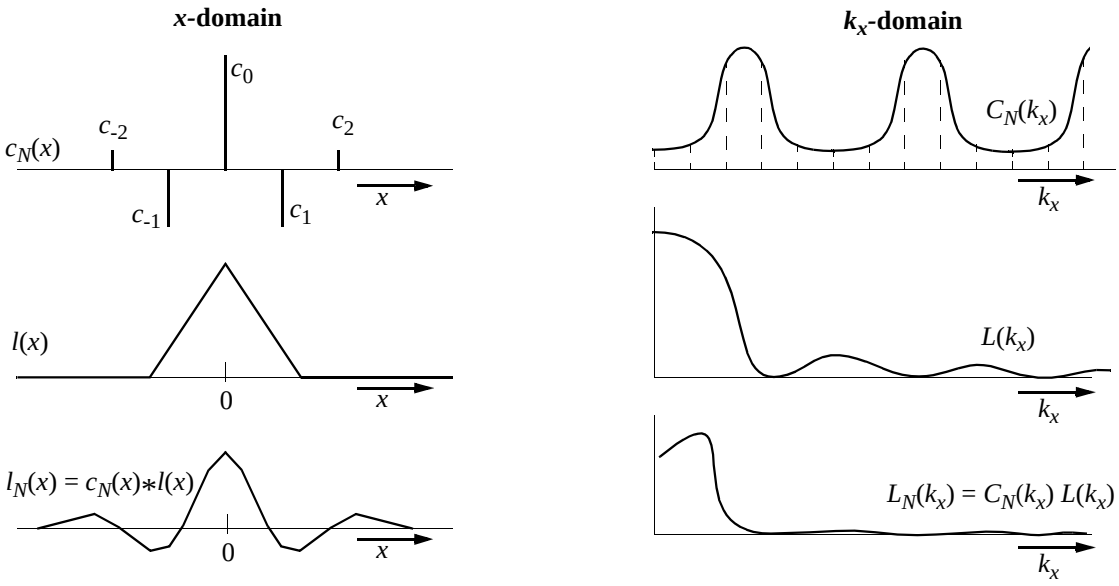


Figure 3.10 Strength function $l_N(x)$ can be constructed by spatial convolution of the strength function $l(x)$ of a single array element with the spatial pulse sequence $c_N(x)$. The coefficients c_n are equal to the gain of the array elements at location $n\Delta x$.

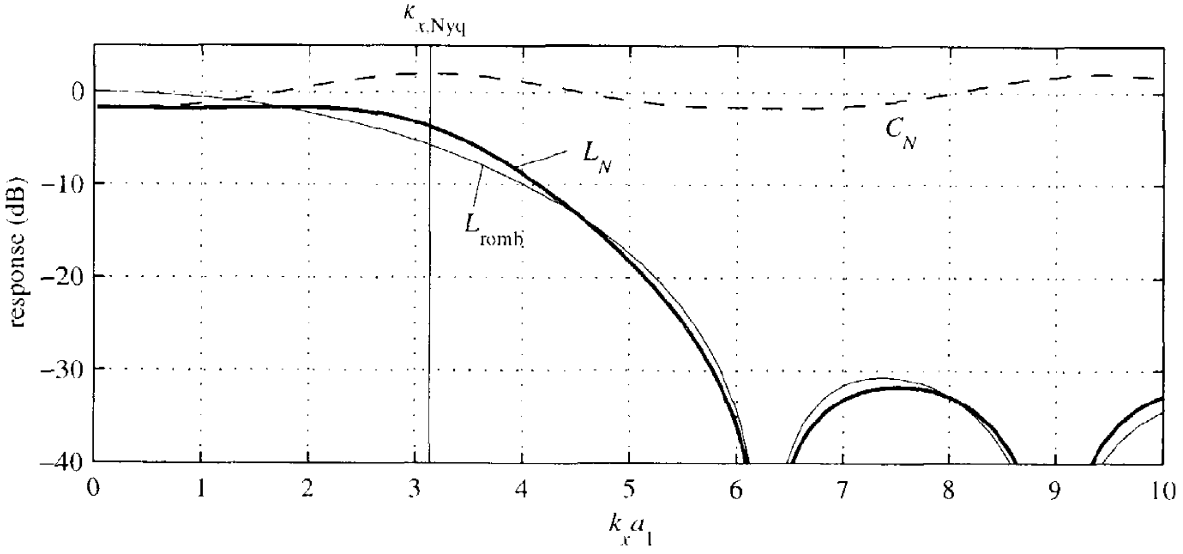


Figure 3.11 Optimization of the pass-band response for the rhomboids with $\Delta x = 2a_1$. The resultant k_x-filter L_N is the product of the directivity function L_{romb} of a single rhombic element and the spatial Fourier transform C_N for weighting coefficients: $\{c_0, c_1, c_2\} = \{1, -0.11, 0.02\}$.

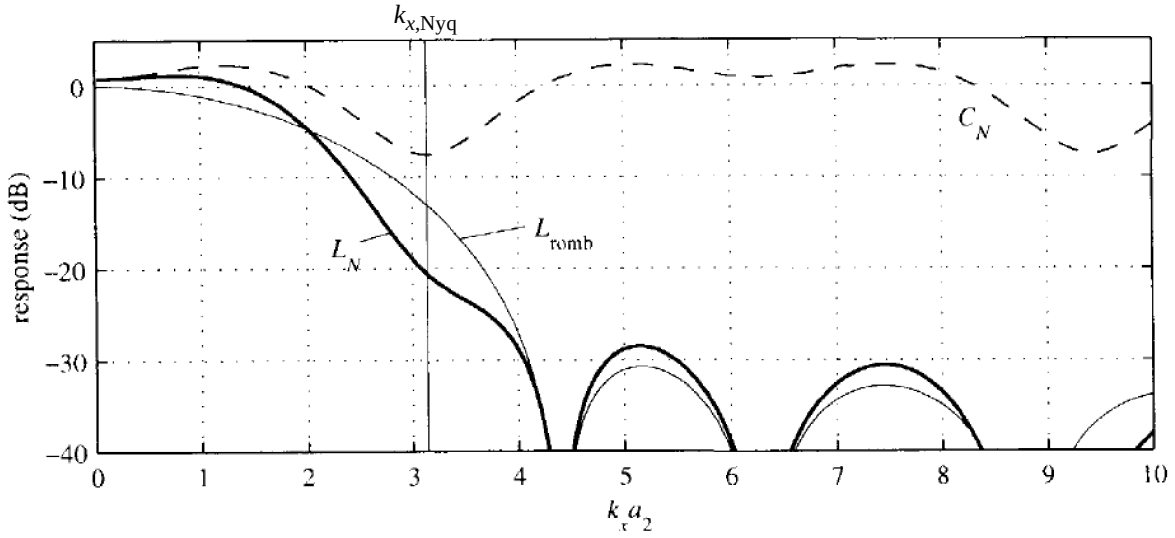


Figure 3.12 Optimization of the stop-band for the rhomboids with $\Delta x = 2a_2$. The attenuation for $|k_x| > k_{x,\mathrm{Nyq}}$ is greater than 20 dB. The resultant k_x-filter L_N is the product of the directivity function L_{romb} of a single rhombic element and the spatial Fourier transform C_N for weighting coefficients: $\{c_0, c_1, c_2\} = \{1, 0.17, -0.12\}$.

For an optimum stop-band attenuation, the rhomboids with $\Delta x = 2a_2$ should be taken, see Figure 3.12. The k_x-filter of this rhomboid is already down –13 dB at the spatial Nyquist frequency, and can be further reduced to less than –20 dB in the whole stop-band. This is achieved by a positive weight at the nearest neighbor and a negative weight at the next neighbor element. Such a configuration of weights produces a wider, but more gently tapered central diaphragm, without increasing the sample interval Δx.

3.4 Design of arrays for audio reproduction

In designing a loudspeaker array for audio reproduction, both the spatial and temporal properties of the loudspeaker elements need to be considered. As for the *spatial* properties, a controlled directional behavior is required, if little spatial aliasing is permitted. The *temporal* conditions can be summarized by stating that the sound quality of the loudspeaker element should be satisfactory within a large frequency range. This condition is formulated here in a subjective manner, and involves a large number of physical properties like distortion, sound power, frequency response and efficiency. High quality loudspeakers for conventional reproduction will generally be applicable for audio reproduction by wave field synthesis as well, but there are some different needs because of the special nature of arrays. In conventional electrodynamic loudspeaker systems, the audio bandwidth is usually covered by two or three drivers with complementary frequency ranges, because it is virtually impossible to construct a single driver covering the whole audio spectrum with enough sound power. Since the sound power in an array system is generated by more than one element, the sound power criterion is moderated

considerably. Only if the virtual source is exactly at the position of an array loudspeaker, all power has to be generated by one driver. By excluding the vicinity of the array from the virtual source area, the power problem can be solved adequately. Practically, at least two or three array elements bear the burden of sound power generation, if a strip of width Δx parallel to the array is excluded from presence of virtual sources. Therefore, small low-power array elements with large frequency ranges can be used in array systems.

If higher sound power is required, two-way or three-way arrays can be constructed. These arrays consist of two or three different drivers that are active in different sub-bands. The upper bound of the bandwidth determines the maximum spacing Δx for that specific sub-band loudspeaker. For example, a two-way array system with crossover frequency f_x requires a spacing between the woofers of $\Delta x_{woofer} = c/2f_x$ maximally.

Two types of (one-way) loudspeaker arrays are investigated in this section: first, an electrodynamic array with (nearly) full range low-power drivers, and second, an electrostatic array with controlled directivity elements. After a description and qualification of both types of loudspeaker arrays, measurements on the synthesized wave fields are presented. Finally it is considered which one of both types of arrays is more suitable for integration in a wave field synthesis system for sound reproduction.

3.4.1 Electrodynamic array

Directivity

Because of nodal cone resonances, the directivity characteristics of an electrodynamic loudspeaker are difficult to control. Nevertheless, the directivity pattern of a rigid piston is a reasonable approximation in the low and mid frequency range. Most commercially available electrodynamic loudspeakers have circular diaphragms with equal directivity in the vertical and horizontal plane. If the horizontal directivity is desired to differ from the vertical directivity, an oval shaped diaphragm is eligible, with the wider dimension aligned in the direction of the highest directivity.

The directivity function of an oval diaphragm is shown in Figure 3.13, together with those of a circle and a rhomboid. Here, the wider dimension of the oval lies in the horizontal plane, which improves the directional behavior in that plane. For each of the three shapes, the directivity in the vertical plane is kept within bounds by restricting the dimensions in that direction. The directivity of the oval piston is higher than that of the circular piston, but lower than that of the rhomboid. The dimensions of the oval diaphragm correspond to those of the loudspeaker shown in Figure 3.14, which is a candidate for the construction of an electrodynamic array.

Three polar patterns of that loudspeaker are shown in Figure 3.15. The measured directivity patterns for 4 kHz and 16 kHz show that the directivity of the loudspeaker lags behind that of the rigid piston model, which must be a result of irregular diaphragm motion. The asymmetrical shape for the highest frequency is presumably caused by the attachment of the voice coil wires at one side of the diaphragm.

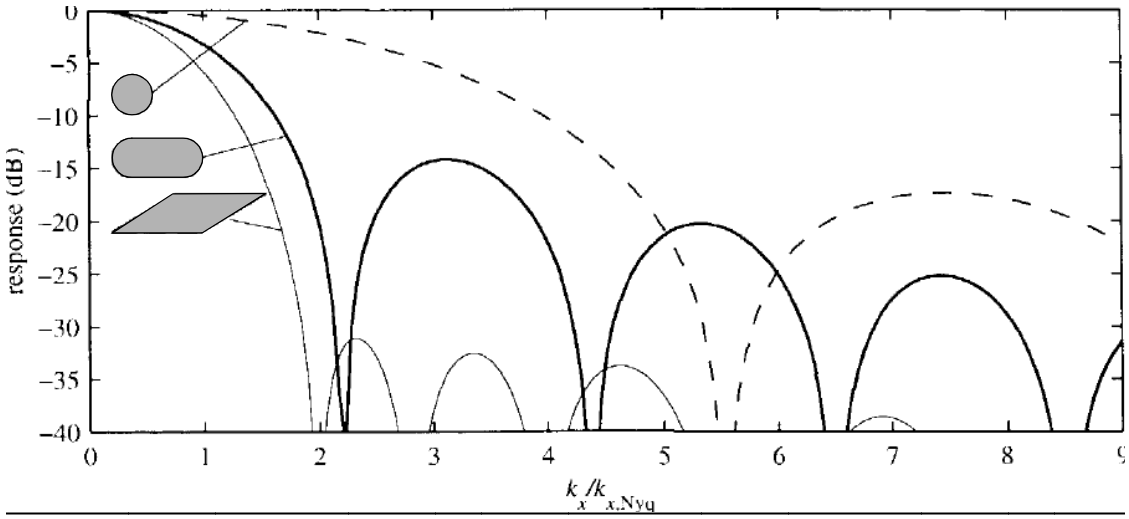


Figure 3.13 Comparison between the directivity functions of a circular, an oval and a rhombic diaphragm. The oval consists of two semicircles, each with radius 0.22 Δx, and a rectangle of width 0.56 Δx, where Δx is the distance between nearly touching adjacent diaphragms. The radius of the circular diaphragm is 0.22 Δx; the rhombic diaphragm has a ratio $a_1/a_2 = 1.43$, where $\Delta x = 2a_1$.

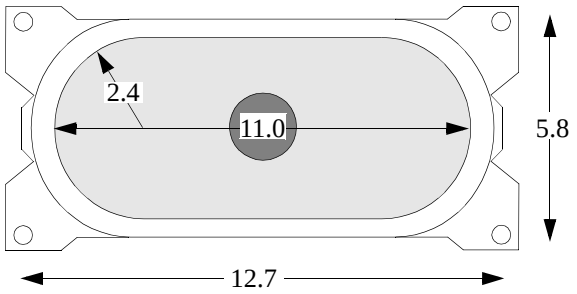


Figure 3.14 Electrodynamic loudspeaker with oval shaped diaphragm (Philips ID 25952/X8). Shown measures are in centimeters.

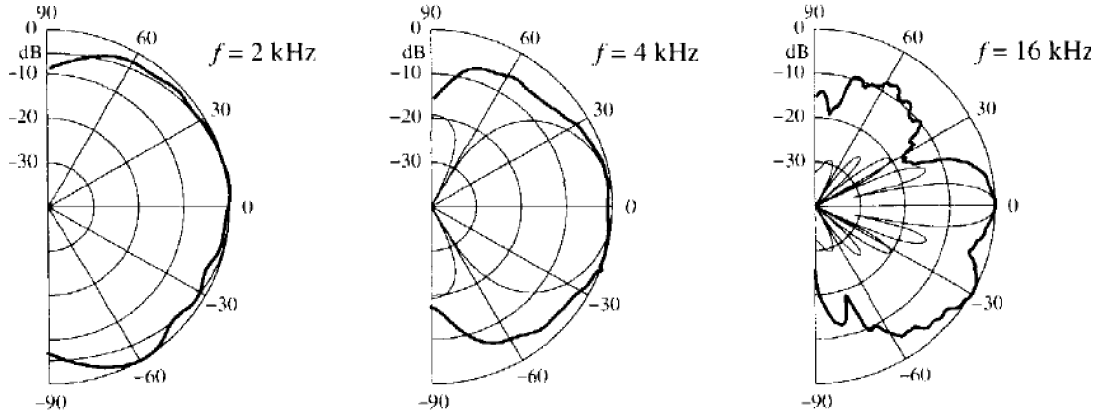


Figure 3.15 Measured (thick) and theoretical (thin) polar patterns for the oval electrodynamic loudspeaker.

From these directivity measurements it is obvious that this type of loudspeaker is not specially equipped for reducing spatial aliasing in the synthesized wave field. However, because of its very reasonable sound quality (tested by concentrated listening), wide frequency range (140 Hz – 20 kHz), low weight, small vertical size and low costs, it was decided to use these drivers in building an electrodynamic array. The design of this array is based on the following considerations concerning the volume of the enclosure in which the loudspeakers are mounted.

Enclosure volume

The loudspeakers can be placed either in separate cabinets or in one large hollow bar. If the latter option is chosen, acoustic coupling from within the bar can be avoided by partitions, creating separate enclosures. The volume (per element) determines the resonance frequency ω_0 and quality factor Q_0. The measured values for this type of loudspeaker in a large baffle are $\omega_0 = 2\pi \cdot 134$ rad/s and $Q_0 = 1.4$ rad m/Ω, respectively. If the loudspeaker is mounted in an enclosure, both the resonance frequency and the quality factor will be higher, since the compliance of the enclosed air, given by (Beranek 1954)

$$C_{AB} = \frac{V_B}{\kappa p_0}, \tag{3.31}$$

where V_B is the volume of the enclosure, $\kappa \approx 1.4$ for adiabatic compression of air, and $p_0 \approx 10^5$ N/m^2, lowers the total acoustical compliance C_A that appears in the denominator of (3.5) and (3.6). Therefore V_B should be large to keep ω_0 and Q_0 low. However, a large cabinet volume raises the dimensions (and weight) of the array, so that a compromise must be made in choosing the size of the cabinet. A practical volume will be that for which C_{AB} approximately equals C_{AS}, so that both compliances contribute equally to the total compliance. With $C_{AS} = 1.3 \cdot 10^{-8}$ m^5/N for this driver, a volume of about 1.8 dm^3 is calculated. Figure 3.16 plots the measured frequen-

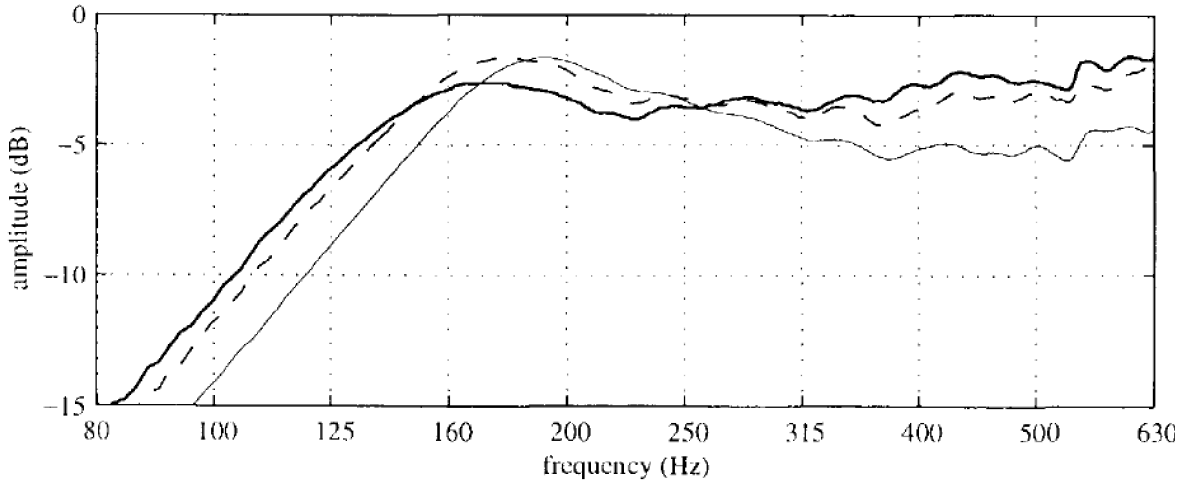


Figure 3.16 Frequency response near the resonance frequency of the loudspeaker of Figure 3.14, as a function of the enclosure volume: 2.0 dm^3 (thick solid); 1.7 dm^3 (dashed); 1.4 dm^3 (thin solid).

cy response as a function of the loudspeaker enclosure volume. As the volume increases, the resonance peak is damped more, while the top position shifts towards lower frequencies. From these results it appears reasonable to keep the volume in the 1.7 – 2.0 dm^3 range.

Loudspeaker spacing

The spacing between the loudspeakers determines the value of the spatial aliasing frequency. From experience with synthesis arrays for sound enhancement, it is known that an aliasing frequency of at least 1500 Hz is favorable for correct sound localization (i.e. Vogel 1993 and Start 1997). This is not surprising: if the wave fronts are reconstructed correctly below 1500 Hz, then the interaural time differences, which dominate localization (Wightman et al. 1992), will correspond to those heard for the original wave fronts. If the aliasing frequency is decreased (by increasing Δx), the localization accuracy decreases as well. Furthermore, in informal listening tests it was noticed that the apparent width of the image increases with increasing spacing between the loudspeakers. This can be understood by realizing that aliased waves, which travel in wrong directions, interfere with the correct waves, thereby causing phase distortion between the ear signals, and hence will degenerate interaural correlation. Potter (1993) showed that the perceived image of a true source broadens with decreasing interaural correlation for frequencies below 1600 Hz. Therefore, the aliasing frequency of an audio reproduction system based on wave field synthesis should not drop far below 1600 Hz. This condition preserves spaciousness for virtual sources, one of the requirements mentioned in Section 1.5.

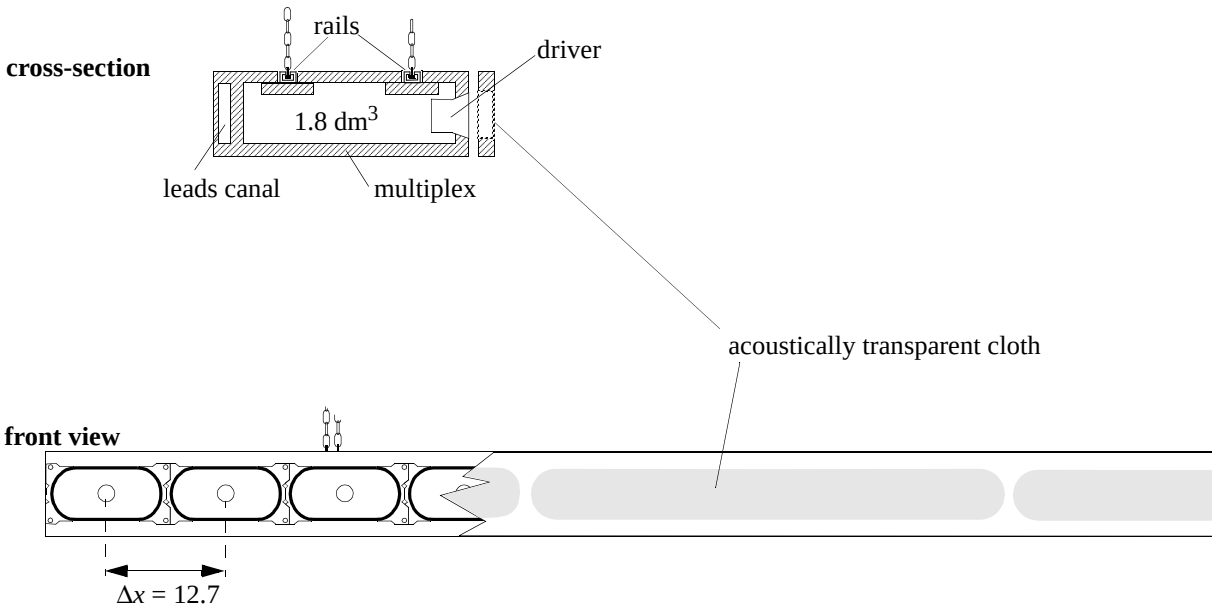


Figure 3.17 Cross section and front view of the 16-channel electrodynamic arraybar. Dimensions of the bar: l×b×h = 203×27×9 cm. Weight: 21 kg. Power handling of single loudspeaker: 3 W rms, giving 90 dB SPL at 1 m.

Based on these considerations, a 16 channel electrodynamic array was built with an enclosure volume of about 1.8 dm^3 per array element, and a loudspeaker spacing of 12.7 cm, which is the minimum spacing between these loudspeakers, see Figure 3.17. This results in a spatial aliasing frequency of 1340 Hz.

3.4.2 Electrostatic array

Because of the promising directivity characteristics, also an electrostatic array-strip with rhombic elements was built. It consists of 16 elements with a spacing Δx = 11 cm (with f_{alias} = 1545 Hz), see Figure 3.18. The rhombic elements have an approximate ratio a_1/a_2 = 1.43. The width in the vertical direction is kept small, 7 cm, to avoid strong directionality in that direction: listeners will not notice strong spectral changes when standing up or sitting down. The width of the elements is set to $2a_1$ as in Figure 3.9a*.

The k_x-filter response can be calculated directly from the directivity patterns via the relation $k_x = 2\pi f \sin(\varphi)/c$. From a theoretical point of view, the measured polar pattern for any frequency may serve as data in calculating the element's k_x-filter, provided that the measured polar patterns are consistent†. Figure 3.19 proves that there is good agreement between the

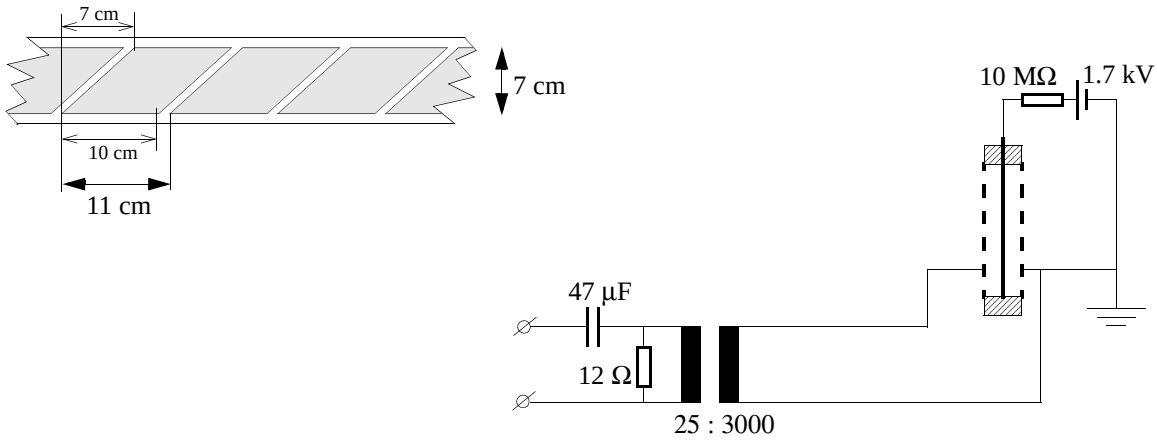


Figure 3.18 Front view and electric circuit of the prototype electrostatic loudspeaker array. The *RC*-filter at the transformer input has a low frequency cut off at 280 Hz. The grounded perforated plate as well as the membrane are made of one piece; 16 separate plates are used as (hot) signal electrodes. The total thickness of the array-strip is 0.5 cm.

* An electrostatic array with the other setting, $\Delta x = 2a_2$, is difficult to construct on condition that its *y*-directionality be small. It would consist of very narrow and fragile elements.

† Polar patterns are mapped to part of the k_x-filter: the polar pattern for *f* Hz is mapped between $k_x = -2\pi f/c$ and $+2\pi f/c$. High frequency polar patterns thus provide more information of the k_x-filter than low frequency patterns. In theory, all mapped polar patterns fit the k_x-filter response exactly.

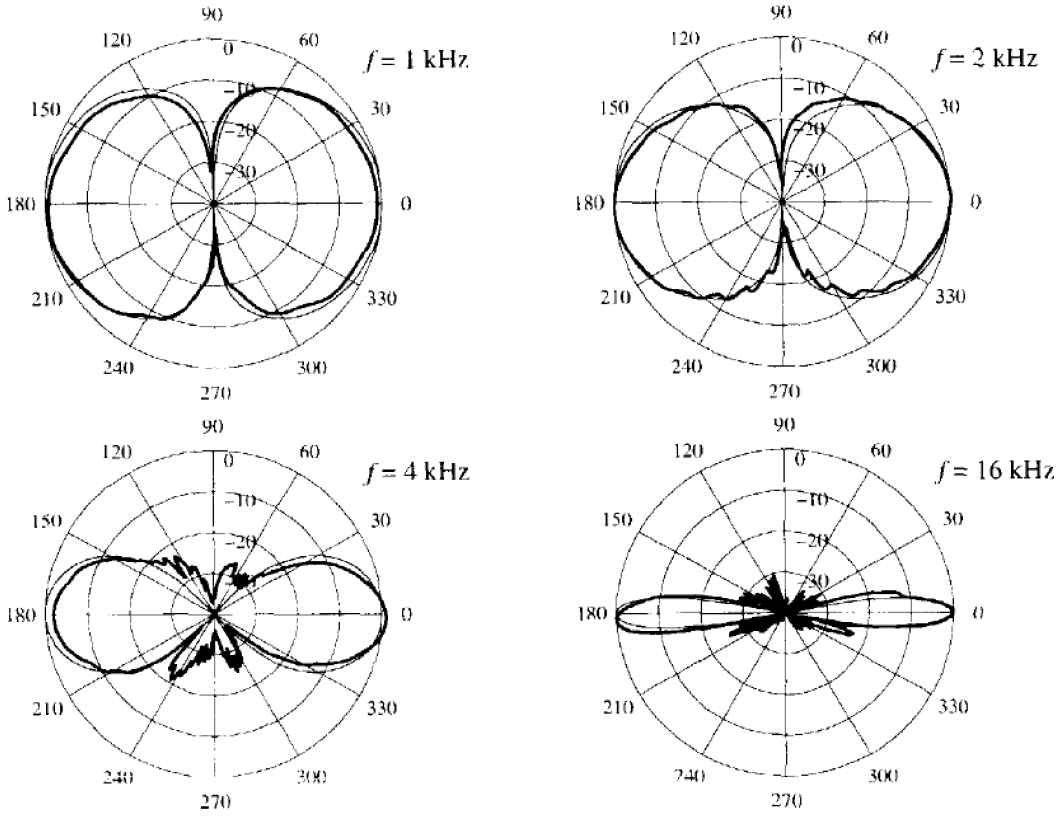


Figure 3.19 Measured (thick) and theoretical (thin) polar patterns for one element of the electrostatic array.

measured and calculated directivity characteristics. It is therefore allowed to estimate the k_x-filter of an element from the measured polar data for only one high frequency component, see Figure 3.20. The filter bears reasonable comparison with the predicted response. The shape of the main lobe is very close to the theoretical shape, except near $k_x = 70$ m^{-1} where a side lobe 'joins' the main lobe. Though the level of the side lobes is higher than predicted, the position of the lobes is as expected. The measurements indicate that the directivity characteristics of simple electrostatic loudspeakers can indeed be controlled reasonably well.

A disadvantage of the high directionality of the rhombic array elements is that the virtual sources must be confined to areas with a limited angle of incidence towards the audience. Otherwise, the reduction of the bandwidth becomes audible. A solution for this side effect of k_x-filtering is the use of curved arrays, see Figure 3.21. The angle of incidence φ at the array, as seen from a listener looking in the direction of the virtual source, determines the perceived bandwidth of that source. If the angle is small, the bandwidth is large and vice versa. The listening area with full bandwidth is therefore small for virtual sources to the side. If the corner is 'rounded', a much larger listening area is created.

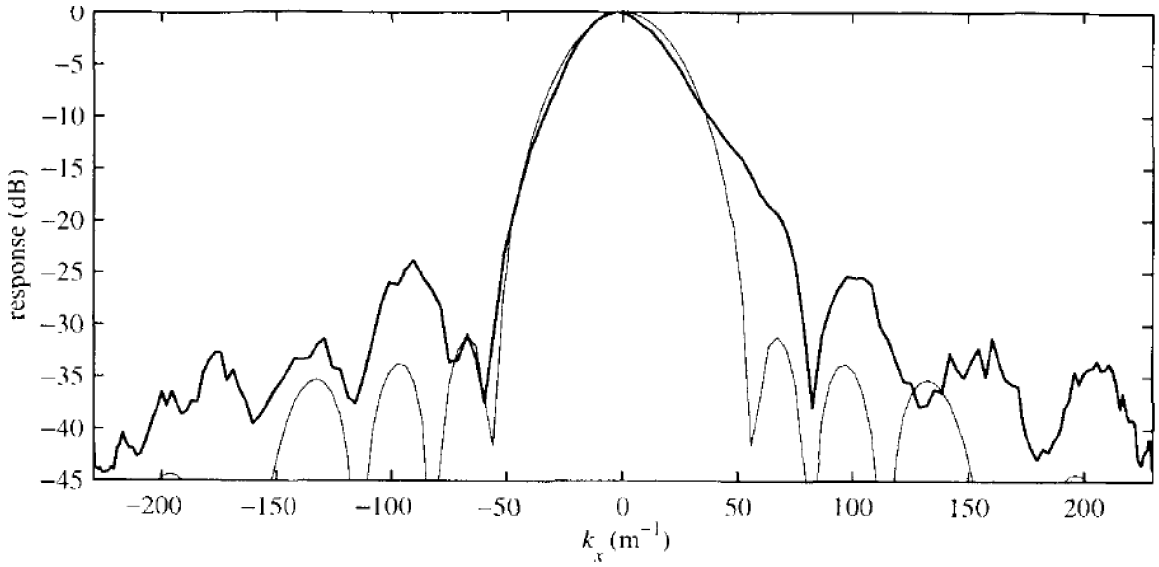


Figure 3.20 Comparison between the k_x-filter of a rhombic element as calculated from measured polar data (f = 12.5 kHz), and the theoretical rhombic response (thin line).

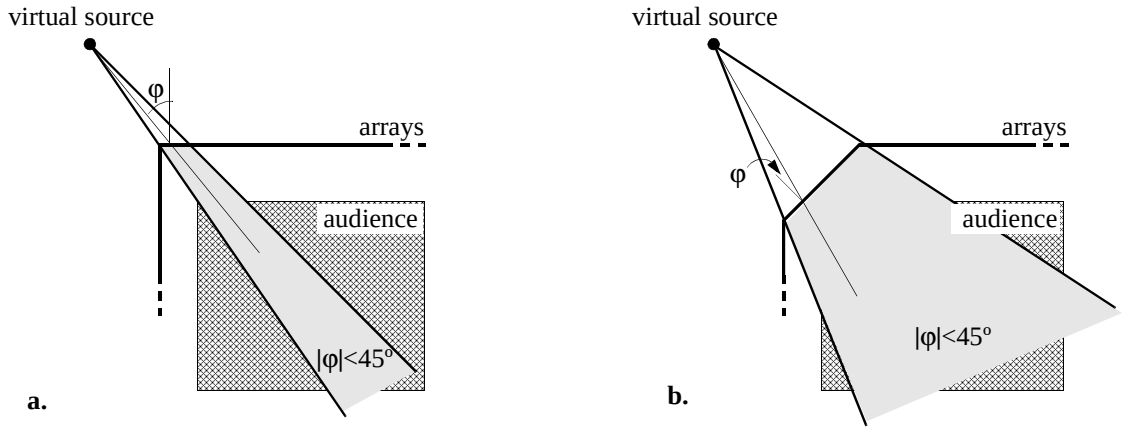


Figure 3.21 a. If the arrays consist of highly directional loudspeakers, only a part of the audience area is radiated in full bandwidth. Here the coverage area for incident angles φ of less than 45° is drawn for a source near the corner of an L-shaped array. **b.** If the corner is canted, a much larger area of coverage is achieved.

3.4.3 Comparative measurements

The frequency response of one element of both the electrostatic and electrodynamic array are measured in an anechoic room, see Figure 3.22. This figure clearly shows the different frequency dependence of electrostatic and electrodynamic loudspeakers. From (3.9) it follows that a flat frequency response is globally expected for an electrodynamic loudspeaker, while according to (3.15) a slope of 6 dB per octave is found for an electrostatic loudspeaker. This is confirmed by the measurements.

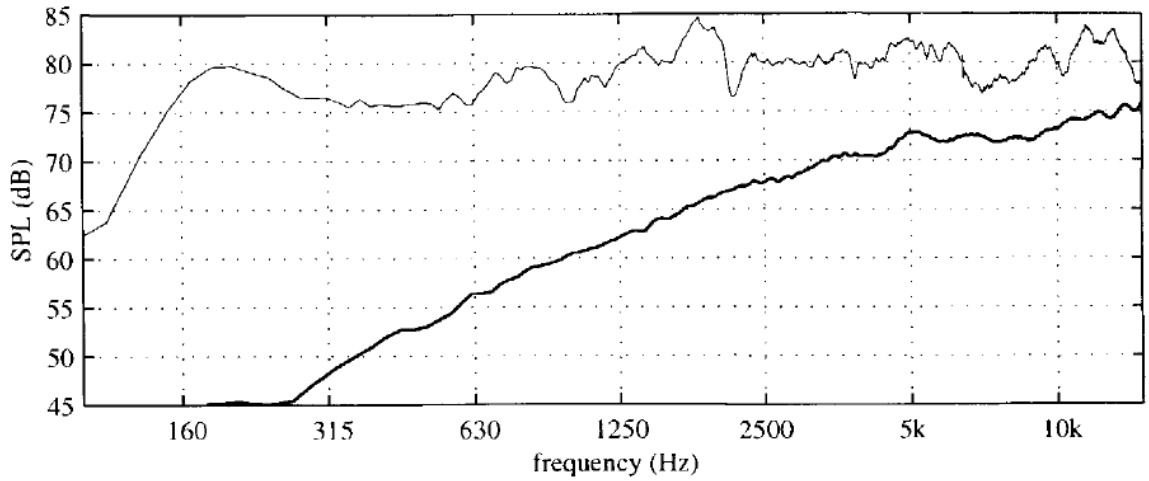


Figure 3.22 On-axis sound pressure level of one electrostatic rhombic loudspeaker (thick) and one electrodynamic oval loudspeaker (thin), driven at the same input voltage level.

Because of the small membrane area (70 cm^2), the sensitivity of the electrostatic element is quite low. Especially the poor bass response can be problematic, because bass equalization will cause audible distortion if the membrane exceeds the linear excursion region. Enlargement of the surface area of the membrane will enhance the bass response, but will also affect the aliasing frequency and/or the directivity in the vertical plane, in an indirect manner. An additional array of (electrodynamic) low-frequency loudspeakers is probably a better solution, however at the expense of a more complicated construction and higher costs.

It should also be mentioned that the sound quality of music produced by the prototype electrostatic array was suffering from irregular sputtering, probably caused by leakage of the polarization charge. A second prototype with a more robust construction exhibited less sputtering, but also had lower sensitivity due to a larger gap between the electrodes. Nevertheless, the (first) prototype was sufficiently stable to explore its behavior in a series of plane wave synthesis measurements.

Figure 3.23 plots a set of measured impulse responses of a plane wave parallel to the arrays. At t = 0.7 ms the main wave front arrives at the microphone array. While the electrostatic array reconstructs the plane wave remarkably well, the response of the electrodynamic array shows an aliasing tail. In a single impulse response taken from this data set, see Figure 3.24a, this is visible as a sequence of separate pulses for the electrodynamic array, while the effect is much weaker for the electrostatic array. This can be understood as follows: the loudspeakers at the right end of the electrodynamic array still radiate much high frequency energy in the direction of the microphone at x = –0.45, so the contribution of each loudspeaker is visible as a separate pulse in the total response. In the frequency domain (Figure 3.24b), the pulse train causes a comb filter spectrum. The dips of the electrostatic response are less pronounced than those of the electrodynamic response.

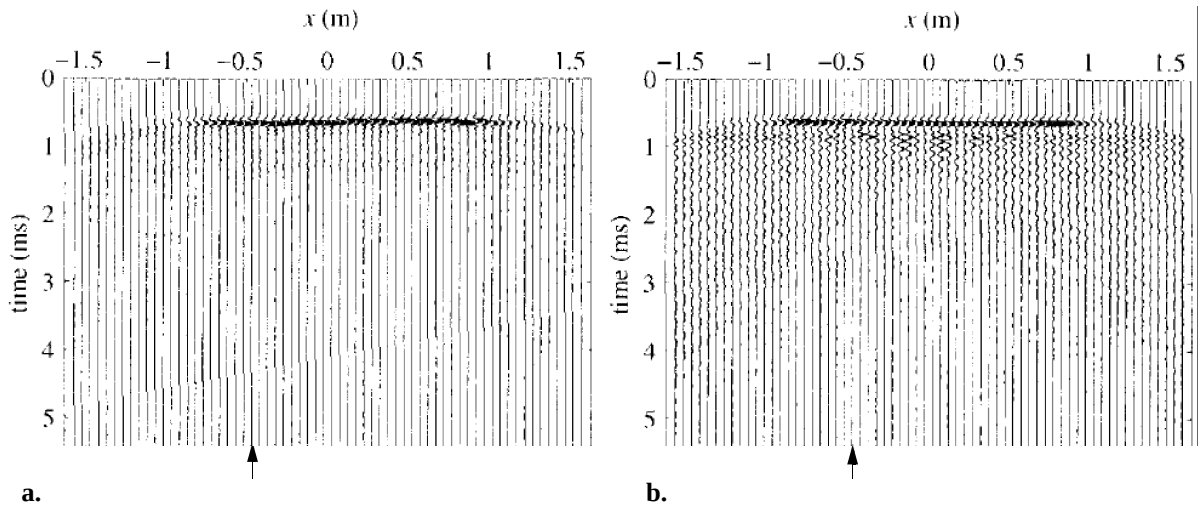


Figure 3.23 a. Measured *xt*-diagram of a plane wave synthesized by the electrostatic array in an anechoic room. The measurement microphone is moved stepwise along a line parallel to the array. The impulse responses are convolved in the time-domain with a Gaussian wavelet. The cut-off frequency of the Gaussian wavelet is 3.1 kHz, which is twice the Nyquist frequency (2.35) for Δx = 11 cm. (Distance of measurement line: 2 m; angle of incidence of the plane wave: 0º; array from x = –0.9 to +0.9 m)
b. Measured *xt*-diagram of a plane wave synthesized by the electrodynamic array. The cut-off frequency of the Gaussian wavelet is 2.7 kHz, which is twice the Nyquist frequency for Δx = 12.7 cm. Clearly the effects of spatial aliasing are visible as wavelets after the actual wave front. (Array from x = –1.0 to +1.0 m)

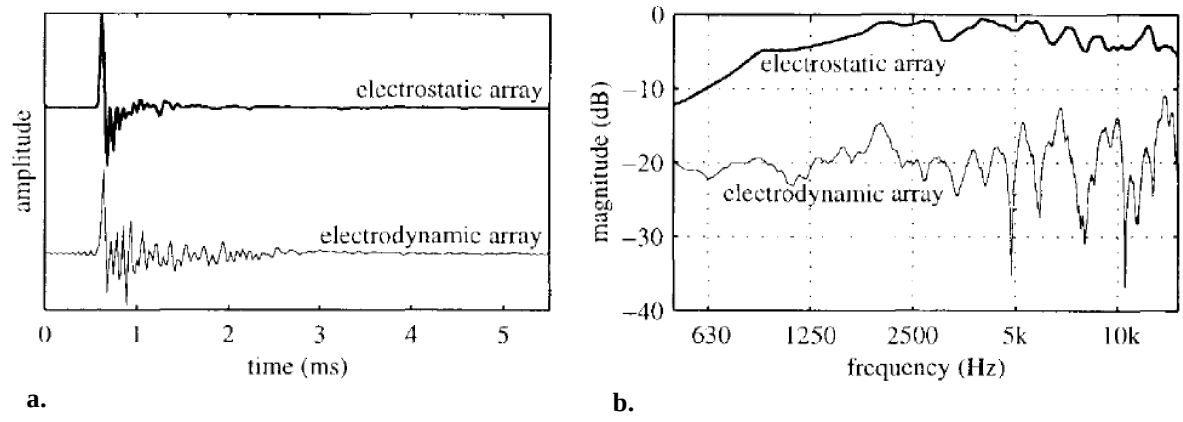


Figure 3.24 a. Impulse responses of the electrostatic and electrodynamic array, reproducing a plane wave. Taken from the *xt*-diagrams of Figure 3.23 at the position of the arrow. **b.** Spectra of these impulse responses (arb. offset).

Whether these effects of spatial aliasing are audible depends on external circumstances like the type of source signal and the listening environment. In a static situation, i.e. motionless listener and source, the aliased wavelets are perceived as a subtle broadening of the virtual source, rather than as coloration. If the listener or the virtual source moves, the comb filter dips will shift gradually, which can be understood from Figure 3.23b where the intervals between aliased wavelets increase as the receiver moves away from the center. These spectral changes are clearly audible in an anechoic room, but tend to drown in the reflections of a normal listen-

ing room. Start (1997) also observed that the coloration (of the virtual source) introduced by a wave field synthesis system for direct sound reinforcement, can be neglected in comparison with the coloration caused by the reflections of a concert-hall.

In this context it should be mentioned that a conventional stereophonic recording suffers from severe coloration during reproduction on two loudspeakers in an anechoic room. Especially if the listener moves off the central axis, a tonal sweep is clearly heard. This effect is explained by Figure 1.16, where a different pattern of comb filter dips is produced at different receiver positions. In a normal listening environment however, it is difficult to distinguish between the sound color of a phantom source in the midst of two loudspeakers and a true source at that position. Seen in this light, the spectral irregularities introduced by a wave field synthesis array with low directivity seem much less dramatic.

Eventually, because of its good sound quality, reliability and wide bandwidth, the electrodynamic array has been selected for the laboratory demonstration system. If the efficiency of the electrostatic array can be enhanced, e.g. by optimizing the construction of the electrodes, the application of these arrays will become feasible too.

3.5 Conclusions

In this chapter, the general properties of electrodynamic and electrostatic loudspeakers have been investigated, as well as their applicability for wave field synthesis. It is found that the driving signals for electrodynamic arrays need a filtering of +3 dB per octave, while for the electrostatic arrays –3 dB per octave is needed, besides an additional phase factor depending on the position of the virtual source.

Since effects of spatial aliasing in the synthesized wave field depend on both the radiation angle and on frequency, a clear understanding of the directional behavior of transducers is of great importance. It is shown how the shape of the diaphragm of a transducer can be characterized by a directivity function, which in its turn defines the k_x-filter that is a tool for reduction of spatial aliasing. The optimization of the shape of the k_x-filter is described in terms of spatial convolution filters, which allows for a generalization of the concept of directivity from a single array element (with an optimally shaped diaphragm) towards subarrays (with optimized weighting coefficients).

An electrodynamic as well as an electrostatic array have been designed and tested. The electrodynamic array consists of small drivers with fairly good sound quality in a wide frequency range, but without controllable directivity characteristics. Measurements on the electrostatic array with diaphragms of optimized shape prove that spatial aliasing can be reduced considerably compared to the electrodynamic array. However, the bass response and the overall sound quality of the electrostatic loudspeakers are still insufficient. Therefore, the electrodynamic array is selected for use in the laboratory wave field synthesis system.

Chapter 4

Concept and Implementation of WFS Reproduction

4.1 Introduction

The concept of sound reproduction, as developed in this chapter, is based on the theory of wave field synthesis (WFS) that goes back to the Kirchhoff-Helmholtz integral expression. Several successive simplifications and approximations of this expression are necessary to construct a realizable concept. The final formulation is held generic to allow the implementation of a wide range of reproduction systems that differ in flexibility, quality and dimensions.

The reproduction concept comprises three phases: recording, storage and/or transmission, and playback (Section 4.2). The recording method is similar to existing multi-channel recording techniques, but an impulse is given towards the development and integration of surround sound microphone techniques. The transmission phase is essentially multi-channel, with an additional coding channel reserved for source coordinates and parameters. At the playback stage, an intelligent processing unit produces signals for the loudspeaker arrays. Besides multi-channel programs, also stereo recordings can be reproduced. An enhanced reproduction method is proposed for these two-channel recordings, offering a wider listening area.

Based on this general concept, a reproduction system is designed and installed in the Delft Laboratory of Seismics and Acoustics. It consists of 8 input channels, a digital signal processing system, 128 outputs and 160 loudspeakers in a closed linear array configuration. This system is described in Section 4.3.

Finally, several applications of the reproduction concept are discussed in Section 4.4.

4.2 General concept

The concept of sound reproduction in this thesis starts from the stipulation that the reproduced sound field should resemble the original one as close as possible. This, however, is not generally assumed by sound engineers, since often the original sound field is considered not to be ideal, or in need of an acoustic compensation of the lacking vision. It will be seen that the final concept presented here offers the desired broad flexibility of control.

Suppose a physically exact copy of the primary sound field is to be generated in the reproduction room, as illustrated in Figure 4.1. Though this ultimate goal may seem impossible to reach, the Kirchhoff-Helmholtz theory in Section 2.2 proofs that there *is* a physical foundation for proper reproduction. The ideal reproduction method is investigated briefly, before dealing with several practical limitations in applying this theory.

4.2.1 Ideal solutions for reproduction

Through the Kirchhoff-Helmholtz integral (2.6) it is stated that the wave field inside a closed surface is completely determined by the pressure and particle velocity of the primary wave field at that surface, as long as the enclosed volume is source-free. With a source-free sub-volume from a concert-hall to be copied, the pressure and particle velocity at the surface of that sub-volume must be known. This knowledge can be acquired in at least three ways:

1. The sound at the surface is recorded with planar arrays of pressure and velocity microphones, and stored at a multi-track recorder. Afterwards planar arrays of loudspeakers with dipole and monopole characteristics are used to re-create the sound field in the (anechoic)

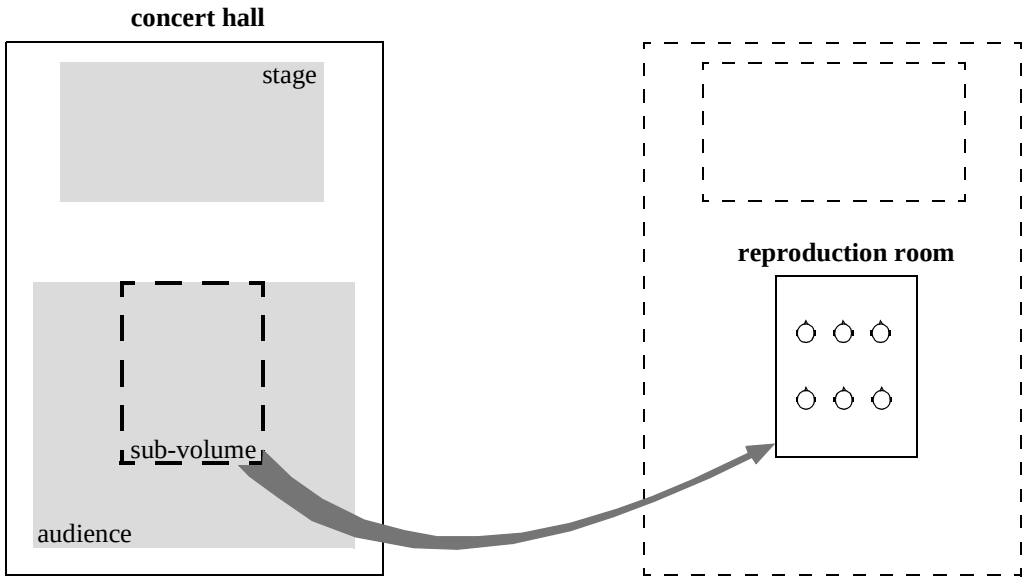


Figure 4.1 The concept of ideal or natural reproduction. The sound field in a certain favorable area of the concert-hall is re-created in the reproduction room.

reproduction room. Snow (1953) already mentioned a similar approach: a linear array of microphones in the recording room is connected to a linear array of loudspeakers in the reproduction room. A transfer function, as found through the Kirchhoff-Helmholtz integral, is however not given in his work, nor does it specify the characteristics of the transducers. Camras (1968) experimented with a sub-volume enveloped by seven front and five surround microphones/loudspeakers. Though he considers the Huygens' principle to be responsible for the wave front reconstruction in the front, again a mathematical description is missing.

2. If the source positions and acoustic properties of the hall do not change during the concert, a considerable simplification of solution 1 is possible. In that case, the transfer functions from the sources to the pressure and velocity microphones at the surface of the sub-volume are time-invariant. By *measuring* these in advance of the event, only the *direct* sound of each source needs to be recorded on a separate track during the concert. The monopole and dipole loudspeaker-feeds are eventually found by convolving the source signals with the measured transfer functions.

3. If the geometry of the hall is known, as well as the acoustic properties of the construction materials, the transfer functions of solution 2 can also be *calculated*. Here again, only the direct sound of each source is recorded. Before reproduction, the mirror images of each source are calculated using the model of the hall, to account for the wall reflections. Diffusion and diffraction should be included in the model. The transfer functions from the original sources to the monopole and dipole loudspeakers are built from the driving functions of virtual sources at the positions of the original sources and their mirror sources.

In theory, each of these solutions should provide an identical copy of the original sound field in the sub-volume. However, the big difference between the methods is the amount of data stored or transmitted (sound tracks, transfer functions, hall model). A large redundancy of recorded information is reduced in going from the first to the second solution. The third solution may be regarded as additional data compression, but its benefit lies in the parametrization of the sound field, based on wave theory of sound fields in enclosures. Since the parametrized sound field is split into source signals and an acoustic hall-model, this solution offers a very flexible method of (re)production: sources can be synthesized at different positions than the original ones, and different hall-models can provide completely different ambiences. Therefore, the third solution will not only comply with the premiss of *natural* sound reproduction stated at the beginning of Section 4.2, but goes further: application of wave field synthesis for virtual reality systems has become possible.

4.2.2 Practical solutions for reproduction

In order to design a realizable reproduction method derived from the third solution, a series of further adaptations is required. While the previous adaptations can be classified as *data compression* steps (no physical information lost), the following adaptations can be

regarded as *data reduction* steps (no relevant perceptual information lost). A division is made between the reduction of the complexity of the transfer functions and the reduction of the number of loudspeakers.

A. Reduction of the complexity of the transfer functions

The third ideal solution comprises the recording of only the direct sound, the acoustic modeling of the enclosure and afterwards calculating the array transfer functions. The acoustic modeling of concert-halls is however still a time-consuming job with varying success, because the cumulative action of partly reflecting surfaces is extremely difficult to model (Dalenbäck et al. 1994). Furthermore, each array element would have a different and complicated transfer function, in practice to be realized by a unique convolution filter per loudspeaker. This would require a very powerful real-time processing system. Therefore a different approach is followed, one that exploits the perceptual restrictions of human hearing.

The sound field in a hall can be divided in three perceptually distinct components that are directly related to the impulse response: direct sound, early reflections and reverberation (Section 1.4). Each of these components has its own spatial attributes that should be preserved (Section 1.5). Therefore, each of these components is treated separately during the three stages of sound reproduction: recording, transmission and playback (Section 4.2.3 to 4.2.5).

B. Reduction of the number of loudspeakers

A large amount of loudspeakers is needed for physically correct reproduction. In practice several simplifications are allowed.

1. Because the previously mentioned planar arrays are necessarily *discretized* instead of continuous, a first approximation is already made: the reproduced sound field is identical to the original sound field only within a certain bandwidth, that is limited by the spacing between the loudspeakers according to the spatial Nyquist frequency (2.35).

2. Instead of monopole *and* dipole transducers, only one of both is needed, if cancelation of the sound field outside the enclosure is of minor importance. In that case, the Rayleigh I or II integral (Sections 2.2.1 and 2.2.2) provides the array driving functions. A mirrored sound field, radiated backwards from the arrays, can be avoided by building the loudspeakers into the walls of the reproduction room.

3. If the planar arrays are replaced by linear arrays, a further reduction of loudspeakers is possible. This is allowed in applications where sound sources and listeners are restricted to the horizontal plane. If the absence of reproduced sound (e.g. reflections) in the median plane becomes noticeable, the construction of layered arrays and a sparsely covered ceiling may be considered, being an intermediate solution between planar and linear arrays (Figure 4.2).

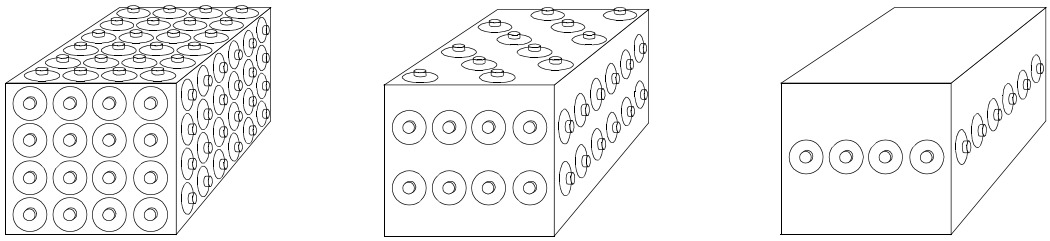


Figure 4.2 Stripping of the planar loudspeaker arrays in the reproduction room. **a.** Fully covered for 3D-reproduction. **b.** Sparsely covered ceiling and layered linear arrays for simplified 3D-reproduction. Two layers and hence two source heights are supported in this example. **c.** Linear arrays in the horizontal plane for 2½D-reproduction: all sources are imaged at one height.

4. If the scenery takes place solely in the frontal plane, and the reflective waves from back or sides are unwanted or unimportant, a frontal linear array will suffice. In this way, the array serves as a window through which the sound entering from the front is imaged. This elementary version of a wave field synthesis array is well-suited for modular construction, because both the planar and (layered) linear extensions can be readily built from it.

A detailed discussion of the reproduction concept is given in the next paragraphs. A preview is shown in Figure 4.3. The stage of recording is drawn at the top. The microphone signals, which ideally contain either only direct sound or only reverberation, are recorded on a multi-track machine or broadcasted directly (transmission stage: middle). Besides these signals, the playback system (bottom) uses geometrical parameters that are transmitted as well. The software program in the playback system is denoted as **W**: the synthesis operator.

4.2.3 Recording

The proposed recording method is related to commonly used multi-track recording methods, in the way that it uses microphones that record the direct sound only, as well as microphones that also pick up the reverberant sound.

Spot microphones

The direct sound of a source is acquired by nearby spot microphones that are usually directional. In stereophony, the spot microphone signals take a minor role in the final two-channel down-mix: they just add some definition and stability to the image created by the main microphone pair. In wave field synthesis, the spot microphone signals are of much more importance, because here it is necessary to create virtual sources with their own source signals. Ideally the direct sound of each source is recorded on an individual track or channel. During playback the source signal is fed to a virtual source of which the position corresponds with that of the original source (if natural reproduction is strived for). Besides the individual source signals, therefore also the coordinates of the sources must be delivered to the reproduction operator **W**.

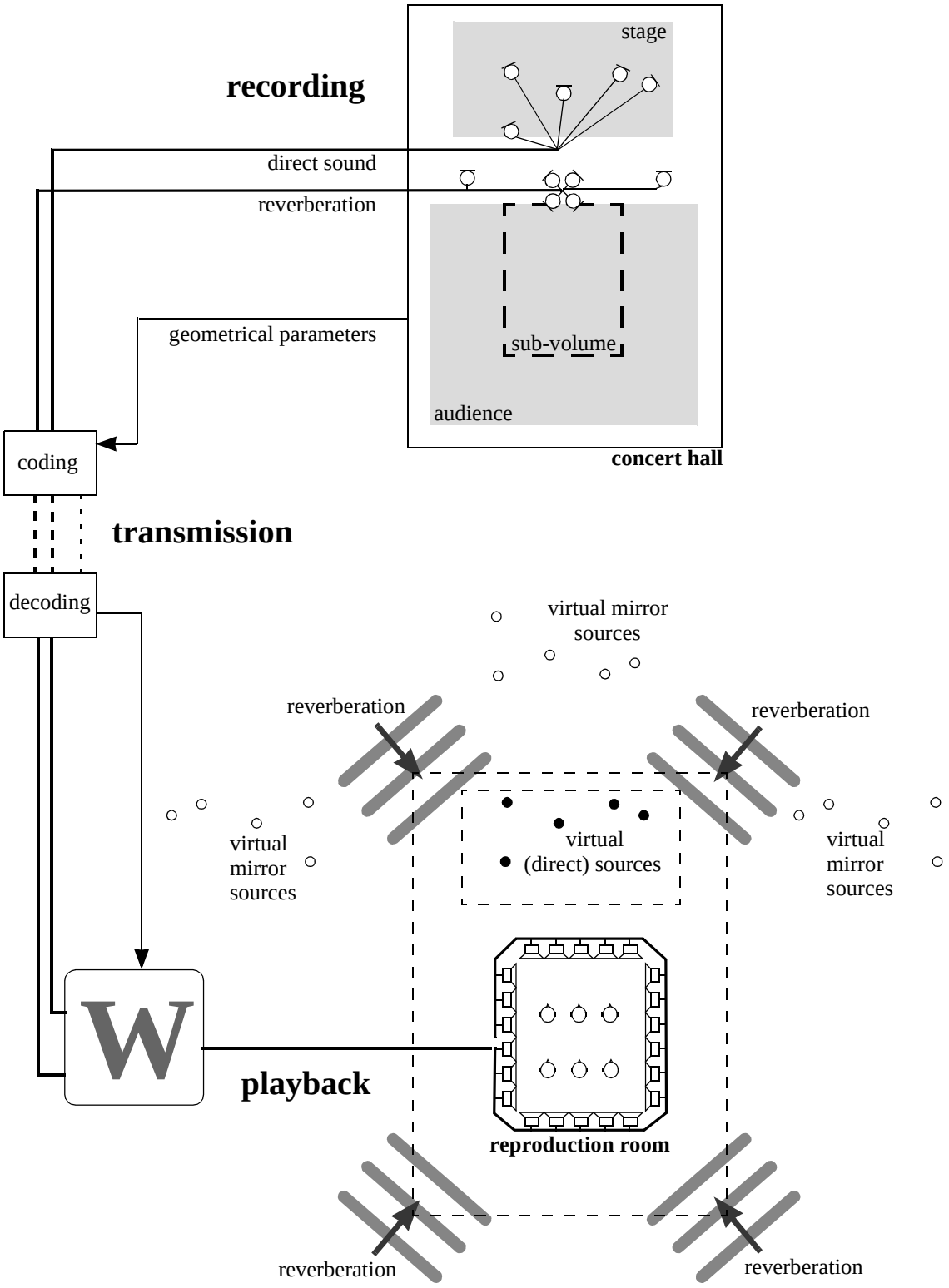


Figure 4.3 General reproduction scheme. The recorded signals, together with parameters for the playback process, are stored and/or transmitted. Operator **W** prepares the array signals for playback.

If many sound sources are active simultaneously, a very large number of spot microphones, sound tracks and transmission channels would be required. To deal with this inconvenience, the concept of *notional sources* is adapted from Berkhout et al. (1993), who developed it for sound reinforcement based on wave field synthesis. A notional source merely consists of a signal to which a source position has been assigned. This idea is illustrated by Figure 4.4. A group of sources, e.g. a string section of an orchestra, can be recorded either with individual microphones (close-miking), or with one microphone at an appropriate distance from the section. In the latter case, the compound microphone signal can not physically be associated with one source. It is therefore assigned to a notional source at the center of the section*. This method reduces many physical sources to less notional sources, at the expense of a slightly decreased spatial accuracy of reproduction, and less flexibility of processing.

From experience with spot microphone signals in wave field synthesis, it is known that crosstalk between the microphone signals has a degenerating influence on the definition of the reproduced virtual sources. Strong crosstalk leads to decreased stability and increased apparent width of the reproduced sources. This phenomenon can be understood by realizing that crosstalk causes the sound of the original source to be distributed among more than one virtual source. Crosstalk can be suppressed by placing the spot microphone closer to the source and arranging larger distances between the sources. Though these remedies are commonly accepted in studio recordings, they may be difficult to defend in live recordings where artistic and esthetic arguments are also to be considered.

Moving sources (e.g. in opera, drama, sports) must be recorded with spot microphones that are kept at a constant distance from the source. The coordinates of the source should be tracked regularly. This way, effects of motion are excluded from the recorded signal, but can be synthesized during playback by making use of the varying source coordinates.

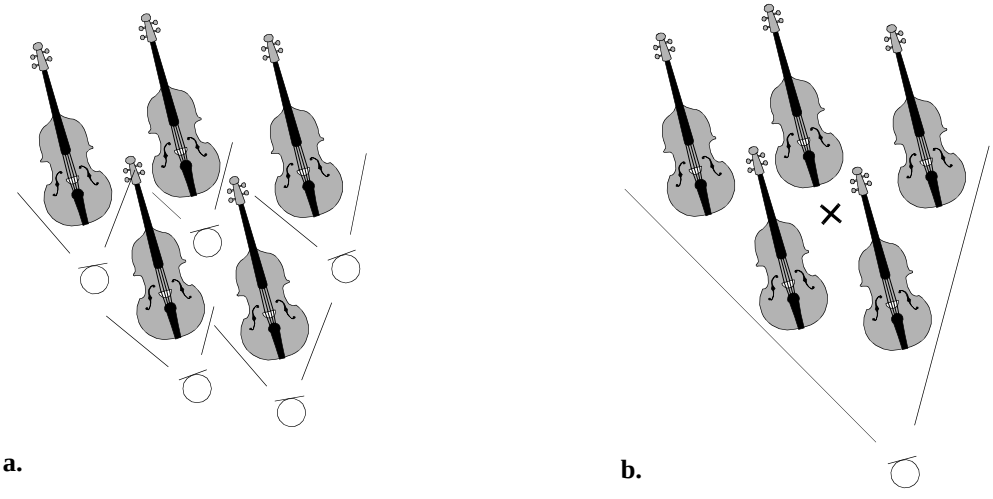


Figure 4.4 a. The sound of each source is recorded by a nearby directional microphone. **b.** The sound of a group of sources is recorded with one directional microphone placed at some distance from that group. The microphone signal can be assigned to a notional source in the center (cross).

* A *notional* source refers to a special situation during recording, while a *virtual* source refers to any input signal of the playback system in wave field synthesis. A notional source is reproduced as a virtual source.

Ambient microphones

In stereophony, the ‘main microphone pair’ (Section 1.6.3) is carefully positioned in front of the ensemble to obtain the desired place of phantom sources within the stereo image, and the desired balance between direct and reflective sound. In many cases these microphone signals can also be used for reproduction by wave field synthesis, see Section 4.2.5. However, in the present concept, the term ‘main’ is not appropriate any more, because imaging for wave field synthesis depends largely on the processing of *spot* microphone signals. Therefore, instead, the word ‘ambient’ is introduced to denote all microphone techniques that capture the reverberant sound for the purpose of wave field synthesis. Speaking in terms of impulse responses, the ambient signals ideally consist of late reflections and reverberation. In practice, also direct sound and early reflections will be recorded, but these will not influence the reproduced image if the direct/reverberation ratio is sufficiently low.

The need for recording techniques of ambient sound has grown strongly since the introduction of surround sound. A few techniques have already been proposed. The ambisonic reproduction method (Section 1.7.3) has encouraged the development of a four-channel coincident microphone system with electronically controllable directivity and orientation (SoundField). Theile (1996) uses a four-channel ORTF-like microphone configuration intended for use in 2/2 or 3/2 surround sound recordings (Figure 1.20b). Both microphone techniques belong to a special class of recording techniques based on the idea that sound must be recorded all around, because it arrives at the listener from all around. Figure 4.5a shows the general idea

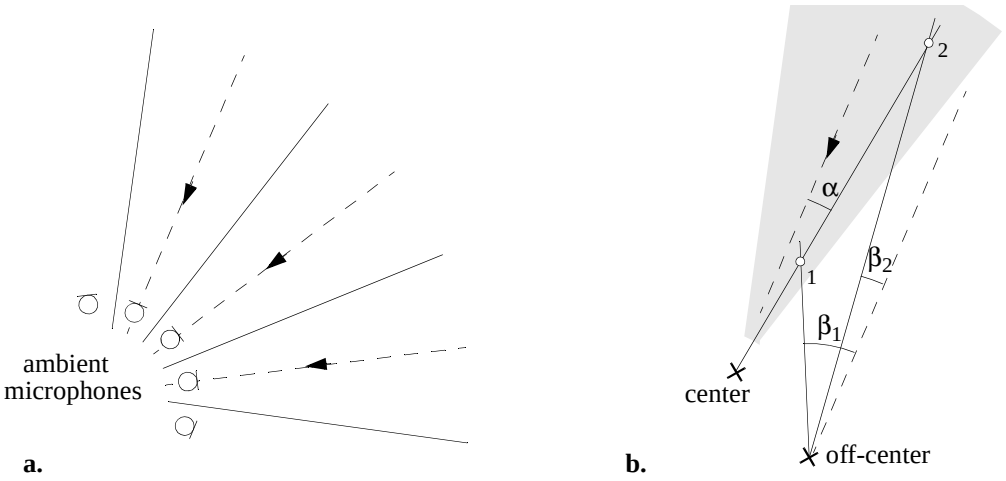


Figure 4.5 **a.** The sound field is recorded by directing the ambient microphones towards angular sectors of the acoustic space. The degree of correlation between the signals is determined by the radius of the circle of microphones and their directivity characteristics. During playback, the microphone signals are reproduced as plane waves along the dashed bisectors. Note that the radius of the microphone circle is zero for the SoundField microphone, and 15 cm in the ORTF-like configuration. With a cardioid directivity pattern, four sectors provide sufficient coverage. **b.** The directional error during playback depends on the listener position. The gray area is the angular sector of one particular microphone. The waves originating from sources 1 and 2 are reproduced as plane waves along the arrowed bisector. For a listener at the center, the directional error of these particular sources equals α. For a listener off-center, the directional error is smaller as the source is further away: $\beta_2 < \beta_1$. Since late reflections can be considered to originate from far-away mirror sources, this reproduction technique yields a fair reconstruction of these waves.

behind these techniques. Clearly, this approach leads to a ‘point solution’: only at the center position a fair reconstruction can be achieved. If, however, the method is used only for reproduction of reverberation, the directional errors at off-center positions are kept within bounds, see Figure 4.5b. Moreover, the human auditory system is not very sensitive for directional errors of individual late reflections, because these reflections are usually too dense to create precisely located auditory events (Blauert 1983). Thus, in principle this microphone technique provides an efficient way to image reverberation.

4.2.4 Transmission

For efficient transmission of data, the signal-flow chain in the reproduction system must be interrupted at the point of the lowest data rate. In contrast with stereophonic reproduction, this point does not coincide with the loudspeaker signals: in wave field synthesis, the number of loudspeaker signals is much higher than the number of microphone signals. This implies that the down-mix coefficients need to be delivered together with the microphone signals.

These coefficients are mainly geometrical parameters which inform the synthesis operator **W** about the source positions (direct sound), the reflection model (early reflections) and the plane wave angles (reverberation) that are used for preparing the array driving functions. The geometrical parameters may also include the directivity characteristics of the sound sources, as these can be synthesized correctly according to driving function (2.27). If the rate of parameter acquisition and transmission is suffiently high, *moving* sources can be synthesized as well.

An example of transmission channel coding of microphone signals and parameters is given in Section 4.2.6.

4.2.5 Playback

Since bare microphone signals are received at the playback stage (instead of loudspeaker signals as for e.g. stereophonic reproduction), an intelligent playback system is necessary to synthesize the sound field. This system, consisting of the synthesis operator **W** and the loudspeaker arrays, can be made as flexible as desired. This aspect may be of particular interest for commercial applications: the complexity of the system can range from a domestic set-up, with fixed source settings and only a frontal loudspeaker array, to a professional theater system with control of sources and their virtual acoustic environment.

A schematic representation of the synthesis operator is given in Figure 4.6. By default, the direct sources are imaged as virtual sources at their original position. Notional source signals, referring to more than one source, are imaged at the center of the original contributing sources. The early reflections are synthesized using the mirror source concept, depicted in Figure 4.7a. Reflective waves can be imaged by creating additional virtual sources at the position of the mirror sources. The reflection coefficients of the walls, usually exhibiting frequency dependence, can be taken into account by feeding the direct sound to wall absorption filters. For this purpose the reflection coefficients must be sent along with the geometrical parameters to the playback system.

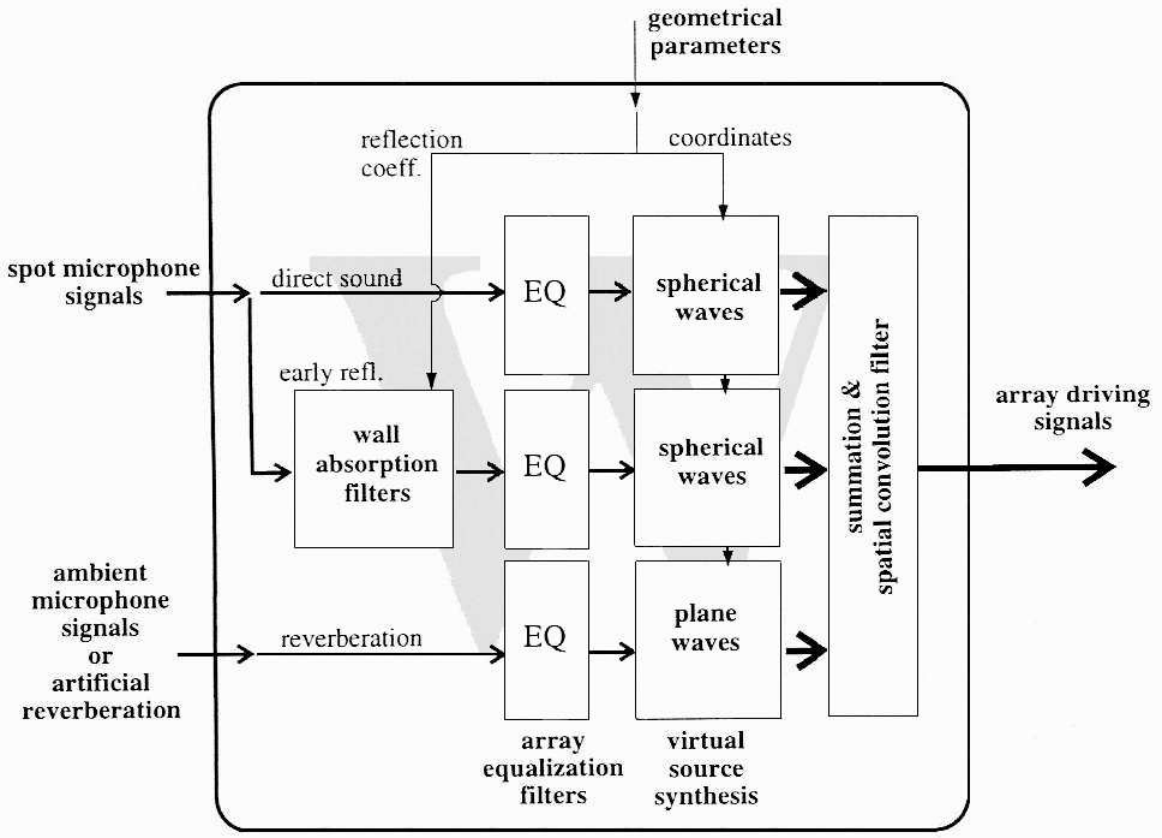


Figure 4.6 The synthesis operator **W**. The spot microphone signals for the direct sound unit are also used as input for the early reflections unit. Direct sound as well as early reflections are reconstructed as spherical waves, while the reverberation signals are reproduced as plane waves.

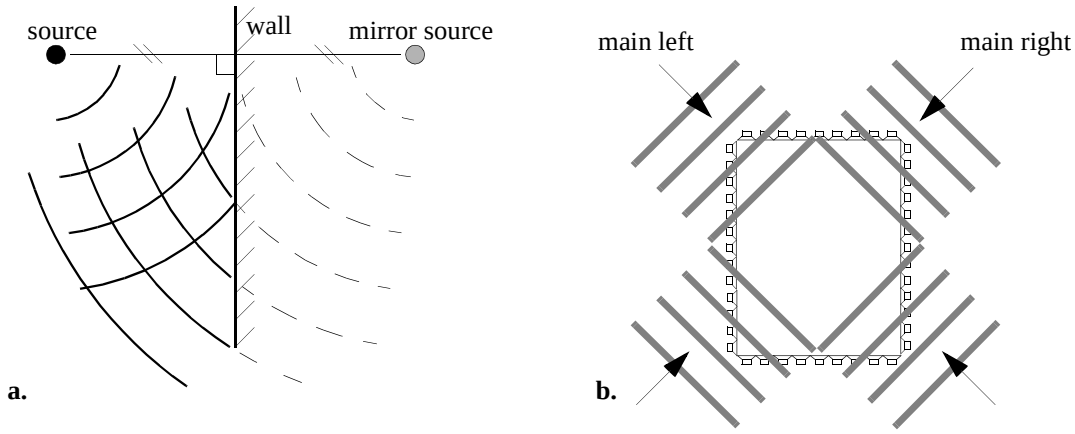


Figure 4.7 **a.** Mirror source model. If sound is reflected by a wall, the reflected waves correspond to that of a fictive mirror source at the other side of the wall. In wave field synthesis, both the source and its mirror source are synthesized as virtual sources. **b.** Plane wave synthesis of ambient microphone signals, including the stereo main pair signals.

The late lateral reflections, that are held responsible for a sense of envelopment (Bradley et al. 1995), are incorporated in the reverberation signals by using the ambient microphone technique described in Figure 4.5. These waves are reproduced as plane waves along the original bisectors (Figure 4.7b). Artificial reverberation may serve as an alternative for recorded reverberation. The stereo 'main' pair signals, if available, can be reproduced similarly (see also the next sub-section).

Returning to Figure 4.6, it can be seen that all signals pass a set of equalization filters before the actual synthesis takes place. The filter shape depends on the loudspeaker type and the source position with respect to the arrays, as argued in Section 3.2. After synthesis, the produced virtual source driving functions (spherical or plane waves) are added and, if necessary, spatially convolved according to Section 3.3.3.

The reproduction of *moving* sources is also possible in the present concept. Moving sources are important in virtual reality applications, because they create action and suspense by drawing the listener into the scenery. In order to allow fast and smooth motion, a time-dependent synthesis operator $\mathbf{W}(t)$ must be used. This requires additional processing power, because the positions of the source as well as its mirror sources are altered continuously.

4.2.6 Compatibility

The compatibility issue is concerned with upward and downward compatibility to stereophony and 3/2 surround sound. In this application, upward compatibility is the possibility of playing back old program material on the new system. Downward compatibility refers to the way in which old systems can cope with program material developed for the new system.

A. Upward compatibility

Upward compatibility is obtained by applying the *virtual loudspeaker* concept. In this concept, some appropriately placed virtual sources are fed by the loudspeaker signals originally meant for stereo or 3/2 surround.

For the stereo material, an equilateral triangle may be arranged (Section 1.6.1), but this would give the usual small listening area of two-channel stereophony. Instead, the stereo signals can be reproduced by means of two far-away virtual loudspeakers. If these virtual loudspeakers are placed at infinity*, two plane waves angled $\pm 30^{o}$ are produced (Boone et al. 1995, Geluk 1994), with several advantages:

1. the direction of the notional sources remains at $\pm 30^{o}$, independent of the listening position;

2. the intensity of the virtual loudspeakers is (almost) independent of the listening position;

3. the reproduced wave field is decoupled from the listening room: the wave fronts cannot be associated with a realizable (point) source position within the room.

* Of course the corresponding infinite delay is omitted and sound level is kept at a reasonable height.

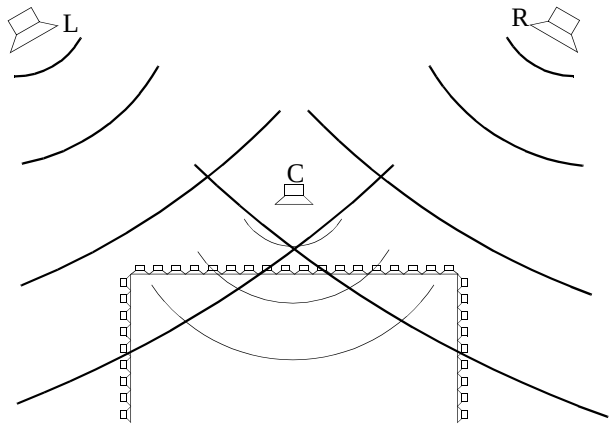


Figure 4.8 Virtual loudspeakers and a center cluster for stereophonic playback.

With these advantages already a less critical sweet spot area is constructed. Nevertheless, the perceived image still shifts if the listener moves to the side, because in that case one signal will reach the listener before the other one does. The central phantom source (often a soloist or a voice-over) will gain stability by adding a virtual center cluster on the symmetry axis between the virtual loudspeakers, as drawn in Figure 4.8. Such a central loudspeaker, which is regularly applied in theater sound systems (Bracewell 1993), is fed by the left and right stereo signals, but at a lower level to avoid coloration. Furthermore, the left and right channels are slightly delayed with respect to the central channel. The precedence effect will then secure correct localization, while (comb filter) coloration is avoided or further reduced.

For 3/2 program material, the virtual loudspeakers of the left, right and surround channels are imaged at infinity, at the standardized angles ± 30º and ± 110º. The center channel can be routed to an appropriate center cluster position.

B. Downward compatibility

Downward compatibility can be achieved by special coding of the microphone signals. The synthesis operator **W** requires the bare microphone signals as inputs, while the stereo and 3/2 surround systems need mixed signals. The stereo signals consist of main microphone signals to which spot microphone signals are added by panning (Section 1.6.3). Since at present the 3/2 recording technique is still under development, a coding scheme for this material will definitely be premature, but an example scheme is given in Figure 4.9. This scheme assumes that together with a stereo 'main' microphone pair, two additional ambient microphones are used (e.g. the reversed ORTF-pair drawn in Figure 1.20a). Inside the mixing console, an encoder keeps track of the routing paths and levels that will serve as a key to recover the original microphone signals for reproduction by wave field synthesis. The output of the decoder can be connected to the synthesis operator **W**. The stripped L and R channels and the original S1 and S2 channels are reproduced as ambient plane waves, while the spot microphone channels are further processed to yield direct sound and early reflections using the geometrical parame-

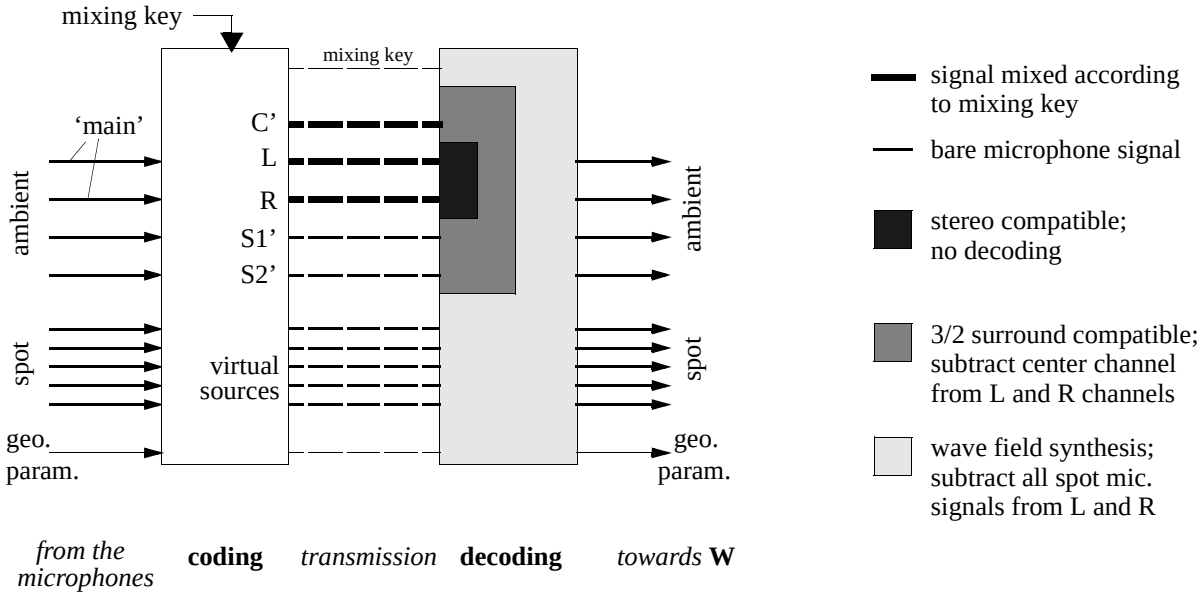


Figure 4.9 Coding scheme for downward compatibility. The primed channels belong to the 3/2 surround system. The central channel C' may consist of one or more spot microphone signals. The L' and R' channels of the 3/2 system are retrieved by setting L' = L – 0.7 C' and R' = R – 0.7 C'. The wave field synthesis channels are recovered by similar, but more complex, information enclosed in the mixing key.

ters. In this example, the C' channel consists of merely spot microphone signals. Since these signals are transmitted individually, the C' channel is not decoded for wave field synthesis.

4.3 Laboratory demonstration system

The playback concept of Section 4.2.5 formed the basis for the design of a demonstration and research reproduction system. This system, that is described below, is installed in the reproduction room at the Delft Laboratory of Seismics and Acoustics.

4.3.1 Overview

Figure 4.10 depicts the playback scheme for the laboratory demonstration system. A digital multi-track recorder provides up to 16 microphone signals that are mixed down to maximally 8 input signals for the DSP-station. This station, that forms the synthesis operator **W**, produces the driving signals for a closed array surrounding the listening area, according to Figure 4.2c. The heart of the station is formed by 8 digital signal processors (DSPs) that control the position of the virtual sources and the direction of the reverberant waves. Both the DSP-station and the loudspeaker arrays are custom-built by Duran Audio, Zaltbommel.

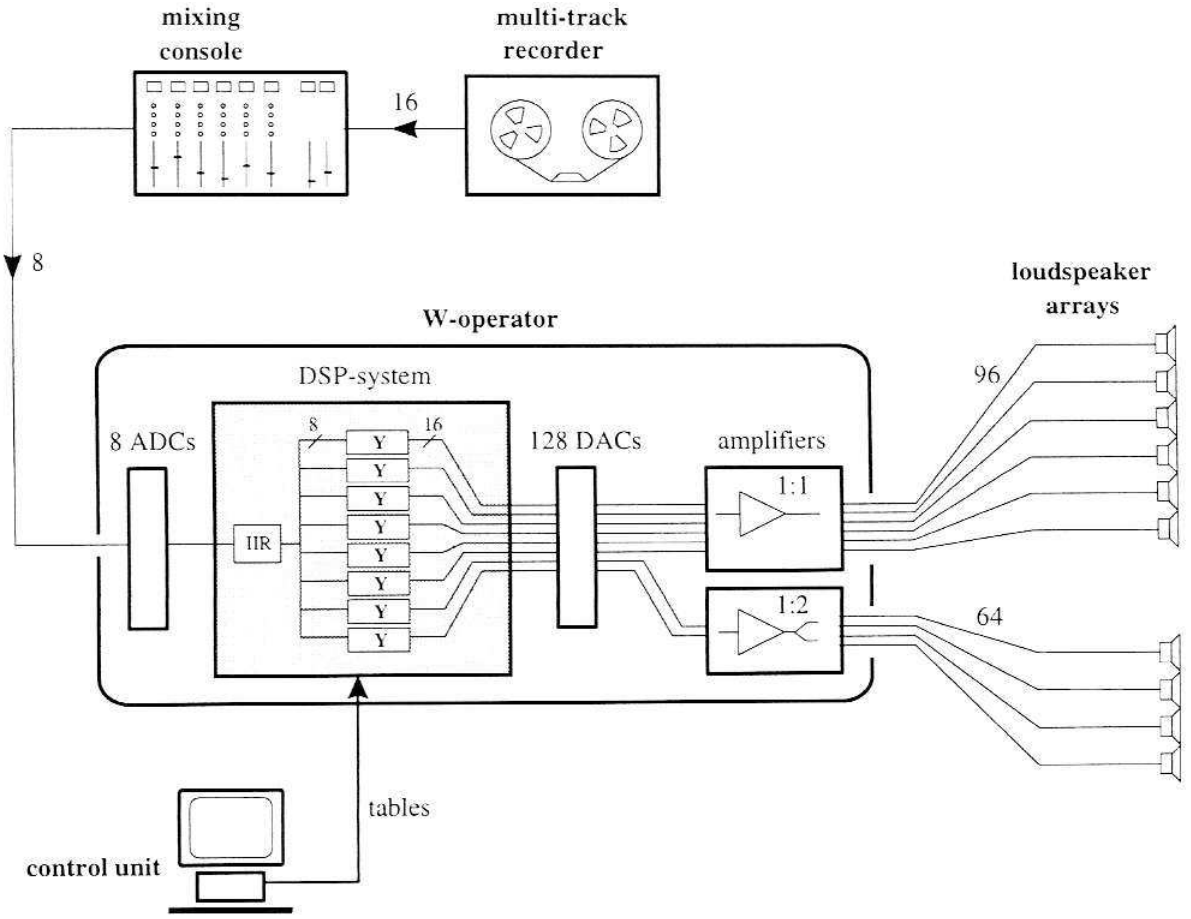


Figure 4.10 Play back scheme of the laboratory wave field synthesis system. The signals from the multi-track recorder are mixed down to 8 notional source and reverberation signals. These are processed in the DSP-station, which is an implementation of the synthesis operator **W**. The 160 output signals are fed to the loudspeakers.

4.3.2 Array configuration

The assessment of the arrays has already been made in Section 3.4.1. A rectangular arrangement of ten array bars, enclosing a listening area of about 24 m^2, has been installed in the laboratory reproduction room, see Figures 4.11 and 4.12. The arrays hang from the ceiling at ear-height (1.65 m), to facilitate listening for a critical audience that is walking through the room to check the spatial properties of the image. One array bar consists of 16 loudspeakers (Figure 3.17) with a frequency range from 140 Hz to 20 kHz.

In general, the ambience of the reproduction room itself should be masked by the reproduced ambience. This means that at least the reverberation time of the reproduction room should be much shorter than that of the recording room. In the laboratory room, a reverberation time of less than 0.5 s is measured above 100 Hz. Reflections are damped by a carpet on the floor and by absorptive panels in the ceiling and along the walls.

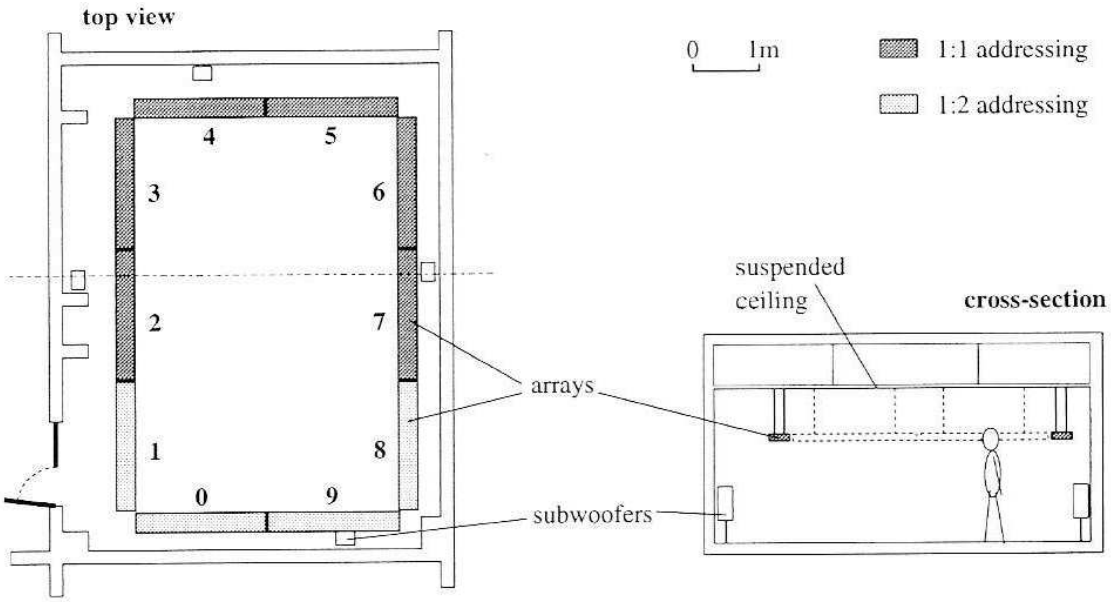


Figure 4.11 Map of the laboratory reproduction room.

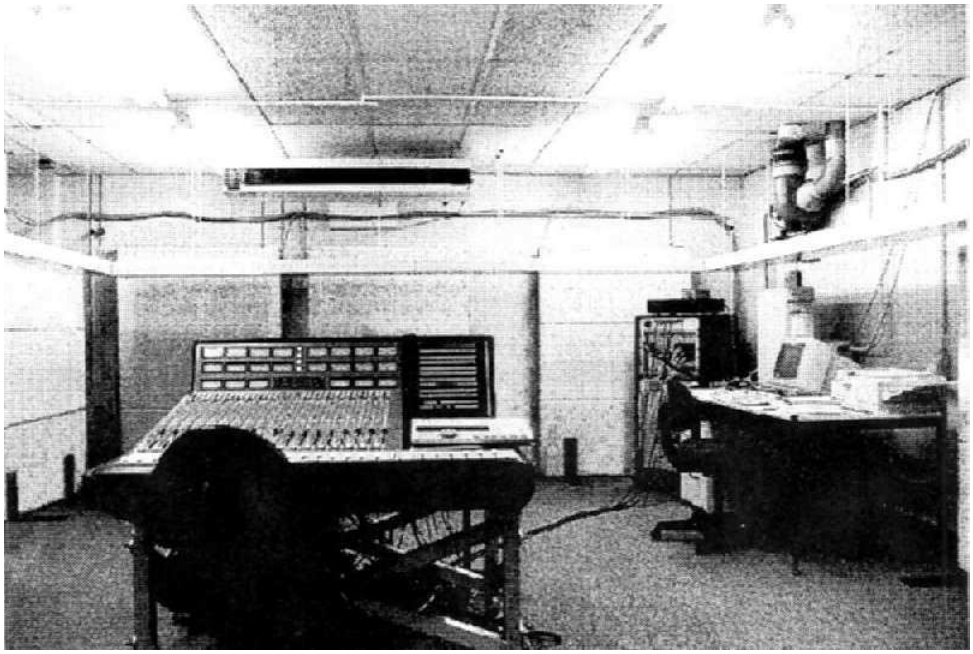


Figure 4.12 The laboratory reproduction room with the mixing console (left), the arrays (nos. 3 to 6), and the DSP-system (in the corner of the room).

The frequency range below 140 Hz is supplied by free standing loudspeakers used as subwoofers (Philips DSS 930, described by Willems 1993). According to the spatial aliasing requirement (2.35), the loudspeaker spacing for these woofers should not exceed 1.2 m. At the frequency range of operation, however, the wave pattern in this room is dominated by standing wave modes, so that proper synthesis is impossible below 140 Hz. Therefore, a more efficient woofer arrangement can be made without affecting quality of synthesis: each side is provided with one woofer, for which the signals are taken from the middle array element of each bar at the respective sides. For example, the front subwoofer receives the low-pass filtered signals (–3 dB at 140 Hz, –24 dB/oct.) of the eighth elements of bar 4 and 5. Thus, the front woofer signal is a filtered addition of two signals, just like the rear woofer signal, whereas the side woofers are each supplied by three signals. As long as the sources keep away about 50 cm from the arrays, the variation in bass response between different source positions remains within an acceptable range.

4.3.3 DSP-system hardware

A DSP-system with 8 input and 128 output channels takes care of the processing of the loudspeaker signals. Each of the 8 floating point DSPs (TMS320C32) reads all 8 inputs and writes 16 outputs. The actual number of virtual sources processed in real-time by one DSP is 4 at 44.1 kHz sampling rate, or 6 at 32 kHz. This process, denoted **Y**, is modular in the sense that the number of DSPs running **Y** is proportional to the number of output channels.

A full coverage of the four sides of the reproduction room requires at least 160 loudspeakers, whereas the DSP-system hardware supports a maximum of 128 outputs. To cope with this discrepancy, it is reasoned that in normal use only reverberation and no direct sound will be reproduced from the rear arrays, and this may be done with less spatial accuracy. Therefore, the following distribution is made: arrays 2 to 7 are addressed by one DSP each, or in other words, each of their loudspeakers is driven by an individual signal (denoted 1:1). The remaining 4 rear arrays are driven in two-element subarrays (denoted 1:2), which may be thought of as a simple form of a spatial convolution filter (see Section 3.3.3). The use of subarrays partly compensates the effects of the lower spatial aliasing frequency (with respect to the front arrays), because it benefits the synthesis of waves with a small angle of incidence with respect to the array.

The DSP-system is table-driven: a control unit (PC 486DX) calculates the synthesis parameters from the (new) virtual source positions entered graphically by the user.

4.3.4 DSP-system software

The gray box in Figure 4.10 shows that two different tasks are carried out by the DSPs: IIR-filtering and array signal processing **Y**. These tasks are related to the generalized monopole driving function (2.30), which can be split in a frequency filter

$$H(\omega) = \sqrt{\frac{\mathrm{sign}(\zeta)k}{2\pi j}}, \tag{4.1a}$$

with $\zeta = z_R/z_S$ as defined in Section 2.3.3, and a loudspeaker signal operator

$$Y_n = \sqrt{\frac{\zeta}{\zeta - 1}} \cos\varphi_n \frac{\exp(\text{sign}(\zeta) jkr_n)}{\sqrt{r_n}}, \tag{4.1b}$$

where r_n is the distance from the virtual source to loudspeaker n. Loudspeaker signal Q_n is now retrieved by

$$Q_n(\omega) = Y_n H(\omega) S(\omega). \tag{4.2}$$

This operation is implemented in the time-domain for reason of efficiency. If a vector of array driving signals is defined as

$$\boldsymbol{q}(t) = \left[q_1(t) \ \ldots \ q_n(t) \ \ldots \ q_N(t) \right]^T, \tag{4.3}$$

with $q_n(t)$ the inverse Fourier transform of $Q_n(\omega)$, then the array driving signals are processed as

$$\boldsymbol{q}(t) = \boldsymbol{Y}(t) * [h_{\text{IIR}}(t) * s(t)], \tag{4.4a}$$

where $*$ denotes time-domain convolution, $h_{\text{IIR}}(t)$ is an infinite impulse response filter yet to be designed, and $\boldsymbol{Y}$ is a vector operator with elements

$$Y_n(t) = \sqrt{\frac{\zeta}{\zeta - 1}} \frac{\cos\varphi_n}{\sqrt{r_n}} \delta(t + \text{sign}(\zeta) r_n/c). \tag{4.4b}$$

First, the design of $h_{\text{IIR}}(t)$ is discussed and next operator $\boldsymbol{Y}$ is implemented for both static and moving sources.

IIR-filter

Ideally, $h_{\text{IIR}}(t)$ is a time-domain version of filter (4.1a), which has a slope of +3 dB per octave, see also Section 3.2.2. This filter is however not applicable *above* the spatial aliasing frequency, where no constructive interference of individual loudspeaker contributions occurs. It appears to be possible to design a single filter that renders a reasonably flat system response for a large range of virtual source positions and a large listening area. The filter, shown in Figure 4.13, is designed by equalizing the average response of a number of virtual sources and receiver positions in the reproduction room. In this option the loudspeaker response is compensated as well. A cardioid microphone was used as receiver to suppress reflections in the measured signals. In agreement with theory, a positive slope of a few dB per octave is visible, apart from a peak at 400 Hz due to a broad dip in the loudspeaker response (compare Figure 3.22). No attempt is made here to match the IIR-filter's phase response to the theoretical frequency-independent value of 45º for virtual sources behind the array, or –45º in front of the array. If desired, these phase shifts can be controlled by applying an appropriate FIR-filter* instead.

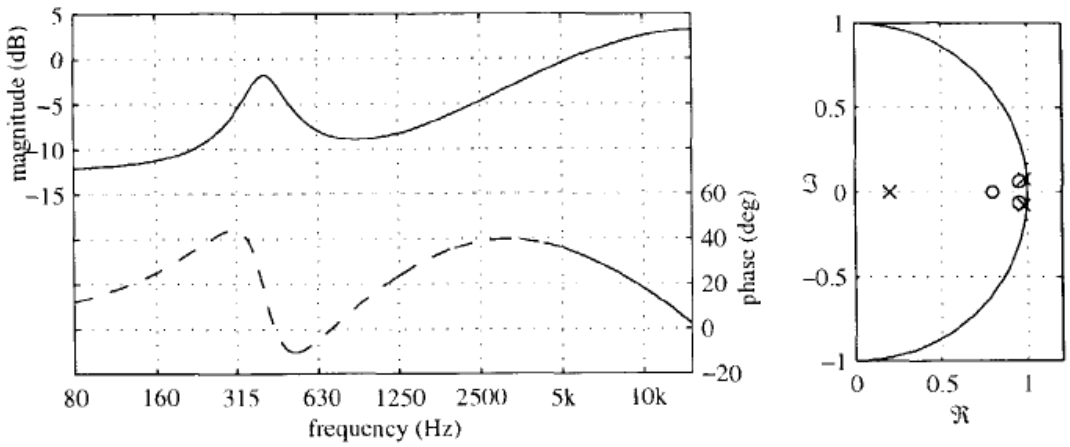


Figure 4.13 Filter design based on the average measured response of virtual sources. The zeros (∘) and poles (×) of the filter in the complex *z*-plane are shown to the right.

Synthesis of static sources

The task of the operator **Y** is to apply the correct delay and weighting coefficients from *M* filtered input signals to *N* output signals. If the input signals are written as a source vector

$$\boldsymbol{s}(t) = \left[s_1(t) \ \ldots \ s_m(t) \ \ldots \ s_M(t) \right]^T , \tag{4.5}$$

then the vector operator ***Y*** can be extended to a matrix operator **Y** yielding array driving signals

$$\boldsymbol{q}(t) = \mathbf{Y}(t) * [h_{\mathrm{IIR}}(t) * \boldsymbol{s}(t)] , \tag{4.6}$$

where $*$ again denotes time-domain convolution, and the elements of **Y** are given by

$$Y_{nm}(t) = a_{nm} \delta(t - \tau_{nm}) , \tag{4.7a}$$

with weighting coefficients

$$a_{nm} = \sqrt{\frac{\zeta_m}{\zeta_m - 1}} \frac{\cos \varphi_{nm}}{\sqrt{r_{nm}}} , \tag{4.7b}$$

and time delays

$$\tau_{nm} = \tau_0 - \mathrm{sign}(\zeta_m) \frac{r_{nm}}{c} . \tag{4.7c}$$

Note that an extra delay $\tau_0 > 0$ has been introduced to avoid non-causality in case $\mathrm{sign}(\zeta_m) = +1$ (for sources in front of the array). The delay and weighting coefficients are calculated in the control unit. The delay values are derived from the distance between loudspeaker and virtual source. The weighting coefficients a_{nm} depend on the position of the reference line *R* (see e.g.

* Finite impulse response filters generally require much more computation power of the DSPs.

Figure 2.7 and Figure 2.17) via the ratio $\zeta = z_R/z_S$. For a *straight* linear array, the reference line at $z = z_R$ is usually chosen parallel to the array in the middle of the listening area. For a linear array with corners, e.g. the rectangular array of Figure 4.11, a single parallel reference line is impossible. A solution is found in applying driving function (2.32b), which permits non-parallel reference lines to be used. By writing $\Delta r/r = \zeta$, the same form is obtained as in (2.30). Figure 4.14 illustrates the assessment of the reference line for cornered arrays.

Not all array bars are active for a certain virtual source. At most, half of the array bars are 'seen' by a virtual source, and in many situations it suffices to process even less loudspeaker signals. Figure 4.15 shows that much processing power can be saved by considering that not all positions within the rectangle will be occupied by listeners. Though a number of six virtual sources per array remains the limit in the laboratory demonstration system, an effective use of processing power makes it possible to reproduce much more virtual (mirror) sources in total, all around the listening space. For this purpose a coding map has been set up manually, see Figure 4.16. The virtual source area is divided in a large number of squares, the field code of which specifies the array bars that are active for a virtual source within that square. Note that no virtual sources can be placed in the blank middle area. This area is excluded for virtual sources, because listeners are not allowed to reside between active arrays and focused sources.

The delay coefficients produced by the control unit are rounded to the nearest integer time sample interval. For example, if a delay of 15.6 samples is calculated for a certain virtual source *m* and loudspeaker *n*, the actual delay is set to 16 samples. At 32 kHz sample rate, the maximum phase error made in this way is 7.5º for the spatial aliasing frequency of 1340 Hz. Synthesis of wave fronts is therefore hardly influenced by this approximation.

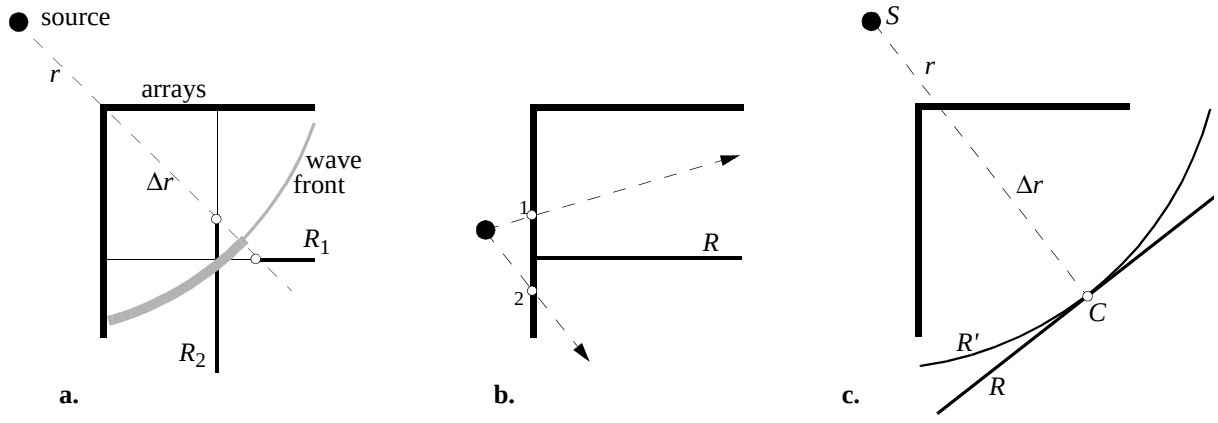


Figure 4.14 Choice of the reference line *R* in a cornered array configuration. **a.** The ratio $\zeta = \Delta r/r$ must be continuous along the arrays to avoid discontinuous amplitudes along the synthesized wave fronts. If two reference lines (parallel to the respective arrays) are used, the ratio $\Delta r/r$ will in general be discontinuous at the corner. **b.** If only one, but fixed, reference line is used, not all stationary phase points at the array are defined: the arrows drawn do not cross the chosen reference line. The loudspeaker driving function is therefore not defined at position 1 and 2. **c.** Each virtual source *S* defines its own reference line, all of which intersect the center *C* of the listening area. Reference line *R* is defined as the line through *C*, perpendicular to the line connecting *S* and *C*. A useful alternative is reference line *R'*, which is the arc through *C*, centered around *S*. In real-time processing, using *R'* instead of *R* speeds up the calculation of $\zeta/(\zeta-1)$. Therefore *R'* is employed in the laboratory system, with *C* at the middle of the rectangular array configuration.

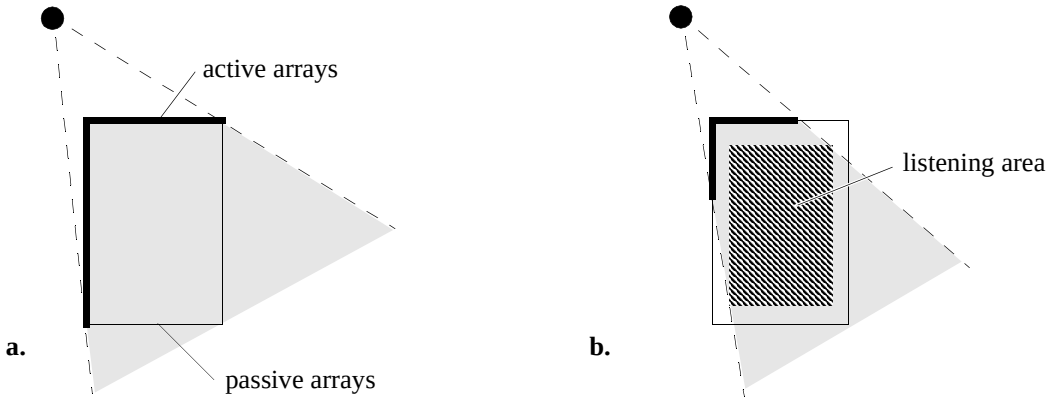


Figure 4.15 **a.** A virtual source near the corner 'sees' half of the arrays. **b.** If the region near the arrays is excluded for listeners, only a small part of the arrays needs to be active.

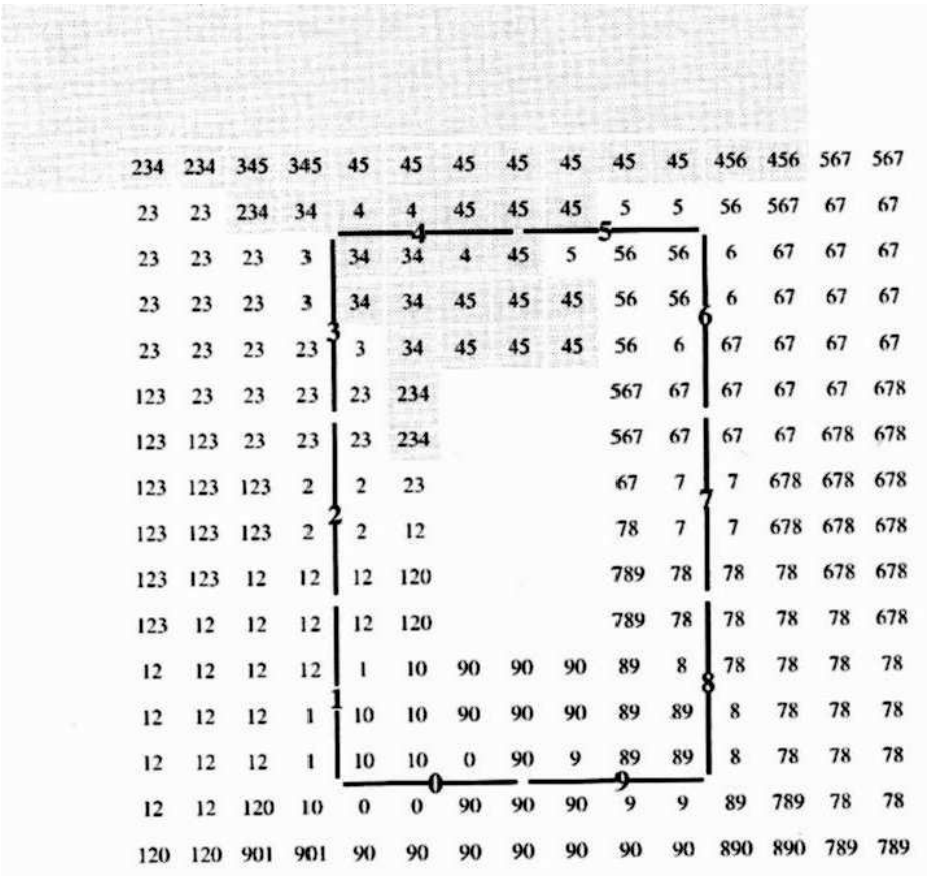


Figure 4.16 Coding map used by the control unit to assign arrays to virtual sources. Each field code consists of three or less digits that indicate the array bars that are involved in synthesis of virtual sources present in that particular field. Virtual sources outside the map are treated corresponding to the nearest field code. As an example, the area belonging to array bar 4 is shaded: only virtual sources within this area are assigned to array bar 4. The coding map, which allows efficient use of processing capacity, is initialized according to the geometry of the array and the desires of the user.

Virtual *mirror* sources are treated by the DSP-system in the same manner as virtual *direct* sources. The mirror source positions and strengths are calculated in the control unit. About 5 to 15 virtual mirror sources, each with a delay time up to 65 ms, can be generated for a single virtual source, depending on the number and positions of the other virtual sources. Frequency-dependent absorption can be incorporated by reserving input channels with adjustable IIR-filters for the mirror source signals.

Synthesis of moving virtual sources

Moving sources require a different treatment by the control unit and DSP-system. Each sample period, an update of delay and weighting coefficients is necessary. For this reason, (4.7b) and (4.7c) become time-dependent (Jansen 1997):

$$a_{nm}(t) = \sqrt{\frac{\zeta_m(t)}{\zeta_m(t)-1}}\frac{\cos\varphi_{nm}(t)}{\sqrt{r_{nm}(t)}}, \tag{4.8a}$$

$$\tau_{nm}(t) = \tau_0 - \mathrm{sign}(\zeta_m(t))\frac{r_{nm}(t)}{c + \dot{r}_{nm}(t)}, \tag{4.8b}$$

where $\dot{r}_{nm}(t) = dr_{nm}(t)/dt$.

Because the control unit can only update the delay and weight tables a few times per second*, the DSPs must take over some calculation tasks from the control unit. An efficient computation method is described by Jansen (1997). The control unit periodically transfers the coordinates and weight tables of certain support points to the DSPs, together with a speed parameter, see Figure 4.17. The coordinates are used to compute $r_{nm}(t)$ along the linearized path. The delays, calculated from $r_{nm}(t)$, are not rounded to the nearest integer like for static sources, because these irregularities become audible as sputtering. This problem is solved by resampling or interpolation between the nearest time samples.

The weight tables along the path are computed through linear interpolation between the weight tables of the support points. This method saves much processing time in comparison to the exact calculation of $a_{nm}(t)$ from the coordinates of the support points.

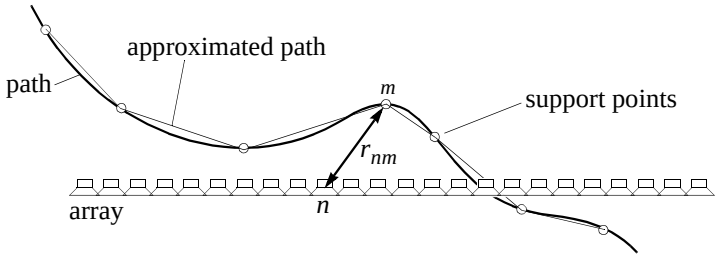


Figure 4.17 A moving source can be processed by interpolating between support points along the source path.

* The rate of table-updating is limited by the time it takes for the control unit to compute a table.

Note that the weights are estimated with much less accuracy than the delays. The difference in precision of calculation is motivated by perceptual considerations. The perceived *direction* of the source is mainly determined by the distribution of delay coefficients along the arrays. If accurate localization is appreciated and round-off errors should not be audible, the computation of delays must be done with great precision. The perceived *sound level* of the source depends on the weighting coefficients. Since level fluctuations of a few dB are hardly audible, medium precision is allowed in generating weight tables. By using linear interpolation between weight tables, smooth level variations along the path are guaranteed.

4.4 Applications of the new concept

The presented laboratory system is one implementation of the concept, but other implementations are also possible. In Section 4.2, already some considerations have been made about simpler as well as more enhanced systems for sound reproduction. In this section, a more detailed discussion is given, where the laboratory system serves as a reference.

4.4.1 Living-room

For domestic use, the design described above is perhaps too sophisticated in software and hardware. A few simplifications are still possible, leaving the overall performance more or less unaffected. Of course, the general purpose DSPs can be captured in a dedicated chip, which will lead to much lower production costs. Also, the number of channels can be restricted by allowing yet a larger spacing of array loudspeakers at the back and sides, if no direct virtual sources are imaged in those directions. Concerning the integration of the arrays into the interior of the living-room, it should be noted that the radiated sound of the arrays should not be shielded by objects in the ‘line of sight’. Built-in arrays (into the walls just beneath the ceiling) or foldaway arrays (towards the ceiling or walls) might be feasible. It should also be mentioned that the exploration of array technology has just begun: special array designs, such as the lightweight transparent electrostatic strips in Section 3.4.2, may further improve chances for the practical implementation of arrays.

The main question for acceptance of the novel reproduction method in the living-room is whether the advantages of the wave field synthesis system outweigh its higher price and its more complicated requirements for installation. The history of sound reproduction proves that compatibility with previous and contemporary reproduction systems is of primary interest in answering this question. Commercial chances are raised if the new system offers a clear improvement in reproducing conventional program material. Preliminary listening tests with the proposed method of reproducing two-channel stereo material (Section 4.2.6) yielded already a striking enhancement of the spatial image. Together with a modular set-up of the arrays and the synthesis operator, this might be the basis for an introduction of the system to the consumer market.

4.4.2 Cinema

In cinemas, loudspeaker arrays usually do not have to compete with furniture, as in a domestic environment. The arrays can be built into the back and side walls, and behind the front screen. With a playback system based on wave field synthesis, surround sound effects are not restricted any more to the position of the surround loudspeakers: sound can be created all around the audience, with a much more realistic spread. Since the input channels of the system consist of discrete source signals with source coordinates, the sound image can be adjusted to the shape or size of the cinema.

4.4.3 Virtual reality theater

A virtual reality system is an interactive device that evokes an image to the user which is very close to reality, by stimulating all his senses in a consistent way. The concept of wave field synthesis lends itself admirably to this purpose, because of its ability to create any desired sound field from a few parameters. Since the listening area is basically unlimited, the user has a large freedom of movement. In this way, several users can use the system simultaneously. The ultimate virtual reality system consists of 2D loudspeaker arrays behind an acoustically and visually transparent screen, producing a full 3D sound image. Especially the possibility to focus sources in front of the array will contribute to a thrilling experience.

4.4.4 Simulators

Closely related to virtual reality systems are (flight-)simulators. Instead of an exciting situation, an unusual situation is recalled here, to which the trainee must react in an appropriate way. Also in these training devices a realistic image must be created, so the requirements are very similar to those for the virtual reality theater.

4.4.5 Acoustic research

Of course, the laboratory system is also extraordinarily suited for room acoustic and psycho-acoustic research. Except the acoustic design of newly built concert-halls, also the effects of renovation of existing halls can be modeled and judged in advance. Because of the listener's freedom to move and the possibility to directly deliberate with others, this way of modeling is far more natural and convenient than binaural modeling, which is proposed alternatively.

Spatial perception research is much facilitated by using the laboratory system. In earlier research, sources were represented by loudspeakers placed at positions allowed by the dimensions of the room. In wave field synthesis, sources are placed under software control at any distance from the subject. Especially, research on the perception of moving sources will become much easier.

4.5 Conclusions

The general concept of sound reproduction by wave field synthesis is developed by considering the natural sound field in its acoustic environment. The wave field inside a certain enclosure within that environment is determined by the Kirchhoff-Helmholtz integral applied at the boundary of that enclosure. Theoretically, a physical copy of the original sound field can be synthesized in the reproduction room. In order to come to a practical implementation of the concept, a number of data compression and data reduction steps is necessary. The proposed reproduction method makes it possible to *reproduce* the natural sound field as well as *produce* an ideal or artificial sound field from the same program material.

Three phases are distinguished in sound reproduction: the recording phase, the transmission and/or storage phase, and the playback phase. In the first phase, spot microphones are used to record the direct sound and so-called ambient microphones are used to capture the reverberant sound. These signals are transmitted or stored together with relevant geometrical parameters telling the playback system how to reproduce the sound field. Eventually, the direct sound is processed for virtual sources. The early reflections are reproduced by assigning the direct sound to virtual mirror sources. The reverberation is created by plane wave synthesis of the ambient microphone signals.

The general concept is realized in a laboratory wave field synthesis system. This system comprises a digital signal processing unit, which is an implementation of the synthesis operator, and a rectangular array configuration surrounding the listeners. Any multi-track material can be reproduced, while the recorded sources may be imaged in their original environment or in a non-existing or yet to be built concert-hall. Sources can be placed anywhere, even within the listening area by means of focusing techniques. Therefore the system is also very interesting for virtual reality theaters and simulators.

Chapter 5

Objective Evaluation

5.1 Introduction

In this chapter, the laboratory wave field synthesis system is evaluated by means of objective measurements. This evaluation consists of two parts: the calibration of the system response, which illustrates the temporal and spatial characteristics of virtual sources, and an objective comparison of the new reproduction system in relation to stereophony and surround sound. Besides well-known qualification methods, such as frequency characteristics and sound pressure levels, also less customary methods are used in determining differences in reproduction quality. The ins and outs of these methods are discussed extensively. Next, these methods are applied to the evaluation of reproduction of direct sound, early reflections and reverberation separately.

5.2 System response

The temporal and spatial performance of the laboratory reproduction system is determined by measuring the impulse response for several virtual source positions. With these data, the equalization filter of the system can be calibrated and the spatial accuracy of reproduction can be verified.

5.2.1 Frequency response

The frequency response of the laboratory system is measured to check the design of the equalization filter and the subwoofer arrangement. By placing virtual sources at different distances, also the 1/*r*-dependence of the sound level is checked. Reflections of the reproduction room are expected to influence only the fine-structure of the measured spectra, except in the low frequency range where standing waves become important.

Figure 5.1 shows the frequency response of six virtual sources in the reproduction room. Virtual sources 1 and 2 were focused in front of the array. The distances of the virtual sources to the measurement microphone are 0.5, 1, 2, 4, 8 and 16 m, respectively. Between real sources at these increasing distances, a sound level reduction of 6 dB is expected. The results show that this theoretical value is approximated fairly well for virtual sources 4 to 6, while it is less than 6 dB for the others. This deviation is a result of the use of *linear* arrays, for which the 1/*r*-law is only valid at the reference position (position *C* in Figure 4.14). In practice, these small deviations will hardly be audible. Besides, the actual levels are easily overruled by the gain control on the mixing console.

The (internal) variation of the responses remains within ±3 dB for frequencies above 300 Hz. Below this frequency, standing wave patterns dominate the results.

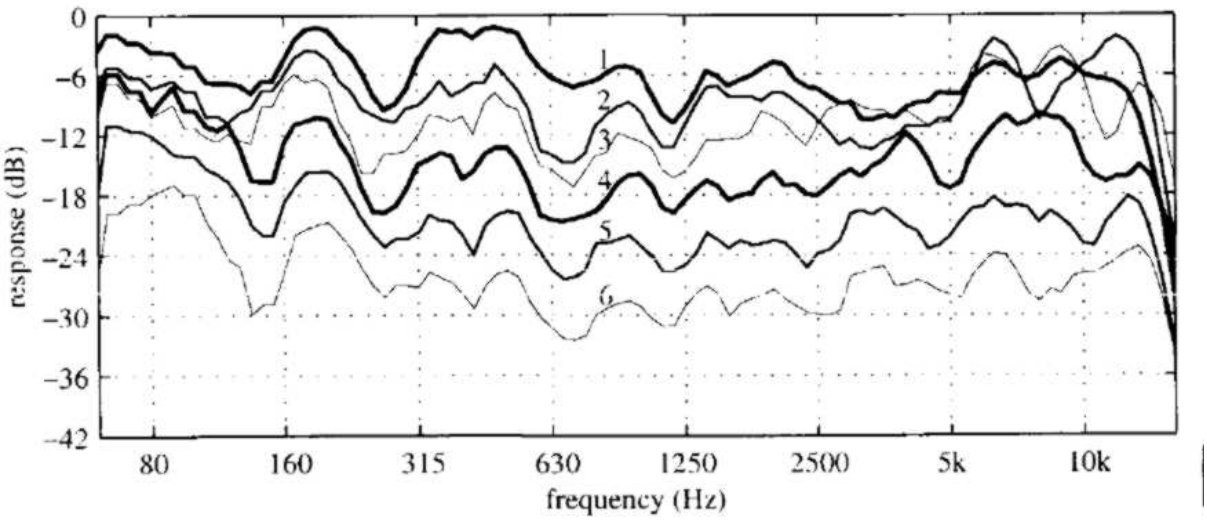


Figure 5.1 System response of virtual sources at positions 1 to 6, as shown on the map to the right. Only the front part of the rectangular array configuration is drawn. The receiver position is at the cross (×). Since the distance doubles between subsequent source-receiver combinations, an offset of 6 dB is expected between subsequent spec-

5.2.2 Spatial response

The spatial behavior of the wave fields of virtual sources is investigated in an anechoic room. The measurement microphone is displaced stepwise along an acquisition line parallel to the array. The measurements and simulations of Figure 5.2 show that the wave fields of

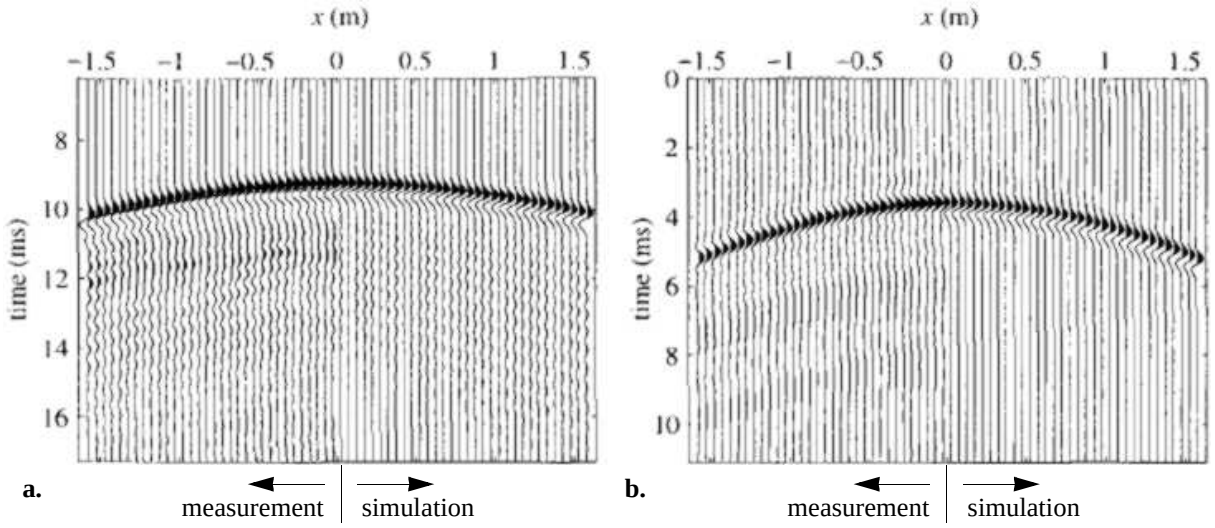


Figure 5.2 Measurement and simulation of a Gaussian wave front of a virtual source in the *xt*-domain. **a.** Virtual source at 1 m *behind* the array. **b.** Virtual source at 1 m *in front of* the array (focused source). The 32 element array is at a distance of 3 m parallel to the acquisition line. Bandwidth: 0.1 - 2.5 kHz.

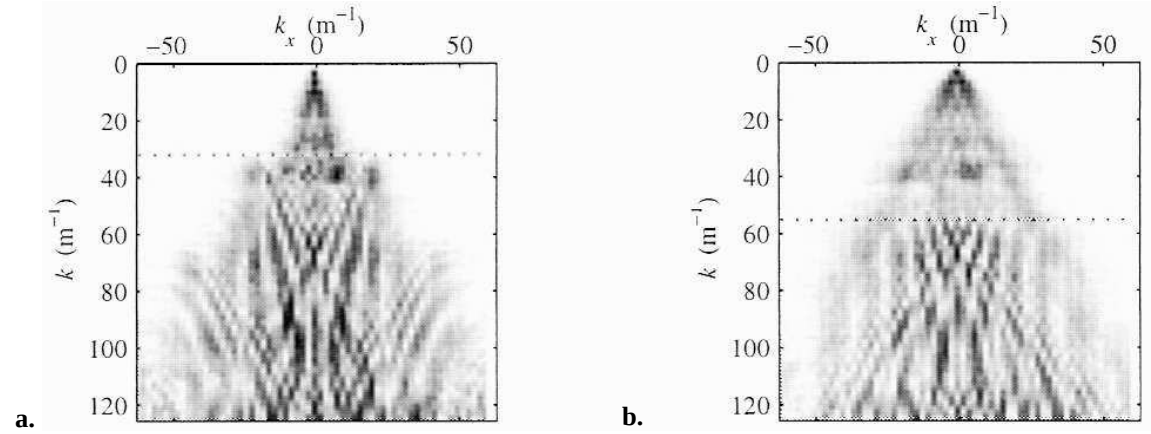


Figure 5.3 2D-Fourier transform or $k_x k$-diagram of the *measured* data of Figure 5.2. **a.** Virtual source at 1 m *behind* the array. **b.** Virtual source at 1 m *in front of* the array (focused source). The interference pattern, seen at high values of k, is typical for aliased wave fields. Bandwidth: 0 - 6.8 kHz.

focused and non-focused virtual sources are well controlled, i.e. their spatial behavior is in good agreement with theory. Small differences can be ascribed to the frequency response and directivity characteristics of the loudspeakers.

Some effects of spatial aliasing are visible for the virtual source behind the array (Figure 5.2a), while they are more or less absent for the virtual source in front of the array. This presumption can be verified by making a plane wave decomposition of the measured data sets, see Figure 5.3. Above the horizontal dotted lines in the diagrams the plane wave components are obviously aliasing-free. Note that the aliased wave fronts appear at higher values of k in the focused wave field, which seems to be in contradiction to the spatial aliasing condition being

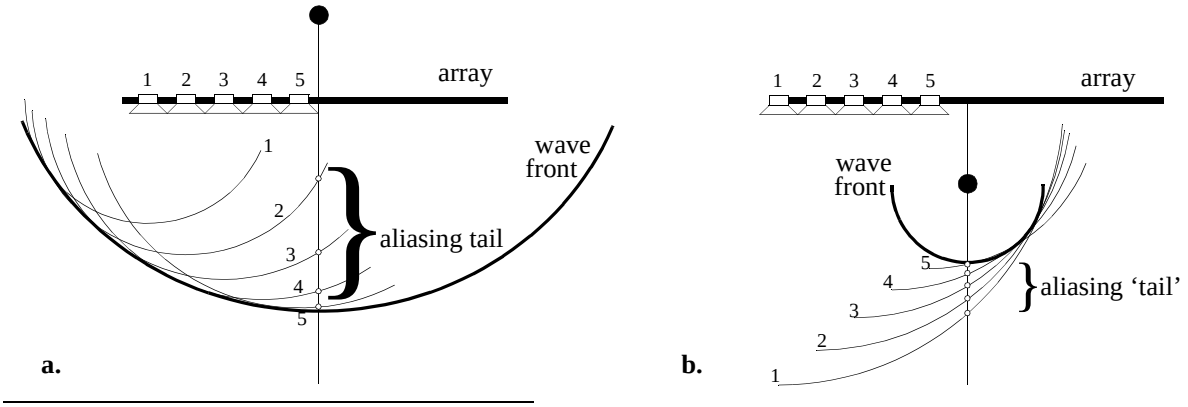


Figure 5.4 Construction of wave fronts. **a.** The contributions of five loudspeakers produce a wide aliasing tail for a virtual source behind the array. The tail is preceded by the synthesized wave front. **b.** The contributions of five loudspeakers produce a dense aliasing 'tail' ahead of the synthesized wave front for a focused virtual source.

the same for both situations. The explanation for this paradox is found in the manner in which the wave fronts are (re)constructed, which is clearly different for focused and non-focused virtual sources. The construction is visualized in Figure 5.4. The contributions of the individual loudspeakers are drawn as thin arcs, while the synthesized wave front is drawn as a thick arc. Near the central axis (perpendicular to the array), the individual wave fronts in Figure 5.4a are too far apart to interfere constructively: they do not cancel each other and will remain as aliased waves. For the focused virtual source of Figure 5.4b, the individual waves near the central axis are much closer, leading to a reasonable cancelation for higher frequencies.

The favorable situation for the focused source is reversed if the outer regions of the listening area are regarded: there, the composing waves are far apart, while for the non-focused sources a dense succession of individual waves is observed. The $k_x k$-diagram of the focused source (Figure 5.3b) does not reveal these outer aliased waves, because the measurement line did not include the outmost parts of the listening area. Hence, the spatial aliasing condition still applies to focused sources, but the effects of aliasing occur in different areas of the synthesized wave field. The different spreading of aliasing products disappears if directive loudspeakers are used: in that case, also the non-focused sources have a central listening area with a higher aliasing frequency, because the outer loudspeakers will not radiate energy towards the central axis.

Because of their innovative application in sound reproduction, focused virtual sources are investigated in more detail in the next subsection.

5.2.3 Focal pressure-distribution

Seen from the viewpoint of sound reproduction, the essence of a focused source is merely that it can be reproduced closer to the listeners than the position of (one of) the loudspeakers, which is often bound to practical restrictions. Since the sound pressure level (SPL) in the focus must be kept within a range safe enough for the human auditory system, it is worthwhile to investigate the SPL distribution near a focused source. Figure 5.5 shows the measured

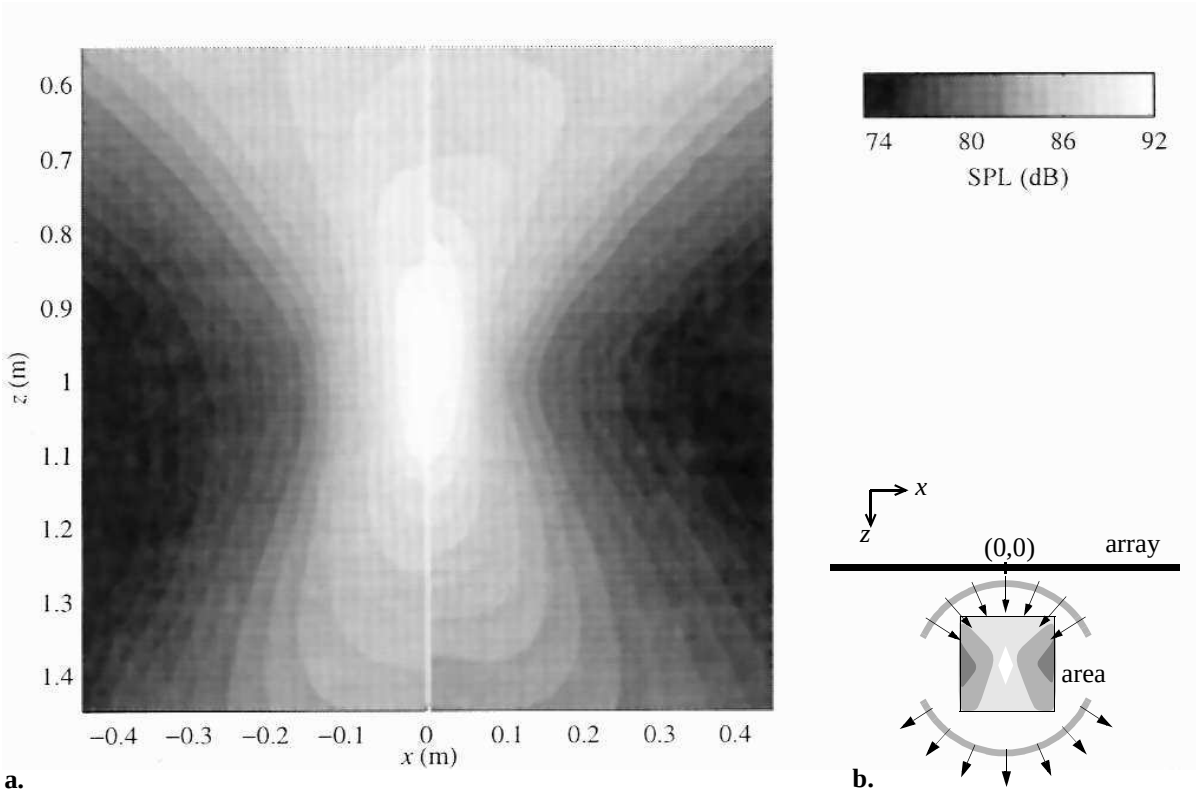


Figure 5.5 Sound pressure levels near a focus at 1 m from a 32 element array (bandwidth 0 - 2 kHz; Δx = 12.7 cm).
a. Left half: measurement in anechoic room, right half: simulation (with monopole loudspeakers).
b. View of the situation. The SPL measurements are taken on a 60 point grid in the left half of the area.

and simulated distributions. The SPL data at the left were taken in a 0.5 m^2 area in the horizontal plane in an anechoic room; the simulation result is shown at the right. For visualization, cubic interpolation is used between the grid points, both for the measured and simulated data set. A very good resemblance is obtained between measured and simulated data, confirming that the sound field produced by the arrays is well-controlled. A maximum pressure level of 91.8 dB is measured in the small white area in the middle. At 0.45 meter beyond the focus (bottom of picture), it is 7 dB less, while to the side a minimum of 73.4 dB is reached.

The focal area is not restricted to the horizontal plane. This can be understood by first considering a linear array of monopole loudspeakers, driven by the focusing operator. Figure 5.6a shows that this array produces a focal *circle* in the plane perpendicular to that array. If loudspeakers with high directivity in the vertical plane are used, most of the energy radiated is concentrated in the intersection point of the focal circle and the horizontal plane in front of the array (Figure 5.6b). The main benefit of this is that strong reflections from floor and ceiling of the reproduction room will be prevented. In addition, a more efficient power consumption is achieved. Both advantages also apply to non-focused sources.

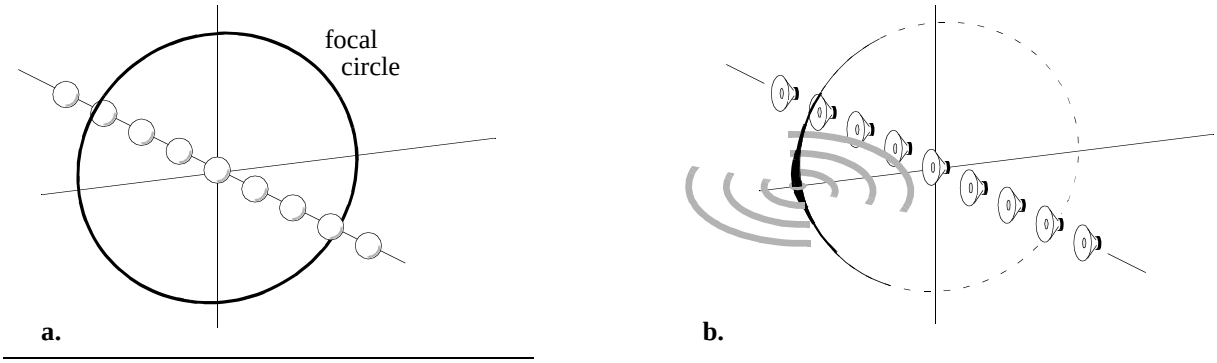


Figure 5.6 Focal shapes for linear arrays. **a.** Monopole elements. **b.** Directional elements. The local thickness of the focal circle indicates the concentration of energy.

5.3 Comparison between reproduction systems

5.3.1 Introduction

In general, it may be stated that any sound reproduction system designed for human listening, must be tested by means of *subjective* experiments. A rating of several subjective parameters for different reproduction systems would be preferable. Such experiments can therefore be classified as subjective *quantification* tests. A number of difficulties have to be considered in setting up these tests. For example, many time-consuming experiments are necessary before statistically significant results are obtained. Also, the results of subjective tests are very sensitive to (unwanted) bias of secondary differences. The subjective *qualification* tests presented in the next chapter attempt to avoid these problems. In this section, an *objective qualification* is made of the laboratory reproduction system in relation to other reproduction systems.

Several investigators have proposed objective measurement techniques to compare different reproduction methods. These measurement techniques have in common that the physical characteristics measured are believed to be related to relevant subjective characteristics. Damaske et al. (1972) and Tohyama et al. (1989) compared the interaural cross-correlation function for different arrangements of loudspeakers. A minimum interaural coherence is assumed by them to correspond with an optimal ‘subjective diffuseness’ for reproduction of reverberation. Mac Cabe et al. (1994) measured the wide band interaural time difference (ITD), interaural level difference (ILD) and interaural cross-correlation coefficient (IACC) to determine the localization performance of two reproduction systems: ambisonics and spectral stereo (a spectral pan pot system). Bamford et al. (1995) calculated the ITD, the ILD and a quantity called ‘the integrated wavefront error’ to compare the image directions for a simulated ambisonics, Dolby Surround and stereo system. Bareham (1996) measured the IACC and ILD in 1/3-octave bands at several locations in the reproduced field of two multi-channel the-

ater systems. The IACC is here assumed to indicate the spaciousness of the sound field, while the ILD should provide information of the image direction.

Caution is needed in interpreting the aforementioned interaural parameters. For a natural sound source in an enclosure, there is usually a well-defined relation between these parameters and the subjective attributes source direction and spaciousness. However, reproduction systems generally do not produce natural sound fields: e.g. a stereophonic phantom source causes several maxima in the cross-correlation function measured at a dummy head. Therefore, in addition, different measurement techniques have been proposed to qualify the reproduced sound fields (Boone et al. 1997). These are listed in Table 5.1, together with the objective and main subjective attributes to which they are related.

Table 5.1 Sound field qualification techniques

Qualification technique	Objective attributes	Subjective attributes
Dummy-head measurements	ILDs and ITDs	Direction
Multi-trace impulse responses	Shape of wave front	Direction and distance
Intensity measurements	Intensity distributions	Direction, spaciousness, diffuseness
Frequency spectrum measurements	1/3-octave band spectra	Coloration

These techniques, which will be explained in the next subsections, can be applied to the direct sound, early reflections and reverberation. For the purpose of comparison, these components are reproduced separately by the reproduction systems under investigation: wave field synthesis, 3/2 surround and (two-channel) stereophony. Note that the IACC is not incorporated in the table as a measure for spaciousness. It has been considered that it is awkward to use this measure in absence of a relevant reference situation. Instead, the intensity measurements are employed to gain insight in the distribution of reflective sound.

The measurement environment is the reproduction room where the wave field synthesis system has been installed (Figure 5.7). The stereo system shares the L and R loudspeakers with the 3/2 surround system. These individual loudspeakers (Philips DSS 930) are placed somewhat lower than the arrays (tweeters at 1.5 m above floor), at a circle of 2.75 m around the origin. The surround loudspeakers S_L and S_R are placed at angles ±125º and a radius of 2.4 m, though the 3/2 standard specifies ±110º with the same radius as the front loudspeakers. This deviation is imposed by the dimensions of the reproduction room. A smaller radius for all loudspeakers would have been possible, but then the listening area for 3/2 surround and stereo would become narrower. It will be shown that no serious affection of the 3/2 surround quality is caused for those situations in which the surround loudspeakers are active.

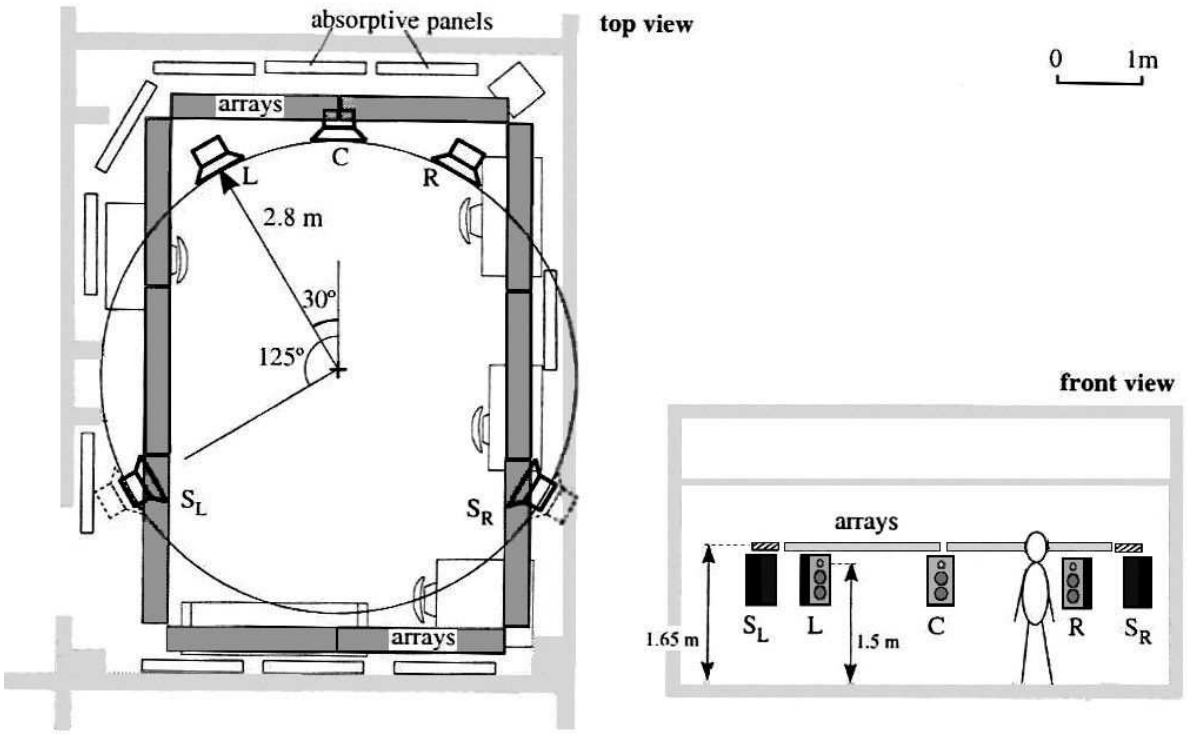


Figure 5.7 Measurement situation for objective comparison between wave field synthesis (arrays), 3/2 surround (L, R, C, S_L, S_R) and stereophony (L, R). The top view also shows the position of the furniture in the room.

5.3.2 Dummy-head measurements

The ILD and ITD are calculated from the impulse responses $s_l(t)$ and $s_r(t)$ measured at the entrance of the left and right ear channel of a human or dummy head. The ILD is given by

$$\mathrm{ILD} = 10\log\frac{\int s_r(t)^2 dt}{\int s_l(t)^2 dt}. \tag{5.1}$$

Since ILDs originate from frequency-dependent diffraction effects around the head, it is customary to determine them in frequency bands. Therefore, the ear signals are filtered in 1/3-octave bands before evaluating Equation (5.1).

The ITD is the travel-time difference between sound arriving at the left and right ear. It can be determined by evaluating the wide band interaural cross-correlation function (Ando 1985)

$$\Phi_{rl}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} s_r(t)s_l(t+\tau)dt. \tag{5.2}$$

The value of τ where Φ_{rl} is at its maximum, represents the time lag (ITD) between the ear sig-

nals, arising from the path length difference from the source to either ear. For a real source in an anechoic environment, the measured Φ_{rl} contains only one distinct peak (Figure 5.8a). This peak occurs in the interval $|\tau| \leq 1$ ms because of the limited path length difference. If the source is present in a reflective environment, multiple peaks occur (Figure 5.8b). Only the peak belonging to the direct sound falls in the range $|\tau| \leq 1$ ms, unless the source is very close (within a few dm) to a reflecting wall. Therefore, the ITD belonging to the direct sound of the source can usually still be determined in a reflective environment.

A different situation exists for sound reproduction systems. Here, the actual sources are the loudspeakers, rather than the phantom or virtual sources that are evoked. If the loudspeaker signals correlate in some degree, the interaural cross-correlation function will contain several local maxima (even if reflections are absent). In some cases, more than one distinct peak may appear within the specified range of $|\tau| \leq 1$ ms. In these cases, the image direction perceived can not be predicted easily. However, because the auditory system applies a frequency analysis to the ear signals before a kind of cross-correlation process is carried out (e.g. Stern et al. 1978, and Raatgever et al. 1986), it makes sense to observe the ITDs in frequency bands. Therefore, just like with the ILDs, the ear signals are filtered in 1/3-octave bands before determining the ITDs. Only the direct sound will be considered: reflections from the reproduction room are windowed off carefully, as a way to account for the precedence effect (Macpherson 1991). It will be shown here that the spectrally unraveled ILDs and ITDs provide interesting information on the frequency dependence of the reproduction quality.

Two sources *without reflections* are imaged by the reproduction systems, see Figure 5.8c. For reference and comparison, the response of a real source in an anechoic room is measured

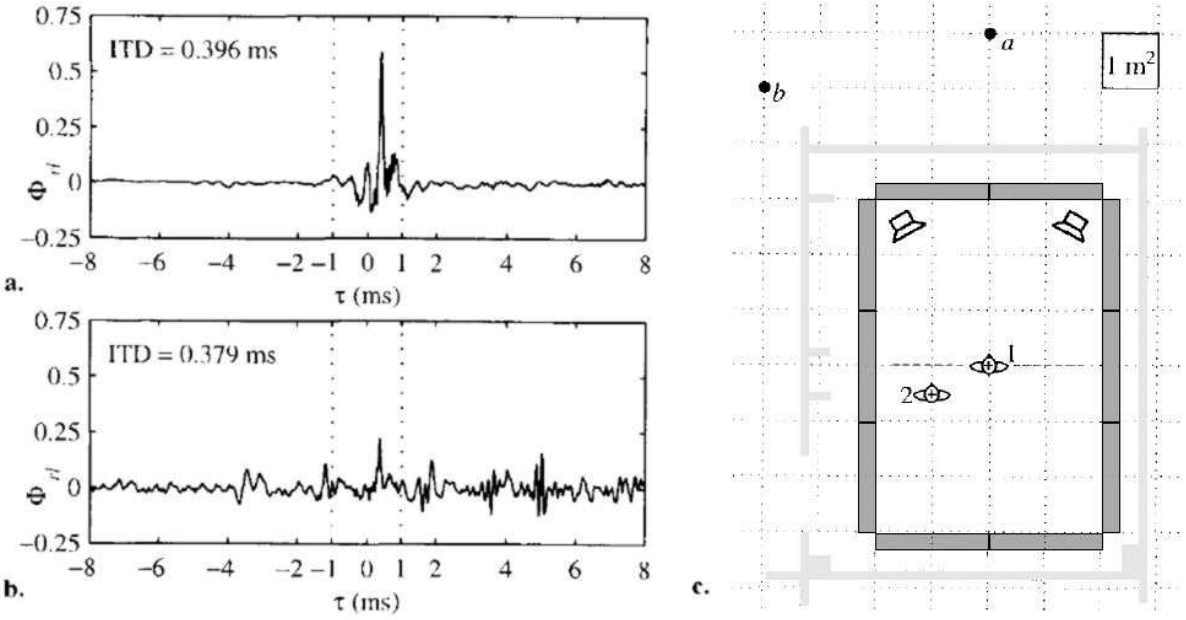


Figure 5.8 a. Normalized interaural cross-correlation function for a real source at 45º in an anechoic room. **b.** Normalized interaural cross-correlation function for a real source at about 45º in the reproduction room of the Delft Laboratory of Seismics and Acoustics. **c.** Two sources *a* and *b* are imaged by the reproduction systems. The dummy head locations are numbered 1 and 2.

as well. Both the stereo and 3/2 surround system are supposed to use the same stereo imaging technique in this case: the central and rear loudspeakers are not in use here for 3/2 surround*. The signals for the L and R loudspeaker are identical for source *a*. The signals for source *b* are obtained by simulating an ORTF-pair at the center of the reproduction room (position 1). For the wave field synthesis system (hereafter: WFS system), sources *a* and *b* are reproduced as virtual sources at the corresponding position. The head and torso simulator is a KEMAR (Burkhard at al. 1975).

Figure 5.9 compares the results for source *a* and listening position 1 of the reference (real source), the stereo system and the WFS system. For this symmetrical listening geometry, the real source was expected to render zero ILDs and ITDs. The non-zero ILDs at the high frequencies are probably caused by asymmetries of the head and torso simulator. The stereo

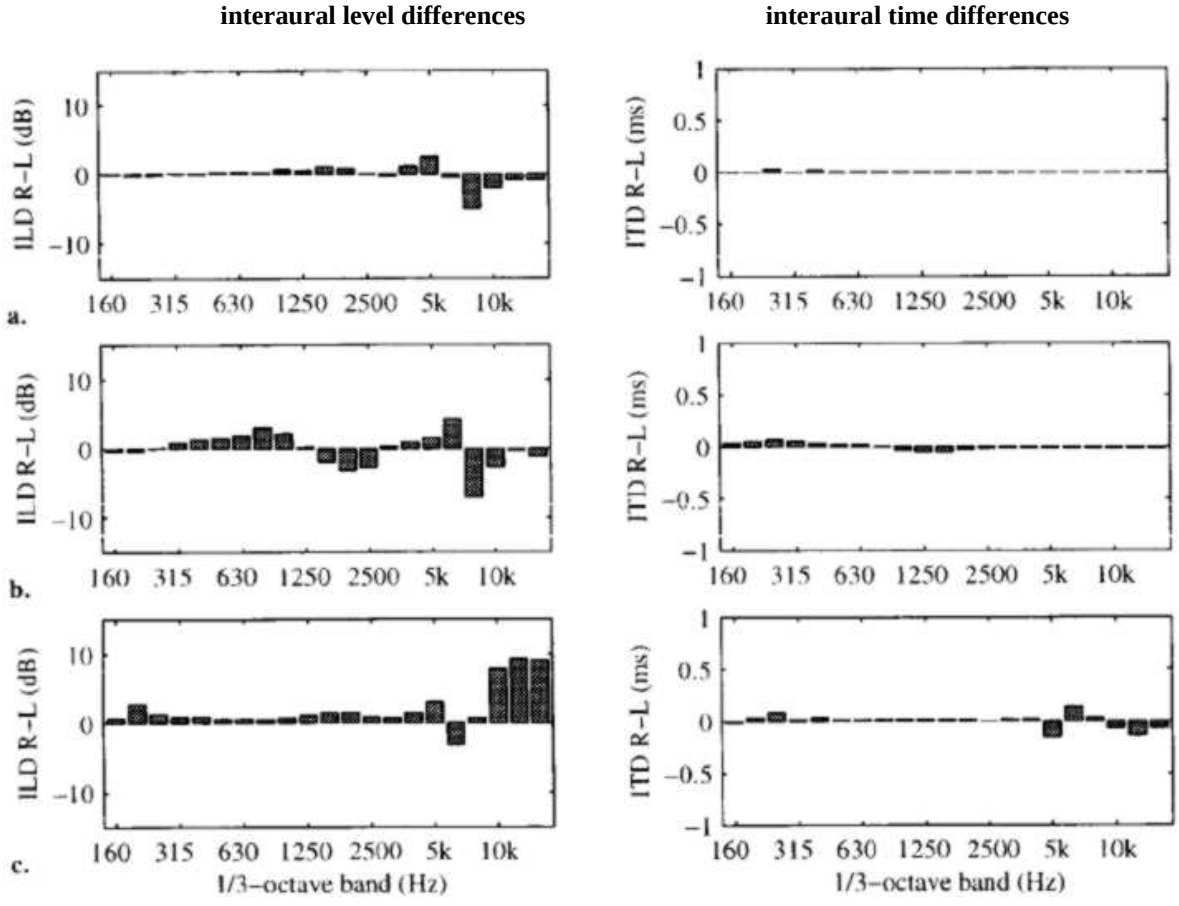


Figure 5.9 ILDs and ITDs for source *a* and dummy head position 1.
a. Real source in anechoic room. **b.** Two-channel stereophony. **c.** Wave field synthesis.

* Source *a* is assumed not to represent a soloist or voice-over, otherwise it could have been imaged by only employing the central loudspeaker of the 3/2 surround system. The C and S loudspeakers are not drawn for this system. Some pioneers of 3/2 surround mixing actually use only a 2/2 surround system for best results, thereby omitting the central loudspeaker (Theile 1996, and Scholtze 1997).

results are very close to the real source results. The WFS system shows deviations in the highest frequency bands which are presumably related to spatial aliasing effects.

For the off-axis listening position (no. 2), some interesting results are obtained, see Figure 5.10. In this situation, the source direction is approximately 10º to the right. The ILDs of the real source grow more or less with increasing frequency. This effect is in agreement with diffraction theory: short wavelengths are attenuated more in bending around a large object. The effective path length, however, is more or less the same for all frequencies, as can be seen from the ITDs being independent of frequency.

The stereo system produces a set of ILDs and ITDs that reveal practically no resemblance with the real source results. Also, at first sight no clear preference for any source direction is suggested from these stereo results, because of apparent incoherence between ILDs and ITDs and different cues for different frequency bands. Nevertheless, if one listens at this position, the stereo image – though broadened – is perceived slightly left of center. Notice that the ITDs between 400 and 1250 Hz, where phase cues are known to dominate localization, are indeed supporting leftward localization.

Both the ILDs and ITDs for the WFS system show a close resemblance to those for the real source. This is not surprising, because this system produces a correct wave field at least up

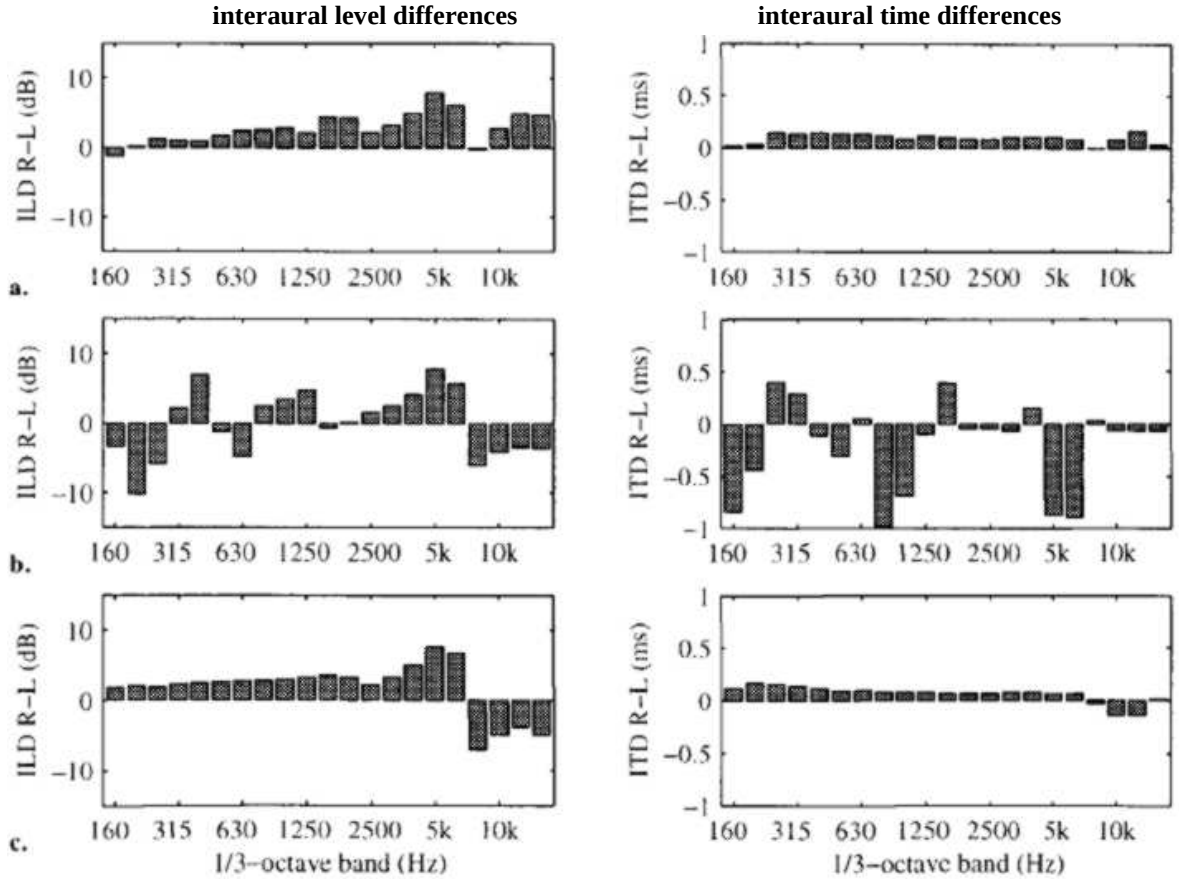


Figure 5.10 ILDs and ITDs for source *a* and dummy head position 2.
a. Real source in anechoic room. **b.** Two-channel stereophony. **c.** Wave field synthesis.

to the aliasing frequency of 1340 Hz (calculated from Equation (2.35) using Δx = 12.7 cm). In this situation, the aliased waves appear not to affect the ILDs and ITDs below the 8 kHz-band.

For source *b*, the dummy head is again placed at the sweet spot position 1. The results are again in favor of the WFS system, as shown in Figure 5.11. The stereo system, here using signals recorded by the simulated-ORTF system*, produces fairly consistent responses. Its ITDs correspond with that of a source nearly 30º to the left†, which indicates that the sound from the left loudspeaker is decisive in this aspect. The high frequency ILDs support this phantom image direction.

It can be concluded that the imaging capabilities of the reproduction systems can be understood reasonably well by examining the ITDs and ILDs in 1/3-octave bands. The method does not claim to give the actual image source direction, but provides results that are consistent with previously gained insight in stereophonic imaging and in wave fields created by WFS.

Concerning the results of the reproduction systems, this qualification technique suggests a constant high quality image direction for the WFS system: for frequencies to about 5 kHz at

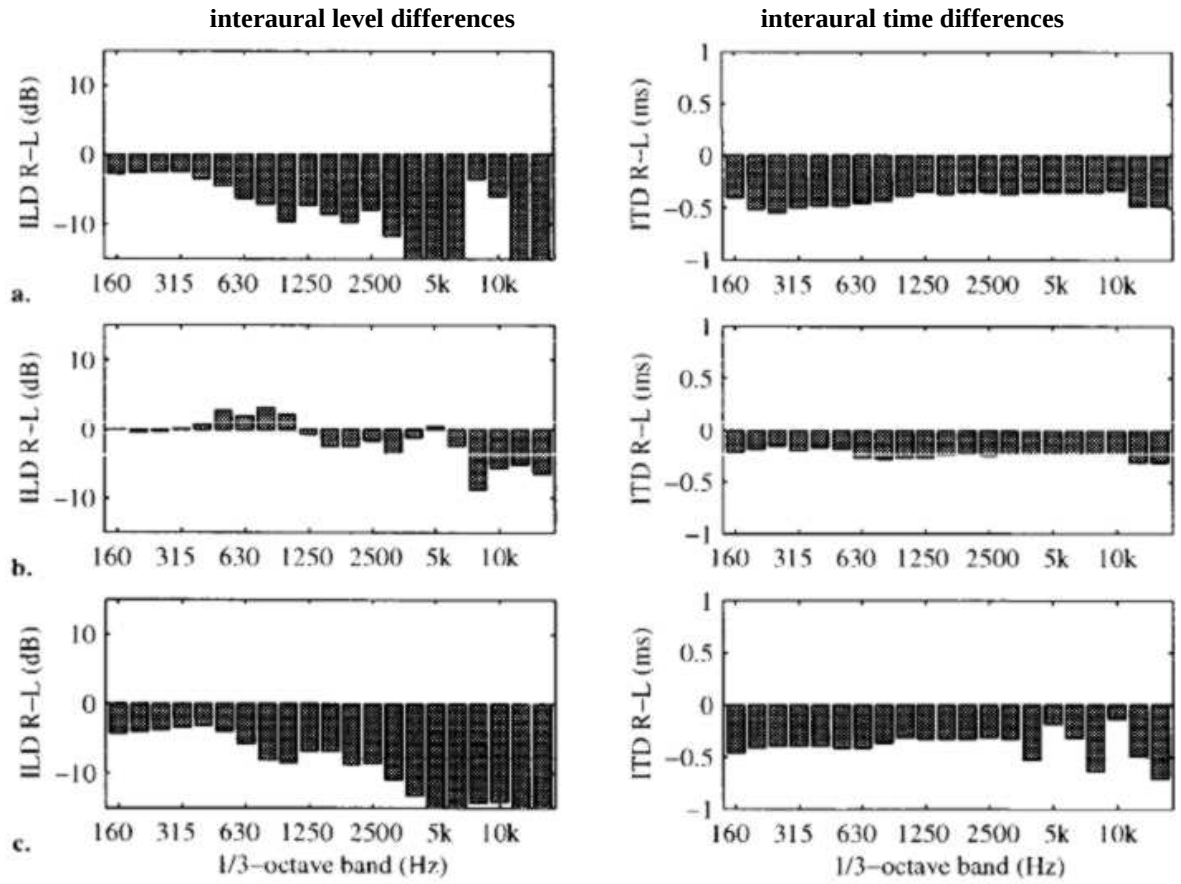


Figure 5.11 ILDs and ITDs for source *b* and dummy head position 1.
a. Real source in anechoic room. **b.** Two-channel stereophony. **c.** Wave field synthesis.

* The ORTF-pair gives 0.3 ms delay and 7 dB level difference between both stereo signals.
† A source direction of 150º is also possible for these ITDs, but can be excluded here for obvious reasons.

least, the ILDs and ITDs are in good agreement with those of the real source in all cases considered. The stereo system works fine in the 'mono' set-up at the sweet spot, but fails at the off-axis position. This precarious result may also be assigned to the 3/2 surround system, since it relies on stereophonic imaging.

5.3.3 Multi-trace impulse responses

The spatial properties of original and reproduced wave fields can be analyzed by means of multi-trace impulse responses (*xt*-diagrams). In this thesis, only the displaying potentials of the method are used, though a thorough spatial analysis is also possible (Berkhout et al. 1997a). Examples of *xt*-diagrams have already been given in Figure 2.24 and Figure 5.2.

The measurement situation is again that of Figure 5.8c, but without the dummy head. The measurement microphone is displaced along the dashed line (parallel to the front array through position 1), where a set of impulse responses is acquired from both the two-channel stereo and WFS systems reproducing direct sound only. As a reference, a simulation of the real sources is provided as well.

The data sets for source *a* are depicted in Figure 5.12. The simulation shows a single wave front crossing the measurement line near t = 18 ms. The stereo system produces two distinct wave fronts followed by ceiling and floor reflections from the reproduction room. The delay between these wave fronts and their directions change along the measurement line, thereby establishing different listening conditions at various positions. The WFS system approaches the original wave front best. The curvature of the wave front is unique for a certain source position, which in this case is precisely located at the true source position, which is outside the actual reproduction room.

The results for source *b* (Figure 5.13) endorse the spatial superiority of the WFS reproduction system. Apart from reflections of the reproduction room, the wave front for this system is identical to the simulated one (within the bandwidth given). Again, the stereo system

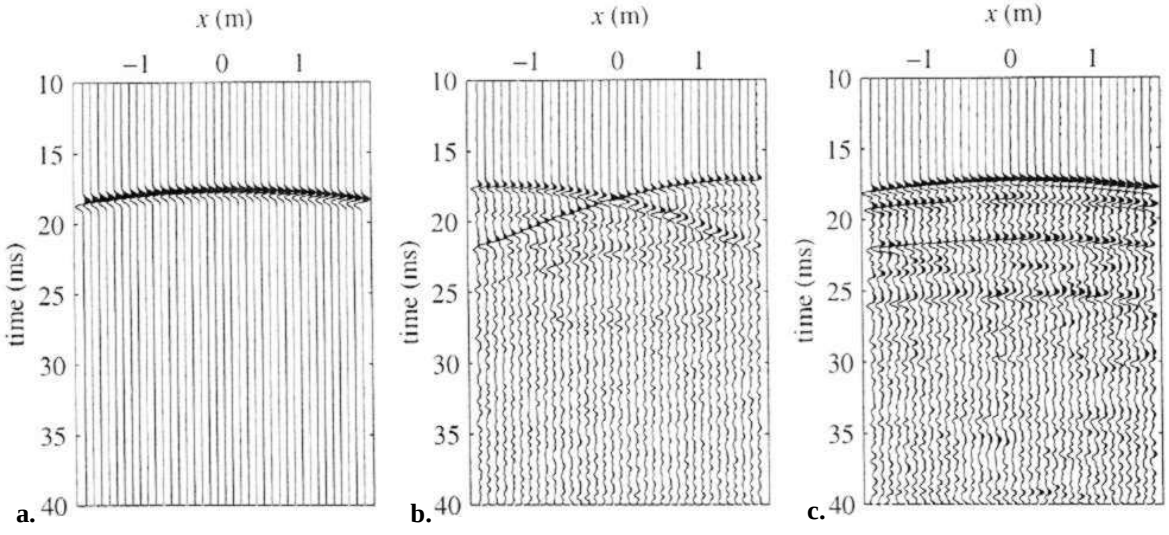


Figure 5.12 *xt*-diagrams for source *a*. The bandwidth of the Gaussian wavelet is 0.1 - 2.5 kHz.
a. Simulated real source. **b.** Two-channel stereophony. **c.** Wave field synthesis.

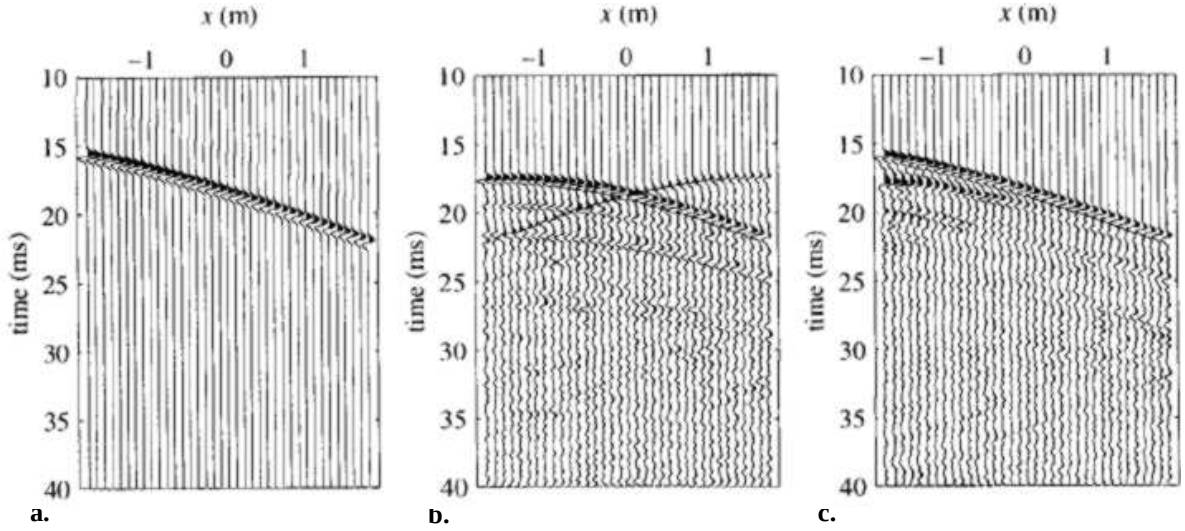


Figure 5.13 *xt*-diagrams for source *b*. The bandwidth of the Gaussian wavelet is 0.1 - 2.5 kHz.
a. Simulated real source. **b.** Two-channel stereophony. **c.** Wave field synthesis.

reproduces the simulated-ORTF signals, yielding two wave fronts rather than one: the responses of both loudspeakers can be recognized immediately.

5.3.4 Intensity measurements

The sound intensity is a vector quantity defined as

$$\boldsymbol{I}(\boldsymbol{r}, t) = p(\boldsymbol{r}, t)\ \boldsymbol{v}(\boldsymbol{r}, t)\,, \tag{5.3}$$

where p is the sound pressure and $\boldsymbol{v}$ is the particle velocity at $\boldsymbol{r}$. The time-averaged sound intensity reads

$$\boldsymbol{I}_{\mathrm{av}} = \frac{1}{T}\int_0^T p(t)\boldsymbol{v}(t)dt\,, \tag{5.4}$$

with T the period of averaging. The sound intensity can be interpreted as the power flux through a surface of unity area. For a monopole source, $\boldsymbol{I}$ is directed radially. This property is useful to determine the source direction in a sound field: in absence of other sources and reflections, the source is located in the direction opposite to that of $\boldsymbol{I}$.

In a reflective environment, the intensity is a resultant of sound entering from all directions. If sound waves are entering equally strong from all directions, e.g. in a reverberation chamber, the intensity approaches zero. Hence, a low value for $|\,\boldsymbol{I}\,|$ is a necessary condition in determining whether a sound field is diffuse[*]. Since the sound intensity depends on the strength of the source, a better condition of diffuseness can be derived from the reactivity index, defined as the intensity level minus the sound pressure level:

$$L_K = L_I - L_P \qquad [\text{dB}] \,. \tag{5.5}$$

The reactivity index is zero for a source in the free field, and negative in a diffuse field.

The sound intensity is measured on a grid in the reproduction room, as depicted in Figure 5.14a. The measurement probe (Figure 5.14b) consists of two calibrated pressure microphones separated by a 12 mm spacer. The probe is placed at a grid point, pointing in the *x*-direction. During 4 seconds of averaging, $\boldsymbol{I}_{x,\text{av}}$ is measured, after which the probe is rotated manually towards the *z*-direction. The intensity vector is calculated by composing $\boldsymbol{I}_{x,\text{av}}$ and $\boldsymbol{I}_{z,\text{av}}$ in the horizontal plane (Figure 5.14c). The vertical component $\boldsymbol{I}_{y,\text{av}}$ is not measured. The accuracy of measurement is guaranteed only in the frequency range between about 250 Hz and 5 kHz, due to the finite-difference approximation used. All measurements presented here employ this frequency range.

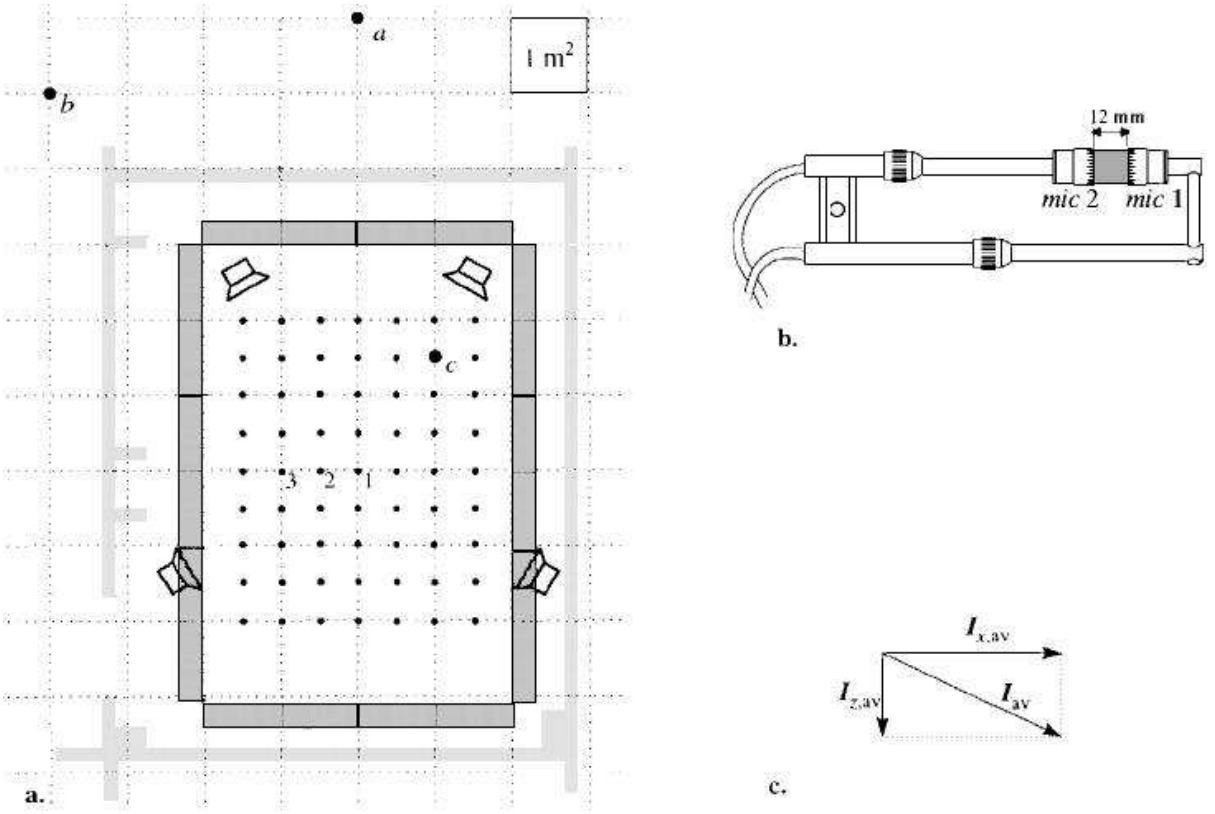


Figure 5.14 a. Two sources *a* and *b* are imaged by each of the reproduction systems. Source *c* is imaged only by the WFS system by means of focusing. Intensity measurements are carried out on a 7×9 grid. (Positions 1 to 3 refer to sound pressure measurements of Section 5.3.5). **b.** The intensity probe (B&K 3545) determines the intensity at the axis through both microphones. **c.** The intensity vector in the horizontal plane is composed of measurements along two perpendicular axes at each grid point.

* Note that $|\boldsymbol{I}| = 0$ is not a sufficient condition: it only requires that sound waves entering from opposite directions should cancel (e.g. two equal plane waves traveling in opposite directions yield $|\boldsymbol{I}| = 0$, but are not diffuse). Caution is therefore needed in interpreting low values of sound intensity as a proof of diffuseness.

The intensity distributions of direct sound, early reflections and reverberation are investigated separately for two-channel stereophony, 3/2 surround and WFS. The imaging technique for direct sound for the 3/2 system is purely stereophonic ('2/0'), as in the previous section. An extension towards 2/2 surround is used for early reflections and reverberation. So again, the center loudspeaker is not incorporated.

Direct sound

The results for the direct sound from source *a* are shown in Figure 5.15. The intensity vectors are drawn at each grid point. The length of an arrow is a measure for the intensity level, as indicated below each plot by the extrema measured. E.g. '76 dB (max)' means that 76 dB is the maximum value of $|\boldsymbol{I}|$ in this plot. Comparison of both reproduced intensity distributions with the simulated distribution learns that the WFS system is closer to the ideal situation. A few disturbances are probably caused by reflections in the reproduction room.

For source *b* in Figure 5.16, the distribution for the stereo system using the ORTF-pair is dominated by the left-hand loudspeaker, which produces a signal power that is 7 dB higher than that of the right-hand one. The direction of the intensity vectors at the rear end of the listening area is hardly influenced by this power difference. The results for the WFS system also show deviations in the rear part of the listening area, though much smaller than those of the stereo system. These artefacts can be ascribed to the deactivation of the rear left-hand array bar, which has been motivated by economical reasons (see Figure 4.15).

Instead of simulating real source *c*, the response of a loudspeaker at this position is measured (Figure 5.17a). This source can only be reproduced by the WFS system. Not surprisingly, the direction of the intensity vectors in the area between focus and contributing arrays is oppo-

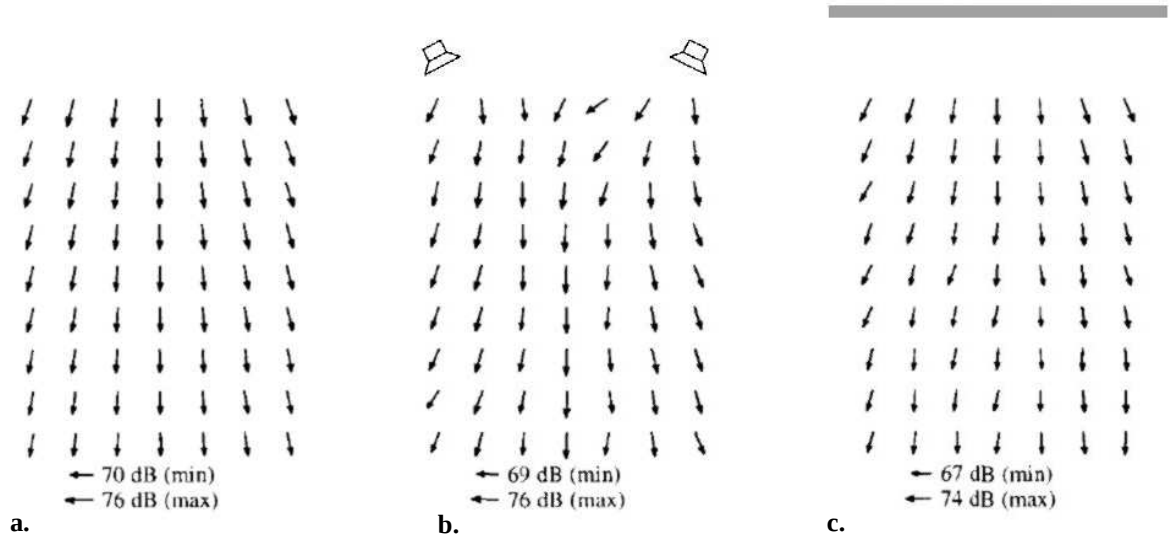


Figure 5.15 Intensity vectors for the direct sound of source *a*. Only the *active* loudspeakers and arrays are drawn. **a.** Simulated real source. **b.** Two-channel stereophony. **c.** Wave field synthesis.

site to that of the real source. At the focus itself, the resultant intensity vector is directed away from the contributing arrays. At some distance away from the focus, the vector field for the WFS system is similar to the ideal field.

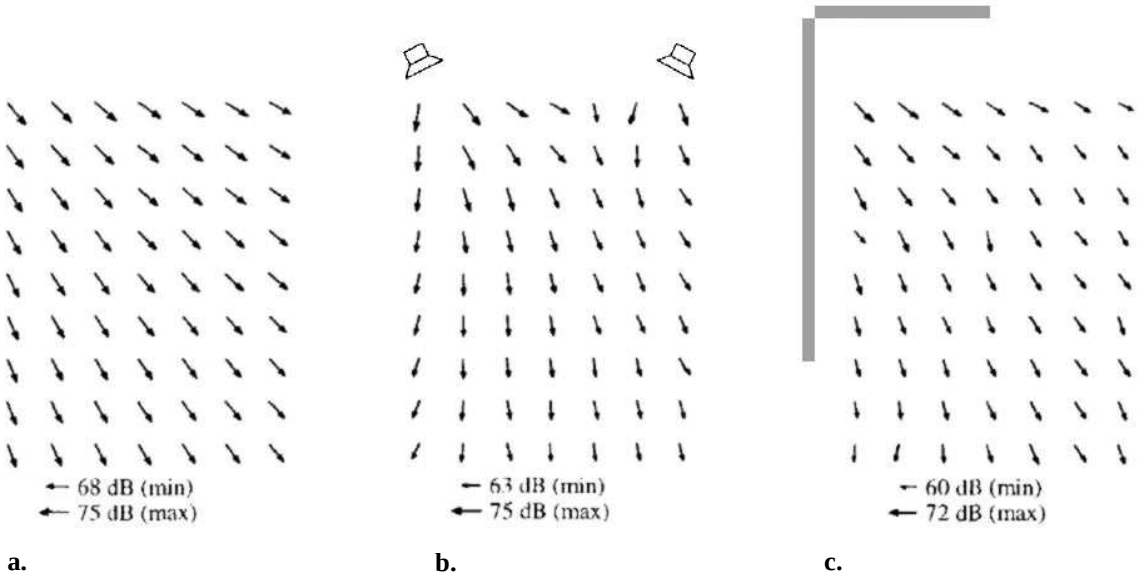


Figure 5.16 Intensity vectors for the direct sound of source *b*. Only the *active* loudspeakers and arrays are drawn. **a.** Simulated real source. **b.** Two-channel stereophony. **c.** Wave field synthesis.

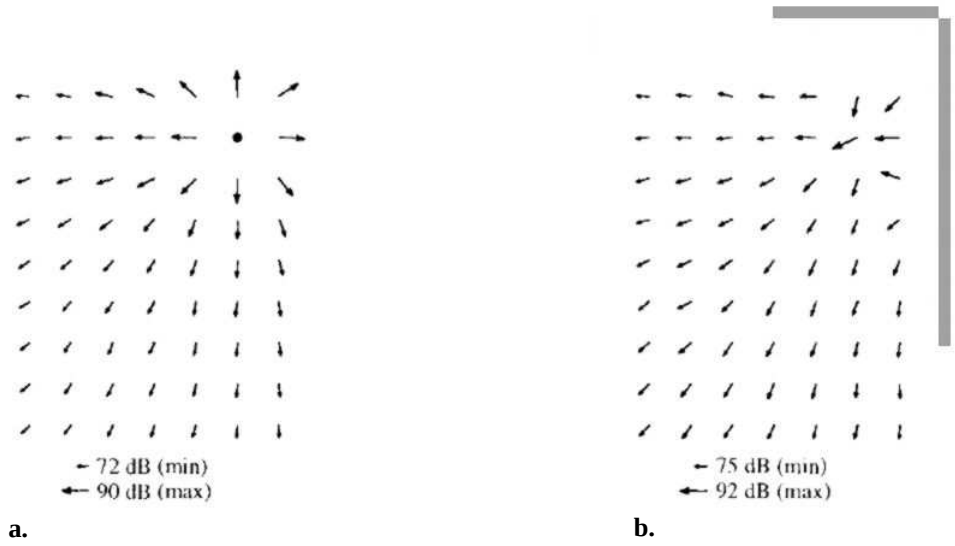


Figure 5.17 Intensity vectors for the direct sound of source *c*. Only the *active* arrays are drawn. **a.** Real source (measured response of a loudspeaker at this position). **b.** Wave field synthesis.

Early reflections

The early reflections of source *b* are modeled in the following manner. A simple virtual hall with a reasonable number of early lateral reflections is designed, see Figure 5.18. The arrival times of the first twelve reflected waves (first and second order) vary between 20 and 120 ms. The two-channel stereo system uses an ORTF-pair at the center (×) to image these reflections. The 2/2 surround system employs an ORTF-quartet (see Figure 1.20b) at this position. The WFS system reproduces the reflections by means of virtual mirror sources.

In interpreting and comparing the energy flows, see Figure 5.19, some anomalies attract attention. The pattern for the stereo system is very similar to that for the monophonic situation of source *a*, shown in Figure 5.15b. Obviously, the temporal fine structure of the loudspeaker signals is lost in the results during averaging – only the total sound power remains, which is more or less the same for both loudspeakers in this situation. These measurements confirm the lack of envelopment in stereophony: all sound power enters from the front.

The influence of the surround loudspeakers on the vector pattern is marginal, but clearly visible (Figure 5.19b). Indeed, the rear loudspeakers are radiating about 3 dB less energy than the front ones, for this particular hall model and ambient microphone technique. Most important however, is that lateral sound energy is missing in this configuration.

For the WFS system, the ‘flow’ of the intensity vectors can be understood by regarding the concentrations of mirror sources around the arrays. The highest power is generated by both virtual mirror sources near the front left-hand corner, because they are closest to the listening area. Other groups of mirror sources assert their influence to a lesser degree. In this reproduction method, the lateral reflections are truly incident from the sides.

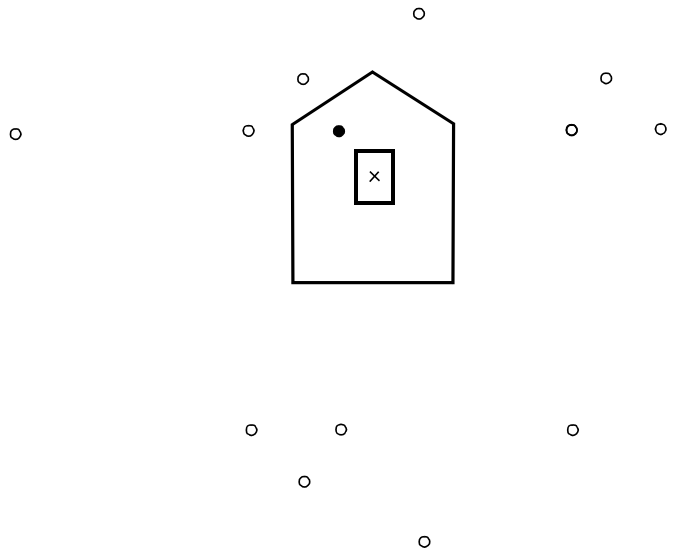


Figure 5.18 Mirror sources (○) from source *b* (•) as a consequence of a large hall (thick lines). The rectangular array configuration is drawn to the same scale. The absorption coefficient of the rear wall is 0.5, while it is 0.2 for the other walls.

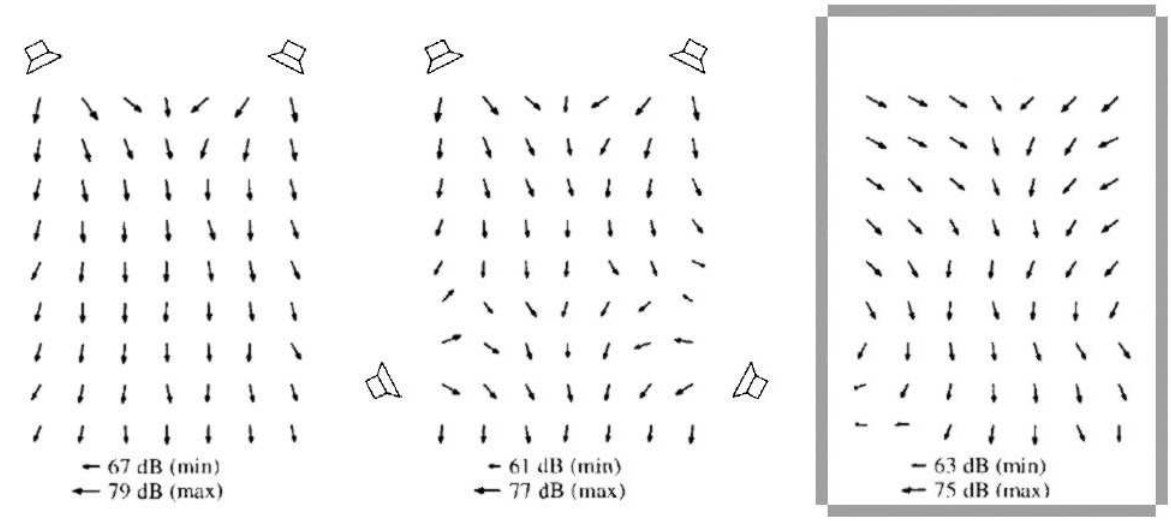


Figure 5.19 Intensity vectors for the early reflections from source *b*.
a. Two-channel stereophony. **b.** 2/2 Surround system. **c.** Wave field synthesis.

Reverberation

The reverberation is reproduced as follows. Two digital reverberation units* supplied four highly uncorrelated reverberant signals, which have no particular relation to the hall-model used for the early reflections. For the stereo system, two of these signals were fed to the loudspeakers. The 2/2 surround system and the WFS system received four signals each. The WFS system synthesized four plane waves, from the directions ± 45º and ± 135º (front direction: 0º). The results are shown Figure 5.20. The reverberation from the stereo system isentirely reproduced from the front. The four channels in the 2/2 system and WFS system disturb this regular pattern.

As a qualification for the diffuseness of the reverberant sound fields, the reactivity index is calculated for these data as well. From the gray levels of Figure 5.21 it is clear that the stereo system produces only very low diffuseness, averaging to L_K = –3.4 dB. The 2/2 system is able to create low reactivity at the sweet spot, but near the corners it is less: the average reactivity index across the listening area is –8.0 dB. It is also interesting to realize that if a center loudspeaker with a reverberant signal is added, thus forming a 3/2 surround system, the conditions for diffuseness worsen due to the breaking of symmetry. In this configuration (not shown), an average of –7.4 dB has been measured. The reactivity measured for the WFS system is spread more evenly across a large area, yielding L_K = –9.7 dB on the average. This large listening area is a benefit of reproducing four plane waves instead of using four individual loudspeakers.

* TC Electronic M2000; modified preset ‘Warm MidSized Hall’ with reverberation time 1.8 s.

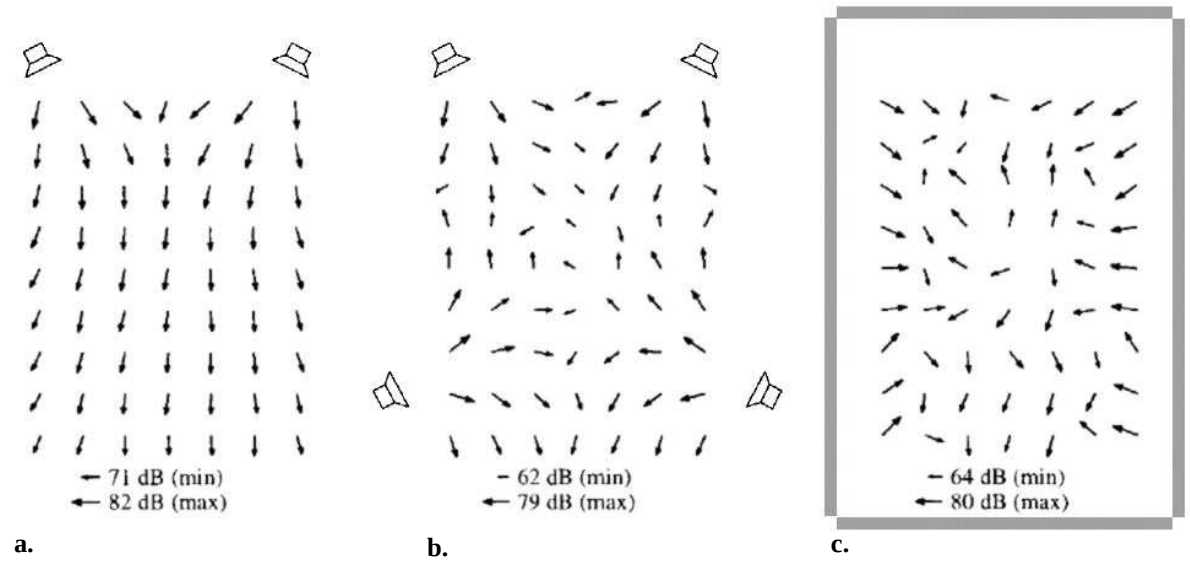


Figure 5.20 Intensity vectors for reverberation.
a. Two-channel stereophony. **b.** 2/2 Surround system. **c.** Wave field synthesis.

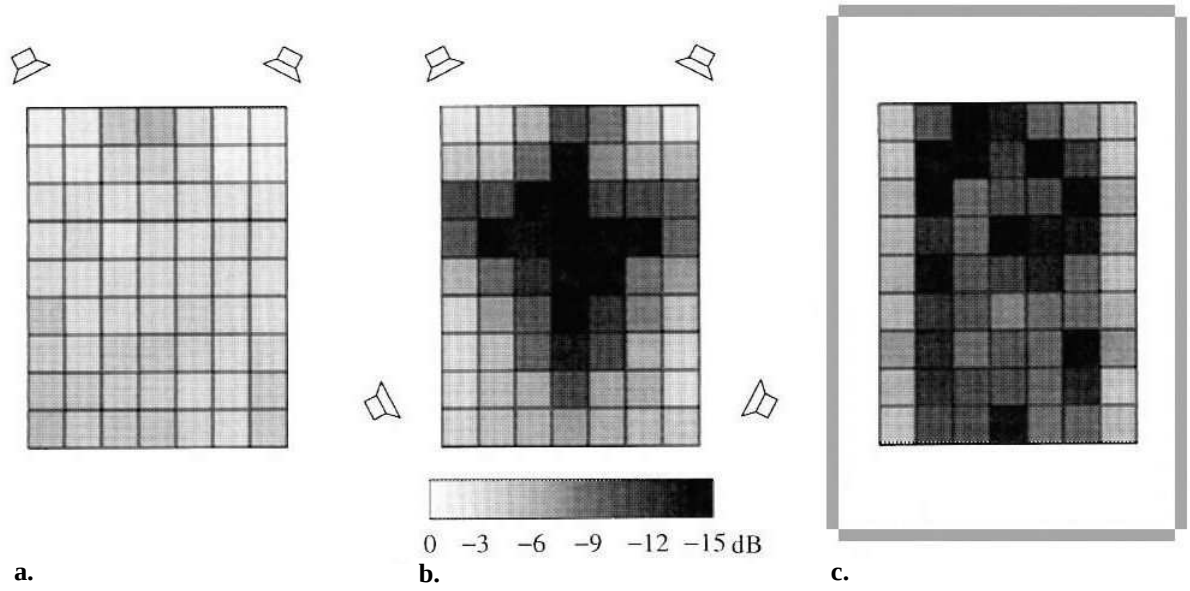


Figure 5.21 Reactivity distribution for reverberation.
a. Two-channel stereophony. **b.** 2/2 Surround system. **c.** Wave field synthesis.

Whether low reactivity really indicates high diffuseness or not, depends on the angular spread of incident sound waves: a large number of waves, arranged in oppositely directed pairs, will produce a diffuse field. In the current qualification of the 2/2 surround system and WFS system, chances were equal because both systems employed four uncorrelated signals. Though this number is restricted by the number of loudspeakers for the 2/2 system, it can be increased easily for the WFS system by adjusting the software. By applying appropriate delays

to the available reverberant signals, a much higher number of perceptually uncorrelated waves can be created. Maat (1997) found that with the present horizontal reproduction array, no more than ten uncorrelated plane waves were necessary to create a perceptually diffuse sound field for trained subjects. For less-trained subjects a lower number of plane waves sufficed.

5.3.5 Frequency spectrum measurements

Together with the intensity data of the previous section, the time-averaged sound pressure levels at the grid points have been collected in 1/3-octave bands. A selection of these data has been made to illustrate some typical spectra for the reproduction of direct sound by two-channel stereophony and WFS. These typical responses are measured at three adjacent grid points close to the stereophonic sweet spot. The choice of the positions is based on the consideration that stereophonic reproduction suffers from comb-filter effects, which depend on the listening position with respect to the sweet spot (Section 1.6.4). In anechoic circumstances, such dips can easily be perceived as coloration. In a reflective environment, the spectral distortion is usually masked by early reflections that arrive within the integration time of the auditory system. Though in the present measurements not only early, but also later arriving reflections are included in averaging, the spectral influence of the latter can be neglected as they possess much less strength. Therefore, the measured spectra can be regarded to be significant in determining whether there will be big differences in sound color perceived at these positions. It will be interesting to compare the differences for stereophony with those for the WFS system.

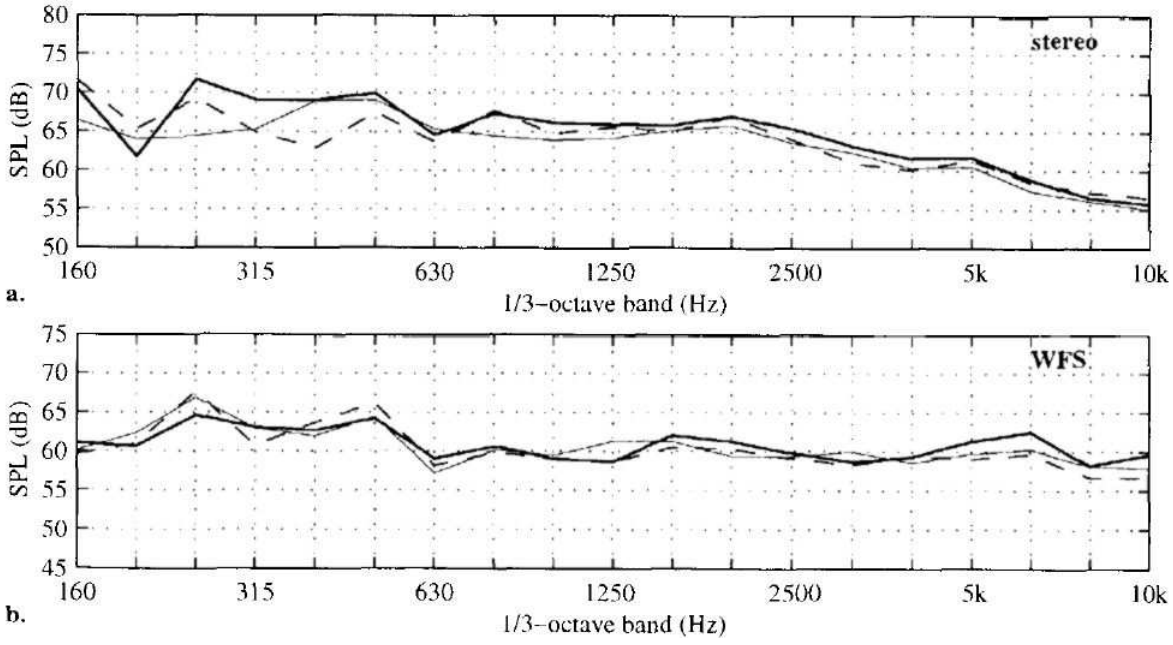


Figure 5.22 Frequency responses of the direct sound of source *a*, measured at positions 1 (thick line), 2 (dashed line) and 3 (thin line) as defined in Figure 5.14a. **a.** Two-channel stereophony. **b.** Wave field synthesis.

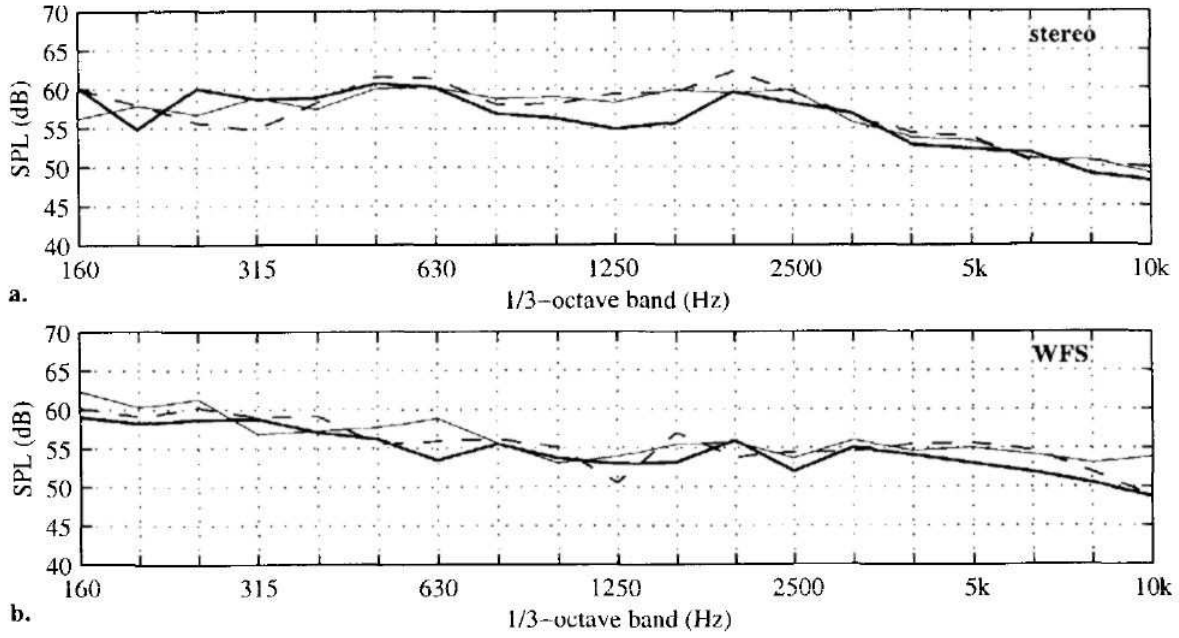


Figure 5.23 Frequency responses of the direct sound of source *b*, measured at positions 1 (thick line), 2 (dashed line) and 3 (thin line) as defined in Figure 5.14a. **a.** Two-channel stereophony. **b.** Wave field synthesis.

Figure 5.22 shows the 1/3-octave band spectra of the reproduced direct sound of source *a*, measured at the grid points marked 1, 2 and 3 in Figure 5.14a. Except for some bigger differences at frequencies ≤ 500 Hz , the spread of the three spectra is within ± 2 dB for both reproduction systems. For source *b*, see Figure 5.23, the spread of the responses is slightly larger ($\pm$ 3 dB), but again both reproduction methods produce comparable results. Similar observations were made at many other grid points, which justifies the conclusion that the frequency responses of both systems, measured in 1/3-octave bands, are equally robust in a large listening area.

5.4 Conclusions

Frequency responses of virtual sources at different distances from the array have been measured. The sound pressure levels of virtual sources at some distance from the arrays obey the 1/*r*-law. However, close to the arrays (< 1 m) the 1/*r*-law does not hold due to the use of linear arrays instead of planar arrays.

The effect of spatial aliasing for focused and non-focused virtual sources appear in different regions of the synthesized wave field for each type of source. This result, that can be explained in terms of wave field reconstruction, favors the frequency response of focused sources near the central axis of its wave field. The sound pressure levels in the focal area reveal

a steep maximum along a line parallel to the array: a level difference of 18 dB is measured between two positions 0.5 meter apart, which is in accordance with theory.

The wave field synthesis (WFS) system is compared with two-channel stereophony and 3/2 surround sound. The image directions, as produced by these systems, are qualified by interpreting the interaural time and sound pressure level differences measured at a dummy head. As expected, the results for stereophonic imaging depend on the listening position, in contrast with the WFS system which creates a large listening area.

A second qualification technique, multi-trace impulse responses, illustrates the spatial properties of the reproduced wave field. The WFS system produces wave fronts that are copies of the original ones, while with stereophonic imaging the separate contributions of both loudspeakers remain.

This is confirmed by a third qualification technique: intensity vector measurements. Also the distribution of early reflections has been visualized by this technique. Though adding surround loudspeakers to two-channel stereophony already greatly progressed the imaging of reflections, a clear lack of incident sound energy still remains to the sides. The advantage of the WFS system in this aspect is that sound can be reproduced from all around the listening area. This system also possesses the capabilities of creating a diffuse reverberant field in a large listening area. With the 2/2 system, only a point solution is offered, while two-channel stereophony cannot generate a diffuse field at all.

Finally, the sound pressure levels of reproduced direct sound have been collected in 1/3-octave bands. The spectra for stereophony and WFS do not depend strongly on the position of measurement. The 1/3-octave band spectra for both systems are of comparable smoothness.

Chapter 6

Subjective Evaluation

6.1 Introduction

The imaging capabilities of wave field synthesis are investigated perceptually in two ways: by localization experiments of virtual sources, and by a listening test in which wave field synthesis is compared directly to two-channel stereophony and 3/2 surround sound. Through localization experiments it is attempted to prove that wave field synthesis is able to reproduce sources at correct positions. The comparative listening tests were carried out to indicate roughly whether the newly-built system is superior to the other systems concerning robustness, flexibility and quality.

6.2 Localization experiments

The localization of virtual sources was compared to the localization of real sources in two sessions of subjective listening tests. The first session (A) investigated the localization of virtual sources *behind* the loudspeaker array, while the second session (B) dealt with virtual sources *in front of* the array. In these experiments, only directional hearing in the horizontal plane was considered. The array consisted of 24 custom-built electrodynamic loudspeakers (Vifa M-110 4" drivers with Δx = 11 cm) described by Vogel (1993) — the new reproduction system was not available yet at the time of these experiments.

A map of the situation is drawn in Figure 6.1. The listener was seated at a distance of 3 m from the array (with numbered loudspeakers), and was free to move his head. The virtual

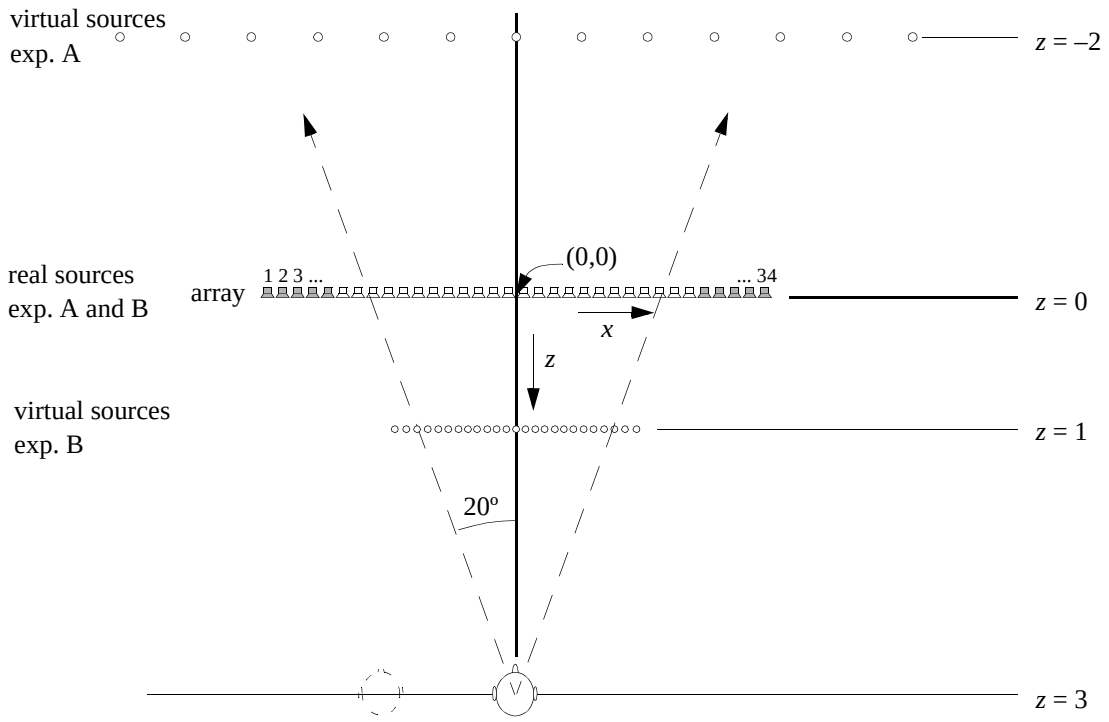


Figure 6.1 Situation map for both of the localization experiments. The origin of the coordinate system is at the middle of the array. The 20º azimuth is relevant for the understanding of effective aperture effects as explained in the text.

sources were synthesized either at 2 m behind the array or at 1 m in front of the array. The real sources were produced by single array loudspeakers. Five dummy loudspeakers were appended to either side of the array to avoid bias by limited choice. In this way, the angular spacing of the loudspeakers varied from 1.6º at the edges to 2.1º in the middle of the array.

Basically, the set-up and data processing method for this source-identification experiment is taken from Hartmann (1983). He investigated the influence of room reflections on the localization of sound sources with the aid of eight loudspeakers placed 4º apart at 12 m from the subject. The results can be described by at least three important statistical quantities. If M is the number of trials per source, S_i the stimulus direction and R_i the response direction for trial i, then the rms error for source k is defined as

$$D(k) = \left(\sum_i (R_i(k) - S_i(k))^2 / M \right)^{1/2} . \qquad (6.1a)$$

The mean response of a certain subject for source k is given by

$$\bar{R}(k) = \sum_i R_i(k) / M . \qquad (6.1b)$$

The standard deviation for source k equals

$$s(k) = \left(\sum_i [R_i(k) - R(k)]^2 / M \right)^{1/2} . \qquad (6.1c)$$

It can be meaningful to average $D(k)$ and $s(k)$ across sources k. Then, these quantities are written as $\overline{D}$ and $\bar{s}$, respectively. Statistic quantities averaged across subjects are notated as <...>.

Because the tested parameters differed between both experiments, their set-up is described in more detail in separate sections.

6.2.1 Virtual sources behind the array (exp. A)

Environment

The test series were carried out in an anechoic room, and repeated in the reproduction room (Section 4.3.2). In the latter room, the z-axis of Figure 6.1 was parallel to the side walls. The array was placed about 1.5 m from the front wall, and directed toward the rear wall.

Stimuli

Frozen white noise bursts with a duration of 1500 ms were used as stimuli. Attack transients were avoided by rise and decay durations of 15 ms. After each stimulus there was a pause of 5 seconds. The input signals were compensated for the average on-axis response of the loudspeakers (similar as mentioned in Section 4.3.4), resulting in a nearly flat spectrum in the range from 100 Hz to 16 kHz. The array was tapered with a cosine-window over the outer 3 loudspeakers, in order to diminish truncation effects.

Variables

Position of the virtual source. There were 13 fixed virtual source positions at a distance of 2 m behind the array, ranging from $x = -3$ m to $x = 3$ m in steps of 0.5 m. The outer positions were expected to be difficult to localize because the subject's position was outside the reconstruction area for these virtual sources (as defined in Figure 2.18).

The spacing Δx of the loudspeakers. Two spacings of the loudspeakers were programed: $\Delta x = 11$ cm and $\Delta x = 22$ cm. The latter was obtained by addressing alternating speakers, leaving the total array length unchanged.

These parameters were varied during one experimental run. There were two trials per source direction. In a separate run, also six real sources were presented. The subjects were not informed of the special nature of this run. Ten subjects participated, all males aged 20 to 30 years without hearing complaints.

Task

The subjects were asked to estimate the direction of the sound by writing down the number of the loudspeaker closest to that direction.

Results

The individual responses $R_i(k)$ of one subject are plotted in Figure 6.2; the solid line represents the true stimulus direction. The responses for central stimulus directions $-20^o < S < 20^o$ are close to the true directions, but they deviate for $|S|>20^o$. The same trend is visible in responses $<\overline{R}(k)>$, which are averaged across trials and subjects (though not across source directions k), see Figure 6.3. Evidently the outer three virtual sources are localized close to the edges of the array. The anechoic and reproduction room responses, as well as the two tested spacings Δx, give similar results. Whether these parameters have a significant influence on localization or not, is investigated by looking at the averages across all sources and subjects.

Figure 6.4a shows the mean and standard deviations across all subjects of statistic $\overline{D}$, the rms discrepancy between the response and true direction of the (virtual or real) source. The outer three virtual sources at either side of the array are excluded here, because these sources were only intended for investigating the finite aperture of the array. In comparing the anechoic and reproduction room results, a little deterioration is noticed for both real sources and virtual sources in the reflective environment. A similar effect is reported by Hartmann (1983), who also studied the localization of broadband noise stimuli without attack transients. Furthermore, the differences in $<\overline{D}>$ for the virtual sources with Δx = 11 cm and Δx = 22 cm are small but significant (68% confidence): $<\overline{D}_{22,\text{anech.}}-\overline{D}_{11,\text{anech.}}> = 0.7 \pm 0.4^o$ and $<\overline{D}_{22,\text{repr.}}-\overline{D}_{11,\text{repr.}}> = 0.4 \pm 0.3^o$.

The reproduction-room results for the virtual and real sources should be compared with some care, because the real sources were closer to the subject than the virtual sources were*. The localization of real sources can be expected to worsen with increasing distance, because the relative strength of the reflections increases. However, this rule does not necessarily hold for virtual sources, because here the reflection patterns depend not only on the position of the virtual sources, but also on the position and directivity of the array. Therefore, the influence of the distance is an unknown factor in comparing these virtual and real sources.

The rms error $\overline{D}$ includes a certain bias, because it is not calculated with respect to the mean response, but with respect to the real source direction. The minimum audible angle (MAA) can be related to $\overline{D}$ for zero bias (Hartmann 1983). For this purpose statistic $\bar{s}$ is introduced: by comparing definitions (6.1c) and (6.1a) it can be seen that $\overline{D} = \bar{s}$ if the mean response $\overline{R}$ matches the true source direction S, i.e. if no bias is present. Whereas $\overline{D}$ can be interpreted as the accuracy of localization, the standard deviation $\bar{s}$ is a measure for the internal blur: if the subject is consistent in his responses (because of a high perceptual resolution), $\bar{s}$ will be low. On the contrary, a large value for $\bar{s}$ corresponds with a large spread in responses. Figure 6.4b gives the mean values of $\bar{s}$ for the same data as Figure 6.4a. The difference between the results for the virtual sources with different Δx are significant in anechoic circumstances, $<\bar{s}_{22,\text{anech.}}-\bar{s}_{11,\text{anech.}}> = 0.4 \pm 0.3^o$, but negligible in the listening room.

* The positions of the virtual sources could not be taken up by the real sources in this experiment, because the former were synthesized outside the listening room.

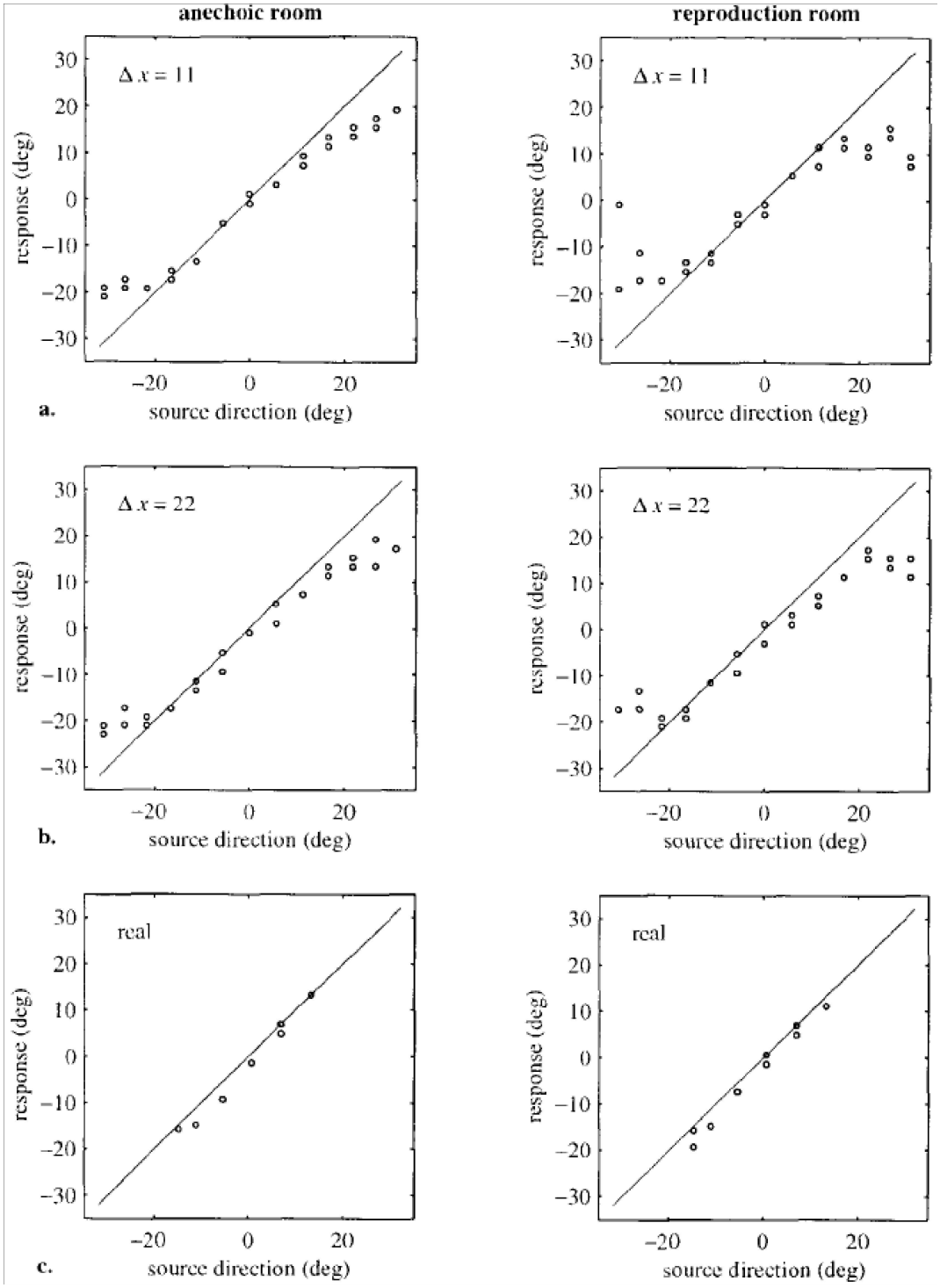


Figure 6.2 Responses $R_i(k)$ of subject PvT for anechoic room (left) and reproduction room (right).
a. Virtual sources, with Δx = 11 cm.
b. Virtual sources, with Δx = 22 cm.
c. Real sources.

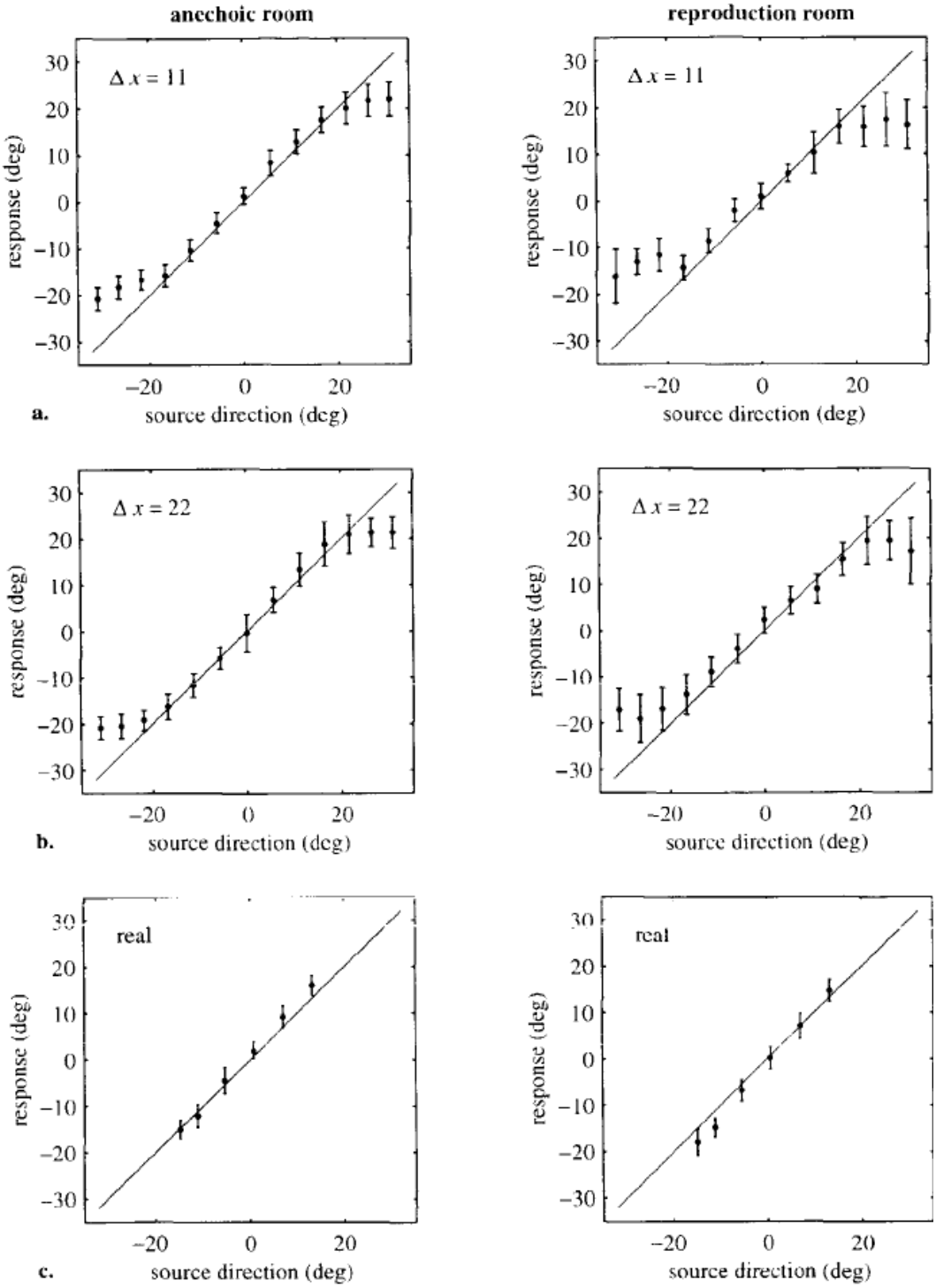


Figure 6.3 Mean responses $\langle \overline{R}(k) \rangle$ across all trials and subjects for anechoic room (left) and reproduction room (right). The error bars are have a width of twice $\langle \overline{s}(k) \rangle$.
a. Virtual sources, with Δx = 11 cm.
b. Virtual sources, with Δx = 22 cm.
c. Real sources.

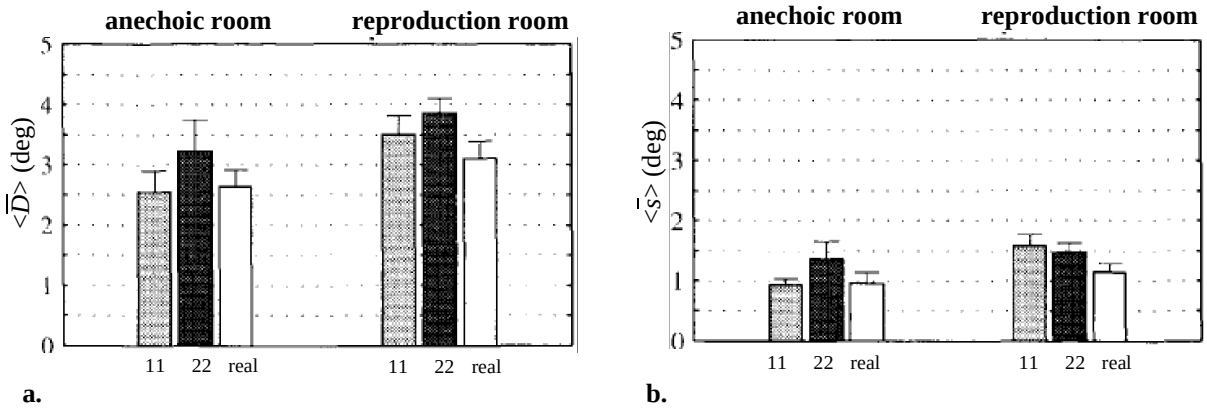


Figure 6.4 Mean results across sources and subjects, for virtual sources (Δx = 11 and Δx = 22 cm) and real sources.
a. Mean rms error $\langle \bar{D} \rangle$.
b. Mean standard deviation $\langle \bar{s} \rangle$.

Discussion

The bending of the response curve in Figures 6.2 and 6.3 for azimuths greater than 20º gives rise to a rule in wave field synthesis, stating that no source can be localized in a direction were no loudspeaker is active. In other words, the listener perceives the space of virtual sources through an acoustic window: the array. This space is restricted by two lines of sight from the listener to positions near the edges of the array (dashed lines in Figure 6.1).

Start (1997) was able to compare the standard deviation $\langle \bar{s} \rangle$ and the MAA measured for virtual sources, by re-analyzing similar source-identification experiments from Vogel (1993) and using his own results for the MAA. Since Vogel did not apply the spacing Δx = 22 cm nor did he use real sources, only a comparison of the virtual source results for Δx = 11 cm was possible. The present work completes this comparison of anechoic results, shown in Table 6.1.

Table 6.1 Comparison of localization experiments in anechoic circumstances.

	$\langle \bar{s} \rangle$ (Verheijen)	**MAA** (Start)	$\langle \bar{s} \rangle$ (Vogel)
real sources	0.96 ± 0.18º	0.79 ± 0.09º	
virtual sources; Δx = 11 cm	0.94 ± 0.10º	0.77 ± 0.11º	1.1 ± 0.1º
virtual sources; Δx = 22 cm	1.37 ± 0.29º	1.57 ± 0.23º	
stimulus type	white noise	white noise	pink noise
bandwidth	0.1 to 16 kHz	0.1 to 8 kHz	0.1 to 8 kHz
subjects × trials	10×14	3×300	3×60
array length	2.7 m	5.4 m	5.4 m
experiment	in situ	KEMAR recordings	in situ

The results support the arguments of Hartmann (1983) that the value for the MAA and the value for $<\bar{s}>$ are of the same order. In addition, it can be concluded that an increased spacing between the loudspeakers causes an increased localization blur. There is no significant difference in localization blur between real sources and virtual sources with $\Delta x = 11$ cm.

6.2.2 Virtual sources in front of the array (exp. B)

Environment

The test series were carried out in an anechoic room only.

Stimuli

Random pink noise bursts with a duration of 2 seconds were used as stimuli. The abrupt edges of the stimuli provided additional localization cues. It was considered that in their localization properties, such stimuli are closer to speech and (rhythmic) music than are cosine-windowed stimuli. A pause of 2 seconds followed each stimulus. Due to hardware limitations, the correction factor $\sqrt{k/j}$ from driving function (2.29b) was not implemented here, which emphasizes the low frequencies. The array was tapered over the outer 4 loudspeakers with a cosine-window to diminish truncation effects. A cut-off filter at 16 kHz limited the bandwidth.

Variables

Position of the virtual source. There were 25 fixed virtual source positions at a distance of 1 m in front of the array, at azimuthal intervals of 2º. The outer virtual sources were expected to be difficult to localize, because they are influenced by array truncation effects.

The bandwidth of the stimulus. The aliased waves of focused sources travel ahead of the actual wave fronts. The aliased waves that arrive earliest are radiated from the outermost array elements (see Figure 5.4b). It was expected that erroneous localization would appear, if only frequencies above the aliasing frequency were produced. To investigate this, the experiment was repeated with high-pass filtered stimuli (2 – 16 kHz).

Position of the subject. Besides the base position at $(x, z) = (0, 3)$, see Figure 6.1, also a position at (–1, 3) was used. Here, only the broadband virtual sources were presented.

In one run, each virtual source was presented four times at random. Two subjects participated, doing each run twice, thus yielding eight trials per source. In a separate run at the base position, also 24 real sources (the array loudspeakers) were produced, with both broadband and high-pass filtered signals.

Task

The subjects were asked to estimate the direction of the sound by writing down the number of the loudspeaker closest to that direction.

Results and discussion

The individual responses $\overline{R}(k)$ of both subjects are depicted in Figure 6.5; the solid line represents the true direction. For the virtual sources, a localization aperture seems to be in effect for about $|S| < 18^o$. This is a bit less than the aperture in exp. A, which is probably a result of the wider taper applied in this experiment. Also, a clear bias for virtual as well as for real sources is present in the results of subject GJ. The results for the high-pass filtered stimuli do not deviate significantly from the broadband results. Apparently, the localization task is not hindered by the (first-arriving) aliased waves from the outer loudspeakers. Because the dense aliasing tail does not exceed a few milliseconds, an integration mechanism in the auditory system may be held responsible for the reasonable accuracy of localization for these focused sources. The response function for the virtual sources is, however, not as regular as for the real sources. This regularity can be expressed by the linear correlation coefficient *r*, being better than 0.999 for the real sources and ranging from 0.991 to 0.993 for the virtual sources (within the aperture).

The responses for the listening position at (–1, 3) are shown in Figure 6.6. At this position, the outermost active loudspeaker is at 38º, while the bending of the response function starts already at about 30º. Like for virtual sources *behind* the array, the effective acoustic window is smaller than the array width.

Figure 6.7 summarizes the statistical parameters of the experiment (only for the base position of the subject). The differences in rms discrepancy $\overline{D}$ and response standard deviation $\bar{s}$ between virtual and real sources are significant in all cases considered. However, a localization accuracy of less than 5º and a localization blur of less than 2º is still a good result for the virtual sources.

6.2.3 Evaluation of both experiments

Due to the large bias for one of both subjects in experiment B, it is awkward to compare $\langle\overline{D}\rangle$ for sources *behind* (exp. A) and *in front of* (exp. B) the arrays. The result of $\langle\bar{s}\rangle$ for *real* sources, however, is approximately equal in both experiment sessions. Therefore it is also allowed to compare the results of $\langle\bar{s}\rangle$ for the broadband *virtual* sources synthesized at either side of the array (Table 6.2), though some reserve is in order here because of the small number of subjects in exp. B. The response standard deviation for focused virtual sources is about 80% greater than for non-focused sources. In an absolute sense, however, there is less difference than 1º.

Table 6.2 Comparison of response standard deviation between experiment A and B.

$\langle \bar{s} \rangle$	**exp. A**	**exp. B**
real sources (º)	0.96 ± 0.18	1.08 ± 0.08
virtual sources (º)	0.94 ± 0.10	1.71 ± 0.08

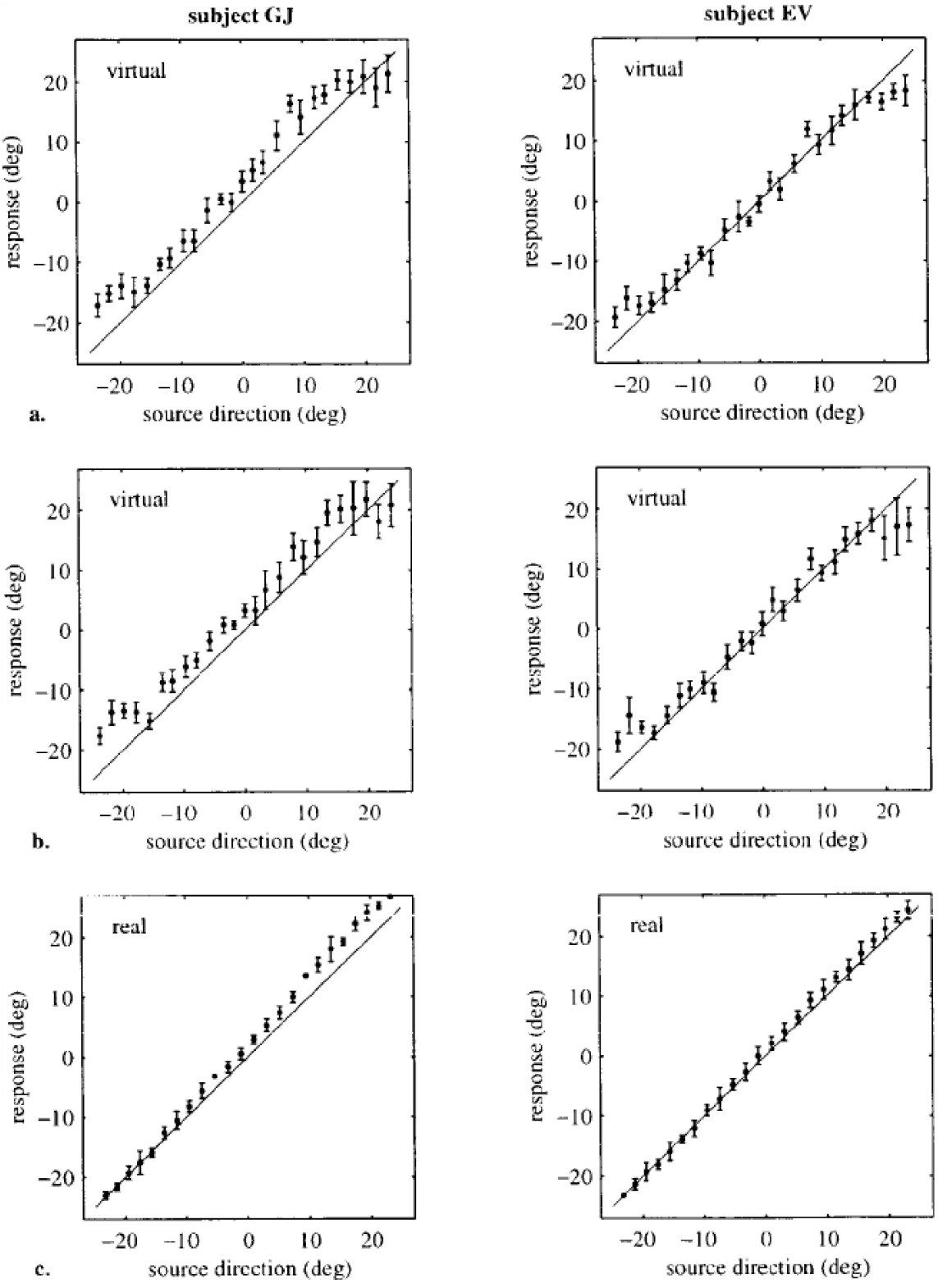


Figure 6.5 **a.** Responses $\overline{R}(k)$ for virtual sources with broadband stimuli.
b. Responses for virtual sources with highpass-filtered (> 2 kHz) stimuli.
c. Responses for real sources with broadband stimuli.

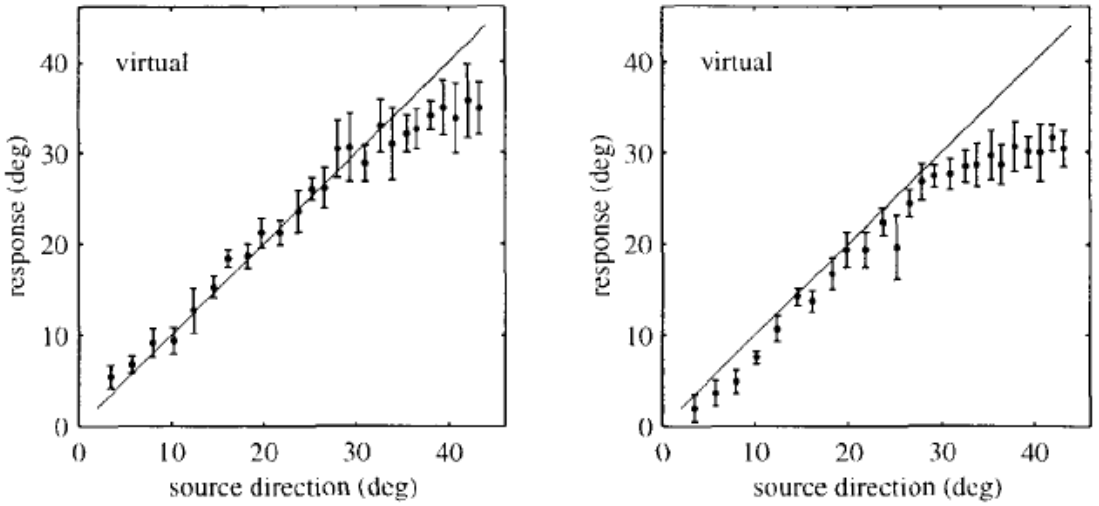


Figure 6.6 Virtual source responses $\overline{R}(k)$ with broadband stimuli for the off-center listening position $(x, z) = (-1,3)$.

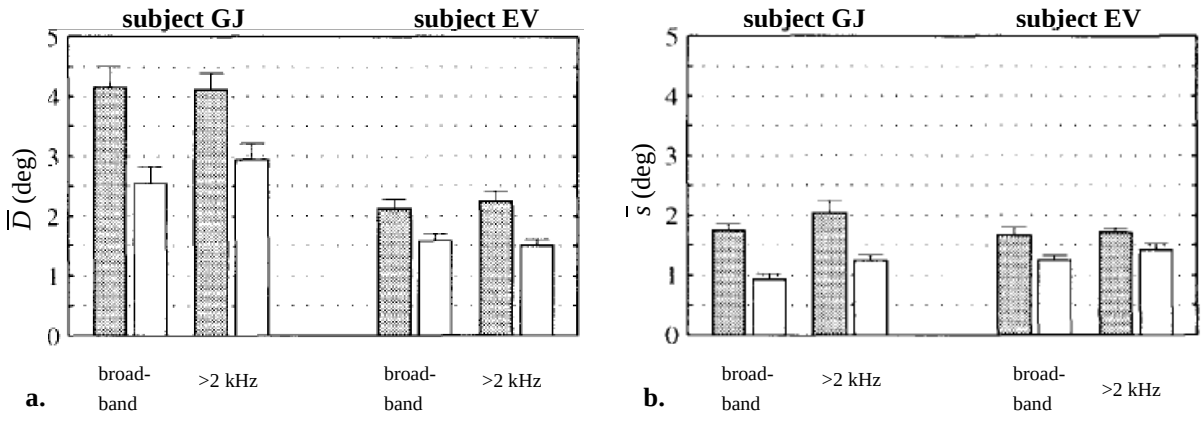


Figure 6.7 **a.** The mean rms error $\overline{D}$ for virtual (gray bars) and real sources (white bars).
b. The mean standard deviation $\overline{s}$ for virtual and real sources.

6.2.4 Distance perception

Besides the direction of the source, the apparent distance of the source is also an important factor in sound reproduction. In contrast with directional hearing, depth perception is a relative sensation with many sources active at the same time. In case of the absence of other sources, the perceived distance depends largely on the loudness of the source, but also on the subject's familiarity with the source signal (Blauert 1983). The distance control of virtual sources in the wave field synthesis system, as incorporated in the synthesis operator (Section 5.2.2), is therefore not the only significant factor in distance judgements. Moreover, the distance resolution of the human auditory system is rather poor. This is confirmed by subjective experiments of Start et. al. (1997), who evaluated the distance perception of real and virtual sources in an auditorium. The subjects, seated 25 m from stage, were asked to judge the depth

level of the sources out of three possible depth levels, 2 m apart. For the real sources, 49% of the responses was correct, while for the virtual sources (behind an array installed at 15 m from the subjects) only 40% of the responses was correct. Both results are not impressively accurate, though they are above chance (33%).

For sources closer than about 1 m to the subject, the spectral influence of the head becomes an important cue for distance estimations. Experiments of Komiyama et al. (1992) suggest that focused sources provide these cues very well. A planar array, consisting of loudspeakers arranged in concentric circles, was used to create focal points on the axis perpendicular to the array, by applying 'appropriate delay times' to the loudspeaker signals in each circle. The experimental set-up was assessed empirically – no mathematical background is given. The purpose of the experiments was to find out whether the apparent distance of the sound images can be controlled with this array configuration. The results of their subjective experiments showed that a stable image can be localized near the focal point, even if the sound level of the focused source was less than that of a reference loudspeaker at the same distance as the focus.

With a linear loudspeaker array instead of a planar array, the geometrical properties of the wave field are different. In informal tests with the laboratory demonstration system it appeared that the spectral cues of focused sources are less pronounced than those of real sources. However, if the listener is allowed to walk towards the focused source, a sense of presence is clearly felt. During many demonstrations with unexperienced as well as experienced listeners, with a saxophone recording being played back as a focused source, the listeners found the focus effortlessly. The feature of placing a soloist in front of the array is also utilized in one of the programs that serve as test material in the next section.

6.3 Comparison between reproduction systems

In this section, an attempt is made to qualify the laboratory wave field synthesis (WFS) system subjectively in comparison to two-channel stereophony and 3/2 surround sound. The previous chapter already investigated the shape of the wave fronts, reflective energy distribution, reactivity index and frequency characteristics of these reproduction systems. The purpose of the tests in this section is to find subjective differences that are related to the reproduction method used by these systems.

6.3.1 Introduction

The difficulty in judging different sound reproduction systems is how to set up a test in which only properties of the system's concept are compared. The influence of program material, the way of recording and mixing, differences in system hardware (loudspeakers, ADCs, etc.) and the placement of loudspeakers must evidently be minimized. In spite of all effort done to set up a fair comparison, it should be realized that the results will certainly be biased by unwanted side-effects. Therefore, only large and obvious differences can be revealed by such a test.

Instead of regarding many subjective criteria separately, only three global criteria are considered:

1. **Flexibility** is a system property that allows the user to control the image.

2. **Robustness** refers to the stability of the image as experienced at different locations in the listening area.

3. **Quality** is a global indication that refers to the spaciousness and naturalness of the image.

The criterion of flexibility is added because it is a novelty of the WFS system made possible by its need of the bare microphone channels as inputs (Figure 4.3). Since stereo and 3/2 surround systems receive and playback the downmix directly, it is awkward to demonstrate this property for these systems.

6.3.2 Experimental set-up

Listening panel

It is reasoned that for this comparison a relatively small number of subjects suffices, provided that they agree unanimously on their findings. Preferably, the subjects must be able to discriminate between concept-related and program-related differences in test material. This requires a group of experienced listeners who are acquainted with each of the reproduction systems under investigation. Therefore, the board that discussed the results of this research project twice a year, was invited for the test. Except the secretary, all of them were professionally involved in sound reproduction and/or acoustics. Those members of the board that were connected to the research group, were excluded from judgements. The participants were:

- a conservatory lecturer of recording engineering;
- a senior recording engineer of a record company;
- a consumer (the secretary of the board);
- a professor of perceptual acoustics;
- two acoustic researchers of a multinational electronics company.

Except the senior recording engineer, all had listened to the WFS system before. Four of them are professionally involved in 3/2 surround. All members are experienced stereo listeners.

Program material

Four music programs with a large difference in musical expression were selected (see also Appendix B):

I. Violin concert in E Minor of Mendelssohn, first 10 minutes;

II. Second Symphony of Mahler, first 4 minutes of the first movement;

III. Conception of George Shearing (jazz), 5 minutes;

IV. Overture of the Concert after Vivaldi; arranged for saxophones, 3 minutes.

The multi-track tapes of these unreleased recordings were mixed for each of the reproduction systems by a graduate student who gained several years of experience in stereophonic mixing for a broadcasting corporation. His acquaintance with the 3/2 and WFS system was much less, but after intensive practicing a fair level was obtained. His task was to strive for a sound image that was virtually the same for each system, without favoring one above the other. The equalizers on the mixing console were not used. The mixed material was taped on a digital multi-track recorder in such a way that it could be reproduced synchronously for each system. The number of channels of the downmix was two for stereophony, five for 3/2 surround and eight for WFS (four virtual sources and four plane waves). The employed bandwidth is 22 kHz for stereo and 3/2 surround, and 16 kHz for the WFS system. Documentation of the recording and mixing sessions is provided in Appendix B.

The position of all loudspeakers during mixing and testing is depicted in Figure 5.7. The rear channels of the 3/2 downmix were delayed 2 ms to account for the smaller distance of the rear loudspeakers to the sweet spot. The experimenter, standing in the rear right-hand corner next to the mixing console, was able to switch directly between the reproduction systems.

Test procedure

Before entering the reproduction room, the participants received a short instruction to inform them of the test procedure and criteria of judgement. The participants attended the test together, but were urged not to talk during the tests. They were encouraged to walk about the listening room, though not to approach the loudspeakers closer than 0.5 m. The music programs were reproduced in the order shown above. The first piece of music was played twice. During the first play, the listeners could become familiar with all audible differences between the three reproduction systems. After that, the experimenter switched at will between each system, thereby calling the name of the system that was just activated.

The experimenter demonstrated the flexibility of the WFS system by changing the source positions and the ambience of program IV in a second play. Initially, the sources were moved far-away from their original positions, while the reverberation lasted. After that, the four sources were focused just in front of the corners of the array and the reverberant signals were turned off, leaving a ‘dry’ atmosphere.

The participants were asked to write down their findings in a table. These quickly written notes served as an aid of memory for the actual evaluation that took place in a meeting room immediately after the test. There, the participants were asked one by one to comment on the performance of the reproduction systems.

6.3.3 Evaluation

Consensus was found among the participants on the following topics:

Coloration

Though coloration was not a criterion by itself, several participants responded spontaneously that the differences in coloration between the systems were clearly audible, especially

between WFS on the one hand and both stereo and 3/2 surround on the other. Probably, these differences are caused by the different type of loudspeakers used for WFS. Presumably also spatial aliasing plays a role in the coloration heard for the WFS system. It is noted that both these causes are hardware-related*. Nevertheless, it was agreed that the coloration differences did not obstruct a comparison based on the criteria listed hereafter.

Flexibility

The demonstrated flexibility of the WFS system (program IV) is appreciated as 'obvious' and 'beautiful'. It is considered that the feature of repositioning sources and changing acoustics is not only an important tool in post-production mixing, but also very interesting for non-professional users.

Robustness

On the average, sources reproduced by the WFS system are found stabler than those reproduced with 3/2 surround. The stereo image is the least robust. An exception to this rule is the Mahler concert (II), in which some instruments are less stable in WFS as perceived at the sides. In program IV, the WFS system was far more robust than the others.

Quality

The WFS system reproduces an image with much more depth than the other systems do. Generally, also the spatial impression created by this system is appreciated above that of others. Especially the live atmosphere (rumble and audience noise) just before the start of the violin concert (I) and the applause after the jazz song (III) is considered very convincing. However, the symphony orchestras (I and II) appear to be a bit detached in WFS. This feeling of distance is probably due to the mixing of the orchestral program material, because, on the contrary, the image of the jazz vocalist (III) seems to be almost touchable in WFS. This may be caused by the fact that her image is placed in front of the arrays: while approaching this image, its sound level increases in a realistic manner. The image of the saxophones (IV) in WFS is also unequaled in naturalness by the other systems.

Many additional comments were reported, but most of these can be ascribed to personal preference concerning ambience, depth and balance. Refining the mix would probably satisfy some of the respondents, but it would again dissatisfy others. Consensus was only reached on the opinions reported above.

Some of the other comments are mentioned here to illustrate the spread in personal appreciation or professional approach of the respondents. Both recording engineers criticized the strength of the spot microphone signals, of which the prominent use is held responsible for a 'double ambience', in some programs of the stereo and WFS system. One of them disapproved of the 'over-acoustical detachment' of WFS in the orchestral programs (I and II); the

* High quality array elements with higher directivity (Section 3.3) are proposed for reduction of coloration.

other admired the natural sound of the WFS system: ‘You can’t hear the loudspeakers any more.’ The latter also criticizes most of the 3/2 surround mixes, in which ‘the center channel is confusing the image too much’. Other respondents did not share this opinion.

Discussion

For the Mahler program (II), the stability of the WFS sources is appreciated better in the middle than at the sides of the listening area. This is caused by the conflicting use of spot and main microphone signals. The spot microphones signals are fed to a limited number of virtual sources, while the main microphones signals, that also contain the direct sound of the spotted sources, are fed to virtual loudspeakers. In the middle of the listening area, the evoked images coincide, but at the sides the virtual stereo image collapses, while the virtual sources persist. Though the use of the main microphones signals does not provide an optimal reproduction quality here, it permits a considerable reduction of source signals and transmission channels.

A much better performance is obtained with the saxophone quartet (IV). Because of the small number of sources, each source could be recorded separately. The crosstalk was kept within bounds by minimizing the distance of the spot microphones to the sources. The ambient waves from the front direction were provided by highly uncorrelated signals with a small direct/reverberation ratio. In this way, the possible conflicting images were avoided.

6.4 Conclusions

The localization experiments confirm that the localization of virtual sources behind the array is as good as the localization of real sources. If the loudspeaker spacing of the arrays is doubled to 22 cm, or if sources are focused in front of the arrays, a deterioration is observed, but the distortion is still small in an absolute sense. The localization blur for virtual sources is of the same order of magnitude as that for real sources.

From a subjective comparison of two-channel stereophony, 3/2 surround sound and wave field synthesis, it appears that the wave field synthesis system is able to create a spatial impression that is appreciated above that of the others. Also, the flexibility of placing sources and changing ambience in this system is considered very positive. The wave field synthesis system is more robust than the other systems, especially if crosstalk is minimized during recording.

While stereophonic recording can appeal to fifty years of experience, the development of recording for wave field synthesis has only just begun. Major improvements can be expected from microphone and signal processing techniques that are able to decompose sound fields into direct sound, reflections and reverberation. This will greatly benefit the flexibility and quality of sound reproduction.

Appendix A

Derivation of the 2½D Focusing Operator

In this appendix the 2½D focusing operator is derived analogous to the 2½D Rayleigh operator given by Vogel (1993). A primary source is situated at $(x, y, z) = (0, 0, z_0)$ between a line of secondary sources and a parallel line of receivers, see Figure A.1. If monopole characteristics for the primary source are assumed, the original wave field of the primary source is given by

$$P(x_R, \Delta z, \omega) = S(\omega)\frac{\exp(-jk\rho)}{\rho} \tag{A.1}$$

with

$$\rho = \sqrt{x_R^2 + (\Delta z_0 - z_0)^2}\,. \tag{A.2}$$

The pressure at R as synthesized by a distribution of monopoles along the x-axis is

$$P(x_R, \Delta z, \omega) = \int_{-\infty}^{\infty} Q_{\mathrm{m}}^{\mathrm{foc}}(x, \omega)\frac{\exp(-jk\Delta r)}{\Delta r}dx\,, \tag{A.3}$$

with driving function

$$Q_{\mathrm{m}}^{\mathrm{foc}}(x, \omega) = A(x, \omega)\exp(+jkr)\,, \tag{A.4}$$

where $A(x, \omega)$ is an amplitude factor yet to be derived. The phase factor $\exp(+jkr)$ describes the phase of a wave traveling in the direction opposite to $\boldsymbol{r}$. Integral (A.3), which is of the form

$$I = \int_{-\infty}^{\infty} f(x)\exp(j\phi(x))dx \tag{A.5}$$

can be evaluated with the stationary phase method, yielding

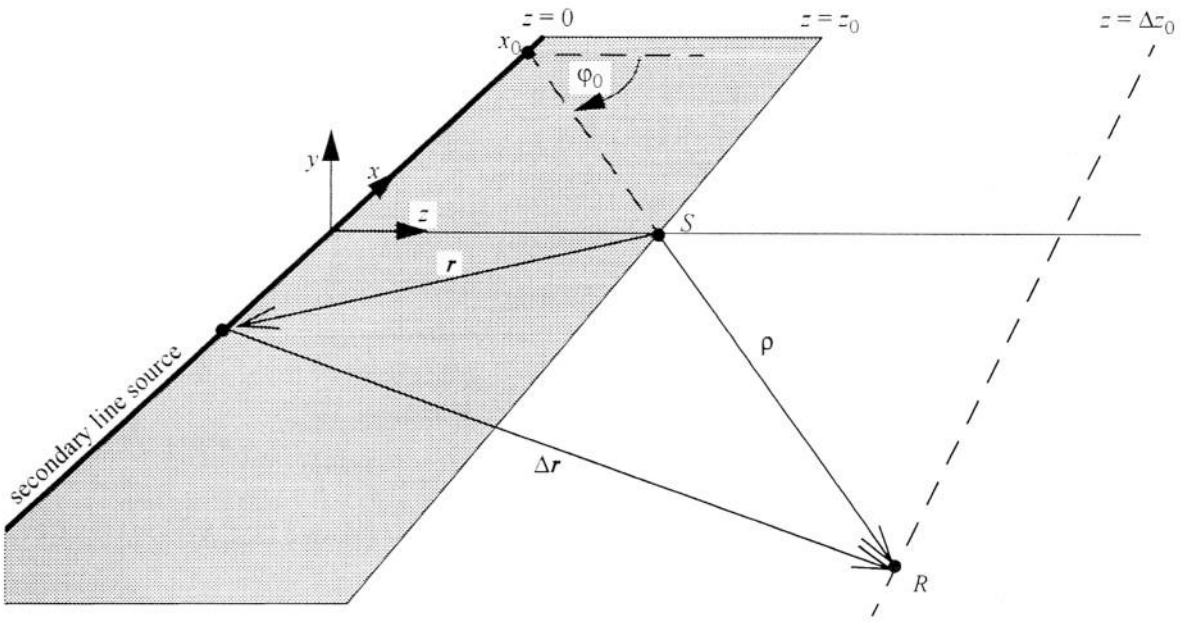


Figure A.1 The wave field of a primary source S is synthesized at $z = \Delta z_0$ by secondary sources at $z = 0$.

$$I \approx f(x_0)\exp(j\phi(x_0))\sqrt{\frac{2\pi j}{\phi''(x_0)}}\,. \tag{A.6}$$

The point x_0 is the stationary point of the phase $\phi(x)$. This point can be found by setting $\phi'(x) = 0$. The approximation is based on the fact that the main contribution to the integral comes from the value of x where the phase is stationary. It follows that

$$\phi(x) = k(r - \Delta r)\,, \tag{A.7}$$

$$\phi'(x) = k\left(\frac{x}{r} - \frac{x - x_R}{\Delta r}\right), \tag{A.8}$$

$$\phi''(x) = k\left(\frac{z_0^2}{r^3} - \frac{\Delta z_0^2}{\Delta r^3}\right), \tag{A.9}$$

and

$$f(x) = A(x, \omega)/\Delta r\,. \tag{A.10}$$

Equating the right-hand term of (A.8) to zero yields

$$x_0 = \frac{z_0}{z_0 - \Delta z_0} x_R\,. \tag{A.11}$$

This stationary phase point lies at the intersection of the x-axis and the produced vector $\Delta \boldsymbol{r}$ (see Figure A.1). Substitution of (A.11) in (A.7) and (A.9) gives

$$\phi(x_0) = -k\rho\,, \tag{A.12}$$

and

$$\phi''(x_0) = k\frac{z_0^2}{r_0^3}\left(\frac{\Delta z_0 - z_0}{\Delta z_0}\right), \tag{A.13}$$

with $r_0 = \sqrt{x_0^2 + z_0^2}$. Equating (A.1) and (A.3), and combining (A.6), (A.10) and (A.11) yields the driving function from (A.4):

$$Q_{\mathrm{m}}^{\mathrm{foc}}(x_0, \omega) = S(\omega)\sqrt{\frac{k}{2\pi j}}\sqrt{\frac{\Delta z_0}{\Delta z_0 - z_0}}\cos(\varphi_0)\frac{\exp(+jkr_0)}{\sqrt{r_0}}\,. \tag{A.14}$$

If the receiver is moved along the line $z = \Delta z_0$, each receiver position will define a unique stationary phase point for which (A.14) is the driving function. This equation may therefore be generalized to*

$$Q_{\mathrm{m}}^{\mathrm{foc}}(x, \omega) = S(\omega)\sqrt{\frac{k}{2\pi j}}\sqrt{\frac{\Delta z_0}{\Delta z_0 - z_0}}\cos(\varphi)\frac{\exp(+jkr)}{\sqrt{r}}\,. \tag{A.15}$$

Similarly, for a distribution of secondary dipole sources the synthesized wave field is given by

$$P(x_R, \Delta z, \omega) = \int_{-\infty}^{\infty} Q_{\mathrm{d}}^{\mathrm{foc}}(x, \omega) jk\cos(\varphi)\frac{\exp(-jk\Delta r)}{\Delta r}dx\,, \tag{A.16}$$

with focusing operator (driving function)

$$Q_{\mathrm{d}}^{\mathrm{foc}}(x, \omega) = -S(\omega)\sqrt{\frac{j}{2\pi k}}\sqrt{\frac{\Delta z_0}{\Delta z_0 - z_0}}\frac{\exp(+jkr)}{\sqrt{r}}\,. \tag{A.17}$$

The driving function is valid for receivers in the xz-plane with $\Delta z_0 > z_0$. It is interesting to investigate the synthesized wave field in the area between secondary sources and primary source. Suppose that Δz_0 in (A.15) is fixed at a certain value $\Delta z_0 > z_0$, which is the usual way to implement the operator in a wave field synthesis system. The synthesized wave field at a position $\hat{r}(\hat{x}, 0, \hat{z})$ with $0 < \hat{z} < z_0$ is found by substitution of (A.15) into integral (A.3)

$$P(\hat{x}, \hat{z}, \omega) = S(\omega)\sqrt{\frac{k}{2\pi j}}\sqrt{\frac{\Delta z_0}{\Delta z_0 - z_0}}\int_{-\infty}^{\infty}\frac{\cos(\varphi)}{\sqrt{r}}\frac{\exp(-jk(\hat{r} - r))}{\hat{r}}dx\,. \tag{A.18}$$

* The subscript 0 is maintained for constants of integration Δz_0 and z_0.

This integral can be evaluated with the stationary phase approximation. The stationary point is calculated as

$$x_0 = \frac{z_0}{z_0 - \hat{z}} \hat{x}, \tag{A.19}$$

giving

$$\phi(x_0) = k\rho \tag{A.20}$$

and

$$\phi''(x_0) = k \frac{z_0^2}{r_0^3} \left(\frac{\hat{z} - z_0}{\hat{z}} \right) \quad (<0), \tag{A.21}$$

so that (A.18) is reduced to

$$P(\hat{x}, \hat{z}, \omega) = -jS(\omega) \sqrt{\frac{\Delta z_0 (z_0 - \hat{z})}{\hat{z}(\Delta z_0 - z_0)}} \frac{\exp(+jk\rho)}{\rho}. \tag{A.22}$$

The amplitude weighting factor is a consequence of fixing the receiver line for the 2½D-operator. The positive sign in the phase function $+jk\rho$ indicates a concave wave front, propagating towards the point where $\rho = 0$, as expected. Note also the constant phase factor $-j$ which represents a frequency independent phase shift of -90°. This factor is lost after the waves have traveled through the focus.

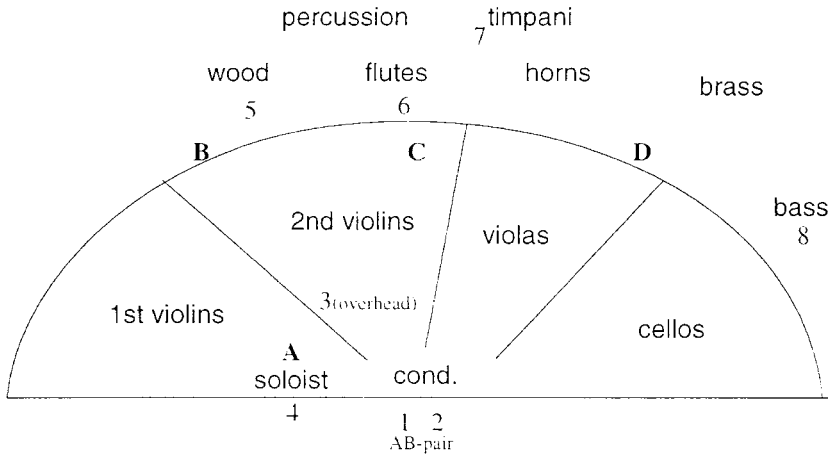
Appendix B

Recording and Mixing of Musical Performances for Subjective Evaluation

I. Violin concert in E Minor of Mendelssohn-Bartholdy

Performed by the Rotterdam Symphony Orchestra K.O.V. conducted by Benno Torrenga, with solo violin by Ilja Grubert. Recorded live in De Doelen (concert-hall), Rotterdam, May 23, 1996, by the Delft Laboratory of Seismics and Acoustics. Figure B.1 shows the recording situation and routing scheme used by mixing.

Since only a few spot microphones were used, the imaging possibilities for wave field synthesis were not fully utilized. The violin however, could be reproduced as a virtual source. Of this source, six virtual mirror sources were synthesized as well, using a map of the floor



Routing scheme

signal	stereo	3/2	WFS
1	L	L	−30°
2	R	R	+30°
3 »»	-	S_L; S_R	±120°
4	LR	LC	**A**
4 »»	L; R	L; R	±30°
5	L	L	**B**
6	LR	RC	**C**
7	LR	C	**C**
8	LR	RC	**D**

»» = via stereo artificial reverberation unit

Figure B.1 Recording map and mixing scheme for the Mendelssohn violin concert. The bold capitals indicate virtual source positions used for playback in the WFS system. Non-bold capitals are loudspeakers of two-channel stereophony and 3/2 surround sound; LC means panning between the left and center loudspeaker. The angles of incidence refer to the plane waves of the WFS system.

level of the hall. Virtual source position **D** was closest to that of the basses (no. 8), within the hardware limits of the WFS system.

Since robustness is a major criterion in the test, the mixes for the 3/2 system were optimized for stable source positions. This was achieved by panning the violin between the L and C in such a way that its image, perceived at the sweet spot, coincided with its position in the main pair. In this way, the direction of that source, as perceived at the sides of the listening area, deviated less than with stereophonic imaging (and panning between L and R).

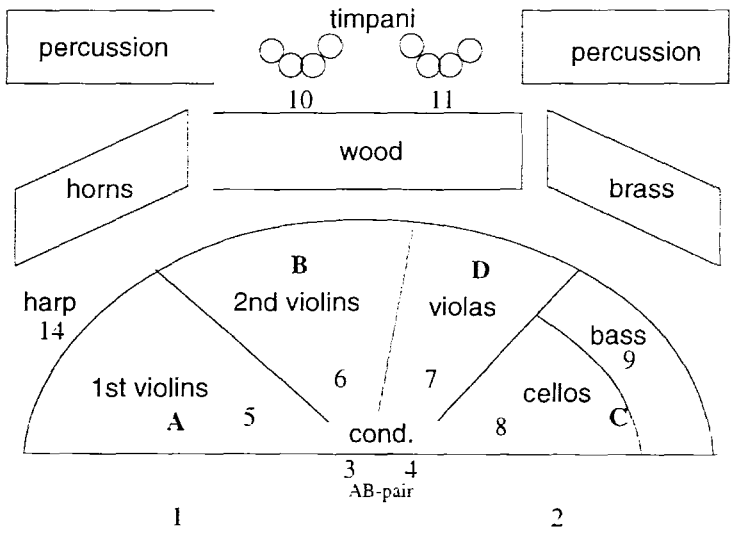
Extra delays were used for each system to compensate for the distance between the spot microphones and either the main microphones (for stereo and 3/2 surround, see Wöhr et al. 1991), or the virtual source positions (WFS). The reverberation program of the hall-effect unit was the same for each of the reproduction systems.

II. Second Symphony of Mahler

Performed by The Hague Royal Conservatory and the Amsterdam Sweelinck Conservatory, conducted by Yoel Levi. Recorded live in the Anton Philipszaal (concert-hall), The Hague, June 30, 1994, by Benno Torrenga and students of the Royal Conservatory in cooperation with the Delft Laboratory of Seismics and Acoustics. Figure B.2 shows a map.

The four processable virtual sources were placed at compromise positions between the eight available notional source signals. Because no soloists were active during the first movement of this symphony, also no virtual mirror sources were synthesized. The wide AB-pair (no. 1 and 2) served as a source of lateral reflective energy.

During recording, a panned sub-mix (no. 10 and 11) had been made from signals of six microphones near the wood-wind and percussion instruments, because the recording was meant originally for stereophonic reproduction.



Routing scheme

signal	stereo	3/2	WFS
1	L	L	−30°
2	R	R	+30°
1 »	-	S_L	−120°
2 »	-	S_R	+120°
3	L	L	-
4	R	R	-
5	L	L	**A**
6	LR	LC	**B**
7	LR	RC	**C**
8	R	R	**D**
9	R	R	**D**
10	LR	LC	**B**
11	LR	RC	**C**
14	L	L	**A**

» = via mono artificial reverberation unit

Figure B.2 Recording map and mixing scheme for the Mahler II concert.

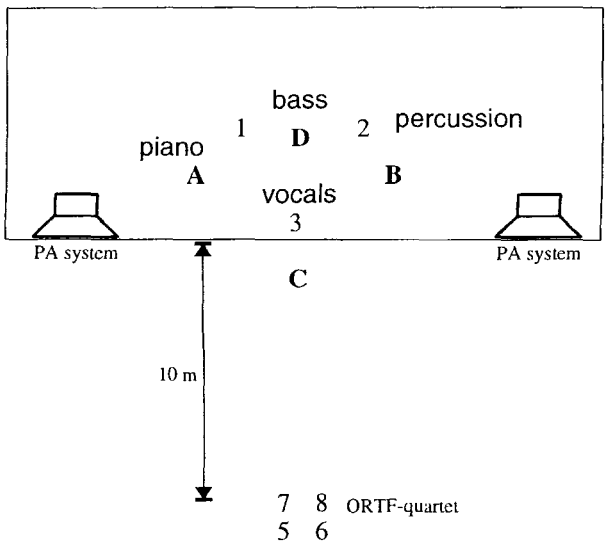
Extra delays were used for each system to compensate for the distance between the spot microphones and either the main microphones (stereo and 3/2), or the virtual source positions (WFS). The reverberation program of the hall-effect unit was the same for 3/2 surround and WFS.

III. Conception of George Shearing

Performed by students of The Hague Royal Conservatory with Sanne van Vliet on vocals. Recorded live in the Kees van Baaren Zaal (theater) in The Hague, February 28, 1997, by Peter Nuyten and students of the conservatory. Figure B.3 shows a drawing of the situation.

A panned sub-mix of several 'close mikes' near the instruments had been made already during recording, because this recording was originally meant for 3/2 surround sound reproduction. This, however, frustrated the imaging with virtual sources for the WFS system. In order to stabilize the image of the double-bass during the improvisation solo, signals 1 and 2 were fed temporarily to virtual source **D** as well. The image of the vocalist was focused at 1.1 m in front of the array. Virtual mirror sources were omitted, because the Public Address System already produced a lot of pseudo-reflective energy by leaking its reinforced sound into signals 7 and 8.

This recording was added to the listening test, primarily because of the live atmosphere and applause after the actual performance. This ambience was recorded using the ORTF-quartet of Theile (1996). Though these microphones are angled ±45° and ±135°, their signals were reproduced by WFS as plane waves from ±30° and ±120°, which gave a smoother envelopment. Note that the 3/2 standard employs similar angles: ±30° and ±110°. Extra delays were used for each system to compensate for the distance between the spot microphones and the ORTF-quartet.



Routing scheme

signal	stereo	3/2	WFS
1	L	L	**A,D**
2	R	R	**B,D**
3	LR	C	**C**
5	L	S_L	−120°
6	LR	S_R	+120°
7	LR	L	−30°
8	LR	R	+30°

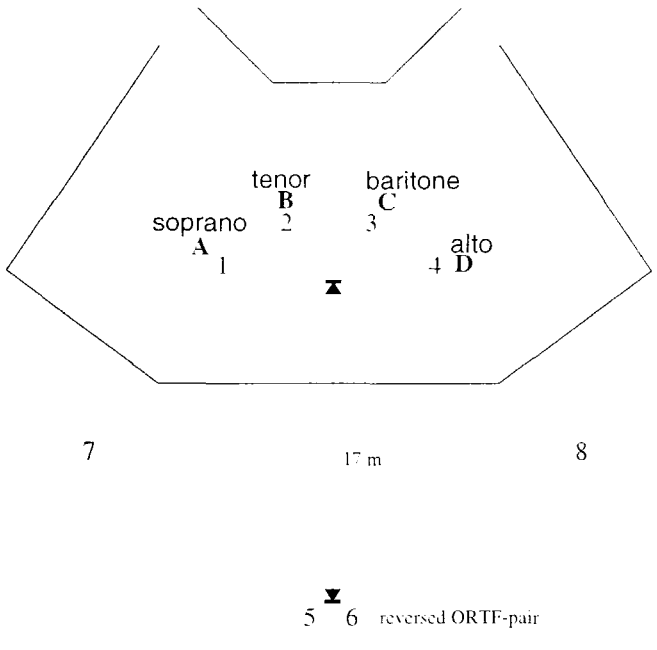
Figure B.3 Recording map and mixing scheme for the jazz concert.

IV. Overture of the Concert after Vivaldi; arranged for saxophones

Performed by the Dutch Saxophone Quartet. Recorded without audience in the St. Martinuskerk (church), Amersfoort-Hoogland, May 10, 1997, by the Delft Laboratory of Seismics and Acoustics in cooperation with Menno van der Veen, Hans Goddijn and Fritz de With. Figure B.4 shows a map of the choir of the church.

The spot microphones were placed as close as possible to the instruments, to minimize crosstalk and to maximize the direct/reverberation ratio. Besides the microphones shown on the map, five more microphones were installed. Among those were a narrow AB-pair and a SoundField microphone. The reversed ORTF-pair actually formed part of an ORTF-quartet, so that the angle between its microphones was 90° instead of 110°. After extensive experiments during mixing, it was decided to use only the shown microphones, which accomplished the best results for all reproduction systems. For example, the stereo image created by the narrow AB-pair (which stood near the center of the arc of instruments) did not mix up smoothly with the spot microphone signals. Also, the front pair of the ORTF-quartet had too low a direct/reverberation ratio for use with spot microphones only. The wide AB-pair that was employed at last, picked up enough reflections from the sides to produce a temporally and spatially complete image in cooperation with the other microphone signals.

Extra delays were used for each system to compensate for the distance between the spot microphones and the ambient microphones.



Routing scheme

signal	stereo	3/2	WFS
1	LR	LC	**A**
2	LR	LC	**B**
3	LR	RC	**C**
4	LR	RC	**D**
5	L	S_L	−120°
6	R	S_R	+120°
7	L	L	−30°
8	R	R	+30°

Figure B.4 Recording map and mixing scheme for the saxophone quartet.

Author Index

Subject Index

References

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

Bamford, J.S. and Vanderkooy, J. (1995), 'Ambisonic sound for us,' preprint 4138 presented at the 99th Convention ofthe Audio Engineering Society in New Vork.

Bareham, J.R. (1996), 'Measurement of Spatial Characteristics of Sound Reproduced in Listening Spaces,'preprint presented at the 101st Convention of the Audio Engineering Society in Los Angelos.

Barron, M. (1971), 'The subjective effects of first reflections in concert-halls - The need for lateral reflections,' *J. Sound and Vib.* **15** (4) p. 475.

Bartlett, B. (1991), *Stereo Microphone Techniques*, Focal Press, Boston, London.

Bennett, J.C., Barker, K. and Edeko, F.O. (1985), 'A new approach to the assessment of stereophonic sound system performance,' *J. Audio Eng. Soc.* **33** (5) p. 314. Reprinted in *An anthology of reprinted articles on stereophonic techniques*, Audio Engineering Society, New Vork (1986).

Beranek, L.L. (1954), *Acoustics*, McGraw-Hill, New York. (Reprinted by ASA, 1986.)

Beranek, L.L. (1996), *Concert and Opera Halls: How They Sound*, Acoustical Society of America, Woodbury, NY.

Berkhout, A.J., de Vries, D. and Boone, M.M. (1980), 'A new method to acquire impulse responses in concert halls,' *J. Acoust. Soc. Am.* **68** (1) p. 179.

Berkhout, A.J. (1982), *Seismic Migration.* Imaging of acoustic energy by wave field extrapolation; A. Theoretical Aspects, Elsevier, Amsterdam.

Berkhout, A.J. (1987), *Applied Seismic Wave Theory*, Elsevier, Amsterdam.

Berkhout, A.J. (1988), 'A holographic approach to acoustic control,' *J. Audio Eng. Soc.* **36** (12) p.977.

Berkhout, A.J., de Vries, D. and Vogel, P. (1993), 'Acoustic control by wave field synthesis,' *J. Acoust. Soc. Am.* **93** (5) p. 2764.

Berkhout, A.J., de Vries, D. and Sonke, J.J. (1997a), 'Array technology for wave field analysis of sound fields in enclosures,' accepted for publication in *J. Acoust. Soc. Am.*

Berkhout, A.J. (1997b), 'Pushing the limits of seismic imaging, Part I: Prestack migration in tenns of double dynamic focusing,' *Geophysics* **62** (3) p. 937.

Blake, L. (1995), 'Digital sound in the cinema,' *Mix*, October issue.

Blauert, J. (1983), *Spatial Hearing*, MIT Press, Cambridge, Mass.

Bleistein, N. (1984), *Mathematical methods for wave phenomena*, Academic Press, New York.

Blumlein, A.D. (1931), British Patent Specification 394,325. 'Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems'. Reprinted in *An anthology of reprinted articles on stereophonic techniques*, Audio Engineering Society, New York (1986).

Boer, K. de (1940), *Stereofonische geluidsweergave*, Thesis, Delft University of Technology.

Boone, M.M., Verheijen, E.N.G. and van Tol, P.F. (1995), 'Spatial Sound-Field Reproduction by Wave Field Synthesis,' *J. Audio Eng Soc*. **43** (12) p. 1003.

Boone, M.M., and Verheijen, E.N.G. (1997), 'Qualification of Sound Generated by Wave Field Synthesis For Audio Reproduction,' preprint 4457 presented at the 102nd Convention of the Audio Engineering Society in Munich.

Bracewell, J.L. (1993), *Sound Design in the Theatre*, Prentice Hall, Englewood Cliffs, New Jersey.

Bradley, J.S. and Soulodre, G.A. (1995), 'The influence of late arriving energy on spatial

impression,' *J Acoust. Soc. Am.* **97** (4) p. 2263.

Burkhard, M.D. and Sachs, R.M. (1975), 'Anthropometric Manikin for Acoustic Research,' *J. Acoust. Soc. Am.* **58** (1) p. 214.

Butterworth, W.E. (1977), Hi-fi - *From Edison's Phonograph to Quadrophonic Sound*, Four Winds Press, New York.

Camras, M. (1968), 'Approach to Recreating a Sound Field,' *J. Acoust. Soc. Am.* **43** (6) p. 1425.

Camras, M. (1985), 'Origins of magnetic recording concepts,' *J. Acoust. Soc Am.* **77** (4) p. 1314.

Ceoen, C. (1972), 'Comparative stereophonic listening tests,' *J. Audio Eng. Soc.* **20** (1) p. 19.

Clark, H.A.M., Dutton, G.F. and Vanderlyn, P.B. (1958), 'The "stereosonic" recording and reproducing system,' *J. Audio Eng. Soc.* **6** (2) p. 102. Reprinted in *An anthology of reprinted articles on stereophonic techniques*, Audio Engineering Society, New York (1986).

Cooper, D.H. and Shiga, T. (1972), 'Discrete-matrix multichannel stereo,' *J. Audio Eng. Soc.* **20** (5) p. 346.

Dalenbäck, B.-I., Kleiner, M. and Svensson, P. (1994), 'A Macroscopic View of Diffuse Reflection,' *J. Audio Eng. Soc.* **42** (10) p. 793.

Damaske, P. and Ando, Y. (1972), 'lnteraural Crosscorrelation for Multichannel Loudspeaker Reproduction,' *Acustica* **27**, p. 232.

Dressler, R. (1993), *Dolby Pro Logic surround decoder principles of operation*, Dolby Laboratories Licensing Corporation.

Eargle, J.M. (1971), 'Multichannel stereo matrix systems: an overview,' *J. Audio Eng. Soc.* **19** (7) p. 552.

Geluk, K.J. (1994), 'A Monitoring System Based on Wave-Synthesis,' *J. Audio Eng. Soc.* **42** (5) p. 389.

Gerzon, M.A. (1977), 'Criteria for evaluating surround-sound systems,' *J. Audio Eng. Soc.* **25** (6) p. 400.

Gerzon, M.A. (1985), 'Ambisonics in multichannel broadcasting and video,' *J. Audio Eng.*

Soc. **33** (11) p. 859.

Griesinger, D. (1996), 'Multichannel matrix surround decoders for two-eared listeners,' *Bericht 19. Tonmeistertagung*, Karlsruhe.

Hartmann, W.M. (1983), 'Localization of sound in rooms,' *J. Acoust Soc. Am.* **74** (5) p. 1380.

Hunt, F.V. (1982), *Electroacoustics*, Acoustical Society of America. (Originally published by J. Wiley, 1954.)

ITU-R (1991), 'Multichannel stereophonic sound system with and without accompanying picture,' recommendation BS 775-1, Geneva.

Jansen, G. (1997), *Focused wavefields and moving virtual sources by wave field synthesis*, Master's thesis, Delft University of Technology.

Kate, W.R.Th. ten (1993), *An overview of analog multi-channel sound systems*, Nat. Lab. Unclassified Report 017/93, Philips Electronics N.V.

Komiyama, S., Morita, A., Kurozumi, K. and Nakabayashi, K. (1992), 'Distance control of sound images by a two-dimensional loudspeaker array,' *J. Acoust. Soc. Jpn.* **13** (3) p. 171.

Lipshitz, S.P. (1986), 'Stereo microphone techniques ... Are the purists wrong?', *J. Audio Eng. Soc.* **34** (9) p. 716.

Maat, I. (1997), *Reduction of perceptual redundancy in reverberation synthesis*, Master's thesis, Delft University of Technology.

Mac Cabe, C.J. and Furlong, D.J. (1994), 'Virtual Imaging Capabilities of Surround Sound Systems,' *J. Audio Eng. Soc.* **42** (1/2) p. 38.

Macpherson, E.A. (1991), 'A Computer Model of Binaural Localization for Stereo Imaging Measurement,' *J. Audio Eng. Soc.* **39** (9) p. 604.

Massa, F. (1985), 'Some personal recollections of early experiences on the new frontier of electroacoustics during the late 1920s and early 1930s,' *J. Acoust. Soc. Am.* **77** (4) p. 1296.

Pierce, A.D. (1991), *Acoustics*, Acoustical Society of America, Woodbury, NY.

Potter, J.M. (1993), *On the binaural modelling of spaciousness in room acoustics*, Thesis, Delft University of Technology.

Raatgever, J. and Bilsen, F.A. (1986), 'A central spectrum theory of binaural processing. Evidence from dichotic pitch,' *J. Acoust. Soc. Am.* **80** (2) p. 429.

Reijnen, A.J., Sonke, J.-J. and de Vries, D. (1995), 'New Developments in Electro-Acoustic Reverberation Technology,' preprint 3978 presented at the 98th Convention ofthe Audio Engineering Society in Paris.

Rife, D.D. and Vanderkooy, J. (1989), 'Transfer-function measurement with maximum-Iength sequences,' *J. Audio Eng. Soc.* **37** (6) p. 419.

Schneider, S. (1996), 'Practical experiences with multichannel postproduction,' preprint presented in the workshop 'Multichannel sound production techniques and technologies,' at the 100th Convention of the Audio Engineering Society in Copenhagen.

Scholtze, O. (1997), senior recording engineer (*Tonmeister*) at Philips Music Group, private communication.

Skudrzyk, E. (1954), *Die Grundlagen der Akustik*, Springer-Verlag, Vienna.

Snow, W.B. (1953), 'Basic Principles of Stereophonic Sound,' *J. SMPTE* **61**, p. 567. Reprinted in *An anthology of reprinted articles on stereophonic techniques*, Audio Engineering Society, New York (1986).

Spikofski, G., Pehrs, J., Rath, H. and Theile, G. (1992), 'Zur Darstellung der akustischen Atmosphäre mit dem 3/2- bzw. 3/4-Stereoformat,' *Bericht 17. Tonmeistertagung*, Karlsruhe.

Start, E.W. (1996), 'Application of curved arrays in wave field synthesis,' preprint 4143 presented at the 100th Convention ofthe Audio Engineering Society in Copenhagen.

Start, E.W. (1997), *Direct sound enhancement by wave field synthesis*, Thesis, Delft University of Technology.

Start, E.W., Roovers, M.S. and Vries, D. de (1997), 'In Situ Measurements on a Wave Field Synthesis System For Sound Enhancements,' preprint 4454 presented at the 102nd Convention of the Audio Engineering Society in Munich.

Steinberg, J.C. and Snow, W.B. (1934), 'Auditory perspective - physical factors,' *Electrical Engineering* **53** (1) p. 12. Reprinted in *An anthology of reprinted articles on stereophonic techniques*, Audio Engineering Society, New York (1986).

Stern, R.M., and Colbum, H.S. (1978), 'Theory of binaural interaction based on auditory-

nerve data. IV. A model for subjective lateral position,' *J. Acoust. Soc. Am.* **64** (1) p. 127.

Streicher, R. and Dooley, W. (1985), 'Basic stereo microphone perspectives - A review,' *J. Audio Eng. Soc.* **33** (7/8) p. 548.

Theile, G. (1996), 'Möglichkeiten und Grenzen der 3/2-Stereo-Aufnahme,' *Bericht 19. Tonmeistertagung*, Karlsruhe.

Thorbecke, J.W. (1997), *Common focus point technology*, Thesis, Delft University of Technology.

Tohyama, M., and Suzuki, A. (1989), 'Interaural cross-correlation coefficients in stereo-reproduced sound fields,' *J. Acoust. Soc. Am.* **85** (2) p. 780.

Vogel, P. (1993), *Application of wave field synthesis in room acoustics*, Thesis, Delft University of Technology.

Vries, D. de, Start, E.W. and Valstar, V.G. (1994), 'The wave field synthesis concept applied to sound reinforcement: restrictions and solutions,' preprint 3812 presented at the 96th Convention ofthe Audio Engineering Society in Amsterdam.

Vries, D. de (1996), 'Sound reinforcement by wavefield synthesis: adaptation of the synthesis operator to the loudspeaker directivity characteristics,' *J. Audio Eng. Soc.* **44** (12) p. 1120.

Watkinson, J. (1989), *The Art of Digital Audio*, Focal Press, London.

Wightman, F.L. and Kistler, J. (1992), 'The dominant role of low-frequency interaural time differences in sound localization,' *J. Acoust. Soc. Am.* **91** (3) p. 1648.

Willems, S. (1993), 'Philips DSS 930: A New Concept in Active Loudspeakers,' preprint 3532 presented at the 94th Convention of the Audio Engineering Society in Berlin.

Wöhr, M., Theile, G., Goeres, H.-J. and Persterer, A. (1991), 'Room-Related Balancing Technique: A Method for Optimizing Recording Quality,'' *J. Audio Eng. Soc.* **39** (9) p. 623.

Zieglmeier, W., Theile, G. (1996), 'Darstellung seitlicher SchalIquellen bei Anwendung des 3/2-Stereo-Formats,' *Bericht 19. Tonmeistertagung*, Karlsruhe.

Summary

Sound Reproduction by Wave Field Synthesis

Sound reproduction comprises the recording, storage and playback of sound (esp. music). Since the 1950s, sound reproduction is mainly based on stereophonic techniques. In two-channel stereophony, the sound is recorded with a pair of *main microphones* placed in front of the source area (e.g. a symphony orchestra). Separate sound sources or groups of sound sources are often recorded by *spot microphones* as well. These signals are mixed with those of the main pair, and are reproduced by two loudspeakers. A listener seated midway in front of these loudspeakers, perceives a sound image similar to that of the original sources concerning color, depth and balance. A listener seated off-center experiences a spatially distorted sound image, that inclines to the nearest loudspeaker. Thus, the listening area for stereophonic reproduction is restricted to a few seats midway between both loudspeakers.

In this thesis a new method of sound reproduction is described, based on the theory of *wave field synthesis*. Through the use of *arrays* of loudspeakers, driven by processed copies of the source signals, it is possible to synthesize wave fronts that are almost identical to those of the original sound sources. Hence, this method of reproduction offers a volume-solution, instead of the point-solution of conventional reproduction methods.

The processing of the original source signals is done by the *synthesis operator*, which is derived from the Kirchhoff-Helmholtz wave field representation. This theory describes the distribution of the pressure and particle velocity on a closed surface as a consequence of (primary) sources outside the enclosed volume. If this surface (of arbitrary shape) is filled with secondary dipole and monopole sources, theoretically a perfect copy of the primary wave field can be obtained inside the enclosure. For the special geometry of a planar distribution of secondary sources, Rayleigh's integrals apply. Then, with the primary sources behind that plane,

either monopoles or dipoles suffice to generate a copy of the primary wave field in front of that plane. In practice, planar arrays of loudspeakers (secondary sources) can be used to synthesize *virtual sources* (primary sources). Another geometrical simplification is allowed if the virtual sources as well as the listeners are present in a horizontal plane. In that case, a horizontal line array of loudspeakers suffices. By means of focusing techniques it is also possible to create sound sources in front of the loudspeakers: the listeners will then perceive a source at the acoustic focus.

Because of the finite distance between adjacent loudspeakers of the array, there will be spectral deviations above a certain frequency in the reproduced wave field: effects of *spatial aliasing*. From theory it arises that the synthesized sound field approaches the original field more closely, if the loudspeakers radiate the sound in a more directional way. The directional behavior of the loudspeakers can be controlled by modeling the diaphragm that radiates the sound. For that reason, best results are to be expected for electrostatic loudspeakers. A prototype array, built according to these discernments, confirms that is possible to synthesize a wave field that is also correct for higher frequencies. However, the efficiency of this prototype array is too low for practical applications in wave field synthesis, for which reason an electrodynamic array has been developed as well.

A 128-channel reproduction system has been built, which includes these electrodynamic arrays surrounding a 24 m² listening area. This system composes the wave field per source as follows: its direct sound (recorded with a spot microphone) is reproduced by a virtual source, its early reflections (computed by a mirror source program) are also synthesized by virtual sources, and its reverberant field (picked up by ambient microphones, or generated by artificial reverberation units) is composed by means of plane waves entering from different directions. Also, fast moving sources (recorded by close-miking) can be synthesized, which is considered to be an important requirement for application in cinemas, simulators and virtual reality theaters.

This reproduction system is evaluated in two ways: by measurements (objectively) and by listening tests (subjectively). In a comparative investigation between wave field synthesis, stereophony and 3/2 *surround sound*, it is shown that the wave field synthesis system offers the largest listening area with spatially correct reproduction of sound. Its sound image is more robust than that of the other systems. Also, this system is able to evoke a very spatial impression, in which the loudspeakers cannot be distinguished any more. Furthermore, the flexibility of the system regarding the manipulation of source positions and ambience is an important innovation.

Edwin Verheijen, Delft University of Technology, 1997

Zusammenfassung

Schallreproduktion durch Wellenfeldsynthese

Die Schallreproduktion umfaßt Aufnahme, Speicherung und Wiedergabe von Schall (insb. Musik). Seit den fünfziger Jahren erfolgt die Schallwiedergabe hauptsächlich mittels stereophoner Techniken. Mit der Zweikanal-Stereophonie wird der Schall von zwei Hauptmikrophonen (*main microphones*) aufgenommen, die sich vor der Schallquelle (z.B. einem Sinfonieorchester) befinden. Einzelne Schallquellen oder Gruppen von Schallquellen werden oft noch zusätzlich von Stützmikrophonen (*spot microphones*) aufgezeichnet. Diese Signale werden dann mit denen der Hauptmikrophone gemischt und über zwei Lautsprecher wiedergegeben. Ein Hörer, der sich so vor den Lautsprechern befindet, daß er zu beiden Lautsprechern den gleichen Abstand hat, nimmt ein Klangbild wahr, das dem Original sehr nahe kommt. Ein Hörer aber, der sich nicht genau in der Mitte befindet, erfährt ein räumlich verzerrtes Klangbild. Deshalb ist eine optimale stereophone Schallwiedergabe nur in dem Raumbereich gewährleistet, in dem beide Lautsprecher den gleichen Abstand zum Hörer haben.

In dieser Dissertation wird eine neue Wiedergabemethode beschrieben, die auf dem Prinzip der Wellenfeldsynthese (*wave field synthesis*) basiert. Hierbei wird das Quellsignal (*source signal*) nach einer geeigneten Verarbeitung mittels einer Reihe von Lautsprechern (*arrays*) wiedergegeben, so daß sich eine Wellenfront ausbildet, die mit der des Originals annähernd identisch ist. Damit kann im Gegensatz zu den konventionellen Methoden im gesamten Raum eine gleichwertige Klangreproduktion erreicht werden.

Die Signalverarbeitung erfolgt mittels eines Syntheseoperators, der sich aus der Wellenfelddarstellung von Kirchhoff-Helmholtz herleiten läßt. Diese Theorie beschreibt die Verteilung des Druckes und der Teilchenschnelle an einer geschlossenen (dreidimensionalen) Fläche als Folge einer primären Schallquelle außerhalb des eingeschlossenen Volumens. Bringt man auf diese Fläche nun genügend sekundäre Dipol- und Monopolquellen, so erhält

man im Inneren der Fläche theoretisch eine vollkommene Reproduktion des ursprünglichen (primären) Wellenfeldes. In dem Spezialfall einer Verteilung der sekundären Quellen auf einer ebenen Fläche kann die Rayleighsche Theorie angewendet werden. Aus dieser folgt für den Fall, daß sich die primären Quellen hinter dieser Ebene befinden, daß sowohl die Monopole als auch die Dipole ein Wellenfeld vor der Fläche erzeugen können, welches eine genaue Nachbildung des ursprünglichen Feldes ist. In der Praxis können ebene Lautsprecherflächen (Sekundärquellen) errichtet werden, die virtuellen Quellen (Primärquellen) erzeugen. Befinden sich die Hörer und die virtuellen Quellen zusätzlich in einer horizontale Ebene, so genügt eine horizontale lineare Reihenanordnung von Lautsprechern. Zugleich ist es mittels geeigneter Fokussierung auch möglich, virtuelle Quellen vor den Lautsprechern zu erzeugen. Die Hörer nehmen dann einen Eindruck einer Quelle im akustischen Fokus vor den Lautsprechern wahr.

Da die Lautsprecher in der dargestellten Anordnung feste Abstände zueinander haben, enstehen oberhalb einer bestimmten Frequenz spektrale Abweichungen im reproduzierten Wellenfeld, sogenannte *spatial aliasing*-Effekte. Die Theorie zeigt, daß das erzeugte Wellenfeld dem Original besser angenähert werden kann, wenn die Lautsprecher eine stärkere Richtwirkung (*directivity*) besitzen. Die Richtwirkung kann durch Form und Gestalt der Membranen festgelegt werden. Dies gelingt am besten mit elektrostatischen Lautsprechern. Mit einer diesen Bedingungen entsprechenden Probereihe hat sich gezeigt, daß auch Wellenfelder erzeugt werden können, die in den höheren Frequenzen stimmen. Da der Wirkungsgrad der elektrostatischen Lautsprecher für praktische Anwendungen der Wellenfeldsynthese noch zu klein ist, wurde zusätzlich eine elektrodynamische Lautsprecherreihe entwickelt.

Mit dem errichteten Wiedergabesystem aus 128 Kanälen wurden diese elektrodynamischen Lautsprecher so um eine Fläche von 24 m² herum angeordnet, daß der Innenraum beschallt werden konnte. Die Erzeugung des Wellenfeldes einer Schallquelle wird folgendermaßen erreicht: der direkte Schall (augezeichnet mit einem Stützmikrophon) wird durch eine virtuelle Quelle erzeugt, die frühen Reflexionen (berechnet durch ein Spiegelquellenmodell) werden ebenfalls durch virtuelle Quellen reproduziert und der Nachhall (aufgezeichnet durch ambiente Mikrophone bzw. künstlich erzeugt) wird durch ebene Wellen aus verschiedenen Richtungen wiedergegeben. Damit können auch schnell bewegte Quellen (aufgenommen durch mitbewegte Mikrophone) für eine Wiedergabe erzeugt werden. Dies hat Bedeutung bei Anwendungen der Schallwiedergabe in Kinos, *Virtual Reality* Theatern oder bei Simulatoren.

Das Wiedergabesystem ist auf zweierlei Weise getestet worden: mittels Messungen (objektiv) und durch Hörversuche (subjektiv). In einer vergleichenden Untersuchung zwischen Zweikanal-Stereophonie, 3/2-Stereophonie und Wellenfeldsynthese hat sich gezeigt, daß die Wellenfeldsynthese in dem größten räumlichen Hörbereich eine korrekte Schallwiedergabe bietet. Weiterhin ist das Klangbild stabiler als bei anderen Systemen, insbesondere wenn die Aufnahmen bereits auf die Anwendung der Wellenfeldsynthese abgestimmt werden. Ferner ist das System imstande, einen derartigen räumlichen Klang zu erzeugen, daß der Hörer nicht mehr zwischen den verschiedenen Lautsprechern unterscheiden kann. Schließlich eröffnet die Wellenfeldsynthese bezüglich der Flexibilität in der Beeinflussung der Position und des Raumeindrucks von Schallquellen neue Möglichkeiten.

(Übersetzung: Bookie Priemer, Wim Speetjens)

Edwin Verheijen, Technische Universität Delft, 1997

Sommaire

Reproduction sonore par la synthèse du champ d'ondes

La reproduction sonore comprend l'enregistrement, le stockage et la restitution du son (surtout la musique). Elle est, depuis les années 1950, essentiellement basée sur l'emploi de techniques stéréophoniques. En stéréophonie conventionnelle l'enregistrement s'effectue en plaçant face à la scène sonore (un orchestre symphonique par ex.) un couple de microphones (*main microphones*). En complément, on a souvent recours à l'utilisation de microphones d'appoint (*spot microphones*) pour enregistrer une source sonore isolée ou un groupe de sources. Les signaux issus de ces microphones sont ensuite mixés avec ceux fournis par le couple puis transmis à deux haut-parleurs. L'auditeur placé dans l'axe des deux haut-parleurs perçoit alors en termes de couleur, de profondeur et d'équilibre, une image sonore identique à l'originale, tandis que l'auditeur excentré observera une distortion spatiale de l'image sonore tendant vers le haut-parleur le plus proche. Ainsi, la zone d'écoute pour une reproduction stéréophonique se trouve restreinte à quelques places situées à égale distance des deux haut-parleurs.

L'objet de cette thèse est de décrire une nouvelle méthode de reproduction sonore basée sur la théorie de la synthèse du champ d'ondes (*wave field synthesis*). Par le biais de réseaux de haut-parleurs (*arrays*), pilotés par les répliques traitées des signaux provenants des sources, il est possible de synthétiser des fronts d'ondes presque identiques à ceux qui proviennent des sources sonores originales. C'est donc une solution volumique qu'offre cette méthode de reproduction là où les autres méthodes fournissent une solution ponctuelle.

Les signaux des sources originales sont traités par le biais d'un opérateur de synthèse, tiré de la représentation du champ d'ondes de Kirchhoff-Helmholtz. Cette théorie décrit les distributions de pressions et de vitesses particulaires sur une surface fermée, induites par des sources (primaires) situées en dehors du volume clos délimité par la surface. Théoriquement, si

cette surface (de forme arbitraire) est entièrement constituée de sources secondaires dipolaires et monopolaires, une parfaite reproduction du champ primaire pourrait être obtenue en tout point du volume intérieur. Dans le cas particulier où la surface considérée est un plan, séparant l'espace en un demi-espace source et un demi-espace d'écoute, la distribution surfacique de sources secondaires peut être décrite par des intégrales de Rayleigh. On montre alors qu'une distribution de sources soit monopolaire, soit dipolaire, suffit pour générer dans le demi-espace d'écoute une réplique exacte du champ d'ondes primaire. Dans la pratique, des réseaux plans de haut-parleurs (sources secondaires) peuvent être utilisés pour la synthèse de sources virtuelles (sources primaires). Une autre simplification géométrique est permise si l'on considère que les sources virtuelles et les auditeurs se trouvent dans un même plan horizontal. Dans ce cas, il suffira d'utiliser un réseau linéaire (antenne) de haut-parleurs. L'emploi de techniques de focalisation rend également possible la génération de sources sonores en face des haut-parleurs, de sorte que l'auditeur perçoive une source au point de focalisation.

Du fait de la distance finie qui sépare les haut-parleurs adjacents d'une antenne, des distorsions spectrales du champ reproduit apparaîssent au-delà d'une certaine fréquence: c'est l'effet de repliement spacial (*spatial aliasing*). En théorie, on sait que le champ sonore synthétisé sera plus proche du champ original si l'on emploi des haut-parleurs plus directifs. Le comportement en directivité des hauts-parleurs peut être contrôlé en adaptant la forme du diaphragme. C'est pourquoi l'on peut s'attendre à de meilleurs résultats avec des haut-parleurs électrostatiques. Un prototype d'antenne, construit selon ces principes, confirme qu'il est possible de synthétiser un champ d'ondes dont la validité s'étend plus haut dans le spectre. Cependant, l'efficacité d'un tel prototype restant trop faible pour les applications pratiques en synthèse des champs d'ondes, c'est une antenne électrodynamique qui a donc été développée.

Un système de reproduction sonore de 128-canaux fût construit, qui inclus les antennes électrodynamiques délimitant une zone d'écoute de 24 m². Ce système décompose le champ d'onde de chaque source comme suit : le champ direct (pris par un micro d'appoint) reproduit par une source virtuelle, les premières réflections (calculées selon le principe des sources-images) reproduites également par des sources virtuelles, et le champ réverbéré (pris par des microphones d'ambiance ou généré par des unités de réverbération artificielle) recréé par le biais d'ondes planes arrivant de différentes directions. De plus, des sources mobiles rapides (enregistrees en champ proche) peuvent aussi être synthétisées, ce qui répond à d'importants besoins dans les applications telles que le cinéma, les simulateurs, ou la réalité virtuelle.

Ce système de reproduction est évalué de deux façons: Par la mesure (objective) et par les tests d'écoute (subjective). Une étude comparative entre ce système, le système stéréo et le système 3/2 *surround sound*, a montré que la synthèse des champs d'ondes offrait la plus large zone d'écoute tout en conservant une reproduction spatiale correcte du son. L'image sonore obtenue est plus stable que pour les autres systèmes. Ce système est aussi capable de susciter une réelle impression spatiale dans laquelle il n'est plus possible de distinguer les haut-parleurs. Enfin, la flexibilité du système concernant la manipulation des sources et des ambiances constitue une importante innovation.

(traduction: Thierry Piccolo, Bernard Lefranc)

Edwin Verheijen, Université de Technologie de Delft, 1997

Samenvatting

Geluidsreproductie op basis van golfveldsynthese

Geluidsreproductie omvat het opnemen, opslaan en het weergeven van geluid (inz. muziek). Sinds de jaren vijftig berust geluidsreproductie hoofdzakelijk op stereofonische technieken. Hierbij wordt het geluid opgenomen met twee microfoons, het zgn. hoofdsysteem (*main microphones*), die vóór het gebied met geluidsbronnen (bijv. een symfonieorkest) zijn geplaatst. Vaak worden aparte geluidsbronnen of groepen daarvan tevens geregistreerd met steunmicrofoons (*spot microphones*). Deze signalen worden gemengd met die van het hoofdsysteem, en afgespeeld over twee luidsprekers. Een luisteraar die midden voor de luidsprekers plaatsneemt, zal een geluidsbeeld waarnemen dat gelijkenis vertoont in klank, diepte en balans met het oorspronkelijke geluidsbeeld. Een luisteraar die niet precies in het midden zit, ervaart een ruimtelijk vervormd geluidsbeeld dat naar de dichtsbijzijnde luidspreker nijgt. Het optimale luistergebied voor stereofonische weergave beperkt zich dus tot enige zitplaatsen op de lijn midden tussen beide luidsprekers.

In dit proefschrift wordt een nieuwe reproductiemethode beschreven, die is gebaseerd op het principe van golfveldsynthese (*wave field synthesis*). Hierbij worden rijen (*arrays*) van luidsprekers op dusdanige wijze aangestuurd met bewerkte versies van de bronsignalen (*source signals*), dat er golffronten worden gevormd die vrijwel identiek zijn aan die van de originele geluidsbronnen. Deze reproductiemethode biedt daarmee een volumeoplossing, in plaats van de puntoplossing die kenmerkend is voor conventionele reproductiemethoden.

De originele bronsignalen wordt bewerkt door een syntheseoperator, die uit de golfveldrepresentatie van Kirchhoff-Helmholtz wordt afgeleid. Deze theorie beschrijft de verdeling van druk en deeltjessnelheid op een gesloten oppervlak, ten gevolge van (primaire) geluidsbronnen buiten het omsloten volume. Wordt dit oppervlak bedekt met secundaire dipool- en

monopoolbronnen, dan is theoretisch een perfecte reconstructie van het primaire bronveld mogelijk in het omsloten volume. Voor het speciale geval waarbij de secundaire bronnen in een plat vlak liggen, kan Rayleighs theorie worden toegepast. Deze stelt dat, indien de primaire bronnen achter het platte vlak liggen, een van beide soorten secundaire bronnen volstaat om een kopie te genereren van het primaire golfveld vóór dat vlak. In de praktijk kunnen vlakke rijen van luidsprekers (secundaire bronnen) worden gebruikt om virtuele bronnen (primaire bronnen) te synthetiseren. Wanneer nu zowel de luisteraars als de virtuele bronnen zich in een horizontaal vlak bevinden, is een lineaire horizontale rij van luidsprekers voldoende. Tevens is het door middel van focussering mogelijk om virtuele bronnen vóór de luidsprekers te creëren: de luisteraars nemen dan een bron in het akoestische focus waar.

Omdat de luidsprekers in de rij een zekere afstand tot elkaar hebben, ontstaan er boven een bepaalde frequentie spectrale afwijkingen in het gereproduceerde golfveld: de zogeheten effecten van ruimtelijke onderbemonstering (*spatial aliasing*). Uit de theorie komt naar voren dat naarmate de luidsprekers een sterkere richtwerking (*directivity*) hebben, het opgewekte geluidsveld het originele veld meer benadert. De richtwerking van de luidsprekers kan worden beïnvloed door het afstralende diafragma te modelleren. Dit lukt het beste met elektrostatische luidsprekers. Een prototype rij van luidsprekers, gebouwd volgens deze inzichten, bevestigt dat het mogelijk is een geluidsveld op te wekken dat ook voor hogere frequenties correct is. Het rendement van dit prototype is echter nog onvoldoende voor toepassing bij golfveldsynthese, zodat tevens een elektrodynamische rij van luidsprekers is ontwikkeld.

Met deze elektrodynamische luidsprekers is een 128-kanaals reproductiesysteem gebouwd, dat een luistergebied van 24 m² omsluit. De geluidsvelden worden in dit systeem per bron als volgt opgebouwd: het directe geluid (geregistreerd met een steunmicrofoon) wordt als een virtuele bron weergegeven, de vroege reflecties (berekend met een spiegelbronnenmodel) worden eveneens als virtuele bronnen gesynthetiseerd, en de nagalm (opgenomen met microfoons in het galmveld) of kunstgalm wordt door middel van vlakke golven uit verschillende richtingen samengesteld. Ook snelbewegende bronnen (opgenomen met meebewegende steunmicrofoons) kunnen worden gesynthetiseerd, hetgeen vooral voor toepassing in bioscopen, simulatoren en *virtual reality* theaters van belang is.

Het systeem is op twee manieren geëvalueerd: met metingen (objectief) en met luisterproeven (subjectief). Uit een vergelijkend onderzoek, waarvan ook stereofonie en 3/2 *surround sound* deel uitmaken, blijkt dat het golfveldsynthese systeem het grootste luistergebied met ruimtelijk correcte geluidsreproductie creëert. Ook is het geluidsbeeld robuuster dan bij de andere weergavemethoden, in het bijzonder als de opnametechniek wordt afgestemd op weergave met gesynthetiseerde golfvelden. Het systeem is bovendien in staat een zeer ruimtelijke impressie te genereren, waarbij de luidsprekers zelf niet meer kunnen worden onderscheiden. Voorts wordt ook de flexibiliteit van het systeem ten aanzien van het beïnvloeden van bronposities en akoestiek beschouwd als een belangrijke vernieuwing.

Edwin Verheijen, Technische Universiteit Delft, 1997