Audio Engineering Society
Convention Paper 5500
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Microphone techniques and directional quality of sound reproduction

Ville Pulkki

[1] *Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, POBox 3000, FIN-02015, Finland*

Correspondence should be addressed to Ville Pulkki (Ville.Pulkki@hut.fi)

ABSTRACT

Different spatial sound reproduction techniques are evaluated using a binaural auditory model. Ear canal signals for different microphone techniques and different loudspeaker reproduction are simulated. Directional auditory cues are calculated and directional quality is discussed. The results of recording techniques for stereophonic listening explain the subjective opinions presented in literature: With coincident microphone techniques directionally fairly stable and consistent virtual sources can be produced, and with spaced microphones more spread and ambiguous virtual sources are achieved. In multichannel reproduction, none of the existing microphone techniques are found to produce good directional quality. Both coincident and spaced microphone techniques produce spread virtual sources.

1 INTRODUCTION

The temporal and spectral structure of sound signal can be captured and reproduced with good accuracy with modern audio technology. In contrary, the reproduction of spatial attributes of sound can not be judged to be accurate in general. Spatial attributes denote here the part of sound perception that is different in different listening rooms, and in different listening setups within one room. Different spatial attributes occur because direct sound, reflections, and reverberation depend on listening setup and listening room acoustics.

Two-channel stereophony [1] is the most used spatial sound reproduction method. The listener perceives all auditory objects appearing on a line between the loudspeakers. The line can be thought to be an acoustical opening to the room where a recording was made. Naturally, using such system a listener can not have equal perception of spatial sound as in the recording room. In 70's there were many attempts to create

2-D immersive sound perceptions by using four loudspeakers in a square around the listener to enlarge the acoustic window [2]. Unfortunately the microphone techniques and analog encoding-decoding from a multi-channel microphone via two channels to four loudspeakers failed to create stable spatial impressions. Ambisonics [3] is the only method that has survived.

In past ten years a five-loudspeaker listening standard (5.1) has been becoming more common. New microphone techniques have been suggested for such loudspeaker systems. No one of techniques have been commonly recognized. There seems to be no implicite way to record spatial sound for multi-loudspeaker systems with existing microphone types.

There has not been reliable ways to measure objectively the perceptual quality of spatial sound reproduction systems. Recently, work has been conducted on measuring the perceptual directional qualities of amplitude panned virtual sources [4]. The use of binaural auditory model has shown to be a reliable tool for the evaluation of direction perception.

In this work the directional qualities of different reproduction methods are measured objectively using the auditory model applied in previous studies. The model is used to evaluate different spatial sound recording methods for stereophonic setups, evaluate first- and second-order Ambisonics in four- and six-loudspeaker setups and evaluate a method to record spatial sound for 5.1 systems.

2 SPATIAL HEARING

Spatial and directional hearing have been studied intensively; for overviews, see for example [5] or [6]. The duplex theory of sound localization states that the two main cues of sound source localization are the *interaural time difference* (ITD) and the *interaural level difference* (ILD) which are caused respectively by the wave propagation time difference (primarily below 1.5 kHz) and the shadowing effect by the head (primarily above 1.5 kHz). The auditory system decodes the cues in a frequency-dependent manner.

The cues resolve in which cone of confusion the sound source lies. A cone of confusion can be approximated by a cone having axis of symmetry along a line passing through the listener's ears and having the apex in center point between the listener's ears. Direction perception within a cone of confusion is refined using other cues, such as spectral cues and effect of head rotation to ITD and ILD. Spectral cues and head rotation are considered to carry elevation and front-back information. The precedence effect [5, 7] is an assisting mechanism of spatial hearing. It is a suppression of early delayed versions of the direct sound in source direction perception, which helps to perceive direction of sound source in reverberant conditions.

In spatial sound reproduction, a sound source may be perceived at a location within the confusion cone other than where it was intended. The most common problems are front-to-back and back-to-front confusions.

Both main cues of spatial hearing are decoded across the audible spectral range; however, the relative importance of the cues is unclear. Wightman et al. [8] have proposed that when the cues are distorted, the auditory mechanism uses the cue which is most consistent. A cue is consistent if it suggests the same direction in a broad frequency band.

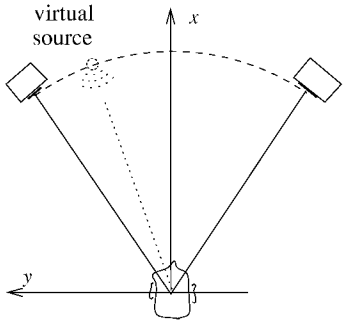

Fig. 1: Standard stereophonic listening configuration.

3 SOUND REPRODUCTION TECHNIQUES

In sound reproduction chain the sound arriving from a physical sound source directly or via reflections or diffraction is captured with microphones and stored. The stored signal is thus the sound source signal convolved with the responses of room and microphone. In listening phase the microphone signals are applied directly or via decoding matrix to loudspeakers. A listener perceives sound and its spatial attributes.

A reproduced sound source is called as a virtual source. A virtual source may appear to a point-like or spreaded location with respective to a listener. If a realistic reproduction is desired, virtual source properties should be equal that appeared in recording room. Also, in realistic reproduction the perception of reverberation should be equal that occurs to a listener in recording room. However, often a realistic reproduction is not even desired, techniques that modify spatial properties of virtual sources and reverberation may be used.

3.1 Stereophonic techniques

Stereophonic sound reproduction system is the most common way to reproduce spatial sound. In it two loudspeakers are placed in front of a listener, as presented in Fig. 1. The loudspeaker aperture is typically 60°. Reproduced sound objects appear between the loudspeakers. If special techniques such as cross-talk canceling [9] are used, virtual sources may appear also at other directions, however, these techniques are not considered in this paper.

Audio engineers have different ways to use this media. Sometimes the sound sources are reproduced point-like in spatial arrangement that they appeared in reality, though compressed to a line between the loudspeakers. Other approach is to hide sound source directions and create a feeling to listener that each source is spreaded evenly between the loudspeakers.

Naturally an immersive reproduction of reverberation is not possible with a stereophonic setup. Typically it is reproduced as appearing evenly between the loudspeakers. Some reproduction techniques are critisized that the reverberation is concentrated mostly to the loudspeakers, not to a line between them.

There exists different methods to record spatial sound to be presented with a stereophonic setup. Polar patterns available for current microphones are of zeroth order (omnidirectional), or of first order (figure-of-eight, cardioid and hypercardioid), as shown in Fig. 2. In this study microphones are considered

to have an equal directional pattern at all frequencies.

A coincident technique denotes a microphone technique in which two or more directive microphones are placed as close as possible to each other, first used in [1]. In the resulting signals there are thus no differences in phase, the only differences are in amplitude. Coincident microphone techniques are therefore equal to amplitude panning [1]. In amplitude panning a sound signal is applied to several loudspeakers with different amplitudes without time differences.

A non-coincident technique denotes a microphone technique in which the microphones are separated in space. This produces also time differences between loudspeaker signals. The directional patterns of the microphones may be of any form.

Some frequently used methods are listed below

- Coincident techniques
  - XY cardioids
  - XY hypercardioids
  - Blumlein
- Non-coincident techniques
  - ORTF
  - Spaced omnidirectional microphones

The microphone techniques can also be divided to techniques in which they are close to sound sources, or far away from the sound sources. In this analysis the far away case is considered. Such techniques are used to also capture the response of the room or hall in which sound sources are. These techniques are used commonly in recording of classical music. Typically the sound sources are in front of the microphones, and the response of the room arrives from all directions.

In following the far-away microphone techniques considered in this paper are described briefly as in [10] and in [11].

XY techniques. Two microphones are placed as close each other as possible, and the polar pattern of the microphones is typically cardioid or hypercardioid. The loudspeaker signals therefore have no time differences, but have different amplitudes. However, when a hypercardioid pattern is used, the signals are in the opposite phase with certain sound source directions. Base angle between microphones varies from 60° to 180°. Virtual sources are perceived quite consistently to one direction independent on the frequency. These techniques capture virtual sources louder from front than from behind. The reproduction of reverberation has been criticized, it may be perceived lacking "air" or "warmth".

Blumlein pair consists of two coincident figure-of-eight microphones with a base angle 90°. As in XY techniques, there are no time differences between loudspeaker signals. The loudspeakers have 180° phase difference when sound source is on left or right of the microphone setup. This method produces consistent virtual sources of sound sources that appear in front or behind the microphone system. The sound sources that are to either side of the Blumlein pair are localized inconsistently. The virtual sources are captured equally loud from all directions, and they are not colored. Blumlein pair have also been criticized about lack of "air" or "warmth".

Spaced microphones are typically separated with distance between 20 cm and few meters from each other. Omnidirectional pattern is commonly used in this technique. The

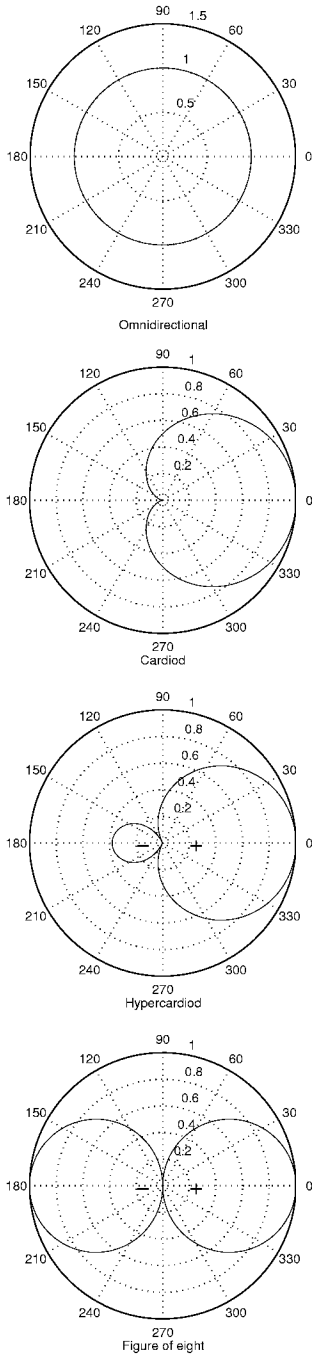

Fig. 2: Basic microphone polar patterns.

captured signals occur at different times in different microphones, the time difference depends on direction of arrival. There might exist also differences of amplitude if the microphones are far away from each other. The virtual sources are localized inconsistently; the localization varies with frequency. However, the sound is often considered to be more ambient", "airy" or "warm" than sound recorded with coincident microphone techniques.

ORTF consists of two cardioid microphones spaced with 17 cm in a base angle of 110°. The captured signals differ both in time and in amplitude. At low frequencies the system equals to the XY cardioid technique, since the distance of microphones is small when compared to wave length. At high frequencies there are also some prominent phase differences involved in microphone signals. It is reported that perceptual qualities of virtual sources and reverberation would be somewhere middle between qualities of coincident and spaced techniques.

3.2 Multichannel systems

There has been many attempts to create microphone technique that would reproduce spatial sound over multiple loudspeakers. Some of them are presented here.

3.2.1 Ambisonics

Ambisonics [12] is a microphone technique that is based on use of the Soundfield microphone [13]. Typically the output of the microphone consists outputs of four different microphones, that are an omnidirectional microphone (signal W) and three figure-of-eight microphones faced towards three coordinate axis (signals X, Y and Z). Audio can be stored as these signals, and the storage format is called B-format. In reproduction the signals are matrixed in a way that the signal applied to each loudspeaker corresponds to a signal that could have been recorded with a hypercardioid or cardioid microphone facing to the direction that corresponds to direction of loudspeaker in listening room.

Typically the signals of a Soundfield microphone are matrixed for a quadraphonic loudspeaker setup, in which the loudspeakers are in directions of ±45° and ±135°. Quadraphonic setup is a two-dimensional setup, thus only signals W, X and Y are needed. If the loudspeaker signals of quadraphonic system are denoted with RF, LF, RB and LB, (L = left, R = right, F = front, B = back) the equations can be written as:

$$LF = 0.707 * W + X + Y \quad (1)$$

$$RF = 0.707 * W + X - Y \quad (2)$$

$$LB = 0.707 * W - X + Y \quad (3)$$

$$RB = 0.707 * W - X - Y. \quad (4)$$

There exists also a modification of this technique in which the multiplier 0.707 of W is replaced with 1. The polar patterns of loudspeaker signals are then cardioids.

A theory of second-order Ambisonics has been proposed [14]. The method is based on a hypothesized second-order Soundfield microphone that would output in addition to W, X, Y, and Z signals also five signals having quadrupolic polar pattern in different orientations. The polar pattern $f(\theta)$ of signals fed to loudspeakers would then have form

$$f = 1 + 2\cos(\theta) + 2\cos(2\theta), \quad (5)$$

where θ is space angle between the frontal axis of microphone and the direction of sound source. It is not known if a 2nd-order Soundfield microphone can be constructed.

3.2.2 5.1 surround

The most widely used multi-channel loudspeaker system is the 5.1 loudspeaker configuration. It has loudspeakers in directions ±110°, ±30° and 0° [15]. It is widely used in cinemas, and is gaining popularity in domestic use also.

There exists various methods to record sound for it. One way is to use Ambisonics techniques, which corresponds to coincident techniques. Some spaced microphone configurations have been also used with 5.1 system. In many cases microphones are in figure of star, facing approximately towards the corresponding loudspeaker direction. Different directional patterns of microphones are used. Special hardware may be provided to mix the signals to loudspeakers or to perform spectral filtering.

3.2.3 Wave field synthesis

When the amount of microphones and loudspeakers is very large (over 100), Wave Field Synthesis [16] can be used. It reconstructs the whole sound field that appeared in recording space to the listening room. It is superior as a technique, but unfortunately it is unpractical in most situations. It is not further discussed in this paper.

4 MODELING VIRTUAL SOURCE PERCEPTION

To simulate the directional perception of virtual sources, a binaural auditory model was used in this study to calculate localization cues for the audio signals arriving to the ear canals.

Some simplifications must be, however, tolerated. In this study we have restricted our scope by eliminating the influence of the precedence effect as much as possible so that it does not have to be modeled. When the model lacks the precedence effect, it gives reliable results only if all incidents of a sound signal arrive within about a one millisecond time window to ears. This can be achieved only in anechoic conditions, since in all rooms there exists reflections and reverberation that violate the 1 ms window. Qualitatively the results are also valid in moderately reverberant conditions. Also, in analyzed setups the microphones can not be separated more than 35 cm, otherwise the loudspeaker signals would violate the window.

The model of auditory localization used in this study consists of following parts:

- simulation of microphone technique
- simulation of ear canal signals in listening phase
- binaural model of neural decoding of directional cues
- model of high-level perceptual processing

Since the use of the model is described elsewhere [17], it is discussed here only briefly.

4.1 Simulation of ear canal signals

Sound reproduction simulation as well as torso and ear filtering simulation in the model approximate the sound signals arriving to listener's ear canals, which is shown in Fig. 3. In this study, the audio signals applied to the loudspeakers are calculated by simulating a microphone technique. The signals arriving to the ear canals from each loudspeaker are computed

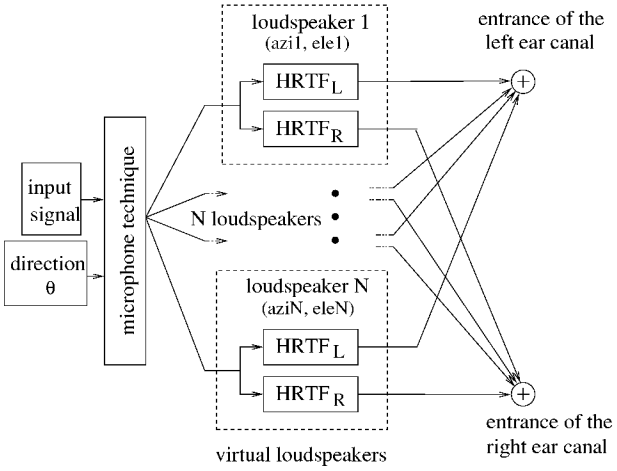

Fig. 3: Simulation of ear canal signals in arbitrary sound reproduction systems.

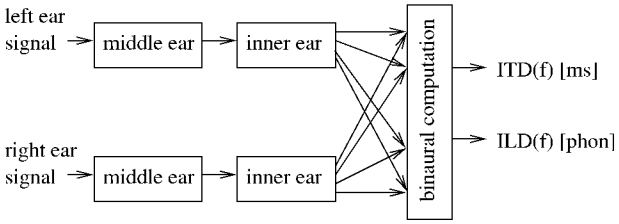

Fig. 4: Binaural model of directional cue decoding.

using digital filters that implement measured head-related transfer functions (HRTFs) of corresponding direction. The arriving HRTF-filtered loudspeaker signals are added together to form ear canal signals.

4.2 Binaural model of directional cue decoding

A schematic diagram for the binaural model of neural decoding for directional cues is presented in Fig. 4. It takes as input the sound signal arriving to the ear canals and computes the decoded frequency-dependent ITD and ILD cues. It models the middle ear, the cochlea, the auditory nerve, and the binaural decoding. The middle ear, cochlea, and auditory nerve models have been implemented based on the HUTear 2.0 software package [18]. The middle ear is modeled using a filter that approximates a response function derived from the minimum audible pressure curve [19]. The cochlear filtering of inner ear is modeled using a 42-band gammatone filter bank [20]. Center frequencies of the filter bank follow the ERB (equivalent rectangular bandwidth) scale [21]. Auditory nerve responses are modeled with half-wave rectification and low pass filtering. The impulse sharpening that occurs in cochlear nucleus [22] is modeled roughly by raising the signal to a power of two.

The binaural computation consists of ITD and ILD decoding. The neural coincidence counting [22] that performs ITD decoding is modeled using the cross-correlation calculation as suggested by Jeffress [23]. The cross-correlations are calculated with a [-1.1 – 1.1] ms time lag range at each ERB band. This produces a function for each frequency band that denotes how the ear signals coincided with different time lags. The time lag corresponding to the highest peak implies the ITD in each frequency band. Due to low-pass filtering of auditory nerve, the ITD corresponds to carrier shifts at low frequencies and envelope shifts at high frequencies.

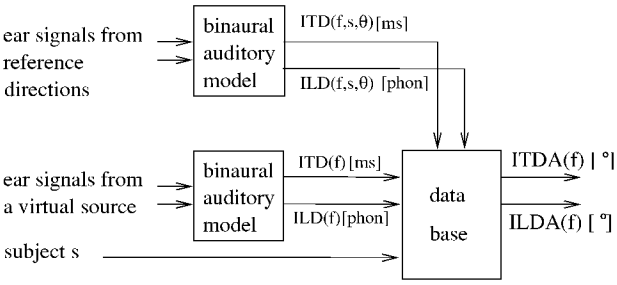

Fig. 5: Functional model of auditory localization.

However, there may exist multiple prominent peaks in the cross-correlation function in some cases. It has been shown that if a peak appears in the same time lag at adjacent frequency bands, it is considered more relevant in localization. To implement this, a second-level coincidence counting unit was also added [24]. The cross-correlation function at an ERB channel is multiplied with cross correlation functions at the next lower and upper ERB band. After this the highest peak is found.

The loudnesses of each frequency band in each ear are calculated using Zwicker's formulae [25]. Due to simplicity, this model is used instead of the more thorough model proposed by Moore [26]. The difference of loudness levels between the ears at each frequency band is treated as ILD spectrum. The loudnesses are summed at each ear and each frequency band to form an estimate the overall loudness of a sound source.

4.3 Model of high-level perceptual stages

Higher levels of human auditory processing produce direction perception as a fusion from ITD, ILD and other cues. These auditory mechanisms are not well known, therefore a physiologically based model can not be applied. However, the modeling of high-level perceptions would be beneficial since the ITD and ILD cues are measured in different scales, which means that they can't be compared directly with each other. Additionally, ITDs or ILDs can not be compared between subjects due to individuality of the cues. If a mapping from the cues to spatial directions that they correspond is formed, the cues can be compared in such ways.

A straightforward method to form a such mapping is a functional model that consists of a database that holds the sound source ITDs and ILDs produced by a sound source at each direction for each individual, which is illustrated in Fig. 5. An auditory cue value that has been measured from a virtual

source is transformed to a direction angle value by database search. Two subsequent values are found between which the cue lies. The resulting direction angle value is interpolated between these two values. The functional model computes frequency-dependent ITD angles (ITDA) and ILD angles (ILDA). These cues present the azimuth angles that the binaural properties of the measured virtual source suggested at each frequency band. Since this study considers only virtual sources in the horizontal plane, the database consists of ITD and ILD values of sound sources at azimuths $\theta = -90°, -80°, \ldots, 90°$.

The cues may behave in an unstable manner in some cases. ILD may behave nonmonotonically with azimuth angles larger than approximately 60°. Nonmonotonic parts of the ILD curves are removed. If a larger virtual source ILD value exists than is on ILD table, the response is extrapolated from previous values. However, the ILDA can not exceed 90°. The ITD values calculated for the database from HRTF-measurements also might be unstable, which would generate error to ITDA estimation. Due to this, the ITD databases were post processed. If one value differed considerably from adjacent values, it was replaced with the mean of values produced by the same sound source at adjacent frequencies. Also, the validity of computed ITDA values was checked and clearly erroneous values were removed. The virtual sources may generate large ITD values that do not correspond to any direction. If at any frequency band the value of a virtual source ITD cue is smaller or larger than any of database ITD values at the corresponding frequency band, the ITDA is not calculated and is considered as a missing value in the data analysis.

4.4 Using auditory model in virtual source perception simulation

The ITDA and ILDA angles are calculated with 6 individual HRTF sets to both side of the listener at 42 frequency channels. The resulting values obtained from left side HRTFs are turned to right side values by inverting the cue value sign. This results 12 estimates of the direction that the computed cue suggests at each frequency band. The mean value and standard deviation are calculated over individuals.

In the results the means and standard deviations of cue angles with microphone systems and different sound source directions are plotted to a same figure. The polarity of the cues is changed to negative in roughly half of the virtual source cues, this is to maintain clarity of the figures. The angle of sound source incident is plotted to vicinity of the curve.

The model have some shortcomings. The ILD of sound sources is nonmonotonic with azimuth at certain frequencies. This yields that ILDA is an unreliable estimate of perceived direction at those frequencies.

The evaluation system is tested by analyzing real sound sources in different directions around the listener. Constant values with frequency should be achieved in ideal case for cue angles, which are shown in Fig. 6. It can be seen that ITDA corresponds well to the direction of sound source. There are minor deviations at large sound source angles. The ILDA values behave consistently with directions below 50°. With angles $> 50°$ ILDA deviates from sound source direction generally, it is roughly correct only at frequencies higher than 4 kHz. The large deviations are caused by nonmonotonic ILD behavior with source direction [5]. This suggests that ITDA can be used in spatial sound analysis generally, in ILDA analysis it should be taken into account that ILD does not have large values between 700 Hz and 4 kHz.

Also, in the figure it can be seen that that sound sources in a same cone of confusion produce equal ITD and ILD values. In this case in same cone of confusion are 0° with 180°, 30° with 150° and 60° with 120°. Clearly the computed cue angle value presents the angle between the median plane and the sound source.

4.5 Directional loudness plot

Microphone techniques reproduce sound sources with different levels depending on sound source direction. To estimate the perceptual loudness of reproduced sound sources, overall loudnesses of virtual sources were simulated with different directions of sound sources. Loudness estimates are calculated for all sound source directions, and are normalized with inverse of maximum value. A mean value over individuals is taken. The resulting values are plotted to a polar plot as a function of sound source direction. The plot is called as a directional loudness plot. The recording and listening space is assumed to be anechoic. This may limit validity of results in reverberant listening conditions. For real sound sources in anchoic conditions it is shown in Fig. 6. The rear sound sources are suppressed mildly, this is because of human ear directivity.

5 SIMULATING THE DIRECTIONAL QUALITY OF SPATIAL REPRODUCTION

In this section virtual source direction perceptions are simulated in various stereophonic and multichannel reproduction methods. The microphone signals are computed by simulating a sound source and a microphone technique in anechoic listening condition to directions $0°, 30°, \ldots, 180°$ around the microphone system. The microphone signals are applied to stereophonic or multichannel listening setup, also in anechoic conditions. Each loudspeaker is simulated with by convolving the simulated loudspeaker signal with a measured HRTF of corresponding direction. This yields ear canal signals, from which the auditory cues are calculated.

The results are shown as a figure that contains the directional patterns of microphones, a directional loudness plot and computed ITDA and ILDA cues. The directional loudness plot describes how loud a recorded sound source appears in reproduction. The cues estimate to where it is localized.

5.1 Stereophonic reproduction

In this section various microphone techniques are simulated that are used with standard stereophonic listening configuration. The loudspeakers in reproduction are in all cases $\pm 30°$ directions.

All simulated microphone systems are symmetric with respect to the median plane, which yields that there exists no differences between microphone signals with sound source 0° and 180° directions. This yields that in stereophonic listening there is no interaural differences, which guarantees that ITDA and ILDA have value zero. Cues for these directions are thus not shown generally.

5.1.1 XY Cardioids

Two cardiod microphones facing to $\pm 45°$ azimuths were considered first. The results of the simulation are shown in Fig. 7. As mentioned earlier, coincident microphone technique is

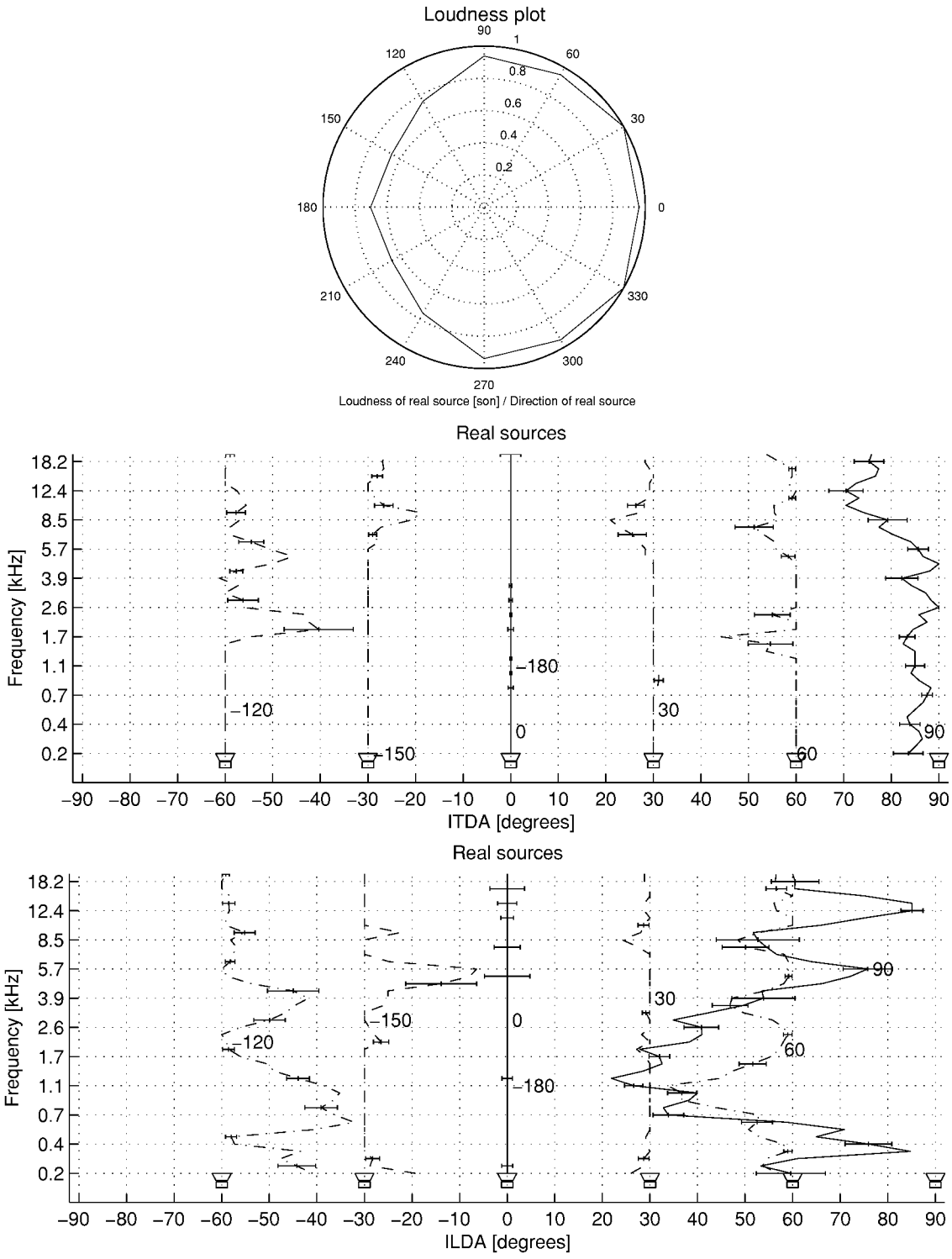

Fig. 6: ITDA and ILDA values measured with real sound sources. Whiskers denote 25% of standard deviation

equivalent with amplitude panning. A typical feature of amplitude panning can be seen in simulated cues. ITDA values are consistent at low frequencies and ILDA values at high frequencies. There are some deviations near 1.7 kHz, as also found in [27].

However, the directions of virtual sources do not coincide with directions of sound sources, virtual source are located between reproduction loudspeakers. The angles of sound source incidents $0° \ldots 150°$ are mapped linearly between 0° and 30°. The method produces relatively consistent cues, which suggests that the virtual sources are localized consistently to one direction, as reported in Sec.3.1.

In the loudness plot it can be seen that this microphone setup reproduces sound sources in frontal directions with greater loudness than in rear directions. In recording of reverberant halls this may not yield desired artistic impression. The level of reverberation and reflections arriving from rear is suppressed because of the directional patterns of microphones. The lateral reflections and reverberation are not suppressed, but they might be localized near the loudspeakers, which may be undesirable.

5.1.2 XY Hypercardioids

The virtual sources were analyzed for a pair of hypercardioid microphones in $\pm 45°$ orientation. The resulting cue angle values are shown in Fig. 8 together with polar patterns of the microphones. Cue values for sound source incident $0°, 30°, 60°$ and 150° are similar as in cardioid simulation. With 90° and 120° sound sources somewhat ambiguous cues are achieved; the cue value is dependent on frequency and there are individual differences.

A reason to this can be seen easily: The microphone signals are in same phase with sound source directions $0°, 30°, 60°$ and 150°, for those cases the cue values have typical behaviour of amplitude-panned sources. With other incidents the loudspeaker signals are in opposite phase. This can be seen as abnormal ITDA behavior at low frequencies. At some cases the values do not exist at very low frequencies, this occurs since the ITD has been larger than any sound source ITD. At high frequencies ITDA values produced with anti-phase signals tend to zero.

The loudness plot shows that this microphone pair reproduces loudnesses of sound sources between $-120°$ and 120° fairly equally. However, in rear sound source directions the loudness drops substantially.

This simulation thus suggests that when recording concerts in halls this microphone system would produce quite consistent localization for frontal sound sources, produce anomalous localization for reflections and reverberation coming from sides and suppress sounds coming from back.

5.1.3 Blumlein pair

A Blumlein pair was simulated, the results are presented in Fig. 9. The polar patterns of Blumlein pair implies that the loudspeaker signals are in same phase with frontal and rear sound sources, and in different phase with lateral sound sources. The simulation result should thus resemble result obtained with hypercardioids. The similarity can be seen in cue angle values: ITDAs of 0°, 30° and 150° sound sources propose quite consistent directions at low frequencies, while lateral incidents 60°, 90° and 120° produce abnormal ITDA cues at low frequencies. With lateral incidents ILDA values behave quite consistently, however differently from ITDA values.

The 150° sound source direction produces cues to opposite side of the median plane. This proposes that sound sources in rear left would be perceived in front right and vice versa. This occurs since in rear a sound source is captured most with a microphone that is connected to the loudspeaker on opposite side of listener.

Differing from the results with hypercardioids, Blumlein pair produces an equal loudness with all sound source directions, as can be seen in loudness plot. Typically this affects reproduction of reverberation, reverberant field will be captured more than with XY techniques. When compared with hypercardioid results, a difference is also that the sound signals captured in antiphase were attenuated, and with blumlein pair they are not. The anti-phase effect is thus more prominent, and it also occurs with a wider span of directions.

5.1.4 Spaced omnidirectional microphones

Two omnidirectional microphones spaced with 17 cm were simulated. The results are shown in Fig. 10. Since this microphone system is symmetric with x- and y- axis, sound sources in only one quadrant need to be simulated. The loudness produced by sound sources in different directions is similar from all directions. However, the cues are very abnormal. ITDA is almost zero at 200 Hz, and changes its value rapidly between 200 and 1000 Hz, after which it stabilizes to some fairly large value. The ILDA behavior is very abnormal also, it oscillates at low frequencies wildly and all directions stabilize near 0° at high frequencies.

This simulation proposes that sound sources are perceived to direction that is dependent on the frequency of sound. Also, a virtual source may be spreaded heavily. Exception to this is naturally for sound signals coming from azimuth 0° and 180° (with all elevations), since the microphone signals are then equal, that produces a virtual source in center point between loudspeakers. However, for reproduction of reverberation, this kind of spreading of directions might be desired; reverberation is then perceived evenly between loudspeakers. This analysis coincides with subjective opinions about recordings made with spaced microphones. In practise the microphones may be farther away from each other than in this simulation. Unfortunately such cases can not be addressed since the precedence effect is not modeled.

5.1.5 ORTF

ORTF microphone placement is somewhere middle between coincident and spaced techniques. Two cardioids are placed within 17 cm distance with $\pm 55°$ angular orientation. Directional cues were simulated for ORTF, results are shown in Fig. 11. In loudness plot it can be seen that ORTF emphasizes strongly frontal sound sources, which is due to polar patterns of microphones. Theoretically it was stated that ORTF technique is equivalent with coicident microphone techniques at low frequencies. This analysis suggests, however, that cues behave differently from XY techniques starting already from 300 Hz. ITDA cues are concentrated near loudspeaker directions at higher frequencies. ILDA cues of sound source directions 0°, $-120°$ and 150° behave quite consistently, while others fluctuate with frequency. There are also quite large individual differences.

In overall, cue values are more ambivalent than as they occur with coincident techniques, and more stable than with spaced

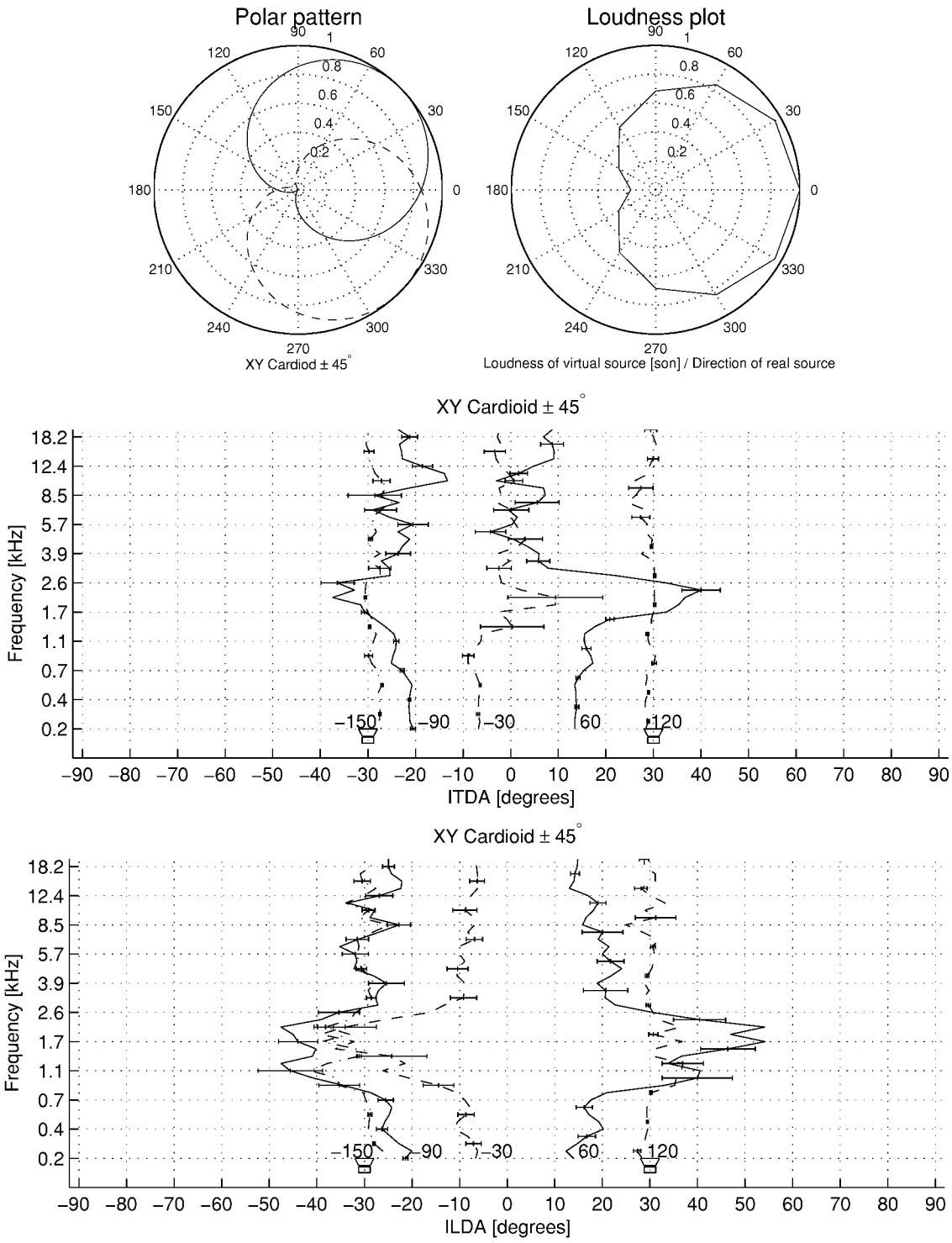

Fig. 7: Directional cues of XY Cardioid microphone technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

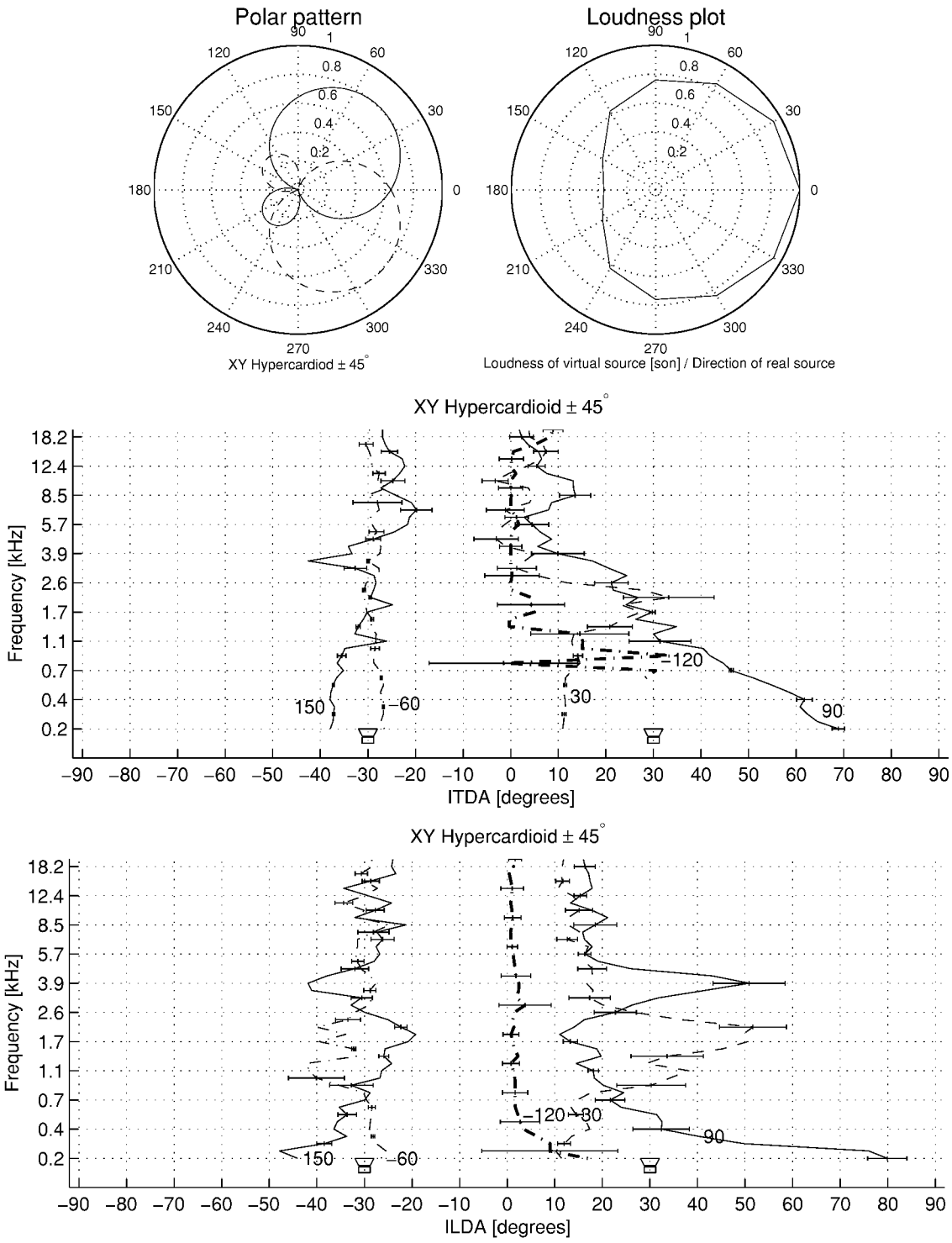

Fig. 8: Directional cues of XY hypercardioid microphone technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

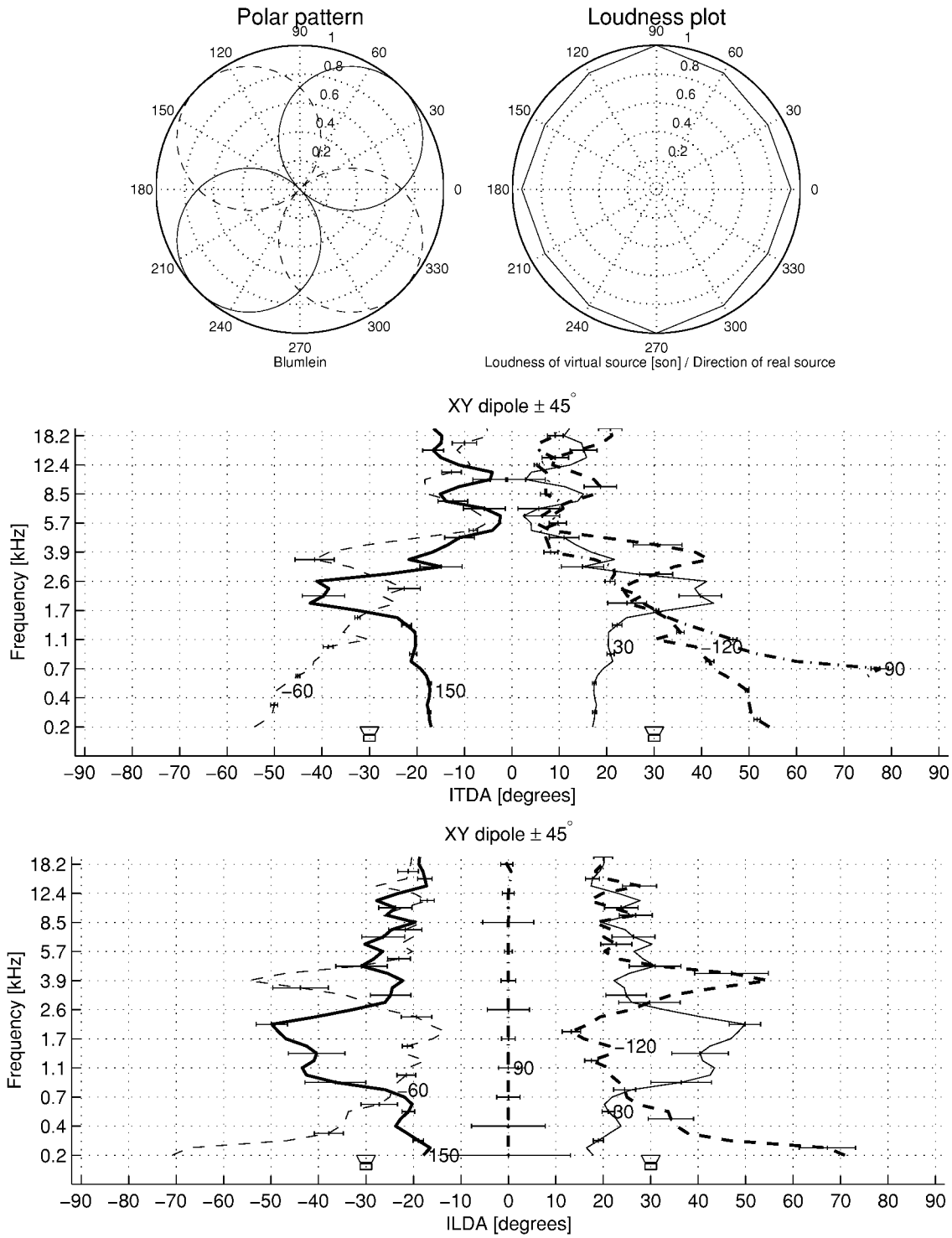

Fig. 9: Directional cues of Blumlein microphone technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

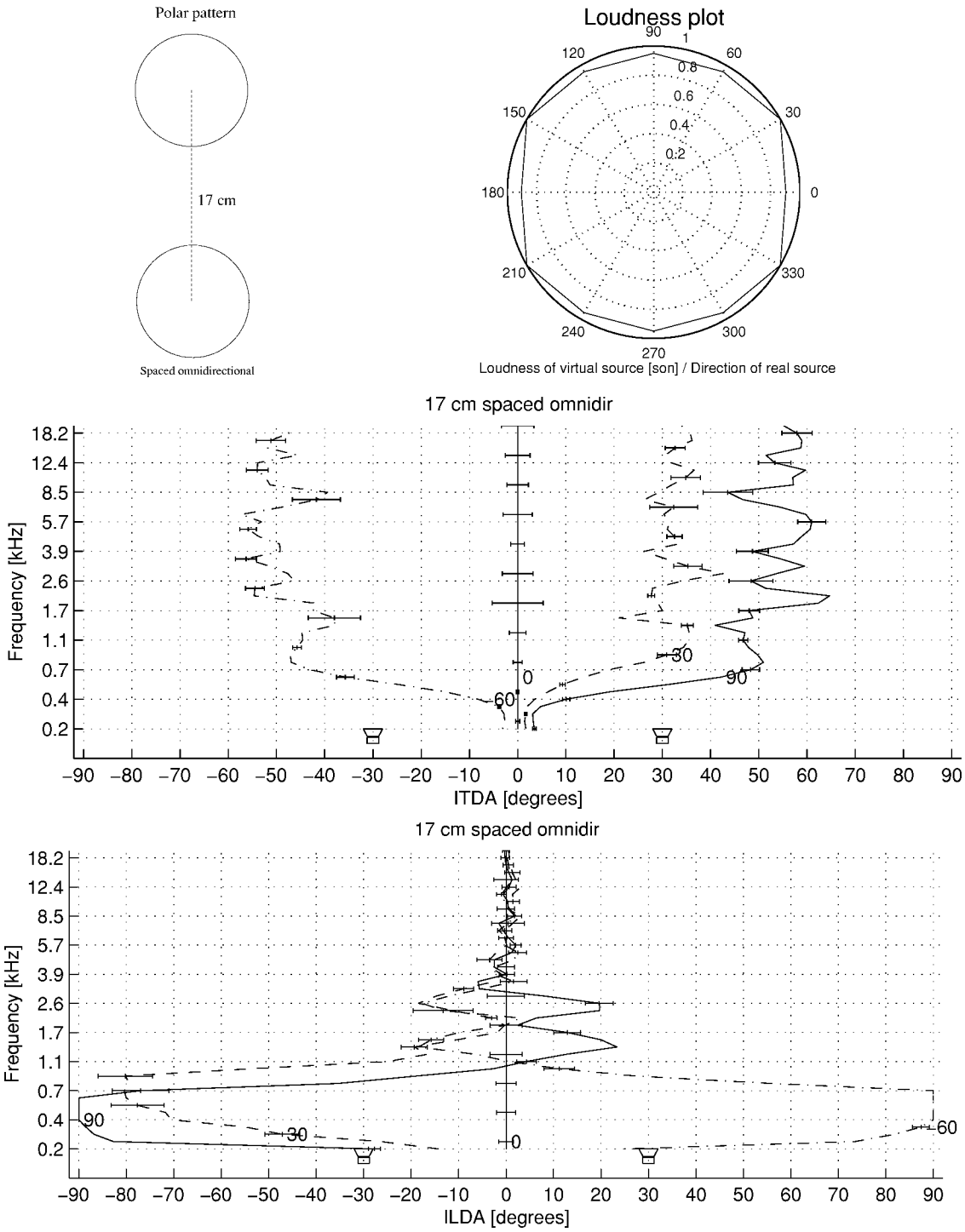

Fig. 10: Directional cues of 17 cm spaced omnidirectional technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

omni techniques. The perceptual qualities of ORTF technique are considered to be between qualities of spaced omni techniques and coincident techniques [11]. The simulation results thus coincide the subjective experiences.

5.2 Multi-channel reproduction

Methods to reproduce spatial sound with horizontal loudspeaker arrays are considered in this section. In these setups there are loudspeakers also behind the listener. In this analysis only ITDA and ILDA cues are considered, which may get values from $-90°$ to $90°$. The ITDA and ILDA cues of virtual sources must then be interpreted with corresponding cone of confusion when sound source direction is over $90°$; e.g. when sound source direction is $120°$, cue values should be $60°$, and for source direction $150°$, cue values $30°$ are optimal.

5.2.1 Ambisonics

The four-loudspeaker horizontal Ambisonics was simulated. It was found that the sound sources in a cone of confusion produced similar cue angle values, which can be understood because front-back symmetry of the loudspeaker setup. Thus the results are shown in Fig. 12 only for sound source directions $0°$, $30°$, $60°$ and $90°$. As with two-channel coincident reproduction methods, ITDA values at low frequencies are fairly stable, however, they deviate from the target value prominently, especially with large sound source directions. ITDA is unstable at high frequencies. ILDA is also generally unstable and deviates from sound source direction prominently, however, the ILDA for more lateral sound source directions have generally greater magnitudes, than for directions nearer the median plane. The stable ITD proposes that virtual sources will be localized relatively stably to one direction. However, their bias towards the median plane predicts that stable virtual sources are not produced to lateral directions.

In the cues there is a large bump between 400 Hz and 3 kHz in ILDA values. This may produce sensation of lateral virtual sources, although the values are larger than any distant sound source can produce, as seen in Fig. 6. Such values may lead to near- or inside-head localization, since they may be produced only with nearby sound sources. The simulation result with hexaphonic setup was similar, and is not printed here. Also, the simulation was repeated with cardioid patterns, which did not produce significant differences.

Second-order Ambisonics was simulated with a hexagonal loudspeaker setup in which loudspeakers are in directions $\pm 30°$, $\pm 90°$ and $\pm 150°$. The results are shown in Fig. 13. The polar pattern is shown only for two microphones, the rest are similar but in different orientations. Low-frequency ITDA suggests quite consistently and accurately the sound source directions, cues at higher frequencies are unstable and biased prominently towards the median plane. ILDA cue does not coincide with sound source direction generally. It seem to be biased towards median plane, especially at high frequencies. Both cues deviate between individuals.

In overall this simulation suggests that 2-nd order Ambisonics produces directional cues relatively accurately at low frequencies, and fails to generate stable cues at high frequencies. There might be some differences between individuals. In recording of concert music it can be assumed that most of sound sources would be localized correctly, since typically instrument sounds contain frequencies below 1 kHz.

When compared to 1st order Ambisonics, it can be seen that ITD values are more correct, and the bump of ILDA values between 400 and 3kHz is missing. This proposes that directional quality is better with 2nd-order Ambisonics than with 1st-order Ambisonics.

5.2.2 A microphone technique for 5.1 reproduction

In recording techniques for 5.1 setup the microphones are often spaced considerably, that generates time differences between signals. The simulation of directional cues generated with this technique is problematic, since the auditory model used does not include the precedence effect. Many of proposed microphone techniques have microphones in farther distance than approximately 35cm from each other, that yields that the precedence effect should be also modeled. Some 5.1 microphone layouts are described in [28].

For this simulation a microphone setup was designed that has enough small distances between microphones, it is shown in Fig. 14 together with its simulation results. It has five cardioid microphones, two of them facing to directions $\pm 90°$ and one to $0°$, separated with 5 cm from the center point. The signals of these three microphones were applied to corresponding frontal loudspeakers. Two microphones were in $\pm 120°$ arrangement separated with 20 cm from the center. Their signals were applied with -6 dB gain reduction to surround loudspeakers.

Loudness plot shows that the microphone setup captures sound sources with fairly equal loudness from all directions. The -6 dB gain reduction was applied to rear loudspeakers, because otherwise the rear sound sources would have been reproduced prominently louder than frontal.

The ITDA behaves fairly consistently at low frequencies, however, the cues fluctuate more than with coincident techniques, and they are compressed roughly between $-30°$ and $30°$. High-frequency ITDA is fairly unstable. ILDA is generally unstable, especially at low frequencies. However, ILDA reaches relatively large values with sound source direction $150°$, which may generate virtual sources also to lateral directions. There are also large individual differences in cues. The simulation suggests that this technique does not produce sound source directions consistently. However, reproduction of reverberation may be satisfying since due to unstable ILDA cues it may be perceived to a large span of directions. It is not known how well this analysis describes 5.1 microphone techniques in overall. Further studies should be conducted on this.

6 CONCLUSIONS

In this study the directional qualities of different reproduction techniques were estimated using a binaural auditory model. The usability of the model was discussed, it was proposed that the model is valid for ITD analysis in general and ILD analysis with some restrictions.

The analysis results for different spatial sound recording systems for stereophonic listening verified the subjective opinions presented in literature. With coincident microphone techniques fairly stable and consistent virtual source can be produced, and with spaced microphones more spread and ambiguous virtual sources are achieved.

The recording techniques for multi-channel sound were found to be behind the technical status of microphone techniques for stereophonic listening when directional quality was taken into account. First-order Ambisonics produces fairly stable ITD

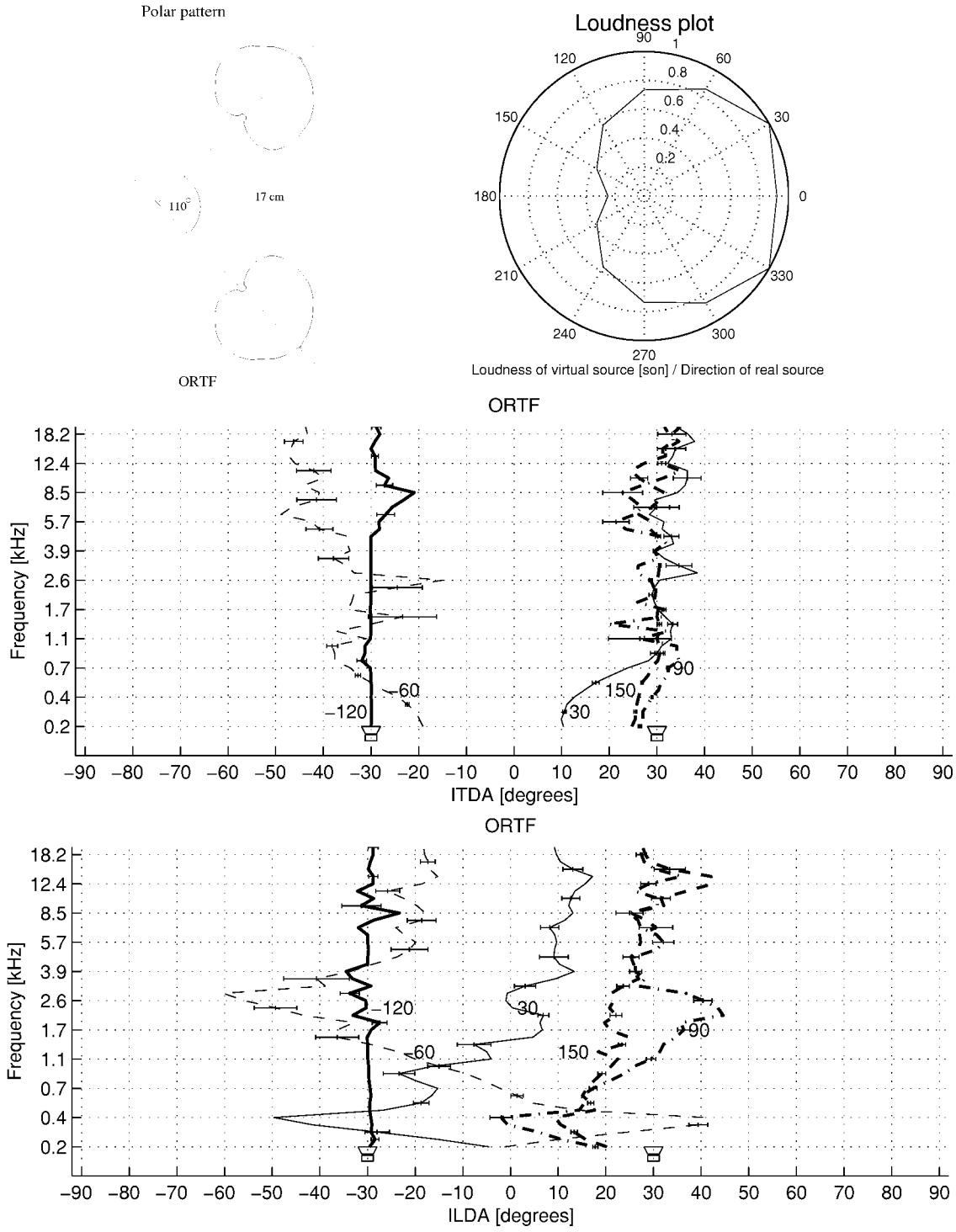

Fig. 11: Directional cues of ORTF recording technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

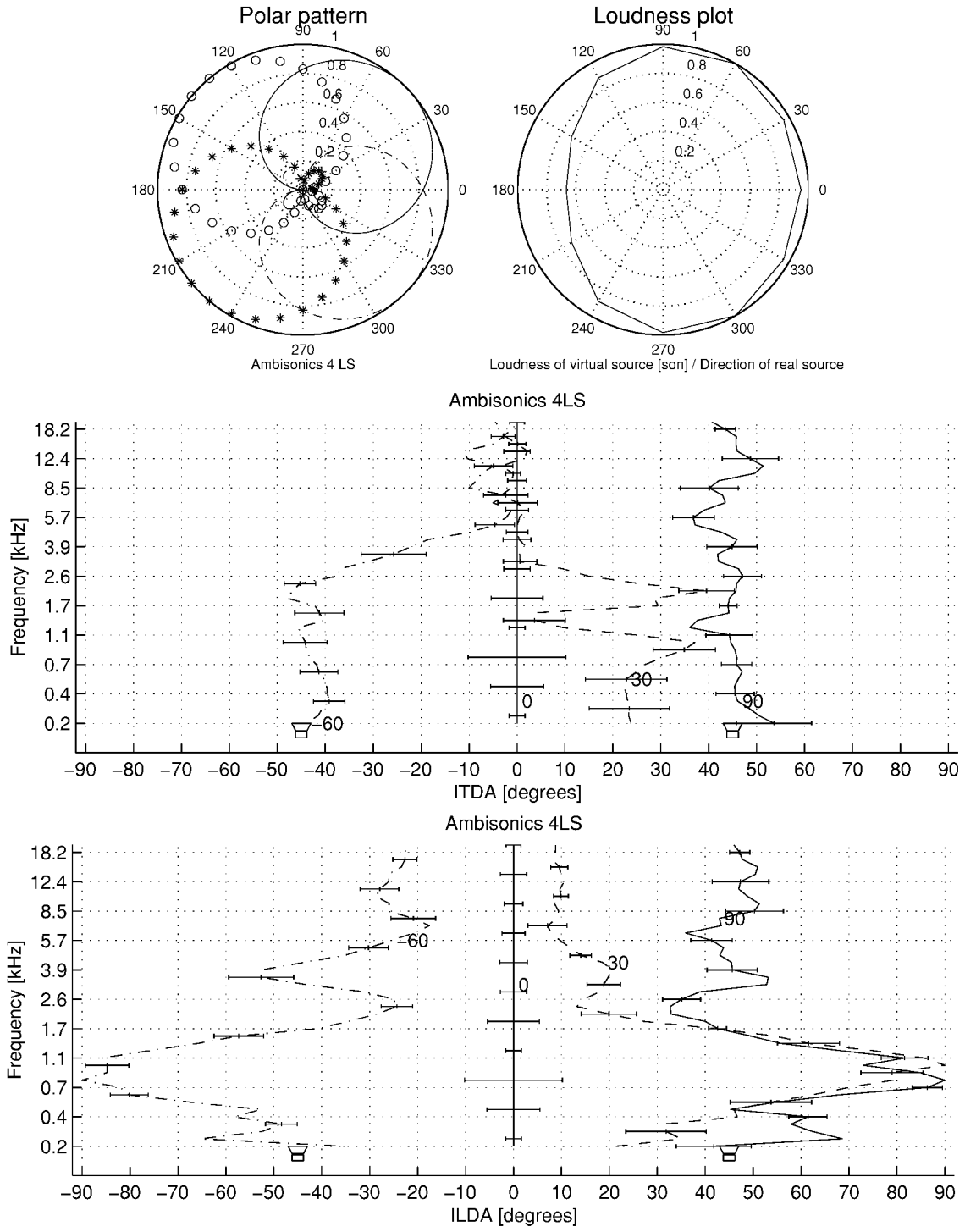

Fig. 12: Directional cues of Ambisonics four loudspeaker reproduction technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

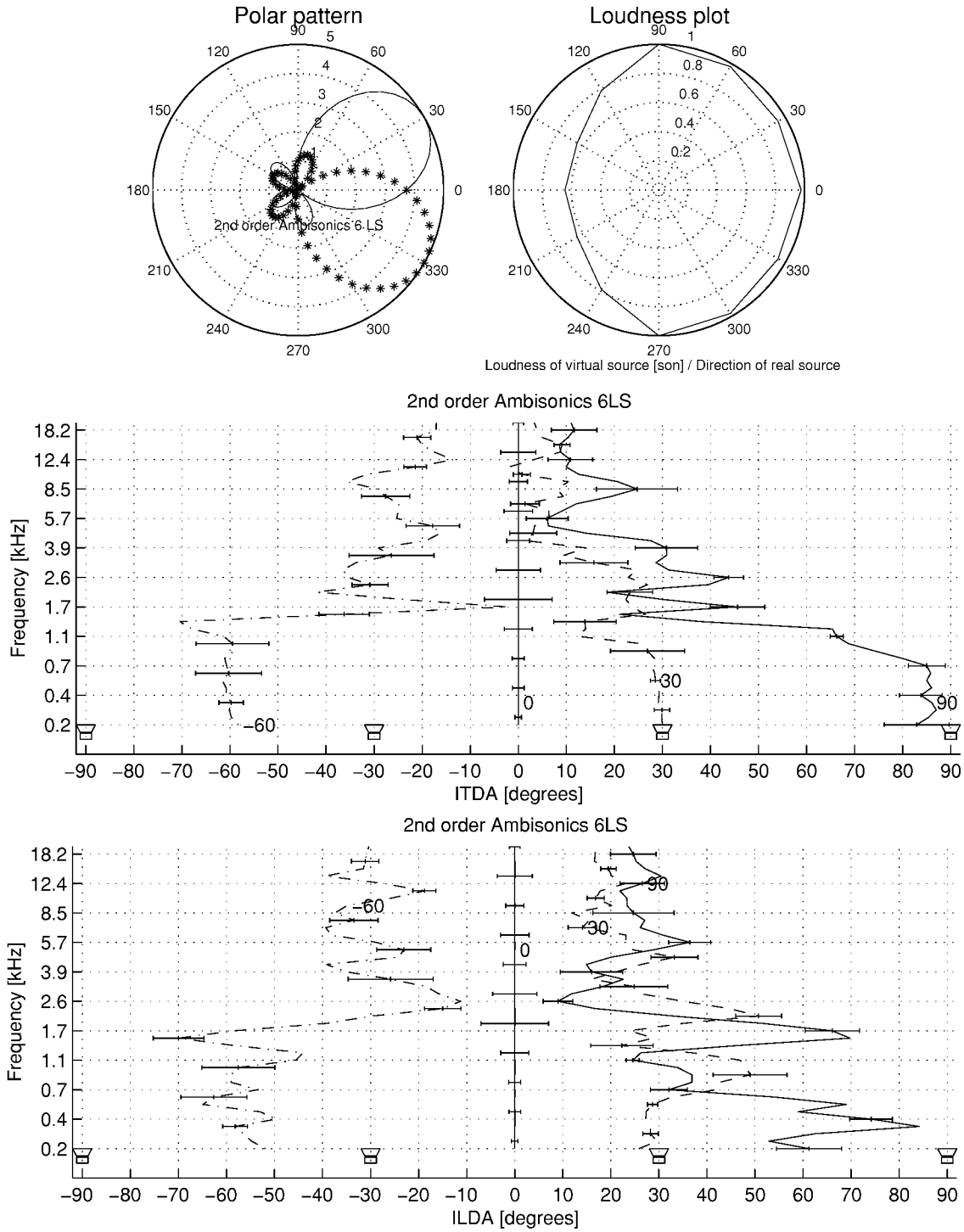

Fig. 13: Directional cues of Ambisonics six loudspeaker reproduction technique in standard stereophonic configuration. The whiskers denote 25% of standard deviation.

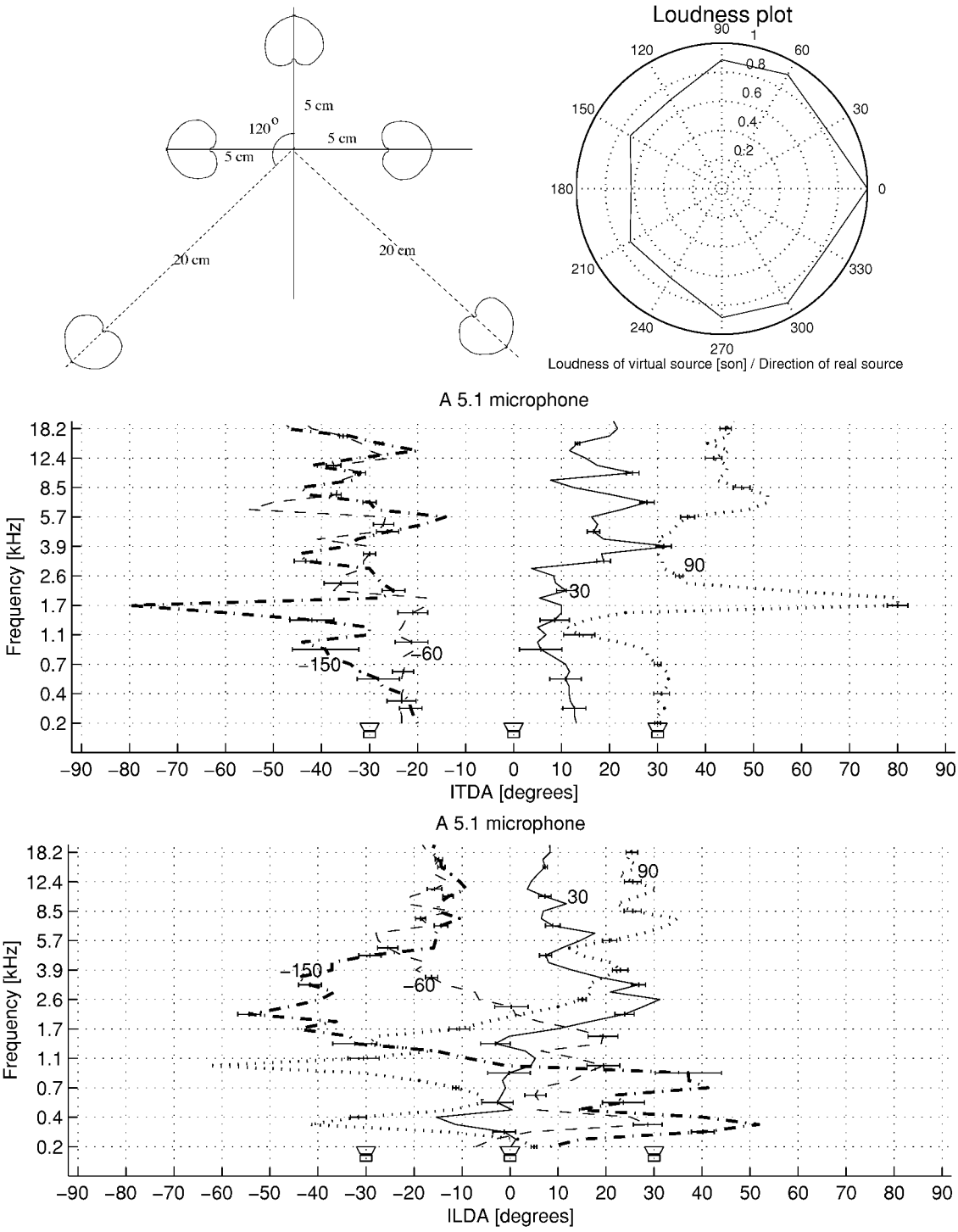

Fig. 14: Directional cues of star microphone recording technique in 5.1 configuration. The whiskers denote 25% of standard deviation.

cues at low frequencies, although biased towards the median plane. There are some unnatural behavior with ILD cues which may cause surrounding perceptions and inside-head locatedness. Second-order Ambisonics seems to be a promising microphone technique, low-frequency cues were stable and unbiased. However, it is not known if a microphone needed in it can be constructed. A sample microphone technique for 5.1 recording was analyzed. It was found that the directional cues generated were not consistent, however, the generated cues might be satisfying in reproduction of reverberation.

ACKNOWLEDGMENT

The work of Dr. Pulkki has been supported by the Graduate School in Electronics, Telecommunications and Automation (GETA) of the Academy of Finland.

REFERENCES

[1] A. D. Blumlein. U.K. Patent 394,325, 1931. Reprinted in Stereophonic Techniques, Audio Eng. Soc., NY, 1986.

[2] G. Steinke. Surround sound - the new phase. an overview. *Paper presented at the 100th AES Convention 1996 May 11-14 Copenhagen*, 1996.

[3] M. A. Gerzon. Periphony: With-height sound reproduction. *J. Audio Eng. Soc.*, 21(1):2–10, 1972.

[4] V. Pulkki. Localization of amplitude-panned virtual sources II: three-dimensional panning. *J. Audio Eng. Soc.*, 49(9):753–767, September 2001.

[5] J. Blauert. *Spatial Hearing, Revised edition.* The MIT Press, Cambridge, MA, USA, 1997.

[6] R. H. Gilkey and T. R. Anderson, editors. *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum Assoc., Mahwah, NJ, US, 1997.

[7] P. M. Zurek. The precedence effect. In W. A. Yost and G. Gourewitch, editors, *Directional Hearing*, pages 3–25. Springer-Verlag, 1987.

[8] F. L. Wightman and D. J. Kistler. Factors affecting the relative salience of sound localization cues. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum Assoc., Mahwah, NJ, YSA, 1997.

[9] D. H. Cooper and J. L. Bauck. Prospects for transaural recording. *J. Audio Eng. Soc.*, 37(1/2):3–39, January/February 1989.

[10] R. Streicher and W. Dooley. Basic stereo microphone perspectives – a review. *J. Audio Eng. Soc.*, 33(7):548–556, July/August 1985. Reprinted in Stereophonic Techniques, AES.

[11] S. P. Lipshitz. Stereophonic microphone techniques... are the purists wrong? *J. Audio Eng. Soc.*, 34(9):716–744, 1986.

[12] M. A. Gerzon. Panpot laws for multispeaker stereo. In *The 92nd Convention 1992 March 24-27 Vienna.* Audio Engineering Society, Preprint No. 3309, 1992.

[13] K. Farrar. Soundfield microphone. *Wireless World*, 85:99–102, 1979.

[14] D. G. Malham. Higher order ambisonic systems for the spatialisation of sound. In *Proc. Int. Computer Music Conf.*, pages 484–487, Beijing, China, 1999. ICMA.

[15] ITU-R Recommendation BS.775-1. Multichannel stereophonic sound system with and without accompanying picture. Technical report, International Telecommunication Union, Geneva, Switzerland, 1992-1994.

[16] A. J. Berkhout, D de Vries, and P. Vogel. Acoustic control by wave field synthesis. *J. Acoust. Soc. Am*, 93:2764–2778, 1993.

[17] V. Pulkki, M. Karjalainen, and J. Huopaniemi. Analyzing virtual sound source attributes using a binaural auditory model. *J. Audio Eng. Soc.*, 47(4):203–217, April 1999.

[18] A. Härmä and K. Palomäki. HUTear – a free Matlab toolbox for modeling of auditory system. In *Proc. Matlab DSP Conference*, pages 96–99, Espoo, Finland, November 1999. Comsol Ltd. http://www.acoustics.hut.fi/software/HUTear/.

[19] B. C. J. Moore. *An introduction to the psychology of hearing.* Academic Press, San Diego, fourth edition, 1997.

[20] R. Patterson, K. Robinson, J. Holdsworth, D. Mckeown, C. Zhang, and M. H. Allerhand. Complex sounds and auditory images. In L. Demany Y. Cazals and K. Horner, editors, *Auditory Physiology and Perception*, pages 429–446. Pergamon, Oxford, 1992.

[21] B. C. J. Moore, R. W. Peters, and B. R. Glasberg. Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am.*, 88(1):132–140, July 1990.

[22] T. C. T. Yin, P. X Joris, P. H. Smith, and J. C. K. Chan. Neuronal processing for coding interaural time disparities. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 399–425. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997.

[23] L. A. Jeffress. A place theory of sound localization. *J. Comp. Physiol. Psych.*, 61:468–486, 1948.

[24] R. M. Stern and C. Trahiotis. Models of binaural perception. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 499–531. Lawrence Erlbaum Assoc., Mahwah, NJ, YSA, 1997.

[25] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models.* Springer-Verlag, Heidelberg, Germany, 1990.

[26] B. C. J. Moore. A model for the prediction of thresholds, loudness, and partial loudness. *J. Audio Eng. Soc.*, 45(4):224–240, 1997.

[27] V. Pulkki and M. Karjalainen. Localization of amplitude-panned virtual sources I: Sterephonic panning. *J. Audio Eng. Soc.*, 49(9):739–752, September 2001.

[28] F. Rumsey. *Spatial Audio.* Focal Press, 2001.