

Directional loudness perception

the effect of sound incidence angle on loudness and the underlying binaural summation

Sivonen, Ville Pekka

Publication date:
2006

Document Version
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Sivonen, V. P. (2006). *Directional loudness perception: the effect of sound incidence angle on loudness and the underlying binaural summation*. Acoustics, Department of Electronic Systems, Aalborg University.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

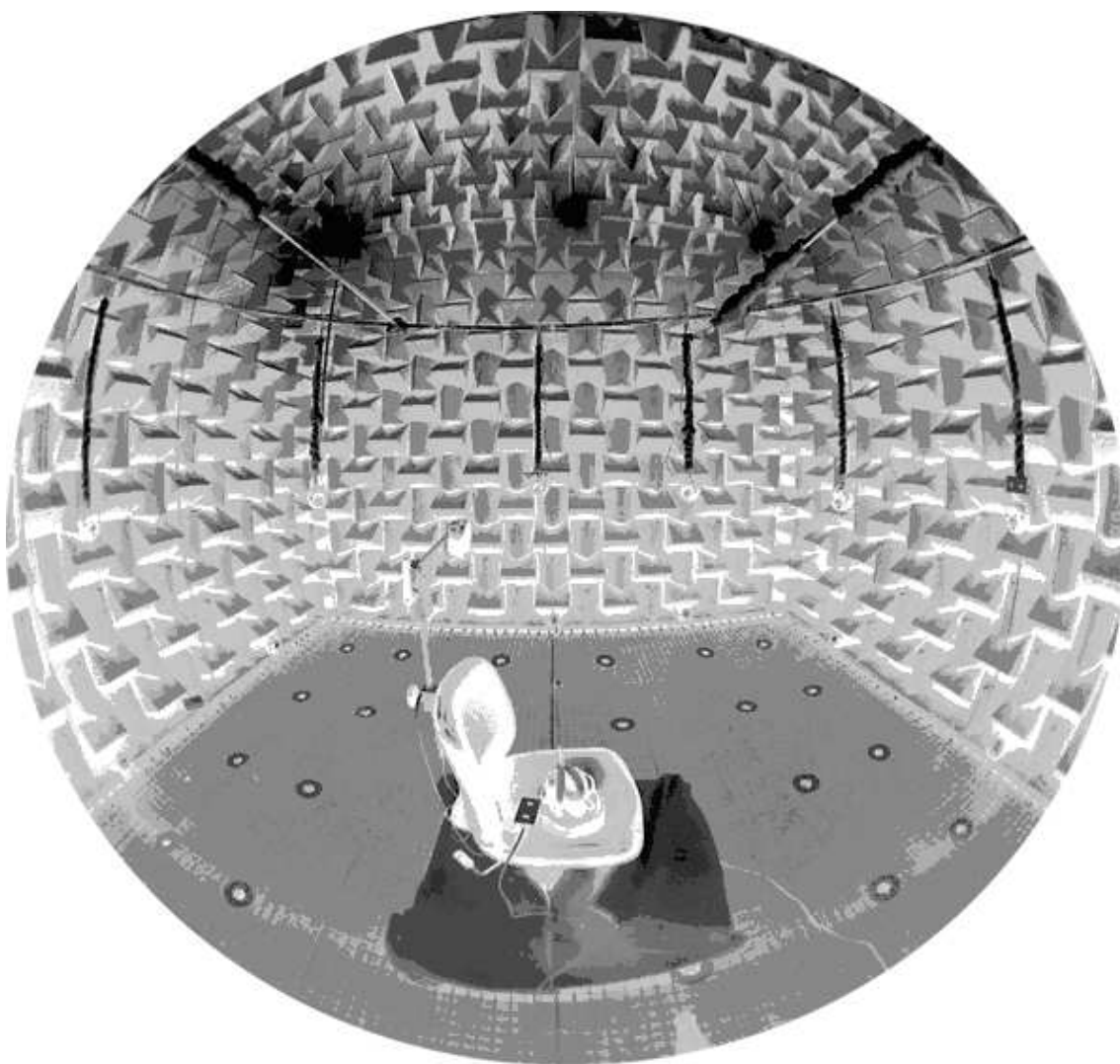
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Directional loudness perception

-the effect of sound incidence angle on loudness
and the underlying binaural summation



Ph.D. thesis by
Ville Pekka Sivonen

Preface

This thesis is submitted to the Faculty of Technology and Science at Aalborg University in partial fulfilment of the requirements for the Ph.D. degree. The work reported in this thesis was carried out between August 2002 and April 2006 in the Sound Quality Research Unit (SQRU) at the Department of Acoustics, Aalborg University. The work was funded by Brüel & Kjær Sound & Vibration Measurement A/S, according to the “Centercontract on Sound Quality” which establishes participation in and funding of the SQRU. The additional participating companies are Bang & Olufsen and DELTA Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

I would like to thank my supervisor Professor Wolfgang Ellermeier for his commitment as an advisor and for the contribution to the work of the thesis. During the process of the PhD study, his guidance in carrying out research, in psychoacoustics and the design and analysis of experiments, as well as in writing scientific papers, has been invaluable. I would also like to thank Pauli Minnaar for discussions concerning head-related transfer functions and binaural technology, and his comments on the manuscripts of the thesis. I am grateful to Finn Kryger Nielsen of Brüel & Kjær for employing me, for his faith in my abilities, and for the granted resources.

Thanks are due to the people of the Department of Acoustics, more specifically to Claus Vestergaard Skipper and Peter Dissing for their help in the laboratory; Florian Wickelmaier with respect to statistical analysis; Sylvain Choisel regarding computerized tools; Jody Ghani for her help in English and discussions both relevant and irrelevant to acoustics; Benjamin Pedersen for his help in translating the summary of the thesis to Danish and all other colleagues and friends for memorable times on and off working hours. I am further indebted to the participants in the listening tests.

I would like to thank the assessment committee, Frederic L. Wightman, Torben Poulsen and Dorte Hammershøi, for their valuable and constructive comments on an earlier version of the thesis.

Finally, I owe the greatest gratitude to my parents and family; my wife Iinu for her everlasting love and support, for her unselfishness and patience to put up with my absence due to long working hours; and to our children Venla and Leevi for all the joy, unconditional love and for their contagious ability to appreciate the simple things in life.

Ville Sivonen
August 2006
Aalborg, Denmark

Table of contents

Preface	1
Summary	4
Sammenfatning (Summary in Danish)	6
Introduction	8
1. Loudness	10
1.1 Listener variables	11
1.2 Measuring loudness	11
1.3 Stimulus variables	13
1.4 Models of loudness	15
2. Binaural loudness of sound fields	16
2.1 Free and diffuse fields	17
2.2 Spatial hearing and HRTFs	18
2.3 Binaural loudness summation	19
2.4 Directional loudness	20
3. Overview of the experimental work	21
3.1 Aim	21
3.2 Organization	22
3.3 Abstracts and interrelations of the manuscripts	22
3.4 General conclusions	24
3.5 Future work	25
References	26
Manuscript 1: “Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation”	30
Introduction	30
Method	32
Results	35
Modeling binaural loudness	39
Discussion	42
Manuscript 2: “Effect of direction on loudness in an individual binaural synthesis”	46
Introduction	46
Method	48
Results and discussion	51
Manuscript 3: “Laterality in binaural and directional loudness”	59
Introduction	59
Method	60
Results and discussion	63
Manuscript 4: “Binaural loudness summation for directional sounds”	72
Introduction	72
Method	73
Results and discussion	76

Binaural loudness for artificial-head measurements	80
Manuscript 5: “Directional loudness and the underlying binaural summation for wideband and reverberant sounds”	84
Introduction	84
Method	86
Results	89
Discussion	92
Appendix: Extended discussion of individual differences	100

Summary

Loudness, the perceived intensity of sound, is the psychological correlate of physical sound pressure level. Models for predicting loudness from acoustical measurements of sound pressure exist, and they conventionally utilize a monophonic signal obtained in the absence of a listener. However, for a thorough understanding of loudness perception in sound fields, the binaural nature of our exposure to sound must be considered. Understanding binaural loudness perception is important for many acoustical applications, from the determination of noise exposure as measured with artificial heads to the spatial quality of sound reproduction systems.

This PhD thesis investigates *directional loudness perception*, i.e., how the incidence angle, from which the sound reaches the listener's ears, affects its perceived loudness. This was achieved in a series of listening experiments, in which loudness matches were obtained between a frontal reference sound source and comparison sources at various locations. In addition to obtaining these directional loudness matches, individual at-ear exposures were determined in order to model the binaural summation underlying the matches. The aim of the modeling was to predict the directional loudness data based on the direction-dependent at-ear exposure, and to serve as a basis for developing a binaural loudness model, applicable to artificial-head measurements.

In the first experiment, directional loudness matches were obtained from listeners for narrow-band stimuli using real sound sources located inside an anechoic chamber. In addition to the loudness matches, the listeners' individual head-related transfer functions (HRTFs) were measured on three different occasions, and then utilized in modeling the binaural loudness summation of their individual at-ear exposures. Even though the effect of the HRTFs were accounted for, individual differences in the manner, in which the listeners integrate the left- and right-ear inputs to a single binaural percept, were still evident.

Therefore, the measured HRTFs were utilized in the second experiment for generating virtual sound sources using individual binaural synthesis, and for obtaining directional loudness matches in that virtual sound field. Despite the use of binaural synthesis, where the sets of HRTFs used for stimulation were precisely known, essentially the same directional matches as in the first experiment were obtained. This, on the one hand, corroborated the validity of the HRTFs obtained for analyzing loudness as a function of incidence angle, while on the other hand, it still left the individual peculiarities in the binaural summation unexplained.

In the first two experiments, directional loudness was investigated both for the horizontal and the median planes. In the horizontal plane, sound sources were, however, only located in the left hemisphere, due to assumed symmetry. Thus, in order to investigate possible laterality effects in loudness judgments, the third experiment investigated the effect of the side of stimulation on loudness matches with the frontal reference. For most of the listeners, the effect of side was insignificant, while in some cases, a slight left-ear advantage was observed.

In the fourth experiment, generic, i.e., artificial-head-based HRTFs were utilized in obtaining directional loudness matches, in an attempt to give each listener the same directional changes in at-ear exposure. Despite essentially the same stimulation for all listeners, the individual

differences in the underlying binaural summation were still evident, the summation factors showing considerable similarities with those observed utilizing individual HRTFs.

Finally, to improve the ecological validity of the stimulation, in the fifth experiment the effect of direction was investigated for wideband and reverberant sounds. The directional effect on loudness was significant also for these stimuli resembling more the sounds our ears encounter in real-life, than the anechoic narrow-band stimuli utilized in the first four experiments.

The mean data from all experiments show evidence for a binaural 'power summation', according to which the maximum binaural gain between monaural and binaural stimulation is 3 dB. Based on this finding, a binaural loudness model, applicable to artificial-head measurements, is presented.

Sammenfatning (Summary in Danish)

Loudness, den opfattede lydstyrke, er den psykologiske parallel til fysisk lydtryk. Modeller til bestemmelse af loudness på baggrund af akustiske målinger af lydtryk eksisterer og benytter konventionelt et monofonisk signal fra én mikrofon i lytterens position. For en dybere forståelse af opfattelsen af loudness i lydfelter må den binaurale natur af vores lydopfattelse tages i betragtning. Forståelse af binaural loudness er vigtig for mange akustiske områder; fra bestemmelse af støjpåvirkning, som målt med kunsthoveder, til den rumlige kvalitet af systemer til reproduktion af lyd.

Denne Ph.D.-afhandling undersøger retningsbestemt loudness-opfattelse, dvs. hvordan lydens indfaldsvinkel i forhold til ørene indvirker på den opfattede loudness. Dette er opnået gennem en række lytteforsøg, i hvilke loudness af lydkilder i forskellige retninger blev matchet til loudness af lyden fra en frontal referencekilde. Udover disse retningsbestemte loudness-matches blev individuelle lydtryk ved øret bestemt for at kunne modellere den binaurale summering, som ligger bag de målte loudness-matches. Målet for modelleringen var at forudsige de retningsbestemte loudness data baseret på det retningsafhængige lydtryk ved øret, og at danne grundlag for udvikling af en binaural loudness-model, som kan bruges med målinger fra et kunsthoved.

I det første forsøg blev lytteres retningsbestemte loudness-matches målt for smalbandsstimuli fra virkelige lydkilder i et lyddødt rum. Udover loudness-matches blev lytternes individuelle "head-related transfer functions" (HRTFs) målt tre gange og derefter brugt til modellering af binaural loudness-summering på baggrund af de individuelle lydtryk ved ørene. Selvom der blev taget hensyn til HRTFs, var der stadig tydelig forskel på, hvordan lyttere integrerer input fra venstre og højre øre til én enkel binaural opfattet enhed.

Derfor blev de målte HRTFs brugt i et efterfølgende forsøg til at generere virtuelle lydkilder ved individuel binaural syntese og til at måle retningsbestemte loudness-matches i det virtuelle miljø. Til trods for brugen af binaural syntese, hvor HRTFs for syntesen var kendt helt præcist, var de retningsbestemte matches så godt som identiske med resultaterne fra det første forsøg. På den ene side bekræfter dette validiteten af de målte HRTFs til bestemmelsen af loudness som funktion af indfaldsvinkel, men efterlader på den anden side særegne forskelle i loudness-summeringen uforklaret.

I de første to forsøg blev retningsbestemt loudness undersøgt i både det horisontale og vertikale midterplan. I det horisontale plan var lydkilder imidlertid kun placeret i den venstre hemisfære på grund af antaget symmetri. For at undersøge mulige laterale effekter i loudness-bedømmelser blev det i det tredje forsøg undersøgt, hvordan side for simuleringen påvirker loudness-matches til den frontale reference. For de fleste lyttere var effekten af side ubetydelig, men i enkelte tilfælde blev der observeret en svag tendens til, at lyttere lagde relativt højere vægt på lyden i venstre øre.

I det fjerde forsøg blev der brugt generelle, dvs. kunsthovedbaserede, HRTFs til at måle retningsbestemte loudness-matches i et forsøg på at give hver lytter de samme retningsbestemte ændringer i lydtryk ved øret. Trods den tilnærmelsesvis samme stimulering for alle lyttere var forskellene i den underliggende binaurale summering stadig tydelige, og summeringsfaktorerne havde tydelige lighedspunkter med dem, som blev målt med individuelle HRTFs.

For at gøre stimuleringen mere virkelighedstro blev retningseffekten for loudness for bredbåndssignaler og signaler med refleksion undersøgt i et femte forsøg. Retningens betydning for loudness for disse lyde, som minder mere om de lyde, vores øre møder i dagligdagen end smalbåndssignalerne i et lyddødt rum, som blev brugt i de første fire forsøg, var også signifikant.

Middeldata fra alle forsøg tyder på binaural “effekt summering”, ifølge hvilken den maksimale “binaurale fordel” mellem monaural og binaural stimulering er 3 dB. På grundlag af dette bliver der præsenteret en binaural loudness-model, som kan bruges med kunsthovedmålinger.

Introduction

The perceived intensity of sensations is a salient attribute in psychophysics, be it vision, hearing, touch, taste or smell. In the sense of vision, the perceived intensity is the brightness of visual objects, while in hearing, it is the *loudness* of sounds.

This PhD thesis investigates a particular aspect of loudness perception; namely, the effect of the spatial location, from which a sound reaches the listener, on loudness. When considered from the listener's point of view, a sound source located in space is at a given *distance* and in a given *direction*.

Intuitively, loudness is bound to be affected by the physical distance between the source and the listener. A person speaking in the open air at a great distance is inaudible, if the sound waves propagating in the air are attenuated below the threshold of hearing, when reaching the listener. When the distance is decreased, the speaker becomes audible at some point. Speaking with the same vocal effort directly to the listener's ear, the speaker is probably perceived louder than when speaking from a distance.

The above reasoning relies heavily on the *physics* of the signals reaching the observer. It does not take the *cognitive* processing into account, that is, how the brain makes use of the signals registered by the sensory receptors. The listener might judge the vocal effort of the speaker, i.e., the sound power the speaker is radiating into the air, to remain constant irrespective of distance. Such a *loudness constancy* with varying distance has been reported, despite profound changes in at-ear stimulation (Zahorik and Wightman, 2001). An analogy for loudness constancy can be found in visual perception, where the size, e.g., the height of a person, is judged constant, irrespective of the distance between the person and the viewer, and of the changes in the size of the person's image on the viewer's retina (for a review, see e.g., Goldstein, 2004).

This thesis will focus on the effect of *direction* on loudness, while keeping the physical distance between the source and the listener constant. Investigating *directional loudness perception*, the thesis aims to answer the following questions:

- (1) Does perceived loudness depend on the incidence angle of the sound, or is loudness constant as function of direction, as might be observed as a function of distance?
- (2) If there is a quantifiable effect of direction on loudness, can it be explained by inspecting directional dependencies of the physical stimulation reaching the listener's ears?
- (3) How are the two signals at the listener's ears integrated to yield a single percept of loudness for directional sounds?

Before going into details, it may be useful to look at the transduction of pressure waves in the air to nerve signals in the brain, and ultimately to an auditory percept, such as loudness. A schematic of the human auditory system, illustrating this transduction is shown in Fig. 1. Acoustical pressure waves in the air are picked up by the eardrums at the end of each ear canal. The transmission of sound from an unobstructed field to the eardrums is largely affected by

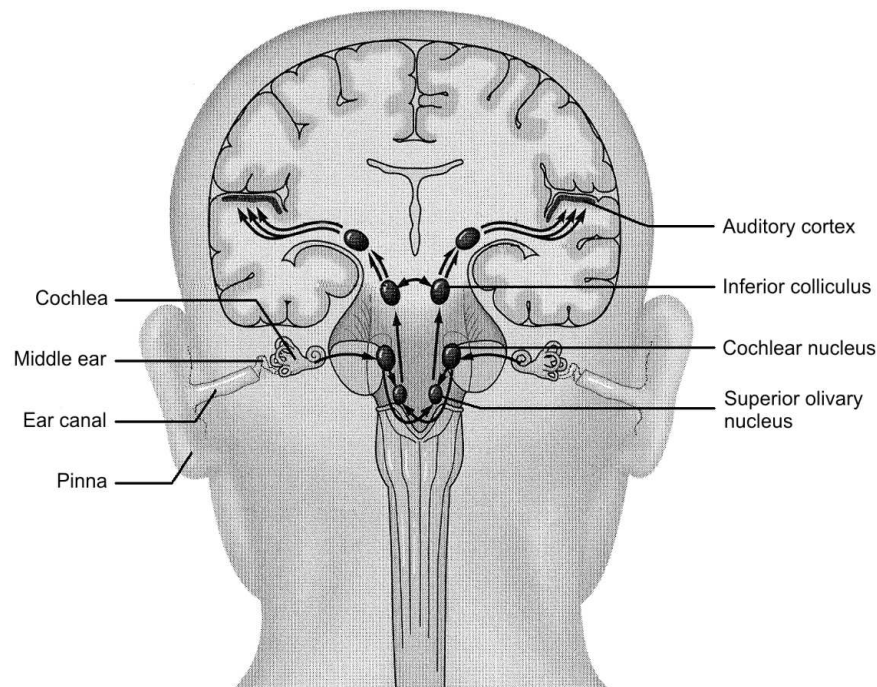


Figure 1: The human head, the outer ear (pinna and ear canal), the middle ear and the inner ear (cochlea), and a schematic of the central auditory pathways, comprising of the cochlear and superior olivary nuclei, the inferior colliculi and the auditory cortices. The figure is adapted from Goldstein (2004), reprinted with permission.

the size and shape of the torso, head, pinnae and the ear canals. The pressure variations cause the eardrums to vibrate, and the bones of the middle ear transmit the vibrations to the inner ear, matching the impedance between the surrounding air and the liquid-filled cochlea. The main function of the cochlea is to transduce the mechanical vibration to electrical signals in the auditory nerve fibers. The auditory nerve fibers then synapse in the cochlear nuclei.

Our auditory system is binaural, i.e., it consists of two ears. The first stage where nerve signals are combined from the two ears is in the superior olivary complex, see the contralateral connections in the superior olivary nuclei in Fig. 1. The two sides are connected also later in the auditory pathway, e.g., via the inferior colliculi and in the brain utilizing the information at both auditory cortices.

This thesis will look into both the *physical* effects of the transmission of sound from various directions to the listener's two ears (strongly affected by the torso, head and pinnae depicted in Fig. 1), and the *psychophysical* effects of listeners' loudness judgments as a function of direction. Physiological aspects concerning where binaural loudness is processed in the auditory pathway are outside of the scope of the thesis.

In section 1, the concept of loudness is introduced. Section 2 deals with binaural and spatial effects on loudness. The experimental work, the general conclusions, and possibilities for future work are summarized in section 3. The experiments carried out are presented in detail in five manuscripts constituting the main body of the thesis, and an extended discussion is appended in the end of the thesis.

1 Loudness

Loudness is the subjective intensity of sound, an attribute which characterizes sounds ranging from quiet to loud. Although sounds are sometimes described as having a certain loudness, loudness is not an intrinsic, physical property of a sound. Rather, loudness is psychoacoustical, depending both on the sound and on the listener.

As a psychoacoustic attribute, loudness is undoubtedly prominent: (1) The concept of loudness is intuitive. Everyday sounds can be judged in terms of their loudness without specific training, and listeners (even preschool children, Werner and Rubel, 1992) can do this fairly reliably. (2) A great amount of research has been accumulated, investigating different aspects of loudness (for summaries, see Scharf, 1978; Zwicker and Fastl, 1999; Moore, 2003). (3) Loudness is the basis of metrics developed for sound-quality evaluation, such as roughness and sharpness (Zwicker and Fastl, 1999). (4) Loudness strongly correlates with other psychoacoustic attributes, such as the perceived quality of sound-reproduction systems (Gabrielsson and Sjögren, 1979; Illényi and Korpassy, 1981), or annoyance caused by noise (Hellman, 1985; Berglund et al., 1990).

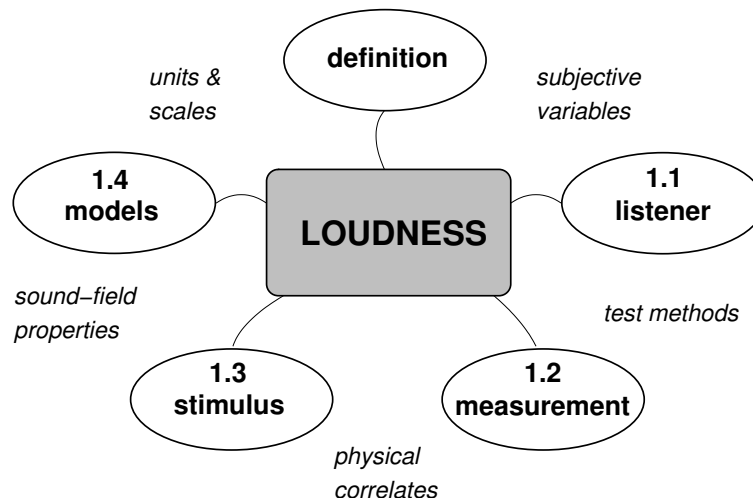


Figure 2: Aspects of the loudness concept.

The concept of loudness may be looked at from several angles which are visualized in Fig. 2, and which will be elaborated in turn in the sections below. Being subjective by definition, loudness resides in the listener (section 1.1). Therefore, it can only be measured (section 1.2) via the listener's response to sounds. The physics of the sound stimulus (section 1.3) can be altered in an attempt to derive functional relationships, and eventually a model (section 1.4), relating the physical properties of a sound to the psychological loudness percept the sound evokes. The categorization of loudness proposed in Fig. 2 is not exhaustive; it will merely be used here for giving an overview on the issue.

1.1 Listener variables

Like any percept (be it auditory or from another modality), loudness is subjective and it only exists "inside the listener's head". Therefore, loudness depends on the individual listener; for example, on his or her linguistic concepts or interpretation of the verbal instructions given by the experimenter. Correlates for individual loudness judgments have been found in personality differences (Reason, 1972), or peculiarities in reporting the subjective intensity of stimuli (Collins and Gescheider, 1989). Reasons for these *cognitive* individual differences may be complex. They might also – to some extent – elude systematic study and thus be outside the control of the experimenter.

Perhaps more systematic individual loudness differences can be observed by looking at the auditory system (as illustrated in Fig. 1). In the transduction of the pressure waves in the air into nerve signals in the brain, individual differences affecting loudness can emerge at various stages: (1) Our anatomical differences result in idiosyncracies in the transfer of sound to mechanical vibration of the eardrums at the end of each ear canal. Thus, the acoustic inputs to the auditory system may be very individual. (2) Physiological variations of the middle or the inner ear affect how acoustical energy is transduced into neural energy. An impairment of the ear, be it conductive in the outer or the middle ear, or sensorineural in the inner ear, greatly affects this transduction and hence, loudness (see e.g., loudness recruitment, Fowler, 1950). (3) After an intense exposure to noise, the ear may also be fatigued, resulting in a temporary shift in hearing threshold and in perceived loudness (McPherson and Anderson, 1971). (4) Finally, since the brain makes use of the nerve signals for judging loudness, any variations in the auditory pathway may have an effect on loudness.

Despite the individual nature of loudness, a number of *generalizations* across listeners can be made, e.g., loudness is closely related to its physical parallel, the sound pressure level (Fletcher and Munson, 1933; Stevens, 1955), as it is to the rate and number of firing neurons in the auditory pathway (for a review, see Scharf, 1978). These general trends allow for fairly consistent measurements and reasonably successful modeling of loudness. It may be stated that of all possible subjective attributes of sound, loudness very likely is the one most accessible to investigation.

1.2 Measuring loudness

Loudness must be measured via the listener's response to sounds. There are several psychoacoustical methods for collecting responses from listeners. The two most commonly used methods for measuring loudness are *scaling* and *matching*.

1.2.1 Loudness scaling

In loudness scaling, numbers corresponding to the perceived intensity of sound are obtained from listeners, for example, by means of magnitude estimation, magnitude production or category rating. In magnitude estimation the listener assigns numbers to sounds that correspond

to their perceived loudness. In magnitude production, the approach is the opposite: the listener must adjust the physical magnitude of the sound to correspond to the numerical values presented. In category rating¹, the listener is provided with a bounded range of numbers or a series of adjectives, and values from these ranges are assigned to the sounds to be judged.

Scaling claims to provide a direct measure of loudness, and based on the work by Stevens (1955), a psychophysical scale of loudness has been defined: The unit of loudness (N) is the *stone*. One stone is arbitrarily chosen to correspond to a loudness level (L_N ; ISO 226, 2003) of 40 phons, being equal to sound pressure level (in dB SPL) at 1 kHz. Furthermore, an increase of 10 phons corresponds to a twofold change in stones, two stones corresponding to a loudness level of 50 phons and four stones to 60 phons etc. Scaling thus also provides a measure of the *slope* of the loudness function.

Although being straightforward and fairly generalizable, loudness scaling can be problematic: (1) Numbers assigned may not warrant inferences on sensation magnitude. A "true" loudness scale may not be obtained by taking numerical judgments at face value, i.e., the validity of the scale may not be justified (Narens, 1996; Ellermeier and Faulhammer, 2000). (2) There may also be numerical response biases, i.e., the listeners having various ways of using numbers. The apparent scale then confounds sensory with judgmental effects (Irwin and Whitehead, 1991), though sometimes biases can be corrected for, e.g., by specific "calibration" tasks (Collins and Gescheider, 1989). (3) Additional problems may be caused by context effects (Marks, 1992), e.g., adapting the response scale to the range of stimuli, or by sequential contrast and assimilation effects (DeCarlo and Cross, 1990), caused by the presentation order of stimuli.

1.2.2 Loudness matching

In loudness matching, a physical property of the sound (its level, duration, or frequency) is varied so that the listener perceives it as equally loud as a standard sound. Loudness matches can be obtained via the classical method of adjustment or using modern adaptive procedures. In the method of adjustment, the listener controls the physical property that is varied, until the stimuli are matched in loudness. Thus, the listener may have unwanted control over the outcome. This can be overcome using adaptive forced-choice procedures, where the experimental procedure has the control over the physical property.

The method of matching is based on the assumption that listeners can compare two sound with respect to loudness, even though they are different in some other dimension: the sounds may differ so much qualitatively that equating them in loudness is very difficult. Furthermore, matching the loudness of two sounds does not provide the experimenter information on how loud the sounds are actually perceived; it only reveals what changes in some physical dimension are required for equal-loudness perception. By comparing matches across experimental conditions, though, sometimes inferences about the growth of loudness can be made, like when interpreting the distances between equal-loudness contours at different frequencies. As loudness scaling, the method of matching is prone to biases due to the listener's inter-

¹Category rating has its origin in the indication of dynamics in musical notation. The two basic indications are *p* for *piano* and *f* for *forte*, denoting softly and loudly, respectively. Indicating more extreme degrees, the basic indications can be combined, i.e., yielding a dynamic range from *ppp* to *fff*.

pretation of the instructions: when asked to match in loudness, listeners could match sounds with respect to some other salient attribute.

Loudness matching, however, is a subject of much less controversy than scaling. Loudness matching does not require the use of numbers, and therefore, avoids the problems (1) of having to infer a latent sensory scale and (2) of having to correct for numerical response biases, as discussed in the section on scaling. Furthermore, by using modern adaptive procedures (Jesteadt, 1980), a match is indirectly derived from a series of paired comparisons where the listener only has to decide which of the two sounds was louder. For these procedures, sophisticated control mechanisms (such as interleaving experimental conditions and randomizing orders) have been devised (e.g., Buus et al., 1997) to deal with the last kind of problems – context and sequential effects – identified in the section on scaling.

To sum up: It appears that results from loudness matching map out equivalences that very closely reflect the sensory processing of sounds, and that the – typically adaptive, forced-choice – methods used to obtain matches are much less fraught with questionable assumptions than are alternative ways to get at loudness with psychophysical means.

1.2.3 Other methods

Correlates of loudness have also been found by using methods other than scaling or matching. The reaction time to a sound has shown a lawful relationship with loudness: Reaction times to equally loud sounds are approximately equal (see Scharf, 1978), and, the louder the sound, the shorter the reaction time (e.g., Kohfeld et al., 1981). Obviously, measuring reaction time is less biased by the listener’s response criteria and the experimental instructions than are loudness scaling or matching. However, nonauditory factors, such as the lower limit of human motor responses of about 100 ms, influence reaction time and possibly obscure the precise relation with loudness.

In case the listener is unable to respond behaviorally, the acoustic reflex of the middle ear or auditory evoked potentials in the brain can be measured. The mapping of these correlates, and of reaction time, to loudness perception is, however, incomplete. Also, they do not measure perceived loudness, rather, the correspondence can only be interpreted via loudness obtained by scaling.

1.3 Stimulus variables

There is a large body of research altering the physical properties of sound stimuli, in an attempt to derive how they affect loudness. The most commonly investigated variables are the sound pressure level², and the spectral and temporal characteristics of stimuli.

For a given sound, loudness is a monotonic function of **sound pressure level**, that is, the higher the sound pressure level, the greater the perceived loudness. The shape of the function

²The effect of sound pressure level is often laxly labeled as an effect of stimulus intensity on loudness. Here, the term *level* will be used, in order to avoid confusion with the vector quantity, *acoustic intensity*: $\mathbf{I} = \mathbf{p} \times \mathbf{v}$ (pressure \times particle velocity).

is based on direct scaling procedures (Stevens, 1955). The standard loudness function for a 1-kHz pure tone, complying with the definition of the sone scale, is plotted in Fig. 3a. The function is a simple power law: $N = 0.01p^{0.6}$, where N is loudness in sones and p is the (linear) sound pressure in μPa (see the solid curve in Fig. 3a). The growth of loudness is steeper between the hearing threshold and about 30 phons, and this can be approximated by a modification of the power law: $N = 0.01(p - p_0)^{0.6}$, where p_0 is the effective threshold in μPa (see the dashed curve in Fig. 3a).

The exponent (0.6) determines the slope of the loudness function, i.e., the greater the exponent, the steeper the function. The value 0.6 is based on data averaged across listeners, although from one person to another, variation in the size of the exponent has been reported (0.4 – 1.1, see Stevens and Guirao, 1964). The slope of the function depends also on the stimulus: For pure tones below 1 kHz, the growth of loudness is steeper, and for white noise shallower, than for the standard function at 1 kHz (see Figs. 5 and 7 in Scharf, 1978, respectively).

In addition to these level effects, loudness is dependent on the **spectral characteristics** of sounds. By varying the sound pressure level of a pure tone at a given frequency, until it is perceived equally loud as a 1-kHz reference tone at a fixed level, *equal-loudness-level* contours have been derived (Fletcher and Munson, 1933; Robinson and Dadson, 1956; ISO 226, 2003). Examples of these contours are plotted for various loudness levels (20, 40, 60 and 80 phons) in Fig. 3b, the contours showing considerable frequency dependencies: at 20 Hz, 100 dB SPL is required for a loudness match with the 1-kHz reference at 40 dB SPL, see the 40-phon contour in Fig. 3b. Also, the contours are level dependent, as they become flatter with increasing loudness level.

Equal-loudness-level contours obtained by matching depict that the growth of loudness as a function of sound pressure level is faster at low frequencies, than at the 1-kHz reference frequency. From hearing threshold to the 80-phon contour, a change of 40 dB is required at 20 Hz, whereas at 1-kHz, the corresponding change needed is almost 80 dB. These results agree well with loudness functions for sounds at low frequencies obtained by scaling (Hellman and Zwislöcki, 1968), suggesting no fundamental conflict between the two measurement methods. Furthermore, the shape of the hearing-threshold curve (the dashed line in Fig. 3b) obtained in a *detection task* is in good agreement with that of the equal-loudness-level contours obtained via loudness matching.

Another spectral effect is the summation of loudness across critical bands (Zwicker and Fastl, 1999). Keeping the sound energy constant, loudness is unaffected by the bandwidth of a sound below the so-called critical bandwidth. Above this limit, loudness grows with increasing bandwidth. The spectral summation of loudness is also somewhat level dependent, being greatest at moderate levels.

Finally, the **temporal characteristics** of sounds have an effect on loudness (Zwicker and Fastl, 1999). For brief sounds up to 100-200 ms, loudness increases with duration, i.e., for a short sound a higher sound pressure level is required for a loudness match with a longer sound. On the contrary, for high-frequency sounds well beyond 1 s, loudness adaptation (a decline of loudness) has been reported (Miśkiewicz et al., 1993). The time constant by which the ear integrates energy is under debate, due to conflicting experimental results (for a review, see

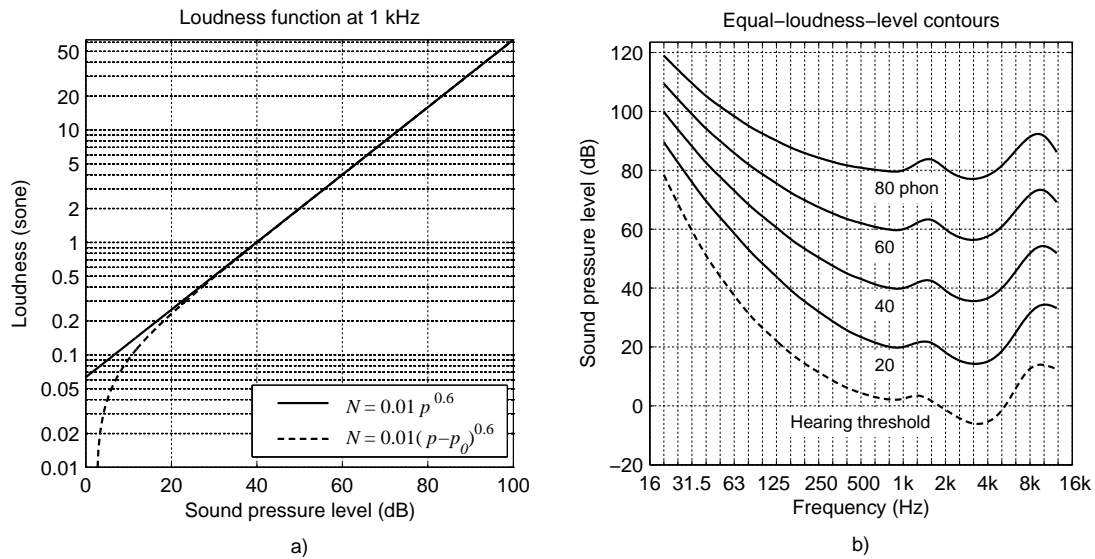


Figure 3: a) The standard loudness function at 1 kHz, plotted as a simple power law (solid line) and corrected for the steeper growth at low sound pressure levels (dashed line). b) Equal-loudness-level contours at 20, 40, 60 and 80 phons, and the hearing threshold for binaural listening in the free field, after ISO 226 (2003).

Scharf, 1978). Alternative explanations to the simple energy- integration for accounting for the temporal effects have been proposed (Viemeister and Wakefield, 1991; Buus et al., 1997).

In addition, the loudness of a pure tone is affected when partially (spectrally) masked by a wideband noise. The masking noise raises the threshold for the tone and makes it appear softer than when the tone is heard in quiet. The partial masking can also be temporal by a masker preceding (forward masking) or succeeding (backward masking) a sound. The loudness function of a partially-masked sound becomes steeper than for the same sound heard alone, without the masker.

1.4 Models of loudness

Modeling auditory perception can be physiological or psychoacoustical. A physiological model is concerned with the different stages of processing in the auditory pathways, determining the relationship between the stimulus and the electrical signals in the auditory pathways, and ultimately their link to perception. A psychoacoustical model analyzes the direct relationship between the stimulus and perception, treating the biological 'hardware' as a black box.

Since collecting subjective responses from listeners to measure the loudness of sounds is impractical for everyday applications, models for *predicting* loudness from the acoustical input have been developed (Zwicker, 1960; Stevens, 1961; Moore et al., 1997), based on the vast amount of psychoacoustical research. The models attempt to compute numerical estimates of loudness based on the acoustical, objective characteristics of sound, as perceived by an average, normal-hearing listener.

The basic stages of all current loudness models are illustrated in Fig. 4. The two left-most blocks account for the transduction by the outer and middle ear. The three right-most blocks include the following: (1) the spectral characteristics of the sound are accounted for by converting to an excitation pattern, which is based the sound’s masking pattern over a wide frequency range, (2) the excitation pattern is then converted to specific loudness, i.e., the loudness per critical (or equivalent rectangular) band, and (3) finally, the overall loudness is computed. Overall loudness is proportional to the area under the specific loudness patterns.

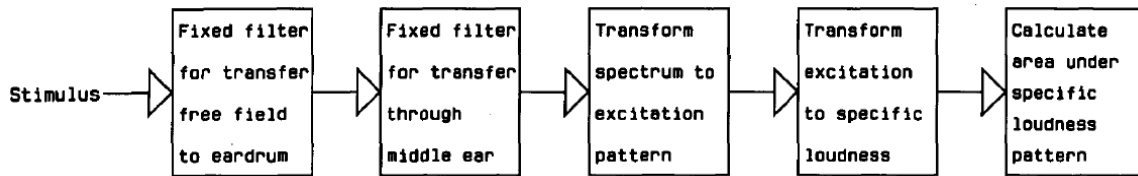


Figure 4: Main (monaural) stages of models used for loudness calculations, taken from Moore et al. (1997).

The above models are based on the long-term spectrum of the sound, and thus are limited to stationary sounds lasting longer than about 200 ms (i.e., after loudness no longer increases with duration). While the level and the spectral effects are accounted for, the temporal characteristics are disregarded by these models, due to the lack of consensus in the research on time constants of the ear. Recently, however, ways of processing time-varying signals for loudness computations have been suggested (see e.g., Zwicker and Fastl, 1999; Glasberg and Moore, 2002).

2 Binaural loudness of sound fields

The loudness models presented above (see Fig. 4) are largely concerned with converting the spectrum of stimuli to a loudness prediction. This section deals with loudness and the properties of sound fields, namely, how sound is transmitted to stimulation at the listener’s ears and what implications the transmission has on modeling loudness.

Sound fields encountered in real life are complex: The source can be a single unit radiating sound from a distinct position in space, it can be a source distributed spatially or consist of multiple sources at different positions, or be virtually of any degree of complexity covering a variety of radiation patterns. The signals from the source propagate in space as sound waves, reflecting from possible obstacles (such as walls), and finally reach the eardrums of the listener. From the listener’s point of view, the direct sound from the source arrives first, then early reflections being reflected once or a small number of times, and at last, the reverberant energy of the space. These time signals are superimposed at the listener’s ears. While the direct sound and early reflections reach the listener from distinct directions, reverberation is typically fairly *diffuse*, i.e., not concentrated to any specific direction.

The acoustical waves in a sound field reach the *both ears* of a listener, the listening being *binaural*. Even in the simplest sound fields, the two at-ear signals typically differ from another.

In auditory experiments utilizing headphones, a stimulation with different signals sent to each earphone (and reaching each ear) is called *dichotic*. A binaural stimulation with the same signal sent to both earphones is called *diotic*. Furthermore, a stimulation of only one ear is called *monaural* or *monotic*. Disregarding the crosstalk from an earphone to the contralateral ear, monaural stimulation in headphone playback is achieved by switching off the other earphone. In a sound field, monaural stimulation occurs only with severe hearing impairment or deafness in one ear, or may be achieved by occluding one of the ears; in practice, the listening in a sound field is binaural and often dichotic. For thorough definitions of sound fields and stimulation modes, see Blauert (1997).

Due to the complexity of real sound fields, listening experiments are typically carried out in relatively simple acoustic environments, e.g., in the *free field*, or in the *diffuse field*. Below, these two types of sound fields are considered with respect to their influence on loudness, and it will be shown that there are deficiencies in applying loudness models to binaural exposures in real sound fields.

2.1 Free and diffuse fields

Loudness models based on Stevens (1961) and Zwicker (1960) are standardized in ISO 532 (1975) as methods 'A' and 'B', respectively. In addition to computational differences in converting spectrum to loudness between the methods, the sound field in Method A is assumed to be diffuse, whereas Method B can also be applied to a free field. The free-field loudness model is applicable to situations where the direct sound is dominant, e.g., in the open air, and the diffuse-field model to sound fields with reverberation, e.g., inside rooms. Method B of ISO 532 (1975), often referred to as 'Zwicker' loudness, can be applied to both types of sound fields and is thus more widely used in practice.

From the listener's point of view, there is a fundamental difference between the two fields. In the free field, the direction from which sound reaches the listener is *frontal*, i.e., only from the direction ahead of the listener in a reflection-free environment. In the diffuse field, sound reaches the listener's ears with *equal* probability from *all directions*, which can be approximated in a reverberant room. The two sound fields thus defined are theoretical extremes of a continuum. Sound fields encountered in practice often lie between the two bounds, consisting of directional sound reaching the listener from sources at different locations and fairly diffuse reverberation caused by the room.

In the standardized models (ISO 532, 1975), a *monophonic* sound pressure, measured with an omnidirectional microphone in the absence of a listener, is utilized for computing loudness. Therefore, the acoustic transfer from the measurement point to stimulation at the listener's ears is part of the model. The model is not concerned with the actual sound signals reaching the listener's ears, but rather, with the properties of the sound field in the absence of a listener. The listener's obstruction of the sound field under measurement is assumed to resemble that of the average acoustic transfer of the loudness model. Furthermore, in Method B of ISO 532 (1975) the assumption as to which sound-field type applies (free vs. diffuse) is left to the person measuring, when using the monophonic sound-pressure signal for loudness computations.

Despite the monophonic acoustical measurement, the standardized loudness models ISO 532 (1975) is based on *binaural* listening. Loudness models in general, as summarized in section 1.4, utilize the long-term spectrum of an acoustical measurement for loudness computations, thus creating a direct relation between a stimulus spectrum and a loudness prediction. In the case of both the free and the diffuse field, the long-term spectra are essentially the same between the two ears of a listener, due to the apparent symmetry of the human body in the interaural axis. Thus, utilizing the left- or the right-ear spectrum (and keeping in mind that the stimulation is binaural) should result in the same loudness prediction.

Should *different* spectra at the two ears then yield different loudness predictions, depending on the ear chosen? The actual at-ear stimulation is not considered in the standardized loudness models (ISO 532, 1975), and thus dichotic stimulation, e.g., with different spectra at the two ears can become problematic for predicting loudness. The models are not concerned with how the signals at the left and the right ears of a listener contribute to overall binaural loudness, and how the at-ear signals should be integrated to predict human loudness perception. In real sound fields, dichotic at-ear signals are the norm rather than the exception, and thus, adjustments to ISO 532 (1975) are called for.

To sum up, two important aspects of loudness are entangled in the standardized models (ISO 532, 1975), which are unraveled here in turn: (1) The physical, acoustical transformation from an unobstructed sound field in the absence of a listener to stimulation at the listener's two ears, and (2) the psychophysical integration of the left- and the right-ear signals into a single percept of binaural loudness.

2.2 Spatial hearing and HRTFs

The physical transformation of sound to the listener's ears has received considerable attention in the research on spatial hearing (for a summary, see Blauert, 1997). Spatial hearing is concerned with the *locatedness* of auditory events, i.e., the position and extent in space of what is perceived auditorily. In this body of research, the direction-dependent transfer of sound to our ears has been investigated as 'sound localization cues', and more recently as the product of *head-related transfer functions* (HRTFs; Shaw, 1974; Mehrgardt and Mellert, 1977; Møller et al., 1995; Blauert, 1997, chap. 5). For a given incidence angle and ear, an HRTF is defined as:

$$HRTF = \frac{\text{sound pressure at the ear}}{\text{sound pressure in the middle of the head with the listener absent}} \quad (1)$$

The sound pressure at the ear in Eq. 1 can be obtained at various positions in the outer ear; at the eardrum, or at the entrance to the open or the blocked ear canal (Møller, 1992). The transfer function can be described either in the time or the frequency domain, often referred to as head-related impulse responses (HRIRs) or HRTFs, respectively.

The sound pressure at the ear, and hence, an HRTF is very dependent on the incidence angle of the sound. In Fig. 5, HRIRs and the corresponding HRTFs are plotted, for measurements at the blocked entrance to the left and right ear canals of the same listener, for two incidence

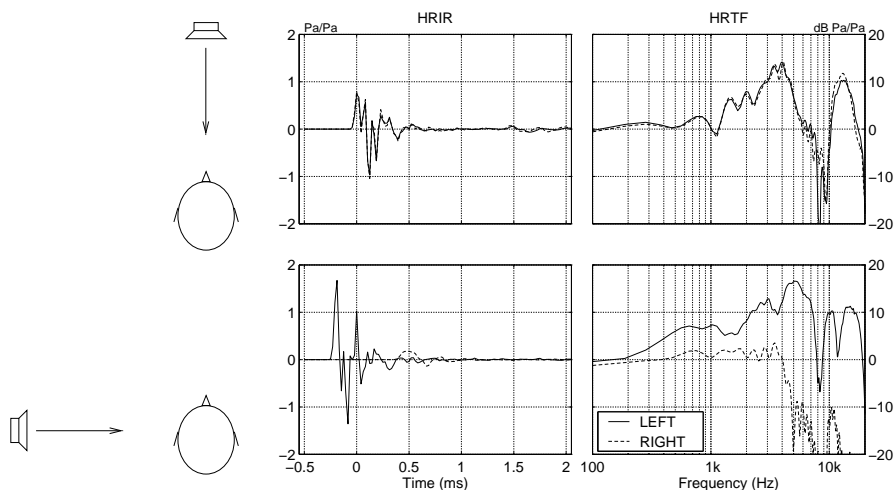


Figure 5: Head-related impulse responses and transfer functions in the time and frequency domains, respectively, measured at the blocked entrance to the ear canal for the same person for two incidence angles in the horizontal plane: 0° in front of and 90° of azimuth on the left-hand side of the person.

angles in the horizontal plane, as measured in an anechoic environment: 0° of azimuth in front of and 90° of azimuth on the left-hand side of the listener. For the frontal direction, which is the case in the free-field loudness model, both the HRIRs and HRTFs are similar across the left and right ears, the stimulation being (close to) diotic. For the direction on the side, interaural differences can be observed both in the time and the frequency domain, the stimulation being dichotic. It is taken that the *directional* changes in the binaural at-ear signals, as illustrated in Fig. 5, enable us to locate sounds in space.

The research in spatial hearing and HRTFs has emerged largely after the development of the first loudness models. Together with the HRTF research, *binaural technology* (Møller, 1992) has also come forth, where the idea is that if the sound pressures at two ears are recorded and later reproduced exactly as they were, the complete auditory experience is reproduced as well. Due to practicality, the sound pressures at the two ears are often recorded using an *artificial head* with a microphone in each ear as a transducer.

The focus is thus shifting from the monophonic measurements of sound pressure in the absence of the listener to binaural, and often dichotic, at-ear signals. Since the standardized loudness models of ISO 532 (1975) utilize a monophonic sound pressure, there is an increasing interest in how binaural measurements of at-ear sound pressure should be used in predicting loudness.

2.3 Binaural loudness summation

While spatial hearing research has concentrated on the acoustical transformation of sound to the listener's ears, studies of *binaural loudness summation* have investigated how the left- and right-ear signals are integrated to yield a single, binaural percept (Hellman and Zwischlocki, 1963; Reynolds and Stevens, 1960; Marks, 1978; Zwicker and Zwicker, 1991). Note that these studies focus on subjective judgments of binaural loudness, and typically do not address

biological or physiological issues as to where and how the interaural information is processed in the auditory pathway (for a review of physiology, see Gelfand, 1998).

A recent loudness model by Moore et al. (1997), the stages of which are depicted in Fig. 4, makes use of the findings of the two branches of research; (1) HRTFs and (2) binaural loudness summation. Firstly, the model facilitates the use of ear-drum sound pressures for loudness computations. This is accomplished by effectively separating the acoustical transformation to the listener's ears from the loudness model using HRTFs. This separation (see the fixed filter for the transfer from a free field to the ear-drum in Fig. 4) appears to be justified, since the agreement in subjective loudness data and objective at-ear measurements of sound pressure between the free and the diffuse sound fields is fair (Kuhl and Westphal, 1959; Robinson et al., 1961). In other words, the loudness difference between the free and the diffuse field is largely due to the first block in Fig. 4, and the processing of the at-ear signal for loudness computations is independent of the sound field type.

Secondly, perfect binaural summation of loudness in sones, i.e., the loudness of a monaural stimulation being half of that of the corresponding binaural stimulation, is implemented in the model by Moore et al. (1997). Thus, the model is able to compute monaural loudness separately for the two ears, and binaural loudness for any (diotic or dichotic) at-ear signals can be computed simply by summing the monaural values. This perfect loudness summation is supported by some headphone studies (e.g., Marks, 1978), while others disagree (e.g., Zwicker and Zwicker, 1991) reporting a smaller loudness difference between monaural and binaural stimulation. The model by Moore et al. (1997) is now standardized in ANSI S3.4 (2005).

In order to control the left- and right-ear signals independently of one another, headphones have been used in the binaural loudness-summation studies. While the use of headphones is beneficial in terms of experimental control, the fact that the acoustical transformation from an unobstructed sound field to the listener's ears is not taken into account can be considered a drawback. The stimuli can easily become unnatural to the listeners, since interaural combinations may be played back, which would never reach the listener's ears in a real sound field. Furthermore, the auditory events of such headphone stimulation are localized inside the listener's head, while loudness models originally have been developed for sounds carrying spatial information. The ecological validity of the stimulation can thus be questioned and it is debatable how these findings relate to the binaural loudness perception of spatial sounds, being affected by the acoustical transformation from an unobstructed sound field to our ears.

2.4 Directional loudness

It thus appears that, despite the large body of research both in spatial hearing and in binaural loudness summation, these two fields of research have not been connected properly. In the studies of binaural loudness summation, the *spatial and directional* aspects have been overlooked, and HRTFs have not been investigated as directional loudness cues.

The two stages of loudness processing, as they are conceptualized here, were first combined in the pioneering work of Robinson and Whittle (1960), who investigated the loudness of sounds coming from various directions in the horizontal, median and frontal planes. In addition to

obtaining loudness matches between different directions, they also determined the effective at-ear sound pressure levels for each incidence angle. When attempting to predict the directional loudness matches from the at-ear levels, Robinson and Whittle (1960) found a “6-dB binaural summation rule” to fit their data best:

$$L_{mon} = 6 * \log_2(2^{L_{left}/6} + 2^{L_{right}/6}), \quad (2)$$

where L_{mon} is the equivalent sound pressure needed for monotonic stimulation to match any binaural (diotic: $L_{left} = L_{right}$, or dichotic: $L_{left} \neq L_{right}$) combination of left-ear (L_{left}) and right-ear (L_{right}) input levels. If, for example, both ears are exposed to 70 dB SPL, the equivalent monotonic SPL turns out to be 76 dB SPL, hence the binaural gain is 6 dB.

Even though real sound sources located in space were used in the investigation by Robinson and Whittle (1960) for modeling the summation of binaural loudness, their findings have not been influential for the refinement of loudness models. Rather, an assumption of perfect binaural summation in sones prevails, roughly corresponding to a 10-dB summation in sound pressure level. Surprisingly, the study by Robinson and Whittle (1960) is the only one on directional loudness, investigating both the effect of sound incidence angle and the binaural summation underlying directional loudness judgments.

Since the work of Robinson and Whittle (1960) more than 45 years ago, several important advances have taken place, which are relevant to the issue, for example: (1) Time-selective acoustical measurements and digital signal processing, such as obtaining impulse responses of a system using the maximum-length sequence (Rife and Vanderkooy, 1989), (2) attention to binaural human exposure to sound and the rise of binaural technology (Møller, 1992), and (3) adaptive procedures for subjective judgments (Jesteadt, 1980). This PhD thesis uses these new technologies to investigate the intersection of the two historically different fields: spatial hearing and loudness perception.

3 Overview of the experimental work

3.1 Aim

This PhD thesis consists of a series of listening tests, where the sound incidence angle was included as an independent, experimental variable. Directional loudness matches to a frontal reference sound source, which in anechoic conditions is compliant with the free-field loudness paradigm, were obtained from listeners. In all listening tests, the setup was calibrated so that the same sound pressure from each incidence angle is measured with an omnidirectional microphone. Thus, loudness predictions with the monophonic, standardized loudness model ISO 532 (1975, Method B) were constant as function of incidence angle. Deviations from this constancy in the subjective matches therefore indicate directional dependency of loudness, and cannot be due to characteristics of the sound source measured in the absence of a listener.

In addition to the subjective loudness matches, the listeners’ individual *at-ear exposures* were determined by means of acoustical measurements for each incidence angle. These objective

measurements were analyzed for directional changes, and were compared with the corresponding changes in the loudness matches of sounds from various directions. Finally, the determined at-ear exposures were utilized to model the *binaural summation* of loudness underlying the directional matches for sounds emanating from various directions in space on an individual basis.

3.2 Organization

The experimental work of the thesis is reported in five manuscripts listed below and attached to this report, referred to as [M1]-[M5]:

[M1] Sivonen, V.P. and Ellermeier, W. (2006). “Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation,” *Journal of the Acoustical Society of America*, 119, 2965–2980.

[M2] Sivonen, V.P., Minnaar, P. and Ellermeier, W. (2005). “Effect of direction on loudness in individual binaural synthesis,” in *Proceedings of the 118th Convention of the Audio Engineering Society*, Barcelona, Spain, Paper No. 6512.

[M3] Sivonen, V.P. and Ellermeier, W. (2006). “Laterality in binaural and directional loudness perception,” in preparation for submission.

[M4] Sivonen, V.P. and Ellermeier, W. (2006). “Binaural loudness summation for directional sounds,” in preparation for submission. Portions of this work were presented in *Proceedings of Euronoise 2006*, Tampere, Finland, Paper No. 177.

[M5] Sivonen, V.P. (2006). “Directional loudness and the underlying binaural summation for wideband and reverberant sounds,” in preparation for submission. Portions of this work were presented in *Proceedings of the 120th Convention of the Audio Engineering Society*, Paris, France, Paper No. 6821.

3.3 Abstracts and interrelations of the manuscripts

Manuscript 1: “Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation”

In [M1], directional loudness matches were obtained for narrow-band stimuli for sound sources located both in the horizontal and the median planes inside an anechoic chamber. The listeners’ individual HRTFs were measured to determine the effective at-ear exposures and these were utilized in the modeling of binaural summation underlying the directional loudness matches.

Directional dependencies of loudness were observed, characterized by the center frequency of the stimuli, which could largely be explained by the individual HRTFs. However, large individual differences were obtained also in the underlying binaural summation, even after the effect of HRTFs on the at-ear sound pressure levels were accounted for.

Manuscript 2: “Effect of direction on loudness in individual binaural synthesis”

Due to the differences between listeners in the binaural summation of the at-ear levels, the individual HRTFs were utilized in synthesizing spatial sounds over headphones in [M2]. In this manner, the sets of HRTFs used for stimulation were precisely known, and they were immune to e.g., small, unwanted head movements. Directional loudness matches were then obtained for a subset of the same listeners and stimuli as in [M1].

Despite the use of binaural synthesis, essentially the same directional matches as in [M1] were obtained. The perhaps more controlled stimulation via headphones did not decrease the individual differences in binaural summation. However, [M2] validated the measured HRTFs, which were used for analysis of [M1] and allowed for the use of individual binaural synthesis in future experiments on directional loudness.

Manuscript 3: “Laterality in binaural and directional loudness perception”

In both [M1] and [M2], only the left hemisphere of the horizontal plane was investigated, due to assumed symmetry between the left and right hemispheres. Therefore, in [M3], binaural and directional loudness matches were obtained similarly for the left and the right sides of the head, in order to investigate possible laterality effects in binaural loudness judgments.

In most of the cases the effect of the side of stimulation was insignificant, although for some listeners a slight left-ear advantage for binaural and directional loudness judgments was observed. The magnitude of this lateral effect, however, was very small compared to the directional HRTF effect on loudness, i.e., directional loudness primarily depends on the transfer of sound to at-ear stimulation, not on which side of the head the source is located.

Manuscript 4: “Binaural loudness summation for directional sounds”

In [M4], generic, artificial-head HRTFs were utilized to obtain directional loudness matches. This was done in order to give each listener the same directional effects in at-ear exposure, and to investigate whether holding this source of individual variation constant would decrease the individual differences in the underlying summation.

Despite the use generic HRTFs, individual differences in binaural summation of at-ear SPLs were still evident and consistent with those earlier obtained with individual HRTFs. Thus, the *mean* data obtained by averaging across listeners were used for describing a binaural loudness model for artificial-head measurements.

Manuscript 5: “Directional loudness and the underlying binaural summation for wideband and reverberant sounds”

Anechoic, narrow-band stimuli were used in [M1]-[M4]. In order to improve to ecological validity of the stimulation, directional loudness matches were obtained for wideband and reverberant sounds in [M5]. The wideband stimuli were played back in binaural synthesis using individual HRTFs, and the reverberant stimuli using individual binaural room impulse responses (BRIRs), measured in a listening room.

Despite the more complex stimuli, the effect of direction on loudness was still observed. Furthermore, the mean data based on this more ecologically-valid stimulation agreed well with the binaural loudness model, as described in [M4].

3.4 General conclusions

In all experiments [M1]-[M5], both for anechoic and reverberant narrow-band, as well as anechoic wideband stimuli, the direction from which the sound reaches the listener's ears had a significant effect on loudness. The directional loudness matches varied over a range of up to 10 dB, showing considerable dependencies on the stimuli and the individual listener. The obtained results were at odds with the notion of loudness constancy; on the contrary these directional effects on loudness could be largely explained by the corresponding changes in individual acoustic transfers to binaural at-ear exposures, as determined by individual or generic HRTFs, or individual BRIRs.

When inspecting binaural summation underlying the directional loudness matches, statistically significant individual differences were obtained in the way the left- and right-ear inputs are combined in a single, binaural loudness percept. These individual peculiarities were robust throughout the whole experimental series, irrespective of the stimuli (narrow- vs. wideband) and the sound field (anechoic vs. reverberant) utilized. Furthermore, when synthesizing directional sounds using the same (generic) set of HRTFs for all participants [M4], similar peculiarities as with individual HRTFs were still evident in the binaural summation.

Putting these unexplained idiosyncracies aside, the mean data of all experiments [M1]-[M5] agree well with a binaural 3-dB (power) summation rule. This summation rule was obtained using the equation published by Robinson and Whittle (1960, see Eq. 2), by relaxing their assumption of a 6-dB summation, and finding the best-fitting binaural gain to account for the present data. Since a modern psychoacoustic procedure to obtain loudness matches, state-of-the-art techniques to measure at-ear exposures, and a wide variety of stimuli were utilized in this thesis, it is the author's belief that the 3-dB summation of the at-ear SPLs will yield a better prediction of binaural, directional loudness than the 6-dB rule proposed by Robinson and Whittle (1960).

The 3-dB summation rule is also at odds with the assumption of perfect binaural summation of loudness in sones. In the present experiments, the summation rule was observed for stimulation where the at-ear signals are due to the physical obstruction of the sound field by the listener, while the binaural sone summation has been reported in headphone studies not incorporating spatial aspects to the stimulation. It may thus be argued that the present results have greater ecological validity, both for predicting loudness in real sound fields and for the application of acoustical measurements using artificial heads.

A binaural loudness model (see [M4]) was developed based on the results. Since the acoustic transfer from an unobstructed sound field is included in the signals arriving at the ears, and diotic or dichotic stimulation can be accounted for by the 3-dB summation rule, the model is able to predict the loudness of any type of a sound field, be it free or diffuse, anechoic or reverberant. The binaural loudness model is applicable to acoustical measurements made with an artificial head, being often used in a variety of applications from the assessment of noise exposure to product sound quality and the fidelity of reproduced sound.

3.5 Future work

The present thesis can be extended in a number of ways. One way could be utilizing other than noise-like signals, such as speech, in investigating directional loudness. Such a real-life stimulus may cause listeners to pay attention to the sound source rather than to the effect the sound radiating from the source has on the listener. In other words, it could be investigated whether cognitive effects, i.e., what use the brain is making of the stimulation, play a role in directional loudness perception, or is loudness purely dependent on the sensory signals arriving at our ears.

The listeners' attention may also be attempted to be manipulated via instructions, asking them e.g., to judge the loudness of the source. Judging the *source* loudness for various distances has shown evidence for loudness constancy (Zahorik and Wightman, 2001), i.e., the perceived loudness of the source remains constant over distances, despite profound changes in at-ear exposure. In the directional loudness paradigm, constancy could be either *physical* or *cognitive*. In an ideal diffuse field, the listener's orientation has no effect on the long-term at-ear spectra, which are the primary inputs to loudness processing. In such a sound field, loudness is constant as a function of direction, due to physics.

The cognitive constancy requires accurate localization of the source. It is known that the visual size constancy requires accurate depth information (Goldstein, 2004). By closing one eye, depth information is degraded and size constancy is decreased. It is fair to assume that for loudness constancy, accurate *auditory* distance localization is necessary. Or, put in another way, poor distance localization decreases loudness constancy, and the perception is based more on the signals at the sensory receptors, not at the source.

Although localization of sources was not specifically required in the work reported in this thesis, informal listening revealed directional localization for the narrow-band stimuli in the median plane of [M1] and [M2] to be poor (this is due to the so-called "directional bands", see Blauert, 1997). In the median plane, the directional loudness matches showed no constancy and agreed well with the corresponding changes in at-ear stimulation. In the horizontal plane, however, the sources were localizable, even with the narrow-band stimuli. In the present series of experiments this did not induce constancy, as might have been expected. It is unclear whether similarly large directional loudness dependencies were obtained, if the source was moving around the listener, or the listener was rotated, while judging loudness. The *dynamic* cues to the auditory system might induce directional loudness constancy, forcing the listeners to base their judgments on the static sound power of the source, rather than the varying signals at the ears.

Furthermore, the effect of the location of the auditory event (Blauert, 1997) on binaural loudness perception could also be investigated. This could be achieved in binaural synthesis by comparing the loudness of stimulation with original and modified interaural cues, which might result in either well or poorly 'auralized' stimuli. The difference between the loudness of loudspeaker and headphone playback, even when receiving the same at-ear signals in both cases, has been debated in the field (Rudmose, 1982; Keidser et al., 2000). For loudspeaker playback the auditory event is outside the head, while for headphone listening without spatial synthesis, more or less inside-the-head percepts are created. This difference in the auditory event may cause differences also in the perceived loudness between the two playback methods.

Also, HRTFs could be decomposed, e.g., by modifying or removing the interaural time difference (ITD), in order to investigate either the effect of the minimum-phase part of an HRTF or the ITD on binaural loudness. Note that the loudness predictions both by the present 3-dB summation model, and the most widely-accepted model by Moore et al. (1997), are based on the magnitude spectra at the ears and other interaural attributes are omitted. Although binaural loudness has been shown to be unaffected e.g., by varying the interaural crosscorrelation (Dubrovskii et al., 1972; Eichenlaub et al., 1996), further experiments for quantifying the effect of various interaural components are justified.

Any of these strategies may help explaining the robust individual differences observed in the present study. However, while the idiosyncracies are troublesome for analyzing the results, it is worth to note that all loudness models are based on mean data, and attempt to predict the behavior of an average listener, disregarding individual differences.

Finally, the present study employed one sound source at a time for loudness comparisons. In real-life, sound fields often consist of multiple sound sources, being considerably more complex than the discrete sources investigated here. The phenomenon of the cocktail-party effect, i.e., the ability of the auditory system to focus on a single sound source in the presence of background noise, has been recognized for some time. The issue of distributed, multiple sound sources and their loudness perception has recently been brought up by Song et al. (2005), investigating the relative contribution of individual sound sources to overall loudness. The findings of such studies, in addition to the ones of the present study, should thus be taken into consideration in developing a universal, binaural model of loudness perception.

References

- ANSI S3.4 (2005). "Procedure for the Computation of Loudness of Steady Sounds," American National Standards Institute, New York, USA.
- Berglund, B., Preis, A. and Rankin, K. (1990). "Relationship between loudness and annoyance for ten community sounds," *Environment International*, **16**, 523–531.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA.
- Buus, S., Florentine, M. and Poulsen, T. (1997). "Temporal intergration of loudness, loudness discrimination, and the form of the loudness function," *Journal of the Acoustical Society of America*, **101**, 669–680.
- Collins, A.A. and Gescheider, G.A. (1989). "The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching," *Journal of the Acoustical Society of America*, **85**, 2012–2021.
- DeCarlo, L. T. and Cross, D. V. (1990). "Sequential effects in magnitude scaling: Models and theory," *Journal of Experimental Psychology: General*, **119**, 375–396.
- Dubrovskii, N.A., Chernyak, R.I. and Shapiro, V.M. (1972). "Binaural summation of differently correlated noises," *Soviet Physics Acoustics*, **17**, 468–473.

- Eichenlaub, C., Chouard, N. and Weber, R. (1996). "On the influence of interaural correlation on binaural loudness," in *Proceedings of Internoise'96*, Budapest, Hungary, 2227–2230.
- Ellermeier, W. and Faulhammer, M. (2000). "Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production," *Perception & Psychophysics*, **62**, 1505–1511.
- Fletcher, H. and Munson, W.A. (1933). "Loudness, its definition, measurement and calculation," *Journal of the Acoustical Society of America*, **5**, 82–108.
- Fowler, E.P. (1950). "The recruitment of loudness phenomenon," *Laryngoscope*, **60**, 680–695.
- Gabrielsson, A. and Sjögren, H. (1979). "Perceived sound quality of sound-reproducing systems," *Journal of the Acoustical Society of America*, **65**, 1019–1033.
- Gelfand, S.A. (1998). *Hearing: An Introduction Psychological and Physiological Acoustics*, Marcel Dekker Inc., New York.
- Glasberg, B.R. and Moore, B.C.J. (2002). "A model of loudness applicable to time-varying sounds," *Journal of the Audio Engineering Society*, **50**, 331–342.
- Goldstein, E.B. (2004). *Sensation and Perception*, Wadsworth, Pacific Grove, CA.
- Hellman, R.P. and Zwislocki, J.J. (1963). "Monaural loudness function at 1000 cps and interaural summation," *Journal of the Acoustical Society of America*, **35**, 856–865.
- Hellman, R.P. and Zwislocki, J.J. (1968). "Loudness determination at low sound frequencies," *Journal of the Acoustical Society of America*, **43**, 60–64.
- Hellman, R.P. (1985). "Perceived magnitude of two-tone-noise complexes: Loudness, annoyance and noisiness," *Journal of the Acoustical Society of America*, **77**, 1497–1504.
- Illényi, A. and Korpassy, P. (1981). "Correlation of loudness and quality of stereophonic loudspeakers," *Acustica*, **49**, 334–336.
- Irwin, R.J. and Whitehead, P.R. (1991). "Towards an objective psychophysics of pain," *Psychological Science*, **2**, 230–235.
- ISO 226 (2003). "Acoustics – Normal equal-loudness-level contours," International Organization for Standardization, Geneva, Switzerland.
- ISO 532 (1975). "Acoustics – Method for calculating loudness level," International Organization for Standardization, Geneva, Switzerland.
- Jesteadt, W. (1980). "An adaptive procedure for subjective judgments," *Perception & Psychophysics*, **28**, 85–88.
- Keidser, G., Katsch, R., Dillon, H. and Grant, F. (2000). "Relative loudness perception of low and high frequency sounds in the open and occluded ear," *Journal of the Acoustical Society of America*, **107**, 3351–3357.
- Kohfeld, D.L., Santee, J.L. and Wallace, N.D. (1981). "Loudness and reaction time: I," *Perception & Psychophysics*, **29**, 535–549.

- Kuhl, W. and Westphal, W. (1959). "Unterschiede der Lautstärken in der ebenen Welle und im diffusen Schallfeld" ("Differences in loudness levels between a plane wave and a diffuse sound field,") *Acustica*, **9**, 407–408.
- Marks, L.E. (1978). "Binaural summation of the loudness of pure tones," *Journal of the Acoustical Society of America*, **64**, 107–113.
- Marks, L.E. (1992). "The slippery context effect in psychophysics: intensive, extensive, and qualitative continua," *Perception & Psychophysics*, **51**, 187–198.
- McPherson, D.F. and Anderson, C.V. (1971). "Relation of Temporary Loudness Shift to Temporary Threshold Shift," *Journal of the Acoustical Society of America*, **49**, 1195–1202.
- Mehrgardt, S. and Mellert, V. (1977). "Transformation characteristics of the external human ear," *Journal of the Acoustical Society of America*, **61**, 1567–1576.
- Miśkiewicz, A., Scharf, B. Hellman R.P. and Meiselman C. (1993). "Loudness adaptation at high frequencies," *Journal of the Acoustical Society of America*, **94**, 1281–1286.
- Moore, B.C.J. (2003). *An Introduction to the Psychology of Hearing*, Academic Press, London.
- Moore, B.C.J., Glasberg, B.R. and Baer, T. (1997). "A model for the prediction of thresholds, loudness and partial loudness," *Journal of the Audio Engineering Society*, **45**, 224–240.
- Møller, H. (1992). "Fundamentals of binaural technology," *Applied Acoustics*, **36**, 171–218.
- Møller, H., Sørensen, M.F., Hammershøi, D. and Jensen, C.B. (1995). "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, **43**, 300–321.
- Narens, L. (1996). "A theory of ratio magnitude estimation," *Journal of Mathematical Psychology*, **40**, 109–129.
- Reason, J.T. (1972). "Some correlates of the loudness function," *Journal of Sound and Vibration*, **20**, 305–309.
- Reynolds, G.S. and Stevens, S.S. (1960). "Binaural summation of loudness," *Journal of the Acoustical Society of America*, **32**, 1337–1344.
- Rife, D.D. and Vanderkooy, J. (1989). "Transfer-function measurement with maximum-length sequences," *Journal of the Audio Engineering Society*, **37**, 419–444.
- Robinson, D.W. and Dadson, R.S. (1960). "A re-determination of the equal-loudness relations for pure tones," *British Journal of Applied Physics*, **7**, 166–181.
- Robinson, D.W. and Whittle, L.S. (1960). "The loudness of directional sound fields," *Acustica*, **10**, 74–80.
- Robinson, D.W., Whittle, L.S. and Bowsher, J.M. (1961). "The loudness of diffuse sound fields," *Acustica*, **11**, 397–404.
- Rudmose, W. (1982). "The case of the missing 6 dB," *Journal of the Acoustical Society of America*, **71**, 650–659.

- Scharf, B. (1978). "Loudness," In Carterette, E.C. and Friedman, M.P. (Eds.), *Handbook of Perception, Vol. IV: Hearing*, Academic Press, New York, pp. 187–242.
- Shaw, E.A.G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *Journal of the Acoustical Society of America*, **56**, 1848–1861.
- Song, W., Ellermeier, W. and Minnaar, P. (2005). "Perceived loudness of spatially distributed sound sources," in *Proceedings of Forum Acusticum*, Budapest, Hungary, 1665–1670 .
- Stevens, S.S. (1955). "The Measurement of Loudness," *Journal of the Acoustical Society of America*, **27**, 815–829.
- Stevens, S.S. (1961). "Procedure for Calculating Loudness: Mark VI," *Journal of the Acoustical Society of America*, **33**, 1577–1585.
- Stevens, J.C. and Guirao, M. (1964). "Individual Loudness Functions," *Journal of the Acoustical Society of America*, **36**, 2210–2213.
- Viemeister, N.F. and Wakefield, G.H. (1991). "Temporal integration and multiple looks," *Journal of the Acoustical Society of America*, **90**, 858–865.
- Werner, L.A. and Rubel, E.W. (1992). *Developmental Psychoacoustics*, American Psychological Association, Washington.
- Zahorik, P. and Wightman, F.L. (2001). "Loudness constancy with varying sound source distance," *Nature Neuroscience*, **4**, 78–83.
- Zwicker, E. (1960). "Ein Verfahren zur Berechnung der Lautstärke, (Procedure for Computation of Loudness)," *Acustica*, **10**, 304–308.
- Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and Models*, Springer, New York.
- Zwicker, E. and Zwicker, U.T. (1991). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," *Journal of the Acoustical Society of America*, **89**, 756–764.

Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation^{a)}

Ville Pekka Sivonen^{b)}

*Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University,
Fredrik Bajers Vej 7 B5, 9220 Aalborg East, Denmark and Brüel & Kjær Sound
& Vibration Measurement A/S, Skodsborgvej 307, 2850 Nærum, Denmark*

Wolfgang Ellermeier

*Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University,
Fredrik Bajers Vej 7 B5, 9220 Aalborg East, Denmark*

(Received 25 July 2005; revised 7 February 2006; accepted 13 February 2006)

The effect of sound incidence angle on loudness was investigated using real sound sources positioned in an anechoic chamber. Eight normal-hearing listeners produced loudness matches between a frontal reference location and seven sources placed at other directions, both in the horizontal and median planes. Matches were obtained via a two-interval, adaptive forced-choice (2AFC) procedure for three center frequencies (0.4, 1, and 5 kHz) and two overall levels (45 and 65 dB SPL). The results showed that loudness is not constant over sound incidence angles, with directional sensitivity varying over a range of up to 10 dB, exhibiting considerable frequency dependence, but only minor effects of overall level. The pattern of results varied substantially between subjects, but was largely accounted for by variations in individual head-related transfer functions. Modeling of binaural loudness based on the at-ear signals favored a sound-power summation model, according to which the maximum binaural gain is only 3 dB, over competing models based on larger gains, or on the summation of monaural loudness indices. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2184268]

PACS number(s): 43.66.Pn, 43.66.Cb [AK]

Pages: 2965–2980

I. INTRODUCTION

There is growing awareness in psychoacoustics that, for a thorough understanding of loudness perception, its binaural nature has to be taken into account. That is true for basic research, namely the construction of general loudness models (e.g., Moore *et al.*, 1997), as well as for applications to audio reproduction systems (Zacharov *et al.*, 2001) or to perceived sound quality (Bodden, 1997). Especially for instrumental loudness predictions based on Zwicker's modeling, as standardized in ISO 532 (1975), the fact that it is essentially monophonic has been regarded as a major drawback. Nevertheless, the adjustments recently made to loudness modeling rest on a fairly narrow empirical data base, which the present study hopes to extend.

To clarify the issues, it may be helpful to distinguish two stages of processing involved when the loudness of a real sound source in space is perceived: (1) the physical transformation of the “distal” stimulus emitted by the source to “proximal” stimuli arriving at the listener's ears, and (2) the neural, psychological, and cognitive process of integrating the two at-ear stimuli into a single percept.

A. Physical (HRTF) filtering

The first stage can be described in purely acoustical terms, namely by applying head-related transfer functions (HRTFs, Shaw, 1974; Wightman and Kistler, 1989a; Møller *et al.*, 1995; Blauert, 1997, Chap. 5). These account for the filtering of the source due to the physical effects of the human torso, head, and pinnae, depending on the incidence angle of the sound. Further along, through the ear canal, the physical sound transmission has been shown to be independent of the direction of the sound source (see, e.g., Hammershøj and Møller, 1996). Thus, the direction-dependent part of an HRTF can be measured at the entrance to the blocked ear canal, and described by (adopted from Møller *et al.*, 1995)

$$\text{HRTF}_{\text{dir-dep}}(\phi, \theta) = \frac{P_2}{P_1}(\phi, \theta), \quad (1)$$

where ϕ is azimuth, θ is elevation, P_1 is sound pressure at the center position of head, and P_2 is sound pressure at the entrance to the blocked ear canal.

In the median plane, the HRTFs of the two ears are fairly similar due to the physical symmetry of the human body in this plane. However, level differences between HRTFs for different directions can approach 10 dB or more over a fairly wide frequency range. By contrast, large interaural time and level differences (ITDs and ILDs, respectively) between the two ears emerge in the horizontal plane, where the ILDs can reach up to 30 dB at high frequencies. HRTFs have been a major research topic during the past

^{a)}Portions of the data were presented at the 147th Meeting of the Acoustical Society of America, New York, May, 2004, and at Internoise, Prague, August, 2004.

^{b)}Electronic mail: vps@acoustics.aau.dk

15 years, but the focus of this research has been on adequate “auralization” or sound localization (Wightman and Kistler, 1989a,b; Bronkhorst, 1995; Møller *et al.*, 1996), not on loudness.

B. Binaural loudness summation

The second stage of processing has been termed binaural loudness summation. It describes how the acoustic inputs to the left and right ear are integrated to yield a single, binaural loudness. Starting from the observation that a sound appears louder when listened to with both ears (i.e., binaurally) than with only one (i.e., monaurally), a number of investigators conducted experiments using headphones, through which different combinations of left- and right-ear sound-pressure levels were presented in order to quantify this effect. The results are often summarized as providing evidence for a binaural-to-monoaural loudness ratio of 2:1, or perfect loudness summation, corresponding to a binaural gain of approximately 10 decibels (e.g., Levitt *et al.*, 1972; Marks, 1978; Schneider and Cohen, 1997), in accordance with the sone scale of loudness. The evidence is far from unequivocal, however, with many studies finding less-than-perfect summation (e.g., loudness ratios of approximately 1.5:1; Zwicker and Zwicker, 1991), and a level dependence of the binaural gain, which appears to increase from approximately 3 dB near threshold to 6–10 dB at high sound-pressure levels (Shaw *et al.*, 1947; Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963).

Interestingly, binaural loudness summation, as conceptualized in this paradigm, has not been investigated with sounds that are likely to reach the eardrums when emitted from a real source in space, i.e., with products of the first (HRTF) filtering stage. Rather, artificial sounds such as tones, or broadband noise, lacking all spatial or directional information have been used, often at interaural level differences (e.g., in monotic-to-diotic comparisons) far exceeding what would naturally occur. Such conditions of stimulation do not yield an externalized sound image, but rather more or less lateralized inside-the-head percepts. Generally, it appears that the considerable literature on binaural loudness summation has contributed more to the development of scaling methodologies than to the auditory issues involved.

C. Loudness of free and diffuse sound fields

For practical purposes, in an attempt to relate the monophonic measurement of a sound field to perceived loudness, two specific types of sound fields have been considered: The free field, where the sound incidence angle is frontal to the listener, and the diffuse field, where the sound is reaching the listener's ears with equal intensity from all directions.

In order to account for the fundamental difference in sound incidence, the standardized loudness model (ISO 532, 1975) has different computation procedures for the two sound fields. The two procedures are based on both objective and subjective data (Kuhl and Westphal, 1959; Robinson *et al.*, 1961; ISO 389-7, 1996): The objective data represent the differences in the at-ear sound pressures between the two sound fields, i.e., investigating only the effect of the first

(HRTF) filtering stage; the subjective data represent the differences in perceived loudness, including effects of both the first and the second stage. Even though the agreement between the objective and subjective data is fair, these investigations do not specify how the two signals at the ears of a listener are summed into a single loudness percept, due to the fact that the stimulation of the auditory system in both sound fields is essentially diotic.

The increasing use of dummy heads for acoustical recordings and measurements, often resulting in dichotic at-ear signals, has led to growing interest in how dichotic at-ear signals should be summed to correspond to the diotic stimulation of the conventional free- and diffuse-field loudness paradigms.

D. Directional loudness

Thus, while studies of HRTF filtering have not explicitly been concerned with the loudness of dichotic sounds, the work on binaural loudness summation appears to lack ecological validity to predict the perception of real sources positioned in space. What remains, then, is less than a handful of studies that have actually investigated directional loudness of real sources in space, taking into account both stages delineated: the physical filtering due to HRTFs, and the ensuing “psychological” summation.

Sivian and White (1933) investigated the effect of direction on hearing thresholds, reporting that at absolute threshold, the binaural minimum audible field is not significantly different from the monoaural one. This implies no or a very small binaural advantage, the ear receiving the higher sound pressure determining the binaural hearing threshold. While the directional HRTF effects are the same at higher sound-pressure levels, extrapolating from a detection task to suprathreshold binaural loudness and to its summation across the two ears may be unjustified.

By far, the most pertinent and complete study investigating directional loudness was published by Robinson and Whittle (1960) more than 45 years ago. The authors used a circular array of 12 equally spaced loudspeakers positioned around the listener seated in an anechoic room to obtain loudness matches between a reference and each test position. Using narrow-band sounds having six center frequencies between 1.6 and 10 kHz, and rotating their apparatus when required, they investigated the horizontal, median, and frontal planes in a sample of 16 to 20 listeners. Using probe-tube microphones they also measured sound-pressure levels at the ears of their subjects, as produced by the same stimuli, thus obtaining crude magnitudes of “HRTFs” for the six test frequencies.

As expected, the average loudness matches showed a strong frequency dependence, with the greatest directional effects (of up to 15 dB; see their Fig. 2) observed at higher frequencies (4–10 kHz). Relating the mean loudness matches to the average at-ear sound-pressure measurements, Robinson and Whittle (1960; see their Fig. 5) found the former to be reasonably well predicted by a “6-dB summation rule,”

$$L_{\text{mon}} = 6 \times \log_2(2^{L_{\text{left}}/6} + 2^{L_{\text{right}}/6}), \quad (2)$$

where L_{mon} is the equivalent sound pressure needed for monotic stimulation to match any binaural (diotic: $L_{\text{left}} = L_{\text{right}}$, or dichotic: $L_{\text{left}} \neq L_{\text{right}}$) combination of left-ear (L_{left}) and right-ear (L_{right}) input levels. If, for example, both ears are exposed to 70 dB SPL, the equivalent monotic SPL turns out to be 76 dB SPL (i.e., a 6-dB binaural gain). Note, however, that there is not sufficient information on the fit of this heuristic (other than what can be judged from visual inspection of their Fig. 5) in Robinson and Whittle's report, or on its feasibility to predict individual subjects' data.

Two more recent studies picked up the issue, though using considerably fewer experimental conditions and subjects, and not including HRTF measurements. Both studies (Remmers and Prante, 1991; Jørgensen, 2002) used wideband noise to obtain loudness matches, thus potentially obscuring a frequency dependence, and obtained much smaller effects than Robinson and Whittle (1960), with directional loudness matches varying by less than 3 dB across incidence angles.

E. Rationale for the present study

It thus appears worthwhile to take up the issue of directional loudness once more. This will be done paying special attention to five methodological issues, which are elaborated in turn:

- (1) Well-defined narrow-band stimuli are needed to investigate the effects of HRTFs and binaural loudness summation. Note that, in Robinson and Whittle's (1960) report, the sounds used were not sufficiently specified beyond stating that they were "below a critical band" (p. 75), and the later studies used wideband noise which might wash out some of the directional effects.
- (2) Given the evidence from earlier headphone experiments showing the binaural gain to increase from approximately 3 dB near threshold to up to 10 dB at high sound-pressure levels, level effects will be taken into account by making measurements at two overall sound-pressure levels.
- (3) With the exception of Jørgensen's (2002) study, classical "method(s) of adjustment" have been used to collect the subjective data. By their transparency, and the explicit control they give listeners over the outcome, these methods are prone to subject-induced biases, such as "correcting" an adjustment due to some expectation. Forced-choice procedures (Levitt, 1971; Jesteadt, 1980), especially when interleaving adaptive tracks for different experimental conditions, are much less susceptible to such biases.
- (4) Advances in the methodology to HRTFs will be brought to the study of directional loudness. Note that Robinson and Whittle's (1960) pioneering study was done before the term HRTF was coined, and that their at-ear measurements of the stimuli actually used merely provide six points along the frequency scale, and thus do not constitute HRTFs as we conceive of them today.

- (5) Since HRTF filtering is known to be highly idiosyncratic, it is likely that with averaged data frequency-dependent directional effects might partially cancel each other, thus underestimating the true effect size. Therefore, a greater emphasis than in earlier studies will be on individual results and analyses.

To sum up, the present investigation will be conducted by having subjects assess loudness in a directional sound field in an anechoic room, and by relating the listening test data both to the distal stimulus given by the sound-pressure level emitted by the active loudspeaker, and to the proximal stimuli given by the participants' at-ear exposure levels as obtained via state-of-the-art HRTF measurements.

II. METHOD

A. Subjects

Eight normal-hearing listeners (between the age of 22 and 46 years; five male, three female), including the second author, participated in the experiment. The subjects' hearing thresholds were determined using standard pure-tone audiometry in the frequency range between 0.25 and 8 kHz with the requirement that none of the thresholds exceed 15 dB hearing level *re*: ISO 389-1 (1998). The five subjects who were not staff members were paid an hourly wage for their participation.

B. Apparatus

1. Loudspeaker setup in the anechoic chamber

The experiment was carried out in an anechoic chamber, which is anechoic above approximately 200 Hz, and has background noise at an inaudible level.

The loudspeaker setup for the experiment consisted of eight identical speakers (Vifa M10MD-39) mounted in hard plastic balls with a diameter of 15.5 cm. A typical frequency response of the loudspeaker can be found in Møller *et al.* (1995).

The loudspeakers were positioned both in the horizontal and median planes. In the horizontal plane, the incidence angles were 30°, 60°, 90°, and 135° of azimuth, and in the median plane the angles were 45° and 90° of elevation. Loudspeakers were also placed ahead and behind the listening position (at 0° and 180° of azimuth with 0° of elevation), where the horizontal and the median planes coincide. Due to assumed symmetry, the loudspeakers were placed only on the left-hand side in the horizontal plane. The distances from the diaphragms of the loudspeakers to the listening position at the center of the setup were 206 ± 4 cm.

The subjects were seated in a chair, the height of which could be adjusted. The chair had a small headrest to restrict head movements of the subjects during the experiment. The subjects' heads and ears were carefully aligned with the center position of the setup by making adjustments to chair height and headrest position using a laser and two video cameras. A photograph of the setup in the anechoic chamber is shown in Fig. 1. The loudspeakers ahead, at 30° and 60° in the horizontal plane, and at 45° and 90° in the median plane are visible in the photograph. The structure suspending the



FIG. 1. The experimental setup in the anechoic chamber.

loudspeakers and the platform (an open metal grid) under the chair were covered with sound-absorbing material.

The subjective responses were collected with a two-button response box. The response box had small lights above the buttons to indicate observation intervals. An enlarged copy of the indicator lights was placed behind and slightly above the frontal loudspeaker to avoid subjects tilting their heads downwards to the response box in their hands.

2. Signal generation and control

All other equipment was placed in a control room next to the anechoic chamber. A personal computer (PC) was used for controlling the experiment and carrying out objective measurements. The PC was equipped with a digital sound card (RME DIGI96/8 PST) with eight audio channels, connected to an external AD/DA-converter (RME ADI-DS8). A custom-made eight-channel attenuator with a 128-dB dynamic range and 0.5-dB step size was used to individually control the level of the eight loudspeakers. The signals from the attenuator were amplified by power amplifiers (Rotel RB-976 Mark II), and then fed to the loudspeakers in the anechoic chamber.

The experiment was run using a program developed in LABVIEW. The program took care of reading session files,

playing back appropriate stimuli, collecting subjects' responses, adapting the attenuator gains according to the responses, and writing the data into result files.

C. Measurements

Acoustical measurements were carried out using the maximum-length-sequence (MLS) system as specified by Olesen *et al.* (2000), with an MLS order of 12, preaveraging of 16, and a sampling rate of 48 kHz. The length of the acquired impulse responses was 4095 samples, which, due to the scarcity of reflections inside the anechoic chamber, was long enough to avoid time aliasing. The measurements were carried out at a level of approximately 70 dB SPL (at 1 kHz), measured in the absence of a listener at the center position of the setup.

First, responses of each loudspeaker [P_1 pressures, see Eq. (1)] were measured at the center position using a 1/4-in. pressure field microphone (Brüel & Kjær type 4136) with 90° incidence to the loudspeaker under measurement. Then, responses of each loudspeaker at each listener's ears [individual P_2 pressures, see Eq. (1)] were measured at the blocked entrance to the ear canal using two miniature microphones (Sennheiser KE 4-211-2), one microphone specifically for each ear. The miniature microphones were fitted inside foam earplugs (E·A·R Classic, halved in length), and mounted flush with the ear-canal entrance. All microphone signals were bandpass filtered between 22.5 Hz and 22.5 kHz by the measurement amplifier used (Brüel & Kjær type 2607 or type 2690 Nexus).

The above measurements were carried out three times: in the beginning, halfway through, and at the end of the experiment. The loudspeaker responses were used to equalize the stimuli for the listening experiment and to obtain reference pressures (P_1) for the HRTF calculations. The responses at each listener's ears were used to obtain individual HRTFs. The HRTF measurement procedure was as described by Møller (1995) with the following exceptions: The subjects were sitting in a chair instead of standing, the anechoic chamber was smaller, and the MLS measurement system was different.

Computation of the HRTFs involved 1024 samples from P_1 and P_2 pressures. First, individual head-related impulse responses (HRIRs) were calculated from P_1 and P_2 including a correction for the differences in the frequency responses of the two types of microphones used in the measurements. These HRIRs included reflections from the loudspeaker setup; therefore, only 140 samples from the HRIRs were used for calculating the final HRTFs. The resulting samples included all reflections from the subjects themselves (and from the chair), but excluded reflections from the other loudspeakers, the loudspeaker suspension, and any other objects inside the anechoic chamber. Note, however, that the excluded reflections were very small compared to the magnitude of the pure HRTFs.

D. Stimuli

The stimuli used for the listening experiment were third-octave noise bands centered at 0.4, 1, and 5 kHz. The length of each stimulus was 1 s.

For generating the stimuli, a 1-s white-noise signal was created, and subsequently filtered using third-octave-band filters at each center frequency. The relative differences in the frequency responses of the loudspeakers were equalized by applying minimum-phase inverse filters based on the direct sound coming from the loudspeakers. Each narrow-band signal was convolved with each of the inverse filters characterizing the individual loudspeakers, resulting in 24 stimuli for each (center frequency \times loudspeaker) combination. Finally, raised-cosine rise and decay ramps of 20-ms duration were applied. The sound files thus corrected were played back at a sampling rate of 44.1 kHz, and with 16-bit resolution in the experiment proper.

The third-octave-band levels of the stimuli were aligned to 64.7 ± 0.1 dB SPL at 0.4 kHz, 64.7 ± 0.2 dB SPL at 1 kHz, and 63.9 ± 0.1 dB SPL at 5 kHz. At the highest possible playback level (75 dB SPL) the levels of the second- and third-order harmonics were more than 37 and 43 dB below the level of the center frequencies of the narrow-band noises, respectively. The distortion was measured to be highest at the lowest center frequency, but it was inaudible for all stimuli. Furthermore, the spectral envelope of the equalized stimuli was verified to be very similar between different loudspeakers.

In the experiment proper, the stimuli were played back at two overall levels as measured at the listener's position; a "low" overall level of around 45 dB SPL and a "high" overall level of around 65 dB SPL. Even though the actual measured sound-pressure levels deviated slightly from these values, note that the misalignment between the loudspeakers was less than ± 0.2 dB at each center frequency.

E. Procedure

The aim of the experiment was to determine how loudness is affected by the sound incidence angle at three center frequencies and two overall levels. This was accomplished by matching the loudness of test sounds emanating from each of the loudspeakers in the setup to the loudness of the same sound coming from the reference loudspeaker positioned in front of the subject at 0° of azimuth and elevation.

1. Adaptive matching procedure

Matches were obtained using a two-interval, adaptive forced-choice (2AFC) procedure (Levitt, 1971) converging on the point of subjective equality (PSE) by following a simple 1-up, 1-down rule. On each trial, the (variable) test sound, and the (fixed) frontal reference were presented in random order, with a 500-ms pause in between. Synchronized with the sounds, two light-emitting diodes were successively lit both on the hand-held response box, and on its larger model in order to mark the observation intervals to be compared. The subject's task was to judge which of the two noises sounded louder by pressing one of the two buttons aligned with the observation-interval lights. The participants

were instructed to judge the loudness of the sounds only, and to disregard any other differences (due to direction, or timbre, for example) they might perceive.

For each adaptive track, the overall level of the frontal reference was fixed to either 45 or 65 dB SPL, as was the center frequency of the sounds to be played, and the test loudspeaker to be matched. The level of the test loudspeaker, however, was controlled by the adaptive procedure: Whenever the subject judged the test sound to be louder than the (frontal) reference, its sound-pressure level was lowered by a given amount; whenever the subject judged the reference to be louder, the level of the test loudspeaker was increased by that same amount. The initial step size was 4 dB; after two reversals (i.e., changes in the direction of the adaptive track) it was decreased to 1 dB. A total of eight reversals was collected in each adaptive track; the arithmetic mean of the last six of them was used to estimate the PSE. Two different starting levels were employed for the adaptive tracks, one 10 dB above, one 10 dB below the level of the reference loudspeaker, thus providing clearly discriminable loudness differences at the outset of each track.

2. Experimental design

For a given overall level, the experimental design required loudness matches to be determined in 44 different experimental conditions. These resulted from the factorial combination of $7(\text{test loudspeakers}) \times 3(\text{center frequencies}) \times 2(\text{adaptive starting levels})$, and additional two conditions of the reference loudspeaker being matched to itself for the 1-kHz center frequency only (using both starting levels) to obtain a measure of the baseline variability of the matches.

Collection of these data was organized as follows: In order to allow subjects to adapt to a given loudness range, "high-SPL" (65 dB; "A"), and "low-SPL" (45 dB; "B") measurements were strictly separated in different sessions, which were counterbalanced following a succession of *ABBA* (respectively, *BAAB*) schemes. The order of the 44 experimental conditions to be investigated at each level was randomized, and subsequently divided into blocks of eight (the remaining four being assigned to the next block, i.e., the following replication of the measurements). Thus, within a given block of trials, eight adaptive tracks were randomly interleaved on a trial-by-trial basis, providing some random sampling of loudspeaker locations, center frequencies, and starting levels. Consequently, it was impossible for the subjects to track the immediate "adaptive" consequences of their judgments, and from their perspective the task was just a succession of unrelated paired comparisons with respect to loudness.

Each listening session consisted of four such blocks (containing eight adaptive tracks each). Completing a block of trials took approximately 10 min. While it lasted, the subjects were instructed to sit as still as possible in order to maintain the alignment with the loudspeaker setup. A short break was taken after each first and third block in a session, and participants were allowed to move their heads and upper body during those breaks, but not to leave the chair. After each second block they had a longer break during which they

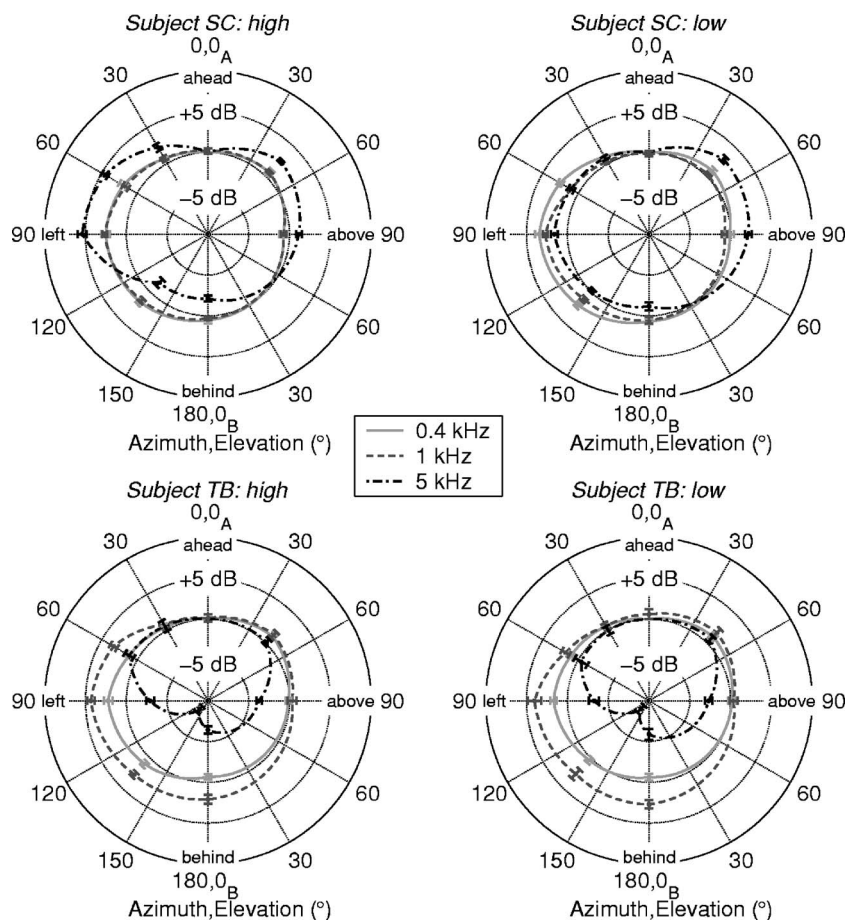


FIG. 2. Directional loudness sensitivities at the two overall levels for subjects SC and TB with 95% confidence intervals of 16 replications. The graphs on the left show the results for the high overall level (65 dB SPL), the graphs on the right for the low overall level (45 dB SPL). Elevations 0_A° and 0_B° are ahead and behind the listener, respectively.

left the anechoic chamber, thus requiring them to realign the seating position upon return. An entire session lasted approximately 1 h.

Since 16 replications of the matches (eight with each of the two adaptive starting levels) were collected per experimental condition, all subjects had to participate in 22 listening sessions. The participants completed a maximum of two sessions per day with a minimum of 1 h between sessions. With three additional sessions reserved for audiometry, HRTF measurements, and practice (one block in each of the high-SPL and low-SPL conditions), the total number of hours amounted to 25 per subject.

III. RESULTS

A. Directional loudness sensitivities

The adaptive procedure matched the loudness of a sound of a given center frequency coming from one of the loudspeakers in the horizontal or the median plane to the loudness of the same sound with frontal incidence. Thus, the raw data from the experiment were the sound-pressure levels (in dB SPL) the loudspeakers would have to be set to, in order to be perceived equally loud as the frontal reference. These raw data were averaged across the 16 repetitions that each participant accumulated in each condition, and normalized by subtracting the result from the fixed level of the respective frontal reference (65 or 45 dB SPL). That way, (relative) directional loudness sensitivities¹ were obtained, positive

values of which indicate loudness enhancement, i.e., a lower sound-pressure level required for that direction to achieve a match with the frontal reference.

1. Individual data

Individual directional loudness-sensitivity curves are depicted for two subjects, SC (upper panels) and TB (lower panels), representing extremes of performance, in Fig. 2. The data are rendered in polar coordinates, though in a particular, asymmetrical way: The left-hand side of each polar graph shows the data for the horizontal plane as the loudspeakers were physically positioned in the setup. On the right-hand side of each polar graph the data are shown for the median plane where the loudspeakers were actually above the subjects. Note that these two planes coincide ahead of and behind the subjects.

For subject SC, loudness matches at 0.4 and 1 kHz vary as a function of sound incidence angle over a range of approximately 3 dB, the subject being most sensitive to loudness for sounds coming from the side, i.e., from 90° to the left in Fig. 2. That holds for both overall levels used. At 5 kHz, by contrast, this pattern is observed at the high overall level only, whereas at the low level the loudness pattern is fairly omnidirectional in the horizontal plane. In the median plane the directional patterns are similar across overall levels.

For subject TB, loudness matches vary over a range of less than 3 dB at 0.4 kHz. At 1 kHz the direction has a larger

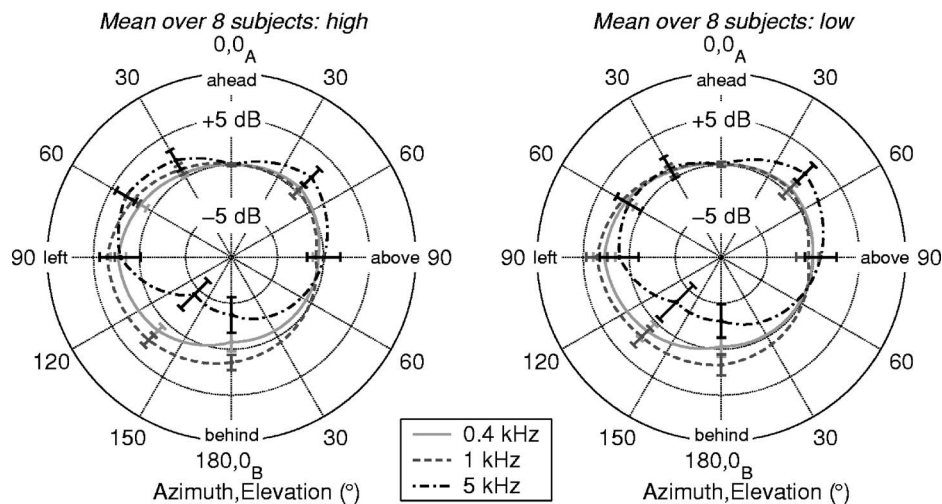


FIG. 3. Directional loudness sensitivities at the two overall levels for means across all eight subjects with 95% confidence intervals of the means. The graph on the left shows the results for the high overall level (65 dB SPL), the graph on the right for the low overall level (45 dB SPL). Elevations 0_A° and 0_B° are ahead and behind the listener, respectively.

effect on loudness, the sensitivity being up to 4 dB higher on the left-hand side than straight ahead. At 5 kHz the directional effect is even more pronounced, the minimum sensitivity at 135° in the horizontal plane being approximately 8 dB below and the maximum being close to the frontal sensitivity. The directional sensitivity patterns for this subject do not appear to be level dependent.

The confidence intervals for the matches of subjects SC and TB in Fig. 2 are small. Average individual standard deviations of the loudness matches across all subjects were 1.0 and 1.2 dB at the high and low overall levels, respectively. All participants adjusted the (identical) 1-kHz frontal test sound to a sound-pressure level close to the (fixed) reference, indicating that there was no systematic bias in the matches. The standard deviation of the identical-direction matches (0.9 dB) was only marginally lower than that of the across-direction matches, suggesting that these were of no greater difficulty.

2. Group data

Figure 3 shows mean loudness sensitivities when data are aggregated across all of the eight subjects. When the listener-specific idiosyncrasies are thus removed, directional loudness sensitivity still varies over some 3 dB at the two lowest center frequencies, whereas at 5 kHz the directional effect is approximately twice as large. Also, the error bars are larger at the highest center frequency due to a wider spread in the individual data. The overall level does not seem to have a marked effect on the patterns when considering the average data: the left and the right panels of Fig. 3 are hardly distinguishable.

The data and the subsequent analyses show that loudness is not constant over sound incidence angles, and the directional loudness-sensitivity patterns change considerably with center frequency, and to a lesser extent, with overall sound-pressure level.

3. Statistical analysis

The significance of the effects observed in the averaged data was confirmed by a $7 \times 3 \times 2$ (directions

\times center frequencies \times levels) repeated-measures analysis of variance (ANOVA) on the means obtained from each subject in each of the experimental conditions.

In addition to a significant main effect of direction, $F(6,42)=28.35$, $p<0.001$, indicating that directional loudness-sensitivity differences persist, even when collapsing across levels and frequencies, all its interactions were highly significant:

- (1) As expected, the direction \times frequency interaction produced the highest F value, $F(12,84)=31.29$, $p<0.001$, confirming that the way in which directional loudness varies is strongly frequency dependent (see Fig. 3). It should be noted that this interaction is also highly significant for each of the eight subjects when statistical analyses are done individually.
- (2) Furthermore, there is a significant direction \times level interaction in the pooled data, $F(6,42)=7.29$, $p<0.001$. Inspecting the average directional loudness sensitivities in Fig. 3, it appears that—ignoring center frequency—the directional effects on loudness are slightly more pronounced at the higher overall level (65 dB SPL).
- (3) More importantly, there is a three-way (direction \times frequency \times level) interaction, indicating that the frequency-dependent directional effects show a different pattern for the two overall levels, $F(12,84)=7.42$, $p<0.001$. This appears to be largely due to the 5-kHz data showing a slightly larger gain in sensitivity in front of the listener, and a slightly larger loss behind when comparing the high with the low overall level (see Fig. 3). Again, this interaction is significant for all of the eight subjects, even though the patterns show strong individual differences (see Fig. 2).

B. Head-related transfer functions

Individual head-related transfer functions were measured to investigate how sound is being filtered from a free field to the subjects' ears, depending on the angle of incidence. As an example, the HRTF magnitude spectra for subject IA from all eight directions are plotted in Fig. 4. Each panel depicts curves for the three separate sets of measurements made at different stages of the experiment. These mea-

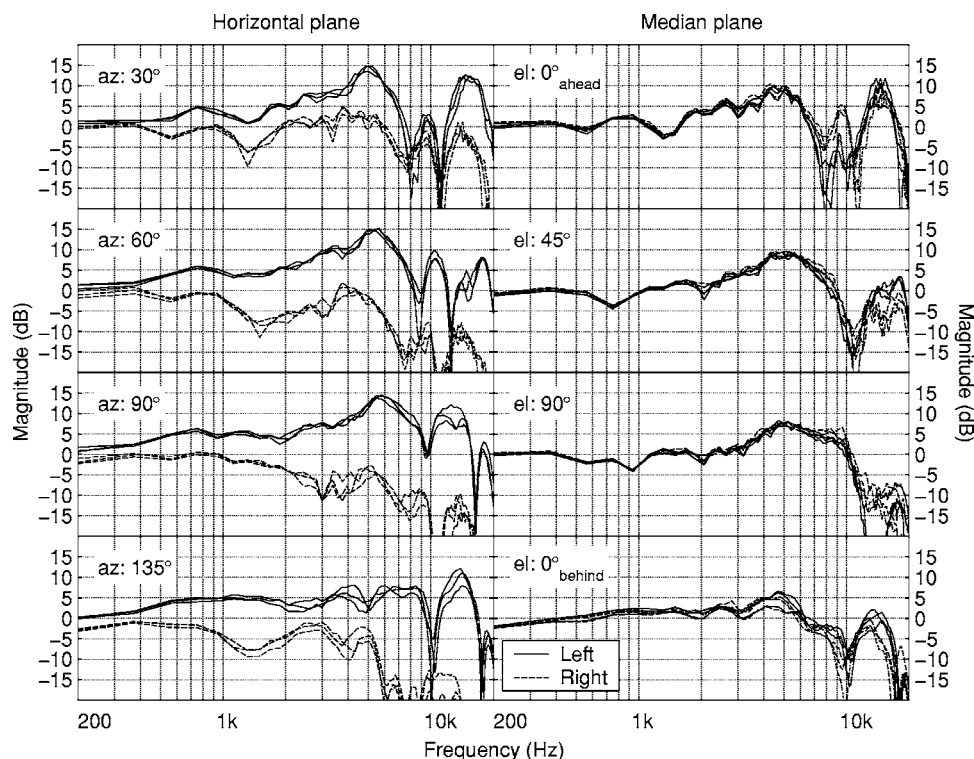


FIG. 4. Three HRTF measurements, performed at different stages of the experiment. The figure shows data for the left and right ears of a single subject (IA), for stimulation from all eight directions. The left panel depicts measurements obtained in the horizontal plane (azimuths of 30°, 60°, 90°, and 135°), the right panel those obtained in the median plane (elevations of 0°_{ahead}, 45°, 90°, and 0°_{behind}). 0°_{ahead} incidence is the frontal (reference) direction.

measurements include individual fitting and positioning of the microphones, aligning of the subjects to the listening position, calibration, and acoustic measurements. As seen in Fig. 4, the measurements are highly repeatable, the variation below 1 kHz on average being within ± 0.4 to ± 0.6 dB (comparable to e.g., Møller *et al.*, 1995).

Figure 4 also shows that the interaural level differences in the median-plane HRTFs are very small up to around 7 kHz. In the horizontal plane, however, HRTFs of the left and right ears differ considerably due to a pressure buildup at the ipsilateral ear and head shadowing at the contralateral ear, especially at high frequencies. For the fairly representative subject whose HRTFs are depicted in Fig. 4, the maximum magnitudes of the ipsilateral (left) ear in the horizontal plane are around 15 dB for azimuths from 30° to 90° (front-left side), while the magnitudes at the contralateral (right) ear are typically below 0 dB.

C. HRTFs and directional loudness

1. Calculating normalized at-ear exposure

In order to investigate the effects of the physical HRTF filtering on the directional loudness matches on an individual basis, the objective HRTF measurements and the subjective loudness data were combined. This was done in order to obtain the actual frequency-specific at-ear exposure, and to evaluate whether the peculiarities of individual HRTFs might account for some of the interindividual variation seen in the directional loudness matches. Note that this analysis was based on the magnitude spectra of the HRTFs, and that the effect of the interaural time difference was disregarded.

The individual HRTFs were averaged across the three repetitions by calculating the mean of the linear magnitude spectra. These means were then converted to the correspond-

ing third-octave-band levels in decibels. Finally, the left- and right-ear SPLs were normalized, for each incidence angle and at each center frequency, by subtracting the respective frontal left- and right-ear levels from them, since the loudness matches were always made to the frontal reference.

2. Relating loudness matches to HTRFs

For each of the eight participants, the normalized at-ear levels and directional loudness matches are combined in Fig. 5 and Fig. 6.

a. Horizontal plane. The combined data for the horizontal plane are plotted in Fig. 5. As seen in Fig. 5, in this plane the individual ILDs reach a maximum of 5 dB at 0.4 kHz, of 12 dB at 1 kHz, and of up to 30 dB at 5 kHz for the calculated third-octave-band at-ear SPLs.

For all subjects, except for subject IA at 0.4 kHz, the subjective directional loudness sensitivities at the high and low overall levels largely fall between the objective at-ear sound-pressure levels. It thus seems that the agreement between the two types of data is fair: For example, by considering the 5-kHz data for subjects TB and WE in Fig. 5, the idiosyncrasies in their at-ear SPLs are reflected in equally individual directional loudness sensitivities. However, the picture is not as clear when considering the two overall levels ("high" at 65 dB SPL and "low" at 45 dB SPL): Generally, the subjective data at the two overall levels are fairly congruous. In some cases, however, the most extreme case being subject SC at 5 kHz in Fig. 5, a clear overall level dependence can be observed.

If loudness were perceived as being constant over sound incidence angles, the subjective directional sensitivity data would follow the 0-dB horizontal in Fig. 5 or, equivalently, the 0-dB circles in Figs. 2 and 3. That would imply loudness

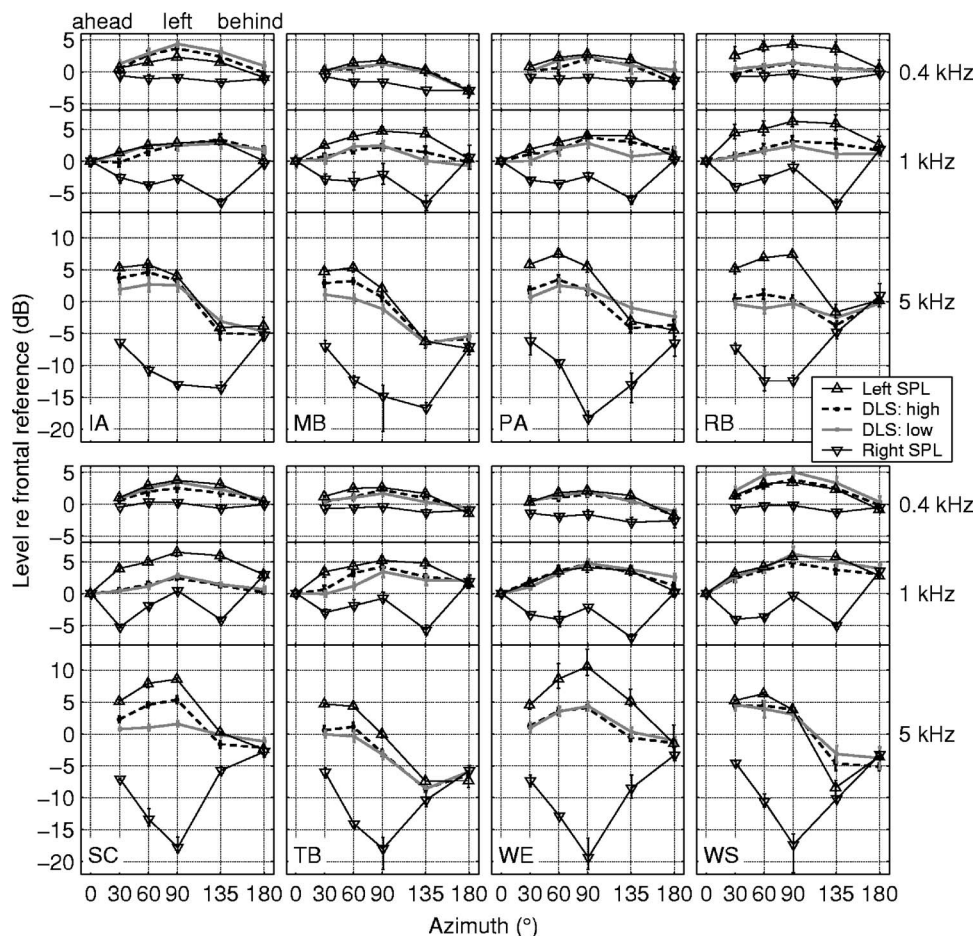


FIG. 5. Horizontal plane: Directional loudness sensitivities (DLS) at the high and low overall level (DLS high: 65 dB SPL and DLS low: 45 dB SPL) with 95% confidence intervals, along with left- and right-ear sound-pressure levels, plotted relative to the frontal reference (see the text).

to be governed solely by the sound-pressure level of the source measured in the absence of a listener, irrespective of the changes in the at-ear sound-pressure levels as a function of sound incidence angle. This does not seem to be the case for any of the data sets.

If, on the other hand, the subjective loudness data always followed the ear with the higher SPL, this would imply

no binaural loudness summation, i.e., loudness would be determined by the ear getting the higher input alone. Evidence for this kind of behavior may be seen in the data of IA, WS, and to some extent in those of WE and PA, though not at 5 kHz.

b. Median plane. In the median plane the ILDs are small, and the two ears are getting approximately the same

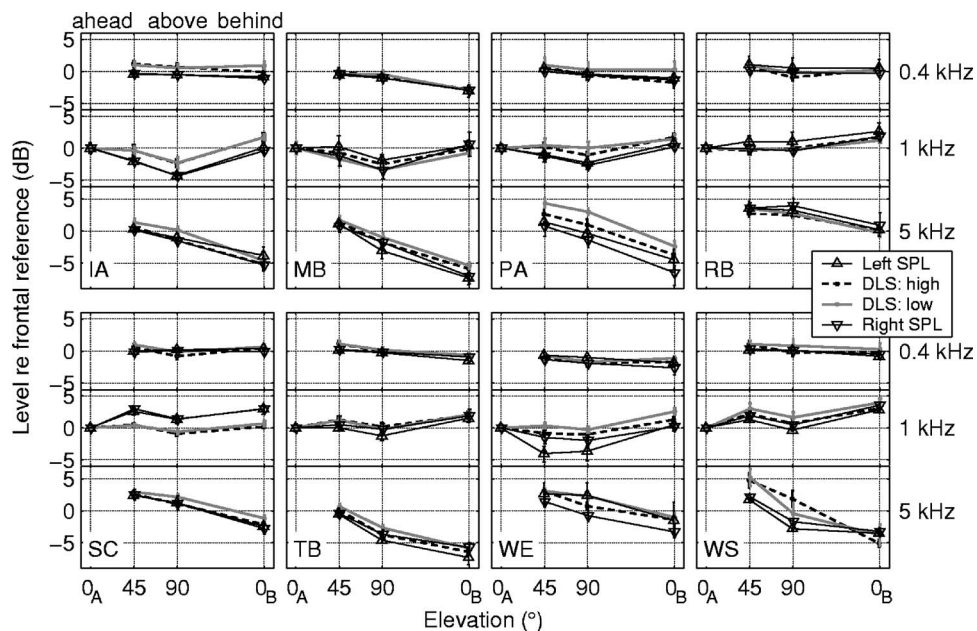


FIG. 6. Median plane: Directional loudness sensitivities (DLS) at the high and low overall levels (DLS high: 65 dB SPL and DLS low: 45 dB SPL) with 95% confidence intervals, along with left- and right-ear sound-pressure levels, plotted relative to the frontal reference (see the text). Elevations 0_A° and 0_B° are ahead and behind the listener, respectively.

input at all sound incidence angles; see Fig. 6. The differences between the ears are largest for subject WE, producing ILDs as large as 3 dB.

The normalized at-ear levels as a function of direction vary over less than 3 dB at 0.4 kHz, by up to 5 dB at 1 kHz, and over a range of almost 10 dB at 5 kHz. In this plane a change in the at-ear SPLs with incidence angle should presumably be reflected in a similar change in directional loudness sensitivity, which is true for most of the subjects. Occasional exceptions from this rule can be seen, however, for example for subjects SC and IA at 1 kHz, and for subject WS at 5 kHz.

c. Summary. Both in the horizontal and median planes, the patterns of the individual directional loudness sensitivities can largely be explained by directional effects the individual HRTFs have on at-ear sound-pressure levels. The way the subjects combine their left- and right-ear SPLs to a single loudness percept is further explored in the next section concerned with modeling binaural loudness.

IV. MODELING OF BINAURAL LOUDNESS

Large interindividual variation was found in subjects' directional loudness sensitivities. As seen in the previous section, these sensitivities exhibit systematic dependencies on the directional variations in individual HRTFs. Thus, a straightforward strategy in modeling binaural loudness is to take the HRTF effects into account, and to relate the physical changes in the at-ear signals—independent of direction—to the corresponding changes in loudness as perceived in a real sound field.

In the median plane, where the loudspeakers were positioned symmetrically with respect to the subjects' left and right ears, the listening situation was close to diotic. In this plane, the sound-pressure levels at the two ears were similar at the elevations under investigation (0_A° , 45° , 90° , and 0_B°); see Fig. 6. In such a situation, the actual amount of summation across the two ears has no effect on binaural modeling. This is due to the fact that the same binaural listening advantage takes effect both for the reference and the comparison to be matched. Note that the same applies for the traditional free- and diffuse-field loudness paradigms.

Dichotic stimulation, with different at-ear levels, thus constitutes the most interesting case for the modeling of binaural loudness. Dichotic at-ear SPLs were observed for the azimuths of 30° , 60° , 90° , and 135° in the horizontal plane (see Fig. 5). At these azimuths subjects typically had to match a dichotic sound to the diotic frontal reference.

Narrow-band stimuli were used in the listening experiment in order to simplify the modeling of binaural loudness, by being able to ignore spectral summation of loudness across critical bands. Also, assuming that perceived loudness is doubled when the listening is binaural (diotic) instead of monaural, a relationship between the psychophysical dimension of loudness (as measured in sones) and its physical correlate, the sound-pressure level (in dB SPL) can be established. By definition, a loudness of 1 sone is produced by a 40-dB SPL, 1-kHz tone, and doubling or halving loudness (in sones) corresponds to a 10-dB increment or decrement in

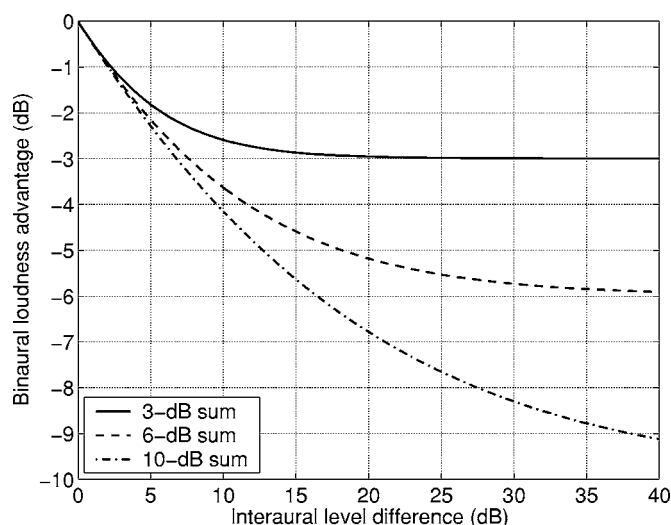


FIG. 7. Binaural loudness advantage as a function of interaural level difference; solid: 3-dB, dashed: 6-dB, and dash-dotted line: 10-dB summation rule.

sound-pressure level, respectively. Due to the shape of the equal-loudness contours (ISO 226, 2003), the increments within the range of sound-pressure levels used in the present experiment are approximately 10.5 and 9.5 dB SPL at 0.4 and 5 kHz, respectively, for a doubling of loudness. Thus, at all three center frequencies (0.4, 1, and 5 kHz) used in the present study, doubling in sones corresponds fairly closely to a 10-dB gain in sound-pressure level.

In order to illustrate how binaural loudness is affected by various interaural level differences, theoretical curves can be obtained utilizing Eq. (2), taken from Robinson and Whittle (1960). It is reasonable to assume that the summation of sound-pressure levels across the two ears is nonlinear, as suggested by Eq. (2): At large ILDs, the ear receiving the lower sound-pressure level presumably has little effect on overall binaural loudness, and the stimulation is effectively monaural. When approaching a diotic situation, however, the signals at the two ears tend to be weighted equally in contributing to overall loudness.

Theoretical curves for three hypothetical binaural loudness-summation rules are plotted as a function of the interaural level difference in Fig. 7. In addition to the 6-dB summation rule adopted from Robinson and Whittle (1960), two other curves were derived by changing the binaural gain factor in Eq. (2): A 3-dB summation rule corresponding to the “power summation” of the linear at-ear magnitude spectra, and a 10-dB summation rule, which for the stimuli used in the present study roughly corresponds to perfect binaural summation in sones.

The different curves in Fig. 7 are normalized so that they all coincide in the origin of the graph: it represents the diotic case with an ILD of zero. As the ILD increases, loudness decreases by different amounts, depending on the summation rule with the “loss” to be read from the ordinate corresponding to the “binaural loudness advantage” achievable by switching from dichotic to diotic stimulation. The 3-dB summation rule fairly quickly converges to the -3 -dB level in the graph: when the ILD increases beyond 15 dB, binaural loud-

TABLE I. Least-squares estimates for the amount of binaural loudness summation [x in Eq. (3), in dB], at the three center frequencies (f_c : 0.4, 1, and 5 kHz) at high (65 dB SPL) and low (45 dB SPL) overall levels. The two right-most columns show the best fits when pooling center frequencies, and the best fits across center frequencies when the data are averaged across subjects (bottom row). Extreme values are marked with stars (see the text for details).

Subject	f_c							
	0.4 kHz		1 kHz		5 kHz		Best fit across f_c	
	High	Low	High	Low	High	Low	High	Low
IA	0.1	0.1	0.1	0.1	0.7	1.5	0.1	0.1
MB	0.3	0.4	3.3	3.5	2.2	4.8	2.4	4.6
PA	1.0	0.4	0.7	1.3	3.1	3.0	2.1	2.4
RB	99.9*	99.9*	4.0	9.3	10.0	18.4*	9.1	17.6*
SC	1.6	0.3	13.1	29.7*	3.1	8.2	3.8	8.5
TB	1.1	2.8	2.1	3.0	3.7	5.0	3.3	4.6
WE	0.1	0.1	0.1	0.3	4.9	4.5	2.8	2.6
WS	0.1	0.1	1.7	0.5	0.1	0.1	0.7	0.1
Median	0.7	0.4	1.9	2.2	3.1	4.7	2.6	3.6
Averaged data							2.6	3.9

ness is no longer affected. At these ILDs loudness is determined by the ear with the higher sound-pressure level alone, and dichotic loudness is 3 dB lower than the corresponding diotic one. With the 6- or 10-dB summation rules, much larger ILDs are required until the curve asymptotes at -6 and -10 dB, respectively. For the 10-dB summation rule, at an ILD of 40 dB (far exceeding the ILDs observed in the present study) binaural loudness still continues to decrease.

A. Individual data

The third-octave-band at-ear sound-pressure levels computed from the HRTFs were used in the modeling, in order to find the best-fitting binaural summation rule to predict the directional loudness-sensitivity data. Robinson and Whittle (1960) reported their average data to support a 6-dB loudness summation across their listeners' ears [see Eq. (2)]. This type of modeling was explored for the present data, but on an individual basis. The modeling was carried out by relaxing the factor 6 in Eq. (2).

To that effect, the optimal amount of binaural loudness summation (x)—assumed to be fixed at 6 dB in Eq. (2)—was *estimated* by minimizing the sum-of-squares of the errors (SSE) between the actual directional loudness sensitivity (DLS) and the sensitivity predicted (L_{mon}) from the changes in at-ear sound-pressure levels using Eq. (3). All 16 (j) repetitions of each condition, and the mean at-ear sound-pressure levels for each of the four horizontal-plane angles of incidence (i ; 30° , 60° , 90° , and 135°) were included in the modeling, which was performed individually for each subject, and separately for the three center frequencies and the two overall levels.

$$\text{SSE} = \sum_{i=1}^4 \sum_{j=1}^{16} \{ \text{DLS}_{\text{high/low},ij} - [L_{\text{mon,comp}_i}(x) - L_{\text{mon,ref}}(x)] \}^2, \quad (3)$$

where

$$L_{\text{mon,comp}_i}(x) = x \times \log_2(2^{L_{\text{left,comp}_i}/x} + 2^{L_{\text{right,comp}_i}/x}), \quad (4)$$

and

$$L_{\text{mon,ref}}(x) = x \times \log_2(2^{L_{\text{left,ref}}/x} + 2^{L_{\text{right,ref}}/x}). \quad (5)$$

In these equations, $L_{\text{left,comp}}$ and $L_{\text{right,comp}}$ refer to the third-octave-band levels for the comparison incidence calculated from the individual left- and right-ear HRTFs, respectively. Likewise, $L_{\text{left,ref}}$ and $L_{\text{right,ref}}$ refer to the corresponding levels for the frontal reference at the left and right ears, respectively.

The subjective directional loudness sensitivities had been normalized to the frontal reference (see Figs. 2 and 5). Therefore, the predictions were normalized as well by subtracting Eq. (5) from Eq. (4) in the minimization of the sum of squares of the errors. Due to this normalization, the overall level (65 vs 45 dB SPL) does not have an influence on the predictions. The possible dependence of binaural loudness summation may nevertheless show up in the subjective directional loudness sensitivities at the high and low overall levels, and may thus influence the estimate of the variable x , the binaural gain estimated from the data. Forty-eight such estimates (for eight subjects, three center frequencies, and two overall levels) for the amount of binaural loudness summation are listed in Table I. The minimization algorithm was limited to a summation value between 0.1 and 99.9 dB.

As was already seen in Fig. 5, the amount of binaural loudness summation varies greatly across subjects, and also within subjects across the three center frequencies. The best-fitting binaural gain estimates roughly fall into three categories: The summation is minor (less than 1 dB) for 19, moderate (from 1 to 10 dB) for 24, and extreme (greater than 10 dB) for 5 out of the 48 cases analyzed. There is a tendency for the summation values to increase with center frequency, but due to the fact that the center frequencies are confounded with variations in ILDs, the comparison may not be fair.

The smaller the amount of binaural loudness summation, the more binaural loudness is determined by the ear getting a higher input. By contrast, the higher the summation value, the more influence the ear receiving the lower sound-pressure level has on binaural loudness. Some extreme values marked with stars in Table I, e.g., subject RB at 0.4 kHz, seem to imply the latter behavior. Closer inspection of Fig. 5, however, reveals that for this subject the directional loudness sensitivity remains close to the 0-dB line, even if the at-ear sound-pressure levels vary over a fairly wide range. As Robinson and Whittle (1960) pointed out, the actual value of the summation parameter (at the natural ILDs in question) does not have a great effect on the directional loudness sensitivities predicted from the at-ear SPLs. For these reasons the minimization algorithm can reach very high summation values (up to the limit of 99.9 dB) when searching for the best fit. However, it is unrealistic that the binaural gain (i.e., the loudness match between monotic and diotic stimulation) for a normal-hearing subject is much larger than 10 dB.

To get a more stable estimate, the amount of binaural summation was also determined by pooling across the three center frequencies; see the two right-most columns in Table I. This was achieved by aggregating the data across center frequencies, and finding the best-fitting summation rule to the aggregated data set. The individual differences are still retained, and the summation values again fall into the three categories defined above.

In order to deal with the variance inherent in the subjective data, a partial F -test (Bates and Watts, 1988, Chap. 3) was performed to investigate whether the subjects summed their at-ear levels in significantly different ways. In a “restricted” model *one* least-squares fit of binaural loudness summation [x in Eq. (3)] common to all subjects was estimated, whereas in a “full” model the summation value was relaxed to estimate *different* parameters for the eight subjects. The data were aggregated across incidence angles, overall levels and center frequencies. The partial F -test showed that the error sum of squares between the subjective data and the estimate was significantly larger for the restricted model having a common parameter for all subjects [$F(7,3064)=211.58$; $p<0.001$]. Therefore, the full model allowing for individually different binaural-gain parameters predicted the data better, and hence, the differences in the way the subjects summed the at-ear levels appear to be significant.

B. Group data

The individual third-octave-band HRTFs and directional loudness sensitivities were averaged across subjects, to make an estimate for the mean data thus obtained. Aggregating over center frequencies, as before, the best fits for the averaged data came fairly close to suggesting a 3-dB summation rule both at the high and the low overall level (see the bottom row of Table I).

Thus far the prediction was entirely based on the at-ear sound-pressure levels at the center frequency of the narrow-band noises used. However, by using a loudness model, the possible spread of excitation to neighboring critical bands

can be taken into account in the modeling. Furthermore, given a relatively large dynamic range, the shape of the loudness function may be better accounted for when using a loudness model.

Therefore, the most widely accepted loudness model by Moore *et al.* (1997) was tested in predicting the present data. This model facilitates the use of eardrum pressures for loudness computations, i.e., using at-ear signals as a product of the HRTF-filtering stage. The model also predicts monaural loudness, by assuming perfect loudness summation in sones between the two ears, and calculating monaural loudness simply as being one half of the binaural, diotic loudness. Dichotic loudness can then be computed as a sum of the two monaural loudness values in sones.

Since the HRTFs of the present study had been measured at the entrance to the blocked ear canal, a direction-independent transfer from the measurement point to the eardrum (mean P_4/P_2) was adopted from Fig. 13 in Hammershøj and Møller (1996). In contrast to the summation rule explored in the previous section, here absolute binaural loudness values were computed. The effects of the HRTFs were taken into account, as before, but now the entire at-ear spectra were included (instead of only using the level at the center frequency). The input data to the loudness model thus were third-octave-band spectra based on the measured stimulus spectrum in the absence of a listener (P_1), combined with the left- and right-ear HRTFs (P_2/P_1), and corrected by the eardrum-to-the-measurement-point transfer function (P_4/P_2).

Monaural loudness values were computed for (dichotic) left- and right-ear signals, subsequently summed, and compared to the loudness produced by the (close to) diotic frontal reference. First, binaural loudness values for each of the frontal reference stimuli were computed, as described above. Then, some values for the comparison directions were computed by varying the level of the P_1 pressures, within the range of ± 10 dB from the frontal reference level, in steps of 0.5 dB. The P_1 sound-pressure levels yielding the binaural loudness values closest to that of the frontal reference were selected. In this way the loudness model was used to find equal-loudness sound-pressure levels for each incidence angle, including the effects of the HRTFs. The inverses of these sound-pressure levels relative to the frontal reference were taken as the directional loudness sensitivities predicted by the model.²

Figure 8 contrasts the predictions made by the loudness model (Moore *et al.*, 1997) with the 3-dB power summation, which fared best in the earlier analysis. Since the effect of overall sound-pressure level on directional loudness was minor for the averaged data (see Fig. 3), only the high-level (65 dB SPL) directional sensitivities are plotted.

For all dichotic situations (horizontal plane, left column in Fig. 8), the 3-dB summation rule predicts the obtained mean loudness-sensitivity data quite well. At each center frequency, the patterns of the 3-dB prediction and the actual matches made are congruous, and only in two instances (at 0.4 kHz, azimuths of 90° and 135° in Fig. 8) do the 95% confidence intervals of the subjective data not include the 3-dB prediction. By contrast, the prediction of the loudness

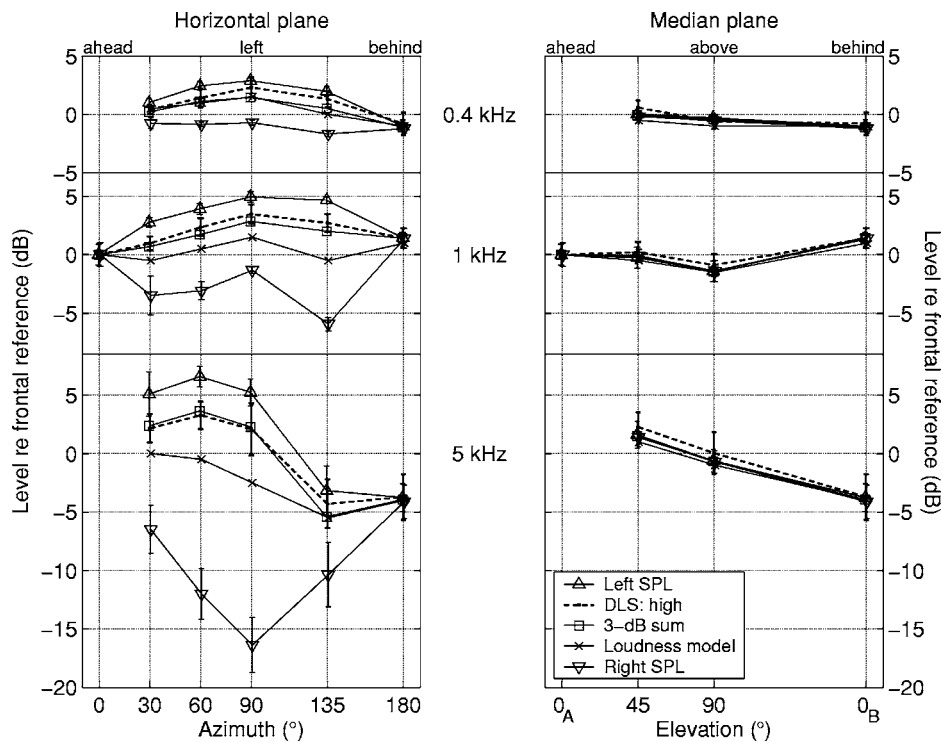


FIG. 8. Average left- and right-ear sound-pressure levels, a 3-dB summation rule, loudness summation in sones, and obtained average DLS at the high overall level (65 dB SPL). The error bars denote the 95% confidence intervals of the means across the eight subjects. The left panel depicts data and predictions for the horizontal plane, the right panel for the median plane. Elevations 0_A° and 0_B° are ahead and behind the listener, respectively.

model markedly deviates from the obtained directional loudness sensitivities, particularly at the two higher center frequencies (1 and 5 kHz). These are the situations in which the interaural level differences range from 6 to over 20 dB. For these ILDs, the prediction is not bracketed by the confidence intervals of the data for seven out of the eight dichotic conditions, the difference between data and predictions reaching up to 5 dB (5 kHz, azimuth 90° in Fig. 8). It thus seems that the 3-dB summation rule of at-ear sound-pressure levels predicts the directional loudness of dichotic sounds considerably better than the assumption of perfect binaural loudness summation in sones.

In the median plane, all five curves (at-ear levels, directional loudness sensitivities, and model predictions) are nearly indistinguishable; see the right panels in Fig. 8. The 95% confidence intervals of the subjective data include both the physical changes in left- and right-ear sound-pressure levels, and the predictions of 3-dB sum and loudness summation in sones. Obviously, the diotic stimulation condition does not provide a critical test for these models.

V. DISCUSSION

A. Comparison with previous work

When comparing the present results to the work of Robinson and Whittle (1960), it may be observed that the average directional effect sizes they obtained are comparable to those measured in the present study: For the incidence angles presented here, the average directional loudness sensitivities Robinson and Whittle (1960) obtained at center frequencies below 6.4 kHz varied from -6.5 to $+5.0$ dB (see their Fig. 2) relative to the frontal reference level. The corresponding range for the average data in the present study is -4.3 to

$+3.5$ dB (see Fig. 3), although the actual stimulus center frequencies used differed somewhat between the two investigations.

In both investigations, direction had a smaller effect on loudness at lower center frequencies, and the effect increased with stimulus center frequency. Qualitatively, this can be explained by the fact that with increasing frequency the physical dimensions of a listener start to obstruct the sound field. The obstruction also becomes more direction dependent at higher frequencies (as can be seen in the sample HRTFs plotted in Fig. 4), and this is reflected in its increasing effect on the directional loudness sensitivities.

The present empirical data collection, however, goes beyond previous work by reporting individual analyses. Consequently, and as expected from research on HRTFs, idiosyncratic directional loudness-sensitivity patterns were found. The individual data also showed that all participants were highly consistent in their judgments, even though the loudness of two sounds coming from different directions, and typically having different timbres, had to be compared.

The consistency in the participants' directional loudness matches provided considerable statistical power. On the one hand, that means that the significance of the major frequency-dependent effects of the direction of incidence on perceived loudness may be ascertained with great confidence. On the other hand, that entails that even small effects on the range of 1–2 dB level will emerge as statistically significant, and thus require further interpretation. That is the case for the effects of overall presentation level,³ and its interaction with the directional and frequency-specific effects.

Comparison of both individual data (e.g., Fig. 2, top row) and of the group averages (Fig. 3) shows a tendency for the frequency-dependent directional effects to become more

pronounced with increasing level. Likewise, small but systematic level effects are found when trying to estimate the amount of binaural gain from the data (Table I). Contrary to what is reported in the literature (Shaw *et al.*, 1947; Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963; Scharf and Fishken, 1970), this gain appears to be smaller at the higher overall level. That may be due to the low-level directional sensitivities being less affected by the ear receiving the greater input than the high-level directional sensitivities (see Fig. 5). Due to the small magnitudes of the overall-level effects, the present authors consider them to be negligible for most practical purposes, at least in the midlevel range investigated here (45–65 dB SPL).

Furthermore, the relatively low binaural-gain parameter derived from the present data is in conflict with the outcome of most of the classical studies (such as Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963; Marks, 1978; Zwicker and Zwicker, 1991, among others) employing headphones, and largely focusing on monotic-to-diotic comparisons. But, note that—apart from other methodological distinctions—a key feature of these earlier studies is that signals may have been generated that would never reach the two ears when being emitted by a real source positioned in space, and fail to produce an externalized auditory event. It is unclear whether the results of the two paradigms (binaural loudness summation versus directional loudness) can be compared directly, since the auditory events produced are so drastically different. The directional loudness paradigm, however, is not only closer to “real-world” stimulation, but also to the application of measuring sound fields using a dummy head, where the signals at the ears of the dummy are due to the physical obstruction in the sound field.

B. Individual differences

Even though tentative general conclusions on the computation of binaural loudness may be drawn from the present data, it is striking how large the interindividual differences in loudness matches (see Figs. 2 and 5), and hence, in directional loudness sensitivity are when comparing the eight listeners participating. The original hope, that *all* of this interindividual variance might be accounted for by the equally large differences in individual HRTFs (e.g., Fig. 4) does not seem to be warranted, as is evident from our analysis of individual “summation rules” displayed in Table I. Obviously, using the actual at-ear sound-pressure levels rather than the levels emitted by the loudspeakers in the analysis still leaves us with considerable residual individual variance.

Several potential reasons for that variance might be explored: An obvious reason may be that the third-octave-band levels derived from the HRTF measurements do not reflect the actual at-ear stimulation well enough. However, the quality of the HRTFs may be examined by contrasting the present results with data obtained in individual binaural synthesis where the directional sound sources are recreated via virtual acoustics, the crucial difference being that the at-ear levels are precisely known in that situation. Performing such an experiment with six listeners from the original sample of eight (Sivonen *et al.*, 2005), we found no appreciable, or

statistically significant, differences between the two sets of data (real vs virtual sound field). Rather, the individual differences remained, leading us to look for factors other than differences in the physical shape of pinnae, heads, and torsos.

A more speculative explanation for the individual differences found might be that the participants exhibited different degrees of “loudness constancy” in our experimental setup. The notion of “perceptual constancy” refers to situations in which a percept remains constant despite profound changes in the physical stimulation affecting the sensory receptors (Zahorik and Wightman, 2001). Typically, loudness constancy is observed when the loudness of a source (e.g., a musical instrument, a human voice) is judged to remain constant, even though its distance to the observer is varied. Stretching this notion somewhat, we might also speak of loudness constancy when listeners judge sounds to be equally loud, despite variations in their angle of incidence (which greatly affects the at-ear stimulation). It might be speculated that observers have learned to deconvolve the signals with their HRTFs in order to infer the loudness at the source.

Do the present data show evidence for loudness constancy defined in this way? The answer is clearly negative: Note that perfect constancy would mean that all of the identical-distance, identical-power sources used in the present experiment should be judged to be equally loud, i.e., the matches should fall on the 0-dB (reference) circle in Fig. 2, or on the 0-dB horizontal in Figs. 5 and 6. That, obviously, is not the case. Nevertheless, subjects might have a tendency to preserve constancy to varying degrees, thus producing different amounts of bias towards the zero line. Potentially, they could do so by using the localization and timbre cues available, as well as the fact that the loudspeakers producing the sounds are in plain view.

The constancy problem is related to that of the “listening attitude” a participant might adopt: In a pioneering investigation of loudness constancy (Mohrmann, 1939), this was operationalized as judging hidden sources at various distances while either adopting a sender attitude (“Sendereinstellung;” p. 155), or a receiver attitude (“Empfangseinstellung”) which yielded appreciably different results. In modern terminology one would refer to judging the distal stimulus vs the proximal stimulus, and in the present situation that would be equivalent to judging the sound power of the loudspeaker as opposed to judging how it affects the listener. It is unclear, however, whether subjects can make that distinction in an anechoic situation, and the present authors know of no published reports implementing the instructional variations required.

Nevertheless, it may safely be said that a “bias” towards constancy can only play a minor role in accounting for the present data. The fact that knowing the individual HRTFs goes such a long way in accounting for the idiosyncrasies seen in the matches argues against constancy being a major factor in directional loudness perception, at least for the synthetic sounds and the anechoic environment studied here.

VI. CONCLUSIONS

- (1) Loudness matches obtained with narrow-band noises in an anechoic environment showed that loudness is not constant over sound incidence angles. Rather, directional loudness sensitivities varied by up to 10 dB in individual, and up to 8 dB in averaged data.
- (2) The directional effects on loudness showed considerable dependency on center frequency, with greater directional effects being observed at higher center frequencies, and to some extent on the overall sound-pressure level of the stimuli.
- (3) Large, but highly reliable individual differences in directional loudness perception were observed.
- (4) The individual patterns of directional loudness could largely be accounted for by the corresponding changes in physical stimulation, as determined by head-related transfer functions (HRTFs).
- (5) These transfer functions were utilized for modeling binaural loudness based on the at-ear sound-pressure levels encountered. A 3-dB binaural loudness-summation ("power-summation") rule predicted the obtained mean data best, but sizable interindividual differences remained, even after the effect of individual HRTFs was taken into account.
- (6) The results can be used for predicting loudness in any type of sound field (be it free, diffuse, or directional, resulting in diotic or dichotic at-ear signals) using a dummy head.

ACKNOWLEDGMENTS

The authors would like to thank Claus Vestergaard Skipper for his help in setting up the apparatus used in the present experiment. Henrik Møller and Dorte Hammershøi are thanked for their advice on how to carry out HRTF measurements. Pauli Minnaar helped in analyzing the HRTFs and with including them in the modeling, Florian Wickelmaier advised us with the statistical analyses. We would further like to thank Brian C.J. Moore for giving us access to and advice in utilizing his loudness model. Finally, we would like to thank the participants for their patience and devotion donated to extensive and at times exhaustive listening tests. This research was carried out as part of the "Centercontract on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and DELTA Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

¹Directional loudness sensitivities are thus defined in loose analogy to the directivity characteristics of microphones (Beranek, 1986, Chap. 6). Being the inverse of the relative sound-pressure level required to produce a loudness match, they are—despite the similarity in terms—not related to sensitivity parameters (such as d') as conceptualized in signal detection theory (Green and Swets, 1988).

²These predictions were made both for the individual and the mean data, essentially yielding the same conclusions. Thus, only the results for the mean data are presented here.

³Even though a 20-dB range may not appear sufficient to investigate the effects of overall level, note that when considering the extra headroom required for the adaptive starting values (± 10 dB), and HRTF effects boosting or attenuating levels by approximately the same amount, the effective range listeners were exposed to in the experiment was quite large, covering what can be handled in a loudness-matching experiment without encountering floor ("too soft") or ceiling ("too annoying") problems.

- Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications* (Wiley, New York).
- Beranek, L. L. (1986). *Acoustics* (American Institute of Physics, New York).
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
- Bodden, M. (1997). "Instrumentation for sound quality evaluation," *Acust. Acta Acust.* **83**, 775–783.
- Bronkhorst, A. W. (1995). "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.* **98**, 2542–2553.
- Green, D. M., and Swets, J. A. (1988). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA).
- Hammershøi, D., and Møller, H. (1996). "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.* **100**, 408–427.
- Hellman, R. P., and Zwischlocki, J. (1963). "Monaural loudness function at 1000 cps and interaural summation," *J. Acoust. Soc. Am.* **35**, 856–865.
- ISO 226 (2003). "Acoustics—Normal equal-loudness-level contours," (International Organization for Standardization, Geneva, Switzerland).
- ISO 389-1 (1998). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," (International Organization for Standardization, Geneva, Switzerland).
- ISO 389-7 (1996). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions," (International Organization for Standardization, Geneva, Switzerland).
- ISO 532 (1975). "Acoustics—Method for calculating loudness level," (International Organization for Standardization, Geneva, Switzerland).
- Jesteadt, W. (1980). "An adaptive procedure for subjective judgments," *Percept. Psychophys.* **28**, 85–88.
- Jørgensen, S. L. (2002). "MATLAB model for loudness perception—directional loudness," Unpublished masters thesis, Technical University of Denmark (DTU), Copenhagen.
- Kuhl, W., and Westphal, W. (1959). "Unterschiede der Lautstärken in der ebenen Welle und im diffusen Schallfeld" ("Differences in loudness levels between a plane wave and a diffuse sound field,") *Acustica* **9**, 407–408.
- Levelt, W. J. M., Riemersma, J. B., and Bunt, A. A. (1972). "Binaural additivity of loudness," *Br. J. Math. Stat. Psychol.* **25**, 51–68.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Marks, L. E. (1978). "Binaural summation of the loudness of pure tones," *J. Acoust. Soc. Am.* **64**, 107–113.
- Mohrmann, K. (1939). "Lautheitskonstanz im Eutfernungswechsel" ("Loudness constancy with changing distance,") *Z. Psychol.* **145**, 145–199.
- Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.* **43**, 300–321.
- Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (1996). "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.* **44**, 451–469.
- Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.* **45**, 224–239.
- Olesen, S. K., Plogsties, J., Minnaar, P., and Møller, H. (2000). "An improved MLS measurement system for acquiring room impulse responses," in *Proceedings of NORSIG 2000, IEEE Nordic Signal Processing Symposium*, Kolmården, Sweden, 117–120.
- Remmers, H., and Prante, H. (1991). "Untersuchung zur Richtungsabhängigkeit der Lautstärkeempfindung von breitbandigen Schallen" ("Investigation of the directional dependence of the loudness of broadband sounds"), *Fortschritte der Akustik-Deutsche Arbeitsgemeinschaft für Akustik (DAGA)*, 537–540.
- Reynolds, G. S., and Stevens, S. S. (1960). "Binaural summation of loudness," *J. Acoust. Soc. Am.* **32**, 1337–1344.
- Robinson, D. W., and Whittle, L. S. (1960). "The loudness of directional sound fields," *Acustica* **10**, 74–80.
- Robinson, D. W., Whittle, L. S., and Bowsher, J. M. (1961). "The loudness

- of diffuse sound fields," *Acustica* **11**, 397–404.
- Scharf, B., and Fishken, D. (1970). "Binaural summation of loudness: Reconsidered," *J. Exp. Psychol.* **86**, 374–379.
- Schneider, B. A., and Cohen, A. J. (1997). "Binaural additivity of loudness in children and adults," *Percept. Psychophys.* **59**, 655–664.
- Shaw, E. A. G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *J. Acoust. Soc. Am.* **56**, 1848–1861.
- Shaw, W. A., Newman, E. B., and Hirsh, I. J. (1947). "The difference between monaural and binaural thresholds," *J. Exp. Psychol.* **37**, 229–242.
- Sivian, L. J., and White, S. D. (1933). "On minimum audible sound fields," *J. Acoust. Soc. Am.* **4**, 288–321.
- Sivonen, V. P., Minnaar, P., and Ellermeier, W. (2005). "Effect of direction on loudness in individual binaural synthesis," in *Proceedings of the 118th Convention of the Audio Eng. Soc.*, Convention Paper No. 6512.
- Wightman, F. L., and Kistler, D. J. (1989a). "Headphone simulation of free field listening. I. Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858–867.
- Wightman, F. L., and Kistler, D. J. (1989b). "Headphone simulation of free field listening. II. Psychophysical validation," *J. Acoust. Soc. Am.* **85**, 868–878.
- Zacharov, N., Tuomi, O., and Lorho, G. (2001). "Auditory periphery, HRTF's and directional loudness perception," in *Proceedings of the 110th Convention of the Audio Eng. Soc.*, Convention Paper No. 5315.
- Zahorik, P., and Wightman, F. L. (2001). "Loudness constancy with varying sound source distance," *Nat. Neurosci.* **4**, 78–83.
- Zwicker, E., and Zwicker, U. T. (1991). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," *J. Acoust. Soc. Am.* **89**, 756–764.



Audio Engineering Society Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Effect of direction on loudness in individual binaural synthesis*

Ville Pekka Sivonen^{1,2}, Pauli Minnaar^{1,*}, and Wolfgang Ellermeier¹

¹Department of Acoustics, Aalborg University, 9220 Aalborg Ø, Denmark

²Brüel & Kjær Sound & Vibration Measurement A/S, 2850 Nærum, Denmark

*Now at AM3D A/S, 9200 Aalborg, Denmark

Correspondence should be addressed to Ville Pekka Sivonen (vps@acoustics.aau.dk)

ABSTRACT

The effect of sound incidence angle on loudness was investigated using individual binaural synthesis. In the synthesis, individual head-related transfer functions and headphone equalization were used. Acoustical measurements were carried out in order to verify the binaural synthesis. In a listening test, narrow-band noises at various center frequencies synthesized from different directions were presented to the listeners. Their task was to match the loudness of these stimuli in an adaptive procedure to a reference noise synthesized with frontal incidence angle. Considerable variation in the directional loudness matches between center frequencies and listeners was obtained. The results were compared to an earlier study using the same experimental design, but stimuli being played back over loudspeakers. The comparison shows that the patterns of directional loudness are well retained in the binaural synthesis, the difference between playback over loudspeakers and the binaural synthesis in terms of directional loudness perception being insignificant.

1. INTRODUCTION

In order to *objectively* measure the loudness of a given sound field, the conventional method requires a single microphone to be placed in a position, where the center of a listener's head would be. The mea-

surement itself, carried out with an omni-directional microphone, is *independent* of the incidence angle of a sound.

Loudness paradigms exist for two types of sound fields: For the *free field*, where the incidence angle is frontal to the listener, and for the *diffuse field*, where the sound is reaching the listener's ears with equal intensity from all directions. Even if an acoustical

*Based on the comments of the PhD committee, minor changes have been made to this manuscript since its publication in the conference proceedings.

signal measured in the absence of a listener was the same in these two sound fields, the signals *at the ears* of the listener would be essentially different. This is due to direction-dependent scattering caused by the torso, head and pinnae of a listener. The scattering can be determined by measuring head-related transfer functions (HRTFs), which represent the filtering of sounds from a free field to a measurement point at the ears of the listener.

Loudness models, e.g. the standardized Zwicker loudness [1] and a revised model by Moore et al. [2], are available for loudness computations for the free and diffuse sound fields. The input signal applied to these models is typically a signal measured in the absence of a listener. Hence, the effect of the frontal or the diffuse-field HRTFs are included in the models. The latter model also facilitates the use of the signals at the ear-drums of a listener for loudness computations. This is achieved by applying inverses of the frontal, or the diffuse-field HRTFs to the loudness model, effectively separating the effect of HRTF-filtering from the model. Thus, the listener's own HRTFs (which are already in the signals at the ear-drums) can substitute the HRTF-effects in the loudness model. Many artificial-head measurement applications make use of this principle for loudness computations.

A major drawback of the free- and diffuse-field loudness paradigms is that they are based on listening with similar long-term spectra at the two ears. Since the acoustical measurement is made with a single microphone, these paradigms do not address the issue of how the two signals (e.g., different spectra) at the ears of a listener are summed to yield a single loudness percept. The common implementation is, that "monaural" loudness is simply a half of the loudness of the corresponding *diotic* (binaural) stimulation in Sones.

In a real sound field, however, the signals at the two ears of a listener often differ from one another, the stimulation of the auditory system being *dichotic*. Dichotic loudness perception has been investigated, but typically using artificial stimuli over headphone playback without *spatial* information, see e.g. [3]. In a real sound field, where sound signals at the ears of a listener are scattered depending on the sound incidence angle, and the individual physical characteristics of the listener, dichotic loudness has not

been fully investigated.

1.1. Investigations of directional loudness

In a directional sound field, the stimulation of the auditory system can be either diotic or dichotic, depending on the incidence angle of a sound. *Directional loudness*, as we define it, refers to studies, where the sound incidence angle is one of the independent experimental variables, and its effect on perceived loudness of test sounds is investigated in a listening test.

In the laboratory, the most straightforward way of implementing a directional listening situation is to use similar loudspeakers placed at various incidence angles equidistantly from the listening position. Furthermore, if the loudspeakers are set up in an anechoic chamber with little or no reflections, direction-dependent effects caused only by the HRTFs influence the stimuli at the ears of a listener. The experimental setup is typically calibrated so that in the absence of a listener the same sound pressure level (SPL) is measured from all incidence angles. Note that the free- or the diffuse-field loudness paradigms would compute the loudness of a setup of this kind to be constant over incidence angles.

Robinson and Whittle [4] investigated the effect of sound incidence angle on loudness with the paradigm described above. Using narrow-band stimuli, the subjective loudness matches (within one stimulus frequency) *between* incidence angles showed considerable dependency on direction, varying over a range of up to 20 dB. In addition to the directional effects, the patterns of matches were dependent on the stimulus center frequency, the variation increasing with frequency. Being well ahead of their time, Robinson & Whittle also measured the effective signals at the ears of the listeners, and correlated the measurements to the loudness matches. However, they only reported the relationship between the directional loudness matches and a summation of the effective at-ear signals *averaged* across listeners. This can be considered a drawback, as the directional changes in the at-ear signals are very dependent on the listener. In addition, proper HRTF-measurement techniques were not available at the time of their pioneering study.

Only few other studies have addressed the issue of directional loudness. Typically, wideband stimuli have

been used, which may wash out possible smaller directional effects [5, 6], and the experimental design has been rather application-oriented in scope, such as achieving an automated level alignment for multi-channel loudspeaker reproduction in reverberant environments [7].

In order to avoid these deficiencies, a study was recently conducted investigating the effect of sound incidence angle on loudness on an *individual* basis [8]. Selected results of that study will be presented here, since they constitute the baseline for the experiment to be reported. In the earlier study [8], eight listeners matched narrow-band stimuli, played back over loudspeakers at various directions in an anechoic chamber, in loudness. In addition to obtaining directional loudness matches, individual HRTFs from each loudspeaker location were also measured.

The obtained subjective loudness matches showed considerable direction-dependency, and the patterns varied over stimulus center frequencies of the narrow-band noises. In addition, the matches also varied considerably *between* listeners, especially at the highest center frequency. For each stimulus center frequency, the listener's patterns of directional loudness matches could be largely accounted for by the corresponding changes in their individual HRTFs. However, some inter-individual variation was still retained, even though the effects of the HRTFs on at-ear sound pressure levels were taken into account.

1.2. Introduction to binaural synthesis

Binaural synthesis is a part of what has been called binaural technology [9, 10]. It refers to *synthesizing* essentially the same signals, as would be received at the ears of a listener in a real sound field, using headphones.

For individual synthesis, HRTFs for each sound incidence angle, and headphone transfer functions (PTFs) at the same microphone positions as for the HRTF-measurements, need to be measured for each listener. The signals at the ears of a listener are then synthesized using the measured HRTFs, while the effect of the headphones on the signals is equalized by using the inverse of the PTFs.

The success of binaural synthesis has been investigated in the localization [11, 12] and the discriminability [13] of real and virtual sound sources.

These investigations show that with binaural synthesis a localization performance comparable to real sound sources is preserved, and that the origin of the sounds (real vs. virtual) is indiscriminable.

In this study, individual binaural synthesis is utilized in obtaining directional loudness matches for the same experimental conditions as in the earlier (loudspeaker) study [8].

1.3. Goal of present investigation

The aim of the present study is to investigate directional loudness using individual binaural synthesis. In the synthesis, effectively the same stimuli, as in the real sound field using loudspeakers [8], are synthesized to the listeners using headphones.

The attempt is to verify the effect of sound incidence angle on loudness, and to compare the directional loudness results with the results of the earlier study [8]. In addition, with the help of the results of the present investigation, the effect of the changes in individual HRTFs on directional loudness perception can be further explored.

The results can also be used in further experiments, when investigating the effect of individual HRTFs on loudness perception. In a real sound field, the HRTFs are dependent on the individual physical characteristics of a listener and can not be changed or modified. In binaural synthesis, however, this is possible. Furthermore, for directional loudness experiments, the control of the stimulation of the auditory system is better preserved in binaural synthesis. In the synthesis, the HRTFs are a part of the stimuli, and the synthesis is immune to (unwanted) head movements. In a real sound field, however, the HRTFs change even with small movements of the listener's head. Therefore, less variance in the subjective data might be exhibited using binaural synthesis.

The experimental method of the present study was kept as similar as possible to the earlier study [8]. The essential difference between the two studies was that sound reproduction over headphones in binaural synthesis, instead of over loudspeakers in the free field, was used.

2. METHOD

2.1. Subjects

Six subjects (between 23 and 47 years of age), all

of whom had participated in the earlier reference study [8], took part in the experiment. The subjects had been audiometrically tested for normal hearing, and their individual HRTFs had been measured [8]. Individual PTFs had also been measured in connection with the HRTF-measurements, see Section 2.2.2 for details.

2.2. Apparatus

2.2.1. Listening test

For the listening test, the subjects were seated inside an anechoic chamber in a chair with a head rest. The free-field loudspeaker setup of [8] was present in the anechoic chamber to visually increase the plausibility of the binaural synthesis. For each listening session, the chair was carefully adjusted so that the center of the subject's head was aligned with the center position of the loudspeaker setup, and that the orientation of the subject relative to the loudspeakers was as in [8].

The test sounds were played back over a pair of circumaural headphones (Beyerdynamic DT-990). The subjects were instructed to adjust the headband of the headphones for a good fit every time they put the headphones on. A two-button response box was used in collecting the subjects' responses to the sounds, and two red lights marking the observation intervals were placed in front of the subjects for informing them about the progress of the experiment. An intercom was set up for communication between the anechoic chamber and the control room outside the chamber. The subjects were monitored via a video camera during the listening sessions.

The experimental setup in the control room consisted of a computer (PC), a high-quality sound card (RME DIGI96/8 PST), a custom-made programmable attenuator, an amplifier (Pioneer A-616) and a passive attenuator. The passive attenuator with -20 dB gain was added to the playback chain in order to decrease the noise floor of the amplifier to an inaudible level when connected to the headphones inside the anechoic chamber. The potentiometer of the amplifier was bypassed via a modification, and the gain set to yield 0-dB overall gain, when combined with the passive attenuator. The signals from the passive attenuator were fed to the headphones via connection panels between the two rooms. The listening test sessions and the respective gains of the

programmable attenuator were controlled by a program from the PC.

The frequency responses of the left and right channels of the playback setup were within ± 0.1 dB up to 10 kHz. The electrical levels sent to the both channels of the headphones with 0-dB gain settings were controlled to be constant throughout the entire listening test.

The test sounds were played back from the PC using 16-bit resolution and a sampling rate of 44.1 kHz.

2.2.2. Measurements of headphone transfer functions (PTFs)

The PTFs were measured in connection with the HRTF-measurements of [8] using two miniature microphones (Sennheiser KE-4-211-2) inserted in earplugs, and mounted flush with the entrance to each ear canal. The HRTFs and PTFs were always measured successively, and hence, the positions of the microphones at each ear were the same for both measurements. Each PTF-measurement was repeated five times, and the subjects were asked to reposition the headphones between repetitions. The PTFs were measured using a maximum-length-sequence.

Examples of the measured PTFs for the left ear are plotted in Fig. 1. In the upper panel, PTFs for five repositionings are plotted for a single subject (subject SC), whose measurements showed typical variation between the repositionings. There is little variation in the shape of the PTF between repositionings, even at high frequencies.

In the lower panel of Fig. 1, individual means over five repositionings are plotted for the six subjects. Clear differences in the sensitivity of the PTFs between subjects are now seen below 4 kHz, however, the shape of the PTFs is similar for all subjects. A characteristic dip around 4.5 kHz can also be seen in each subject's PTFs. Above 5 kHz, the location of the peaks and the dips is dependent on the subject, and the individual PTFs are very scattered.

The five repeated measurements of the PTFs were used to calculate inverse headphone transfer functions (IPTFs). IPTFs were calculated individually for each subject, as well as for the left and right sides of the headphone. In each case, an "average

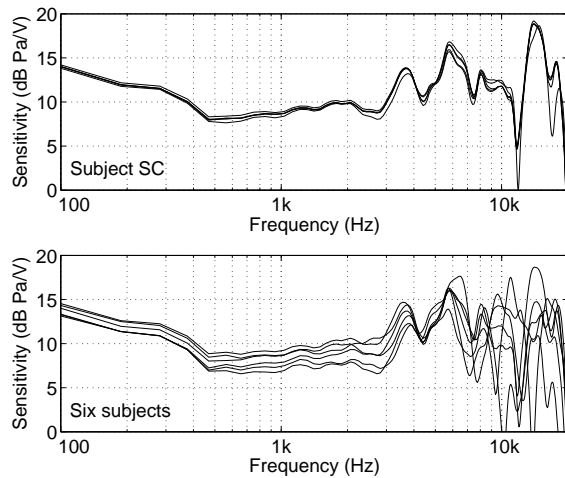


Fig. 1: PTFs (Beyerdynamic DT-990) at the blocked entrance to (left) ear canal: separate curves of five repositionings for subject SC (upper panel), and individual means over five repositionings for six subjects (lower panel).

PTF” was determined by calculating the mean amplitude in the frequency domain. This average PTF was then smoothed (in the frequency domain) by applying a moving average filter corresponding to the Equivalent Rectangular Bandwidth (ERB) [14]. The average smoothed PTF was converted to the time domain by considering it as minimum phase. This PTF was then used to calculate the IPTF by using the method of fast deconvolution with regularization, as described in [15]. The IPTF impulse response was reduced in length to 64 samples and the DC-value was adjusted to give proper low frequency equalization. In this way, a minimum-phase FIR-filter was obtained for every subject and headphone channel individually.

2.3. Verification measurements of binaural synthesis

The same two microphones, as in the headphone measurements, were used in verification measurements of the individual binaural synthesis prior to the listening test. The microphones were connected to a small battery-driven custom-built microphone preamplifier. The playback chain in the verification measurements was the same as for the listening test, with the exception of an external D/A-

converter (RME ADI-DS8) connected optically to the sound card, which was used in order to avoid hum in the measurements. The same device was used as an A/D-converter for the microphone signals.

Before and after each subject’s verification measurements, the responses of the miniature microphones were checked by measuring the same sound field simultaneously with one of the microphones and a 1/4-inch reference microphone (B&K 4136). The reference microphone has a flat response up to 20 kHz, and it had also been used as a reference in [8].

Repeated measurements with five repositionings were carried out, in order to investigate the effect of headphone positioning on the binaural synthesis. The subjects again repositioned the headphones between repetitions.

Binaurally synthesized white noise was used as measurement signals at a level corresponding to a free-field level of 75 dB SPL, which was the maximum playback level used in the listening test. The noise floor of the microphones (inside the anechoic chamber, including all electronics of the setup) was measured to be 40 dB SPL, and hence, the verification measurements were carried out using an adequate signal-to-noise ratio.

The main results of the verification measurements of the binaural synthesis are shown in Section 3.1.

2.4. Stimuli

Third-octave-band noises were used as stimuli in the listening test centered at 0.4, 1.0 and 5.0 kHz. The stimuli were binaurally synthesized from eight incidence angles: In the horizontal plane from azimuths of 0°, 30°, 60°, 90°, 135°, and 180° (incidences on the left-hand side from ahead to behind), and in addition, in the median plane from elevations of 45° and 90° just as in [8]. Note that azimuths of 0° and 180° coincide between the two planes. The length of each stimulus was 1 s, including 20-ms raised-cosine rise and fall ramps in the beginning and the end of a stimulus, respectively. The stimuli were played back at an overall level corresponding to 65 dB SPL, when measured in the free field in the absence of a listener.

The third-octave-band noises were then convolved with the individual HRTFs and IPTFs resulting in 24 different stimuli. In addition to the stimuli for

the listening test, white noise (of 10-s duration) from all eight incidence angles was synthesized similarly for each subject. The binaurally synthesized white noise signals were used in the verification measurements mentioned above. All signal processing and preparation of the sound stimuli was done in MatLab.

2.5. Procedure

An adaptive, two-interval two-alternative forced-choice loudness matching procedure [16] was used in the listening test. Separately for each center frequency, directional loudness matches were obtained between the synthesized frontal reference (at 0° of azimuth and elevation) and each of the comparison sound incidences angles in the horizontal and median planes.

Loudness matches for 21 conditions (three center frequencies \times seven comparison incidence angles) were collected. In addition, at one center frequency (1 kHz) the loudness of the frontal reference was compared to itself as a baseline to investigate the subjects' ability to make proper loudness matches. Eight replications for all 22 conditions were collected.

On each trial, the subjects had to compare the loudness of a pair of successive sounds, one from the frontal reference and another from a comparison incidence angle, and to respond, which of the two sounds within the pair was louder. The order of the reference and the comparison incidence angles was randomized. Responses were collected with the two-button response box, the left button representing the sound played during the first interval and the right button the sound played during the second interval in the pair.

The level of the frontal reference corresponded to 65 dB SPL measured in the free field. The starting level of an adaptive track was either +10 or -10 dB from the level of the frontal reference, corresponding to 75 or 55 dB SPL in the free field, respectively. For half of the adaptive tracks the starting level was +10 dB and for the other half -10 dB. An adaptive track was terminated after eight reversals in the track, and the last six reversal levels were averaged for estimating a loudness match.

One block consisted of eight adaptive tracks. The tracks were interleaved in a block so that the sub-

jects were not able to follow the course of their previous responses in the block. The order of the adaptive tracks in the blocks was randomized over center frequencies, sound incidence angles and starting levels.

One listening session consisted of four blocks, and the duration of a session was approximately an hour including breaks between the blocks. The first session included a practice block and two actual blocks, and in total, each subject participated in six 1-hour sessions.

3. RESULTS AND DISCUSSION

Results from the verification measurements and from the listening test of the present study, including comparisons with the earlier loudspeaker study [8], are plotted in Figures 2–5. In these graphs, one panel is dedicated to one subject, indicated by the subject initials on top of the panel.

3.1. Verification of binaural synthesis

The binaurally synthesized white noise recordings, recorded at the blocked entrance to the ear canal for each test subject, and from all eight incidence angles, were analysed in order to verify the success of the binaural synthesis. The effect of the headphones had been equalized for the synthesis, and thus, the recordings were essentially comparable to the same white noise being recorded from loudspeakers with perfectly flat responses at the same incidence angles.

Synthesized individual head-related transfer functions were computed from the recordings by a complex division with the original white noise signal (i.e. the signal *before* the binaural synthesis) in the frequency domain. The calibrated recordings were made at a level corresponding to 75 dB SPL in the free field, and the relative level the original white noise was set to this level before the complex division.

The individual HRTFs measured in the *real* sound field and implemented in the binaural synthesis, and the corresponding synthesized HRTFs measured in headphone reproduction are plotted in Fig. 2 for one (frontal) of the eight incidence angles. The solid gray lines are the HRTFs in the real sound field, and the dashed black lines are the HRTFs in the binaural synthesis. The HRTFs for the left and the right ears are plotted separately.

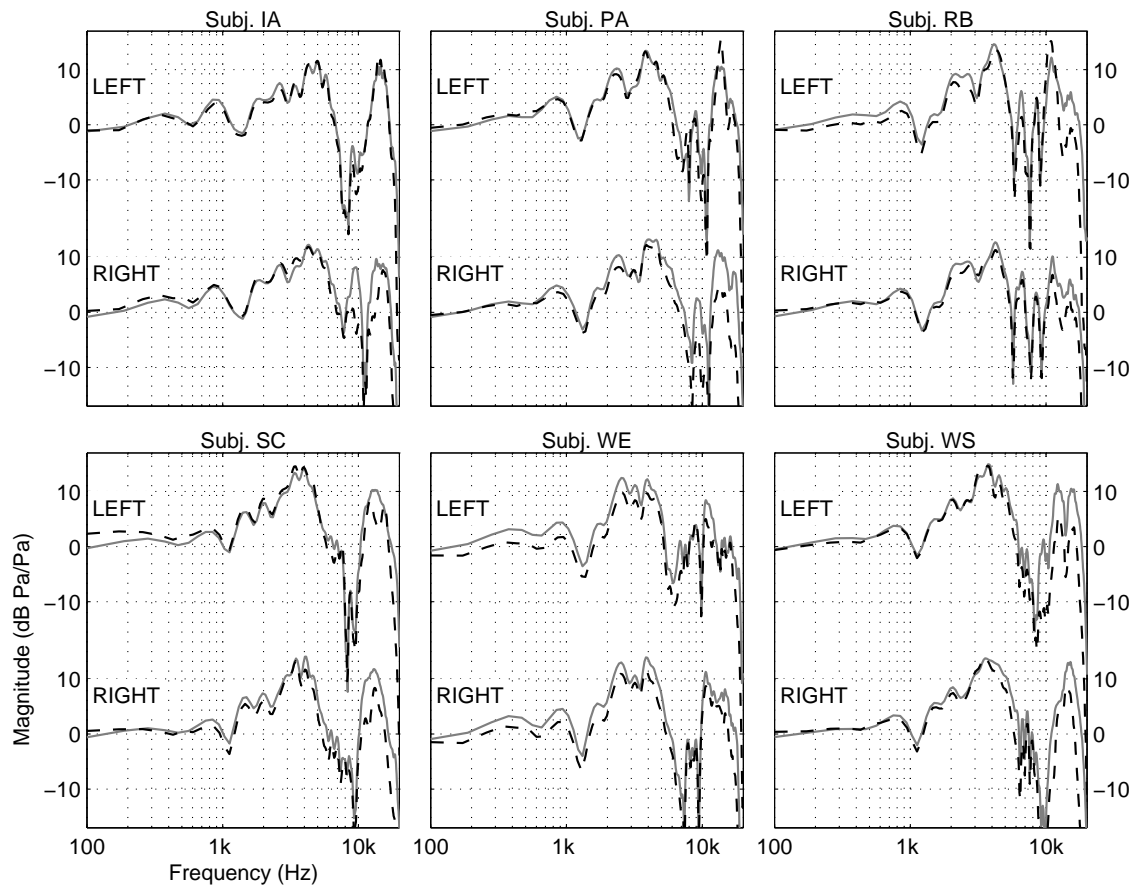


Fig. 2: Individual frontal (0° azimuth and elevation) head-related transfer functions measured in a real sound field (solid gray lines) and in binaural synthesis (dashed black lines).

Generally, the agreement between the HRTFs measured over loudspeakers and synthesized over headphones is good. The characteristic peaks and dips of the HRTFs are well reproduced for all subjects, even at the highest frequencies above 10 kHz. However, for all subjects at frequencies above approximately 8 kHz, the two curves deviate from one another. Below that the differences are small, except for subject SC (bottom-left panel), where for the left ear the synthesized HRTF does not converge towards 0 dB at low frequencies. Also, for subject WE (bottom-middle panel), the real-sound-field HRTFs have slightly higher values across the whole frequency range, although symmetrically at both ears. Note, however, that in the actual listening test, narrow-band sounds were used within

a frequency range, where there is good agreement between the two curves.

The level differences in decibels between the real-sound-field and the synthesized HRTFs within an incidence angle (i.e. the difference between the two curves in Fig. 2, now for each incidence angle) are plotted with third-octave-band resolution in more detail in Fig. 3. These differences were calculated by converting all HRTFs first to the corresponding third-octave-band levels, and then subtracting the real-sound-field levels from the synthesized levels independently for each incidence angle. As seen in Fig. 3, the difference is not constant over frequency, but substantial fluctuation around the 0-dB line can be seen, especially at higher frequencies. Generally, the differences are within ± 2 dB for all subjects

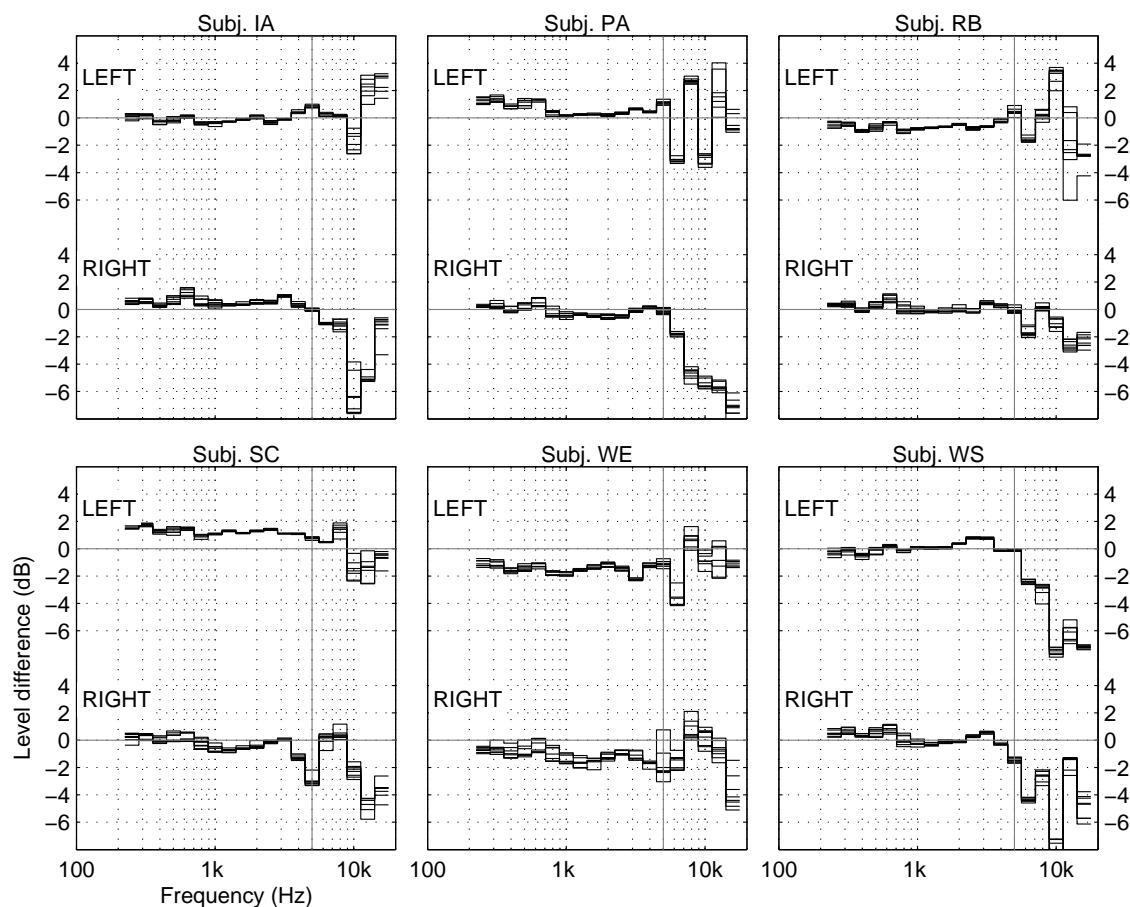


Fig. 3: Level differences on third-octave-bands between individual HRTFs measured in a real sound field and in binaural synthesis for eight sound incidence angles. See the text for details.

at frequencies below 5 kHz (a vertical gray line in the panels). Within each subject, the variations are smaller, on the order of ± 1 dB.

More importantly, the differences *between* the eight incidence angles (i.e. the eight curves for the left and the right ears in each panel) are much smaller. The curves are, except for the right ear of subject WE, very congruous up to 6 kHz. This implies that the success of the binaural synthesis was independent of incidence angle, at least at the stimulus center frequencies used in the listening test of the present investigation. This is important, since the aim of the listening test was to study the effect of sound incidence angle on loudness judgments, and thus, the variation caused by the HRTFs must not be con-

founded with the loudness data as a function of incidence angle.

The fluctuations over frequencies seen in Fig. 3 are thus due to the implemented inverse headphone transfer functions and inaccuracies in the verification measurements. The former could be due to inaccuracies in the measured PTFs and the implementation of the subsequent inverse filtering, the latter due to inaccuracies in the microphone and headphone positioning, and acquisition and analysis of the verification measurements.

In the listening test of the present study, loudness comparisons were only made within one center frequency. Thus, it was important in the synthesis for the (binaural) loudness judgments, in addition to

congruous levels over incidence angles, that the relative levels at the left and the right ears were reproduced correctly within each center frequency. That seems to be the case in Fig. 3, except for subject SC, where the left ear in the binaural synthesis is getting a higher input than the right ear, compared to the inputs in the real sound field. At 5 kHz, an additional interaural level difference on the order of one or two decibels for all subjects appears to be due to the binaural synthesis.

3.2. Loudness as a function of incidence angle

The subjects matched the loudness of a sound from a comparison incidence angle to the same sound with frontal incidence. The level of the frontal reference corresponded to a free-field level of 65 dB SPL. The raw data from the experiment were relative sound pressure levels for the comparison stimuli matched for loudness at each center frequency to the frontal reference.

If a match for a comparison incidence angle produced a gain of e.g. -5 dB from the frontal reference, a 60-dB corresponding free-field sound pressure level for that incidence angle would be needed in order for it to be perceived equally loud as the frontal reference. This would imply that the subject is *more sensitive* to loudness at this angle of incidence by 5 dB. In this manner, *directional loudness sensitivities* were calculated as inverses of the gains of the matches relative to the frontal reference for all conditions. A positive sensitivity value denotes a loudness increase and a negative value a loudness decrease for a given incidence angle.

3.2.1. Horizontal plane

Directional loudness sensitivities in the horizontal plane are plotted in Fig. 4 individually for each subject. The data are plotted separately for the three center frequencies. The x-axis is the incidence angle in the horizontal plane (azimuth in degrees), 0° being the frontal direction, 90° on the left, and 180° behind the subjects. The directional loudness sensitivity is plotted on the y-axis in decibels relative to the level of the frontal reference. The results from the earlier experiment in the real sound field [8] are plotted along with the results of the present study.

If matches were along the 0-dB line in Fig. 4, that would indicate that loudness is constant over incidence angles. As already seen in the data in the free

field ([8]; gray solid lines in Fig. 4), this is not the case. The same applies for the present data obtained via binaural synthesis (dashed black lines).

Different directional loudness sensitivity patterns can be observed, both between center frequencies, and between subjects. At 0.4 kHz (top graphs of each panel in Fig. 4), the directional effect on loudness is on the order of 3 dB. At 1 kHz (middle graphs of each panel), the directional loudness sensitivity changes by up to 5 dB. At these two center frequencies, sounds coming from the side are perceived louder than the frontal reference, the sensitivity curve being inversely U-shaped with a maximum sensitivity at 90° of azimuth.

At 5 kHz (bottom graphs of each panel in Fig. 4), the effect of the sound incidence angle on loudness is more pronounced. Now, the incidence angles on the side are not only perceived up to 5 dB *louder* than the frontal reference, but sometimes *softer* by the same amount, see e.g. the data for subjects IA and PA at 135° of azimuth. The variation in the shape of the directional loudness-sensitivity curves between subjects is also larger at this center frequency.

The agreement in the directional loudness data between listening over loudspeakers [8] and to the synthesized stimuli presented over headphones is fair in Fig. 4. The individual shapes of the directional loudness patterns in the real sound field are well reproduced in the binaural synthesis of the same stimuli. At 0.4 and 1 kHz in the lower panels (subjects SC, WE and WS), the inverse U-shaped curves are very similar between the earlier study [8] and the present experiment. In the upper panels (subjects IA, PA and RB), a slightly larger variation can be seen between the two experiments. Note that in both experiments all subjects were able to match the baseline condition of the frontal reference very closely to the 0-dB line.

At 5 kHz, where for subjects IA, RB, and WE the directional loudness-sensitivity patterns are very individual, the different patterns are well preserved in the binaural synthesis. The discrepancies between the two experiments are rather random, and for 64 of the 72 measurement points (azimuths of 30°, 60°, 90° and 135° × three center frequencies × six subjects) the 95%-confidence intervals are overlapping. Five of the eight non-overlapping conditions are at 5 kHz.

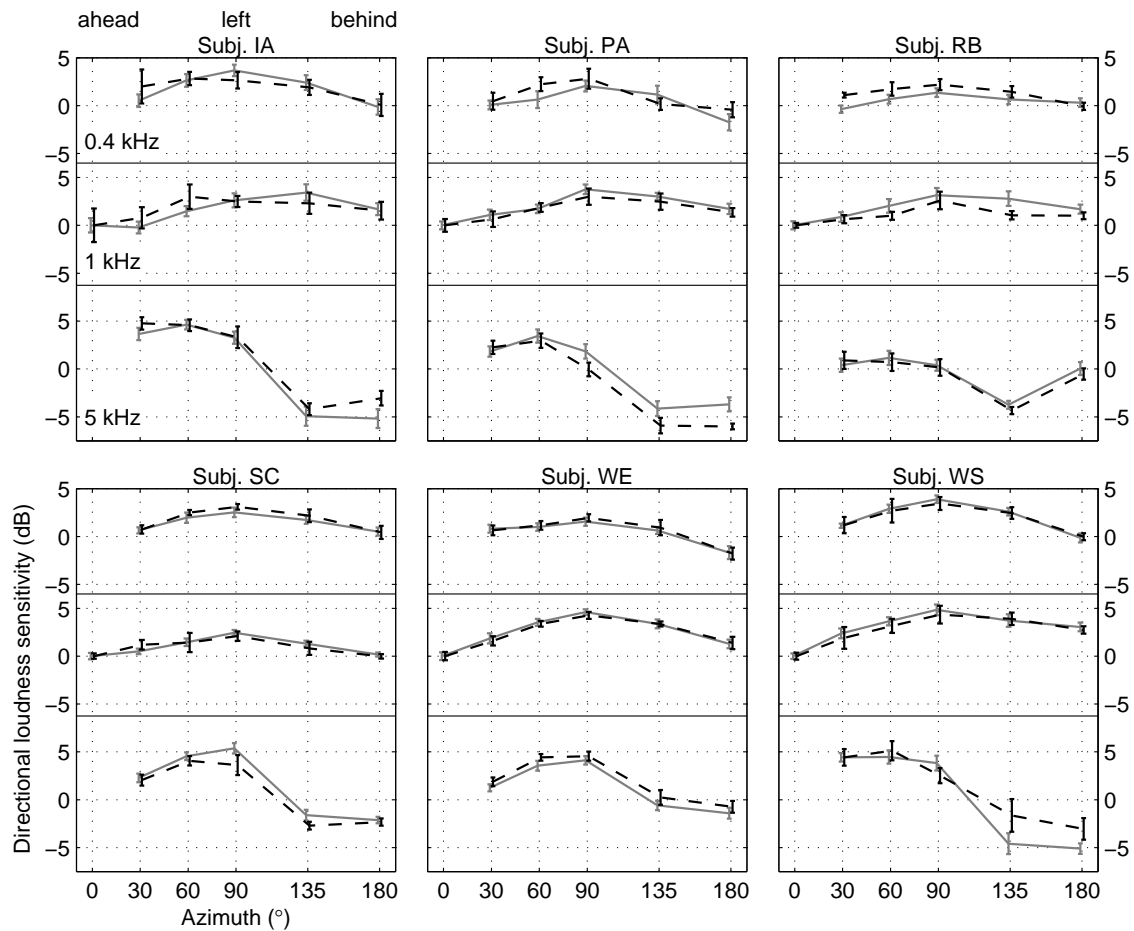


Fig. 4: Individual directional loudness sensitivities (dB) in the horizontal plane at the three center frequencies: Data collected in the real sound field are re-plotted from [8] (gray solid lines); data obtained after binaural synthesis are from the present study (dashed black lines). The error bars denote the 95%-confidence intervals of the loudness matches. See the text for details.

The differences between the directional loudness sensitivities in the real sound field and in binaural synthesis for azimuths 30°, 60°, 90° and 135° range from 0.0 to 2.9 dB, the average absolute difference being only 0.6 dB.

3.2.2. Median plane

Directional loudness sensitivities in the median plane are plotted in Fig. 5. Here the x-axis denotes the elevation in degrees. Note that 0° and 180° of elevation coincide with the same azimuths in the horizontal plane, and these data points are the same as those plotted in Fig. 4.

The directional loudness sensitivity patterns in the median plane are clearly different from those in the horizontal plane. At 0.4 kHz, the sound incidence angle has little or no effect at all on loudness. Only for subject WE (top graph in the lower-middle panel in Fig. 5), the directional sensitivity curves are different from zero for all incidence angles. At 1 kHz, loudness changes more as a function of elevation, the matches varying over a range of 5 dB. At 5 kHz, there is a trend for directional loudness sensitivity to increase as a function of elevation for all subjects, with loudness matches varying over a range of up to 10 dB.

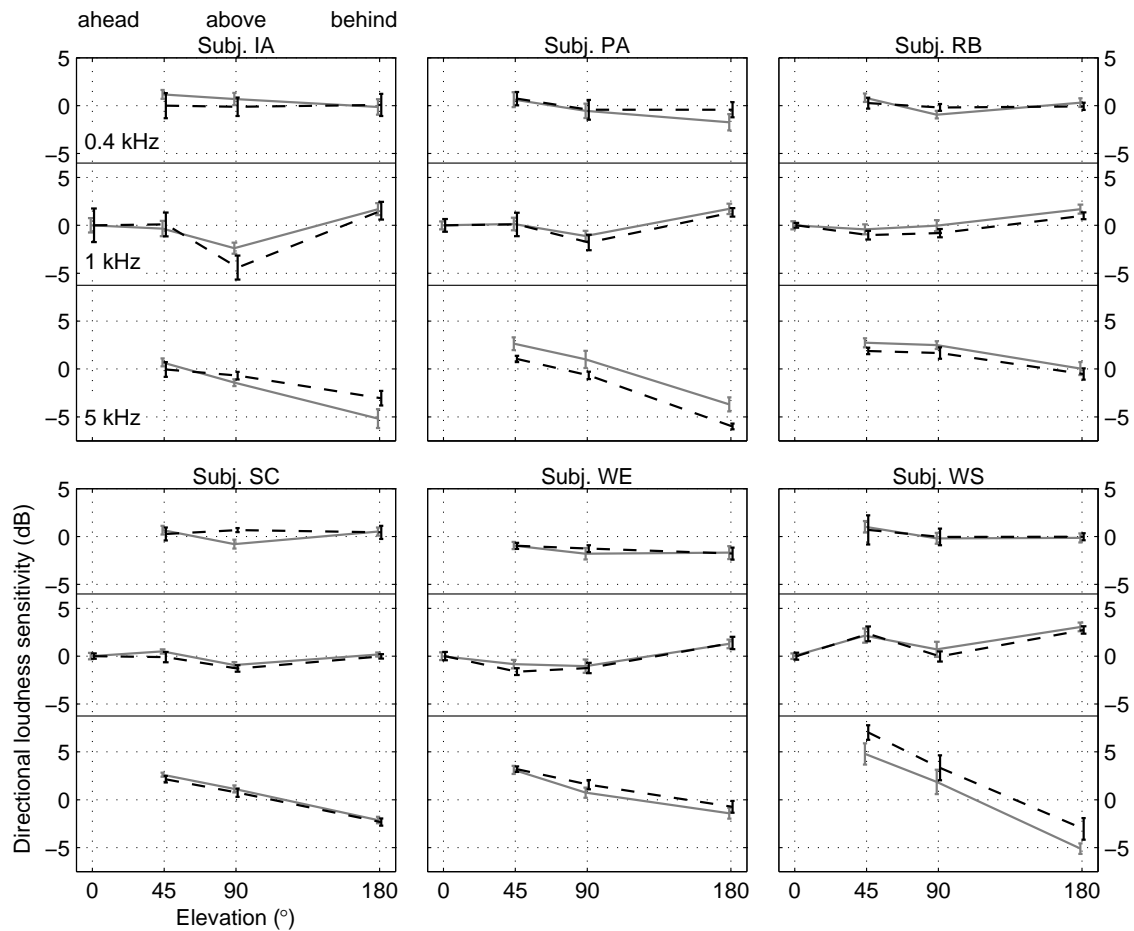


Fig. 5: Individual directional loudness sensitivities (dB) in the median plane plotted as in Fig 4. Gray solid lines: matches obtained in the real sound field; dashed black lines: stimulus presentation using binaural synthesis. The error bars denote the 95%-confidence intervals of the loudness matches.

At the two lowest center frequencies (top and middle graphs of each panel in Fig. 5), the directional sensitivity patterns are very similar between the real sound field and the binaural synthesis, except for two instances at 90° elevation for subjects IA and SC at 1 and 0.4 kHz, respectively. At 5 kHz, the similarity is retained in most cases, but some systematic deviations can also be seen, such as for subject PA and subject WS (the bottom graphs of the upper-middle and lower-left panels, respectively).

As seen in Fig. 3, the variations in the third-octave-band levels are generally larger between the synthesized and the real-sound-field HRTFs at higher

center frequencies. Thus, some differences in loudness perception of the narrow-band stimuli could be expected. For nine out of 60 measurement points (elevations of 45°, 90°, and 180° × three center frequencies × six subjects, and the baseline condition for all subjects) in Fig. 5, the confidence intervals are not overlapping, seven of these condition being at 5 kHz.

The differences between the directional loudness sensitivities in the real sound field and in the binaural synthesis for elevations 0°, 45°, 90° and 180° range from 0.0 to 2.3 dB, the average absolute difference again being only 0.6 dB. Note that these deviation

are on the same order as in the horizontal plane.

3.2.3. Statistical analysis

The 95%-confidence intervals that can be seen in Figures 4 and 5 are slightly larger in the present study than in [8]. This is due to 8 replications for each condition instead of 16 in [8]. The standard deviations of the replications, however, between the real sound field and the binaural synthesis, were very similar.

A three-factor repeated-measures analysis-of-variance (ANOVA) (7 sound incidence angles \times 3 center frequencies \times 2 “reproduction modes”) was carried out on the means obtained from each subject, in order to inspect whether the directional loudness differences between the two reproduction modes (real sound field vs. binaural synthesis) were statistically significant.

As in [8], the ANOVA returned a significant interaction of sound incidence angle \times center frequency [$F(12, 60) = 22.05; p < 0.001$]. This was to be expected, since variations in the patterns of directional loudness matches across center frequencies were obtained for both reproduction modes. However, the factor of sound reproduction over loudspeakers vs. headphones never reached statistical significance, either as a main effect, or in interaction with other factors (p -values > 0.10). This indicates, that - based on mean data, as plotted in Figures 4 and 5 - the two reproduction modes do not exhibit any systematically different loudness sensitivity patterns.

4. CONCLUSION

Individual binaural synthesis was employed in investigating the effect of sound incidence angle on loudness. Measurements verifying the acoustical signals at the ears of the listeners in the binaural synthesis were carried out.

Directional loudness matches of binaurally synthesized sounds were obtained in a listening test. The matches were compared to the results obtained earlier [8] in a listening test with the same experimental design in a real sound field. The comparison shows that the frequency- and the subject-dependency of the directional loudness matches were well retained in the individual binaural synthesis. A statistical analysis confirmed that the difference in the subjective directional loudness matches, between listening

in a real sound field and in individual binaural synthesis, was insignificant.

The difference in loudness perception between loudspeaker and headphone reproduction has been debated in the field. Even though this study did not directly compare the loudness of a sound from a loudspeaker to the same sound from headphones, *directional* loudness perception of loudspeaker playback in a real sound field and headphone playback in binaural synthesis was compared, and confirmed to be essentially indistinguishable.

5. ACKNOWLEDGEMENT

This research was carried out at the Sound Quality Research Unit (SQRU) at the Department of Acoustics, Aalborg University, Denmark. The unit receives financial support from Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration, as well as from the Danish National Agency for Industry and Trade (EFS), and the Danish Technical Research Council (STVF). Thanks are due to the colleagues and staff of the Department of Acoustics at Aalborg University, and to the partners in the Sound Quality Research Unit for technical support and scientific advice during the investigation. We would further like to thank the participants for their effort and motivation to perform well in an extensive experiment.

6. REFERENCES

- [1] E. Zwicker and H. Fastl, *Psychoacoustics - Facts and Models*, Springer-Verlag, New York, 1999.
- [2] B.C.J. Moore, B.R. Glasberg, and T. Baer, “A model for the prediction of thresholds, loudness, and partial loudness,” *J. Audio Eng. Soc.*, vol. **45** (1997), 224–239.
- [3] E. Zwicker and U.T. Zwicker, “Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution,” *J. Acoust. Soc. Am.*, vol. **89** (1991), no. 2, 756–764.
- [4] D.W. Robinson and L.S. Whittle, “The loudness of directional sound fields,” *Acustica*, **10** (1960), 74–80.

-
- [5] H. Remmers and H. Prante, "Untersuchung zur Richtungsabhängigkeit der Lautstärkeempfindung von breitbandigen Schallen," *Fortschritte der Akustik - DAGA '91* (1991), 537–540.
 - [6] S.L. Jørgensen, "Matlab model for loudness perception - directional loudness," Unpublished master's thesis, Technical University of Denmark, Copenhagen, Denmark 2002.
 - [7] N. Zacharov, S. Bech and P. Suokuisma, "Multichannel level alignment part II: The influence of signals and loudspeaker placement," in *Proceedings of the Audio Engineering Society 106th Int. Conv.*, San Francisco, USA, 1998 September 26–29, preprint 4816.
 - [8] V.P. Sivonen and W. Ellermeier, "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation," *J. Acoust. Soc. Am.*, vol. **119** (2006) no. 5, 2965–2980.
 - [9] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. **36** (1992), 171–218.
 - [10] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, MIT Press, Cambridge, Massachusetts, 1997.
 - [11] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. **85** (1989) no. 2, 858–867.
 - [12] D. Hammershøi, "Binaural technique – a method for true 3D sound reproduction," Ph.D. thesis, Aalborg University, Denmark, 1995, ISBN-87-7307-516-7.
 - [13] E.H.A. Langendijk and A.W. Bronkhorst, "Fidelity of three-dimensional-sound reproduction using a virtual auditory display," *J. Acoust. Soc. Am.*, vol. **107** (2000) no. 1, 528–537.
 - [14] B. C. J. Moore, *An introduction to the psychology of hearing*, Academic Press, London, 1997.
 - [15] O. Kirkeby, P. A. Nelson, H. Hamada and F. Orduna-Bustmante, "Fast deconvolution of multichannel systems using regularization", *IEEE Transactions of Speech and Audio Processing*, vol. **6** (1998), no. 2, 189–194.
 - [16] W. Jesteadt, "An adaptive procedure for subjective judgments," *Percept. Psychophys.*, vol. **28** (1980), no. 1, 85–88.

Laterality in binaural and directional loudness

Ville Pekka Sivonen^{1,2} and Wolfgang Ellermeier¹

¹*Sound Quality Research Unit, Dept. of Acoustics, Aalborg University, 9220 Aalborg, Denmark*

²*Brüel & Kjær Sound & Vibration Measurement A/S, 2850 Nærum, Denmark*

Abstract

The influence of the side of stimulation on loudness perception was studied using two paradigms. In Experiment 1, monaural-to-binaural loudness matches for stimulation from a frontal sound source were obtained from six listeners, separately for their left and right ears. In Experiment 2, a subset of five listeners matched the loudness of sounds having incidence angles either $+90^\circ$ or -90° of azimuth to the same sound with frontal incidence. The loudness matches were obtained for narrow-band stimuli at three different center frequencies (0.4, 1 and 5 kHz) at an overall level corresponding to 65 dB SPL measured in the free field. The stimuli were played back to the listeners via headphones in individual binaural synthesis. An adaptive, two-interval, two-alternative, forced-choice procedure was utilized in obtaining the loudness matches. The results show that for some listeners, the side of stimulation has an effect on the monaural-to-binaural matches of Experiment 1 and the directional matches of Experiment 2. For these listeners, a lower sound pressure level for the left (than for the right) side is required for a match with the frontal reference. When inspecting the overall results across listeners, however, the effect of laterality is insignificant.

I. INTRODUCTION

Loudness and its binaural summation has been the subject of a number of investigations (see e.g., Reynolds and Stevens, 1960; Hellman and Zwisch, 1963; Marks, 1978; Zwicker and Zwicker, 1991). In these studies, headphones were used, allowing for controlling the stimuli independently at the listener's two ears. While this control is desirable for quantifying the binaural input, the stimulation (and hence, the auditory event) can easily become *unnatural* to the listeners. By varying the two at-ear signals independently, interaural combinations may be played back, which would never reach a listener's ears in a real sound field.

By contrast, Robinson and Whittle (1960) and more recently, Sivonen and Ellermeier (2006), investigated binaural loudness summation for sounds of different incidence angles in a *real* sound field, resulting in various natural at-ear exposures. Thus, directional loudness matches were obtained for sound sources located in space, producing externalized auditory events. The at-ear exposures, measured as sound pressure levels in Robinson and

Whittle (1960) and determined by the listeners' individual head-related transfer functions (HRTFs) in Sivonen and Ellermeier (2006), were utilized in modeling the binaural summation underlying the observed loudness matches.

The sound sources in Sivonen and Ellermeier (2006), however, were placed only on the left-hand side of the listeners due to the assumed symmetry in the sagittal plane, i.e., between the left and right hemispheres. The symmetry is fairly apparent in the *physical* transfer from a free-field to the listener's ears, and can be quantified by HRTFs. In Robinson and Whittle (1960), both hemispheres were investigated, but the reported results were averaged across the two sides. Furthermore, symmetry in *binaural loudness summation* was assumed, i.e., no differential weighting of the left- and right-ear inputs was assumed.

There have been occasional observations of laterality effects both in absolute thresholds (e.g., Emmerich et al., 1988), and with respect to loudness (e.g., Schneider and Cohen, 1997), suggesting a slightly greater sensitivity and contribution to fused binaural loudness of the right ear, respec-

tively. Such laterality effects might thus have created some of the large individual differences observed in the modeling in Sivonen and Ellermeier (2006), if the listeners weighed their left- and right-ear exposures in an idiosyncratic manner. But, note that the studies of Emmerich et al. (1988) and Schneider and Cohen (1997) did not include measurements of at-ear exposure, so they might confound differences in headphone-to-eardrum coupling with ‘neural’ sensitivity differences.

Therefore, the aim of this study is to investigate (1) whether the same SPL is perceived equally loud when delivered either to the left or to the right ear, and (2) do the two ears contribute equally, or is one side dominant in contributing to binaural loudness.

The present study makes use of individual binaural synthesis, where sound source located in space are synthesized to listeners over headphones, using the listeners’ HRTFs. Individual binaural synthesis has shown to yield essentially the same directional loudness matches as listening to real sound sources (Sivonen et. al, 2005). In Experiment 1, the listeners matched the loudness of monaural stimulation to a binaural reference separately for the two ears, by switching one channel of in the binaural playback. In Experiment 2, directional loudness matches to a frontal reference were obtained from the same listeners for a sound source either at $+90^\circ$ or -90° of azimuth (i.e., at both lateral sides) in the horizontal plane, in order to investigate possible laterality in binaural loudness judgments.

II. METHOD

A. Subjects

Six listeners, between 23 and 47 years of age, participated in the experiments. The listeners had normal hearing, as determined by standard pure-tone audiometry, and their individual HRTFs and headphone transfer functions (PTFs) had been measured in connection with an earlier experiment (Sivonen and Ellermeier, 2006). All six listeners participated in Experiment 1 of the present study, whereas only five of them participated in Experiment 2.

The listeners’ hearing thresholds are plotted in Fig. 1, which shows that none of the thresholds exceed 20 dB of hearing loss re ISO 389-1 (1998). However, interaural variation up to 25 dB in the

thresholds can be seen, see subject WS at 8 kHz in Fig. 1. Also, the thresholds of subject WE are fairly asymmetric between the two ears across the tested frequency range.

B. Apparatus

The listeners were seated inside an anechoic chamber for the listening test. Loudspeakers were visually present at the intended incidence angles (at a distance of 2.0 m), in an attempt to increase the plausibility of the binaural synthesis. The chair the listeners were seated in was positioned and rotated so that the center of the listeners head was aligned with the center position of the loudspeaker setup, and the listeners were facing the loudspeaker in front of them.

The stimuli were played backed over headphones (Beyerdynamic DT-990). The playback chain was as reported in Sivonen et. al (2005), consisting of a computer (PC), a sound card (RME DIGI96/8 PST), a programmable attenuator, an amplifier (Pioneer A-616) and a passive attenuator. The listeners’ answers were collected with a two-button response box with a light above each button. A model of the response box with similar lights was placed behind the frontal loudspeaker to avoid listeners tilting their heads towards the box in their hands.

Audiometry was performed using a Madsen Orbiter 922 audiometer with TDH-39 headphones. Thresholds were measured using a ‘bracketing’ methodology (ISO 8253-1, 1991) with a 5-dB step size.

C. Stimuli

Third-octave-band noises centered at 0.4, 1.0 and 5.0 kHz were used as test stimuli. The length of each stimulus was 1 s, including 20-ms raised-cosine rise and fall ramps in the beginning and at the end of the stimuli, respectively. The stimuli were prepared for three sound incidence angles in the individual binaural synthesis: $+90^\circ$, 0° and -90° of azimuth in the horizontal plane, i.e., directions at the left-hand side, in front and at the right-hand side of the listener. The locations of the synthesized sound sources, relative to the listener, are depicted in Fig. 2. The stimuli were played back at an overall level corresponding to 65 dB sound pressure level (SPL), measured at the center position of the setup in the absence of a listener

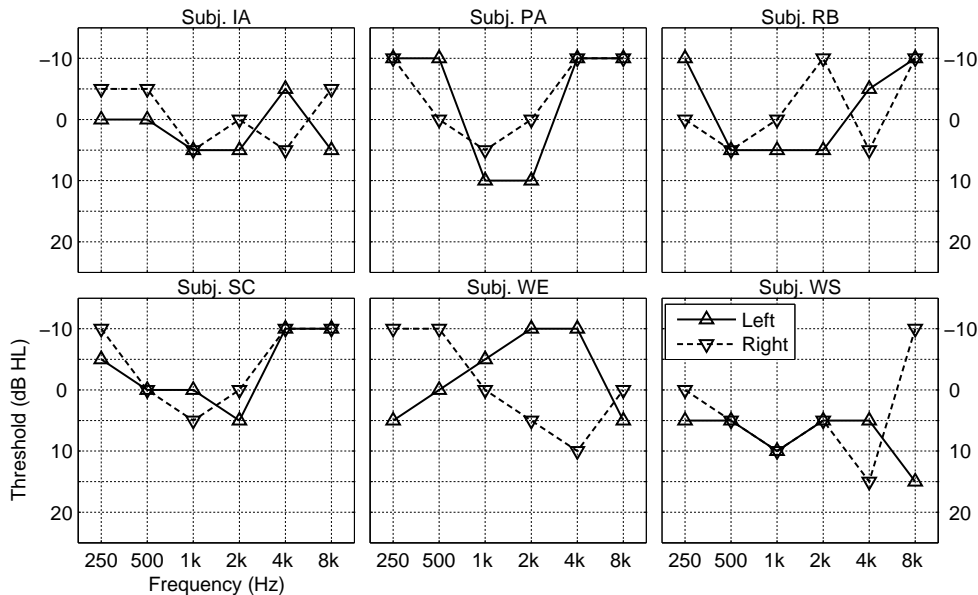


Figure 1: Left- and right-ear hearing thresholds (dB hearing level, re ISO 389-1, 1998) determined with pure-tone audiometry for six listeners.

(i.e., in the free field) using an omni-directional microphone.

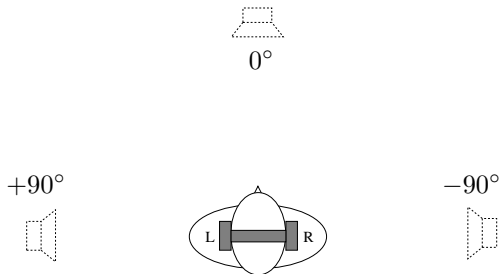


Figure 2: The location of the sound sources synthesized in the present experiments; at $+90^\circ$ and -90° of azimuth at the left-hand and right-hand side and at 0° of azimuth in front of the listener in the horizontal plane.

The stimuli synthesized with the frontal incidence angle were used as references for the loudness matches. In Experiment 1, the loudness of a *monaural* frontal stimulation, either at the left or the right ear, was matched with the *binaural* frontal reference. Thus, loudness matches between monaural and binaural listening to a frontal sound source were obtained, effectively simulating the effect of blocking one ear in free-field listening. In Experiment 2, the loudness of *binaural* stimulation by a synthesized sound source either emanat-

Table I: The reference (REF) and the comparison stimuli (COMP) at each center frequency ($f_c = 0.4, 1$ or 5 kHz) for loudness matches. In Experiment 1, a monaural stimulation from a frontal sound source either at the left (L) or at the right ear (R), and in Experiment 2, a lateral, binaural stimulation either at $+90^\circ$ or at -90° of azimuth, is matched with the binaural frontal reference (0° of azimuth).

	Exp. 1		Exp. 2	
REF	0_{bin}°		0_{bin}°	
COMP	$0_{mon,L}^\circ$	$0_{mon,R}^\circ$	$+90_{bin}^\circ$	-90_{bin}°

ing from $+90^\circ$ or -90° of azimuth was matched with the reference, obtaining directional loudness matches for the two lateral incidence angles with the frontal reference. The comparison modes of both experiments are listed in Table I.

In addition to the (distal) overall level of 65 dB SPL, the individual HRTFs affected the (proximal) stimulation at the ears of a listener. The magnitude spectra of the left- and right-ear HRTFs utilized in the present investigation are plotted in individual panels for each listener for the three incidence angles in Fig. 3. HRTFs for the incidence angles of $\pm 90^\circ$ were not measured for subject WS, and thus, he only participated in

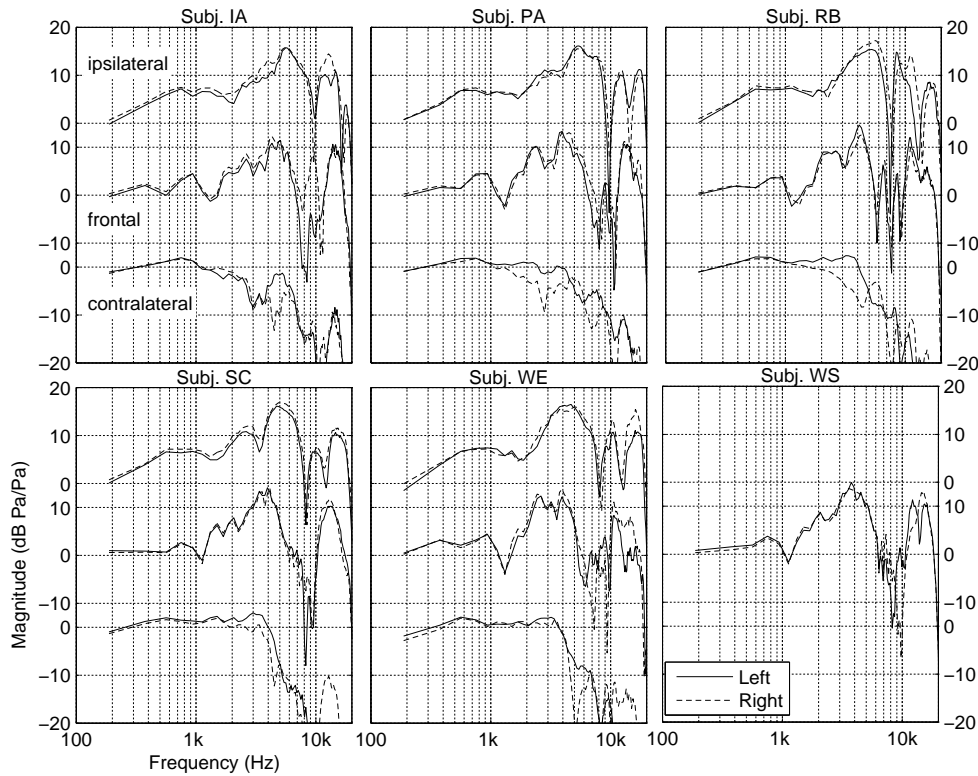


Figure 3: HRTF magnitude spectra for six subjects and for three incidence angles in the horizontal plane: the frontal HRTFs are measured at an azimuth of 0° , the ipsilateral HRTFs for the left and right ears at $+90^\circ$ and -90° of azimuth, respectively, and the contralateral HRTFs for the left and right ears at -90° and $+90^\circ$ of azimuth, respectively. Note the discontinuity of the ordinate.

Experiment 1 utilizing the frontal direction. For the other listeners, the ipsilateral HRTFs (left ear for $+90^\circ$, right ear for -90° , see Fig. 2) have levels between 0 and 15 dB, except for the highest frequencies, whereas for the contralateral HRTFs, the levels are around or below 0 dB. The levels of the frontal HRTFs are typically between the corresponding levels at the ipsi- and contralateral ears.

D. Procedure

An adaptive, two-interval, two-alternative, forced-choice procedure (Jesteadt, 1980) was used in the listening test, both for Experiments 1 and 2. The procedure was implemented as reported in Sivonen et. al (2005), and for the most part as in Sivonen and Ellermeier (2006).

On each trial, the listeners were played back a pair of sounds, the reference and a comparison in a randomized order, and their task was simply to judge, which of the two sounds within the pair was louder. The lights of the response box were lit with

each sound interval, and the listeners indicated the louder interval by pressing the respective button of the box. Loudness matches were obtained separately at each center frequencies for four different comparison modes (Experiment 1: monaural left or right; Experiment 2: $+90^\circ$ or -90° of azimuth, see Table I). Eight replications of each condition were collected, resulting in 96 adaptive tracks in total.

Each adaptive track was run until eight reversal were obtained, the step size being 4 dB before the second reversal and 1 dB for the rest of the track. The levels of the last six reversals were averaged for estimating the loudness match. In Experiment 1, since monaural stimulation is generally perceived softer than the corresponding binaural one, the starting levels of the adaptive tracks were asymmetric: for half of the tracks, the starting level was -4 dB and for the other half $+16$ dB from the level of the frontal reference. In Experiment 2, in which the stimulation was always binaural, symmetric starting levels either -10 or $+10$ dB from

the level of the reference were used.

One listening block consisted of eight adaptive tracks. The order of the tracks was randomized across center frequencies, comparison modes and starting levels. In addition, the tracks were interleaved within a given block, for the purpose of making the experiment less transparent for the listeners. One listening session consisted of four blocks and the duration of a session was approximately one hour including breaks between the blocks. The order of the blocks for Experiments 1 and 2 was counterbalanced across listeners, so that three listeners started with a session with “1221” order of the blocks, while two listeners started with “2112” order. For the next session, the order was swapped. Note that one listener participated only in Experiment 1, and hence, received only blocks of type “1”. Before the first listening session, each listener participated in a practice block, which was of the same type as their first ‘real’ block. In total, the duration of both experiments was three hours per listener.

III. RESULTS AND DISCUSSION

A. Experiment 1

In Experiment 1, the listeners matched the loudness of a sound from the synthesized frontal source heard monaurally to the same sound heard binaurally. These *monaural-to-binaural* loudness matches were obtained at each center frequency for each listener, separately for their left and right ears.

1. Distal monaural-to-binaural loudness matches

The raw data obtained in Experiment 1 were the sound pressure levels the synthesized sound source would have to be set to for monaural listening, to be perceived equally loud as the binaural reference. These *distal* SPLs were as measured at the center position of the listener’s head with the listener absent, the binaural reference being at 65 dB SPL.

The means across repetitions of these monaural distal loudness matches are plotted in Fig. 4, separately for each listener. The matches range from 63 to 71 dB SPL, exhibiting considerable dependence on the individual listener: For subject PA, the matches are always close to 70 dB SPL, while

for subject IA, for example, the matches are between 63 and 68 dB SPL, showing a marked dependence on the center frequency of the stimuli. A similar tendency for a frequency dependence can also be observed for subjects SC and WE, and for the left-ear matches of subject WS.

As expected, most of the monaural matches are *above* the 65-dB horizontal in Fig. 4. The highest matches at around 71 dB SPL (see subjects SC and WE) denote a need for increasing the SPL of the synthesized source by 6 dB for monaural listening to be perceived equally loud as the binaural reference. Thus, a binaural advantage of at maximum 6 dB is observed, the average advantage being 3.4 dB in Fig. 4. Unexpectedly, for subject IA, some means are actually *below* the level of the binaural reference, suggesting a binaural disadvantage for loudness.

Despite the scattering of the matches across listeners and center frequencies, the 95-% confidence intervals are overlapping at all instances for the left- and right-ear matches plotted in Fig. 4. This suggests that the monaural loudness matches are independent of the ear stimulated (i.e., left vs. right), even though on average, a 0.4 dB higher SPL is required for the right-ear matches. The average absolute difference between the ears is 1.1 dB and deviations of up to 3 dB can be observed, see subject IA at 0.4 kHz in Fig. 4. The slight asymmetry seems to favor the left ear in the monaural loudness matches for most listeners, i.e., a lower distal SPL is required for the left ear for a match with the binaural reference.

A two-factor [2 ears (L or R) \times 3 center frequencies], repeated-measures analysis of variance (ANOVA) for the distal monaural sound pressure levels, carried out on an individual basis, returned a significant main effect of center frequency ($p < 0.001$) for subjects IA, SC and WE. This can be seen in Fig. 4, where for these listeners, the monaural matches are frequency dependent. The effect of ear (L or R) was significant for subject IA ($p = 0.002$) and close to significant for subject WS ($p = 0.09$), these being the subjects with the largest absolute ear differences, even though of opposite sign as may be seen in Fig. 4. The 95-% confidence intervals plotted in Fig. 4 show considerable differences between listeners, ranging from 0.5 to 5.0 dB, and being 1.9 dB on average. Thus, large confidence intervals, e.g., for subject RB, most likely mask any potential ear differences in the statistical analysis, even though the absolute

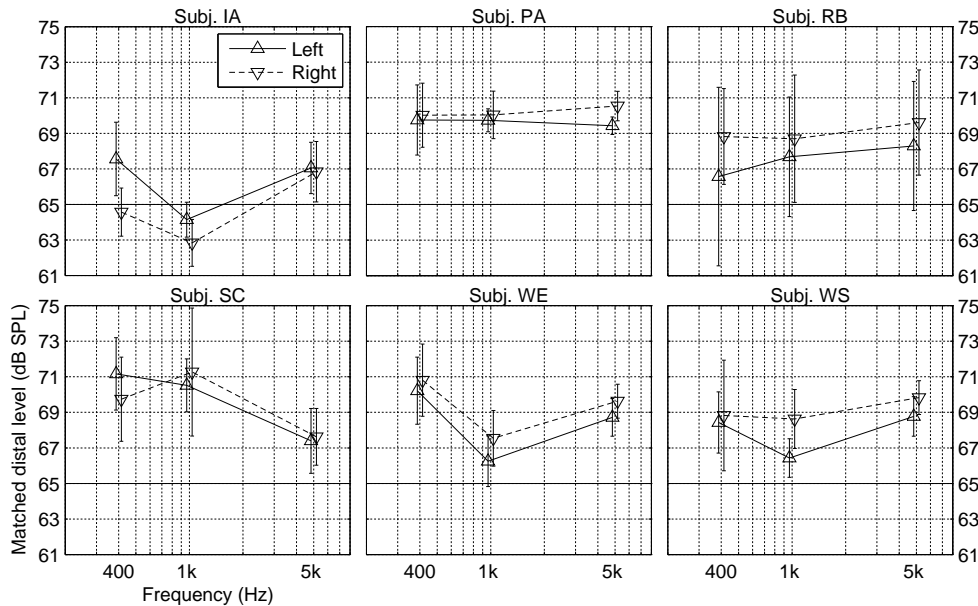


Figure 4: Experiment 1: Monaural distal sound pressure levels (dB) with 95-% confidence intervals for the left and right ears, matched to a binaural reference at 65 dB SPL for a frontal sound source at three center frequencies (0.4, 1 and 5 kHz).

difference between the means for the two ears may be on the same order as for subjects IA and WS. The two-way (ear \times center frequency) interactions were insignificant ($p > 0.1$) for all subjects.

2. Proximal data

Since significant ear difference emerged in some cases when analysing the distal matches, the effect of the physical transformation from the distal stimulation in the absence of a listener to the 'proximal' stimulation at the listener's ears was considered as well. Thus, it was investigated whether the ear differences observed in Fig. 4 could be explained by a corresponding difference in the magnitude spectra of the frontal left- and right-ear HRTFs, see Fig. 3. To that effect, the monaural distal matches were combined with the third-octave-band level of the HRTF of the corresponding ear at each center frequency, and equal-loudness left- and right-ear SPLs, both matched to the same frontal reference, were obtained.

The means across repetitions for these monaural proximal SPLs are plotted in Fig. 5. The effect of the HRTFs as a function of center frequency is clear: While at 0.4 and 1 kHz frontal HRTFs typically have levels between 0 and 5 dB, at 5 kHz the levels are between 5 and 10 dB, see Fig. 3. Thus,

the proximal matches in Fig. 5 are boosted SPLs (up to 81 dB SPL) for the 5 kHz center frequency.

When inspecting the proximal matches for the two ears, including the HRTFs in the analysis does not seem to have improved the agreement between the left- and right-ear curves. The interaural differences observed for the distal matches e.g., for subjects IA and WE in Fig. 4 are still retained in the proximal data of Fig. 5. On the contrary, the average absolute difference between the left- and right ear proximal matches is 1.3 dB as opposed to 1.1 dB for the distal matches. Also, the 95-% confidence intervals do not overlap in two instances, see subjects PA and WE at 5 kHz in Fig. 5. Only for subject RB, at 5 kHz, are the proximal matches clearly closer to one another than in the distal data.

Carrying out individual statistical analyses (ANOVA) on the proximal matches confirms that the alignment between the ears has not improved: the main effect of ear was significant for subject WE ($p = 0.003$) and subject PA ($p = 0.005$), and interactions between ear and center frequency were significant for subjects IA and PA ($p = 0.04$ and $p = 0.02$, respectively). These effects can be visually inspected in Fig. 5, where a lower sound pressure level at the left than at the right ear is required for loudness matches with the reference for

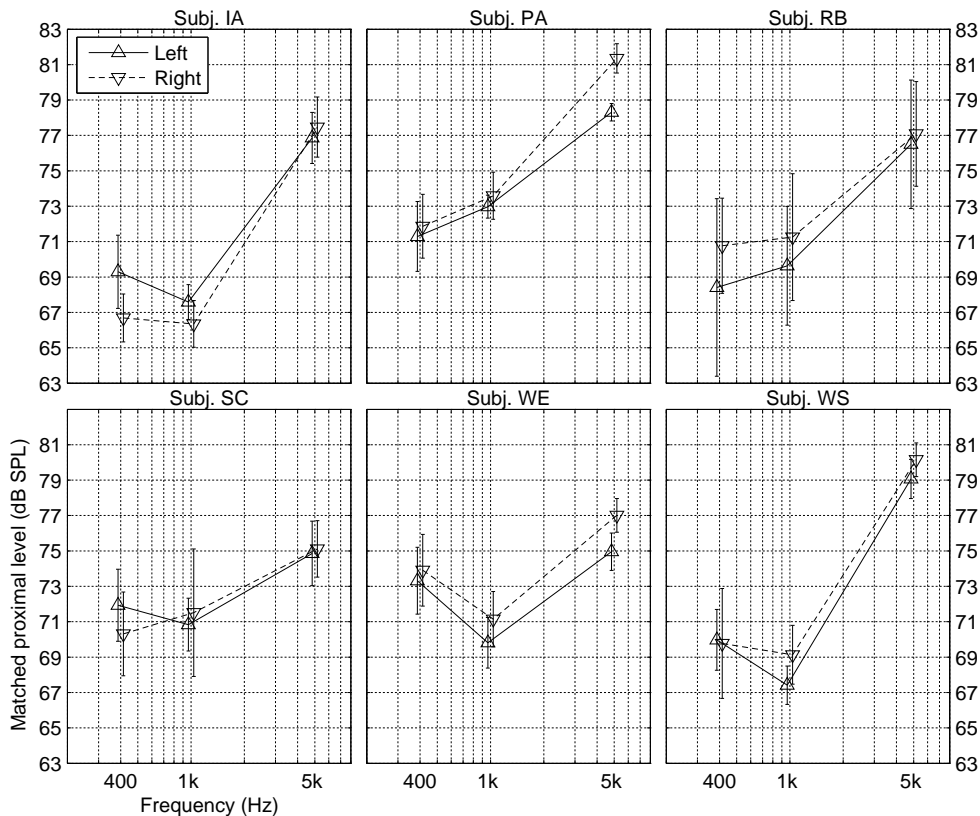


Figure 5: Monaural proximal sound pressure levels (dB) with 95-% confidence intervals for the left and right ears, matched to a binaural reference at three center frequencies (0.4, 1 and 5 kHz).

subject WE at all center frequencies, and for subject PA at 5 kHz. For subjects IA at 0.4 kHz, by contrast, a lower SPL is required for the right ear, displaying a lateral advantage inverse to what was observed for subjects PA and WE. As expected, the main effect of center frequency was significant for all listeners for the proximal data.

3. Summary

To sum up Experiment 1, the effect of ear on the distal matches was significant only for subject IA, exhibiting an advantage of the right ear on loudness, i.e. a lower sound pressure level being required at the right ear than at the left ear for a loudness match with the binaural reference, see Fig. 4. Although the effect is significant only for one subject of the present study, note that the observation is in line with the findings of Emmerich et al. (1988) and Schneider and Cohen (1997), also reporting a right-ear advantage on loudness.

When analysing the proximal matches, i.e., ac-

counting for the physical differences between the ears in the transfer of sound from a frontal source to the blocked ear canal, the effect of ear is significant also for subjects WE and PA. This effect, however, is of the opposite sign: for these subjects, there appears to be an advantage of the left ear on loudness (see Fig. 5), being at odds with the findings of Emmerich et al. (1988) and Schneider and Cohen (1997).

Note that the supra-threshold effects of ear (see Figs. 4 and 5) are relatively small compared to the interaural asymmetries in the hearing thresholds of Fig. 1. In Figs. 4 and 5 the effect size is on the order of 1–3 dB, while at threshold, much more striking asymmetries are observed, of up to 25 dB between the ears (see Fig. 1). This may be due to the different resolution between the threshold determination and the adaptive matching, or to the different frequencies probed in the two procedures. Most likely, however, it is due to the phenomenon of recruitment (Moore, 2003, chap. 4), where any sensitivity differences found at absolute threshold

tend to get smaller with increasing level, thus making it difficult to compare the sensitivity differences with loudness differences at supra-threshold levels.

B. Experiment 2

In Experiment 2, the listeners matched the loudness of a sound coming from a synthesized source either at $+90^\circ$ or -90° of azimuth to the same sound with frontal incidence. In contrast to Experiment 1, the listening was always binaural, for obtaining *directional* loudness matches at each center frequency for each listener, separately for the two lateral sound sources.

1. Distal directional loudness matches

The raw data obtained in Experiment 2 were the sound pressure levels the synthesized lateral sound sources would have to be set to, to be perceived equally loud as the frontal reference. These *distal* SPLs were as measured at the center position of the listener's head with the listener absent, the binaural reference being at 65 dB SPL.

The means across repetitions for the directional loudness matches are plotted in Fig. 6, separately for each listener. The matches range from 59 to 65 dB SPL, in all but one instance, the sound sources at $+90^\circ$ and -90° being required lower levels than the frontal reference to match its loudness. Changes as a function of center frequency in the directional loudness matches, and differences between the listeners' patterns of matches can be seen, although they are not as striking as in the monaural-to-binaural loudness matches plotted in Fig. 4.

When inspecting the differences between the two sides, only for subject PA at 5 kHz and for subject WE at 1 and 5 kHz are the 95-% confidence intervals not overlapping, while for all other cases, the directional loudness matches are close for the two sides, deviating by 1 dB or less from one another. To confirm, a two-factor [2 sides ($+90^\circ$ or -90°) \times 3 center frequencies] ANOVA was carried out on these distal matches, separately for each listener. The ANOVA returned a significant main effect of side for subject WE ($p < 0.001$) and an interaction between side and center frequency for subject PA ($p < 0.001$). These effects can be seen in Fig. 6, where for subject WE at all center frequencies, and for subject PA at 5 kHz, lower sound pressure levels for the $+90^\circ$ azimuth (left-hand

side) than for the -90° azimuth (right-hand side) are required for loudness matches with the frontal reference. Note that this outcome agrees quite well with the statistics of the proximal monaural-to-binaural loudness matches and the data plotted in Fig. 5. In addition, the main effect of center frequency was significant for subjects PA and WE ($p < 0.001$) and for subject RB ($p = 0.04$), denoting that the matches were frequency dependent, see Fig. 6.

It is worth noting that the 95-% confidence intervals for the directional loudness matches (see Fig. 6) are smaller than for the monaural-to-binaural loudness matches (see Fig. 4). Excluding subject WS, who did not participate in Experiment 2, the confidence intervals in Experiment 1 ranged from 0.5 to 5.0 dB, being 2.0 dB on average, whereas the corresponding range in Experiment 2 was from 0.2 to 3.6 dB, being 1.4 dB on average. This suggests higher consistency in the matches of Experiment 2 than in Experiment 1. Note that in Experiment 2, the auralization was always binaural, and the listeners' task was to compare the loudness of sound sources at various incidences. In Experiment 1, the monaural stimulation may have sounded unnatural to the listeners¹ compared to the binaural frontal reference, hence the differences between comparisons being more dramatic.

2. Proximal data

Since some the lateral effects were observed in the directional loudness matches when considering distal SPLs, the same strategy as in Experiment 1 was explored: Again, HRTFs were taken into account, in order to investigate whether a possible side difference could be accounted for by interaural differences in the physical transformation from the sound to the listener's ears. Note that this is largely analogous to analysing the monaural-to-binaural matches of Experiment 1, although now, *two* at-ear levels for each binaural, directional loudness match are obtained. Thus, the distal matches were combined with the third-octave-band levels of the HRTFs at each center frequency obtaining left- and right-ear SPLs.

The means across repetitions for the binaural proximal SPLs are plotted for each listener in Fig. 7. When the distal matches (see Fig. 6) are combined with the HRTFs, the resulting levels at the ipsilateral ear are around or above 65 dB SPL, while the levels at the contralateral ear range from 37 to 62 dB SPL. The interaural level differ-

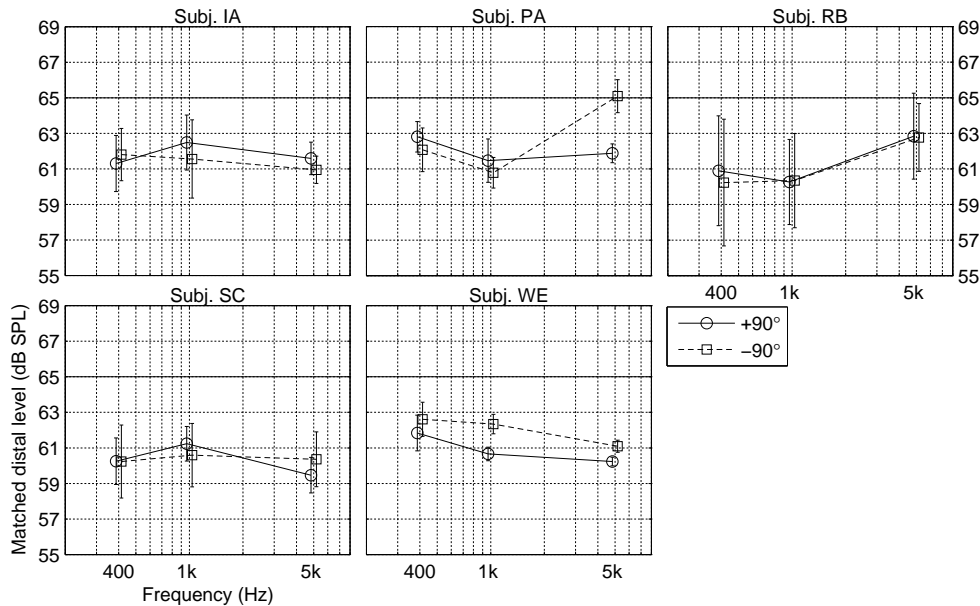


Figure 6: Distal directional loudness matches with 95-% confidence intervals (dB SPL) for $+90^\circ$ and -90° of azimuth (on the left and on the right side of the listeners), matched to a frontal reference at 65 dB SPL in the horizontal plane.

ences (ILDs) increase with center frequency, being greater than 30 dB at maximum, see subject WE at 5 kHz.

The binaural combinations of left- and right-ear SPLs, plotted for the two lateral sound sources at $\pm 90^\circ$ of azimuth in Fig. 7, were perceived equally loud at each center frequency. Thus, congruence of the two ipsilateral on the one hand and the two contralateral curves on the other hand implies no laterality in the directional loudness judgments. The 95-% confidence intervals of the means in Fig. 7 are overlapping for the two lateral sides at the two lowers center frequencies (0.4 and 1 kHz) for all listeners except for subject WE.

At 5 kHz, interaural differences both at the ipsi- and contralateral ears can be observed for all listeners but subject RB. Since both at-ear levels have an influence on binaural loudness judgments, the differences do not necessarily denote laterality: For subjects WE and IA, where the left-ear levels at 5 kHz are higher at the ipsilateral ear, the case is the opposite at the contralateral ear, see Fig. 7. Furthermore, in such an instance the absolute difference between the two ipsilateral and the contralateral curves may of a different magnitude: Presumably, the effect of the ear receiving the greater SPL input is larger, and hence, a smaller change at that ear (than at the ear receiving

the lower input) has an influence on overall, binaural loudness.

However, for subject PA and SC for the sound source at -90° of azimuth, a higher SPL *both* at the ipsi- and contralateral ears is required for a loudness match. This suggests laterality in the directional loudness judgments, favoring the left-hand side, i.e., the source at $+90^\circ$ of azimuth in this particular case.

3. Directional loudness sensitivities and binaural summation

In order to quantify a possible laterality in the directional loudness matches, the binaural summation underlying the directional loudness judgments was modeled. The modeling was carried out as reported in more detail in Sivonen and Ellermeier (2006), separately for the sound sources at $+90^\circ$ and -90° of azimuth.

The distal directional loudness matches were converted to *directional loudness sensitivities*, by subtracting the distal matches, the means of which are plotted in Fig. 6, from the frontal reference level of 65 dB SPL. Therefore, the lower the distal match, the greater the loudness sensitivity, and the louder the sound perceived at the particular instance. The loudness sensitivities thus derived

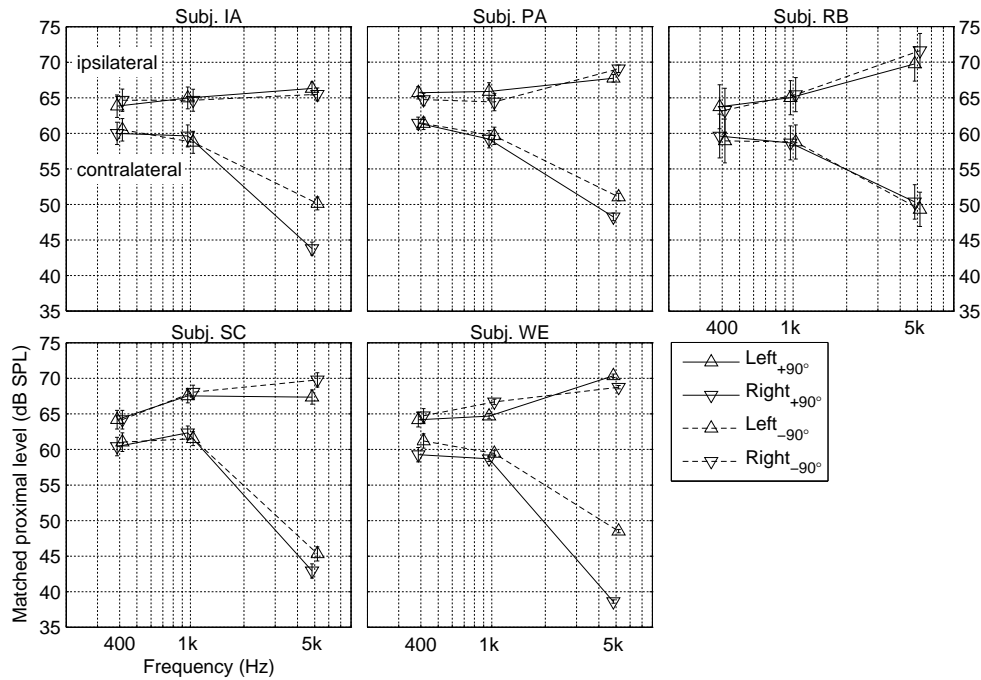


Figure 7: Binaural proximal sound pressure levels with 95%-confidence intervals (dB) for $+90^\circ$ and -90° of azimuth (on the left and on the right side of the listeners), matched to the same frontal reference in the horizontal plane.

are plotted along with the corresponding changes in at-ear levels in Fig. 8. The at-ear levels are plotted relative to the frontal levels, derived from individual HRTFs, by subtracting the frontal reference level from the at-ear levels of the comparison incidence angles separately for the left- and the right-ears at each center frequency.

Note that unlike the distal loudness matches, the directional loudness sensitivities can be directly compared to the objective HRTF-effects. For most listeners, the subjective data are between the objective at-ear levels in Fig. 8. For the two lowest center frequencies, the subjective data seem to largely follow the ear with the higher SPL, whereas at 5 kHz with larger ILDs, the ear with the lower SPL seems to influence the loudness sensitivity.

The best-fitting binaural gain was estimated to predict the directional loudness sensitivity patterns from the corresponding changes in the individual HRTFs. The estimates were derived as in Sivonen and Ellermeier (2006), separately for the two lateral sides and individually for each listener while pooling over center frequencies. The best-fitting gain constants are listed in Table II. For subject IA, the binaural summation values are

close to 0 dB for both lateral sides, i.e., denoting no binaural gain, loudness being determined by the ear with the higher SPL alone (see Fig. 8). For the other listeners, larger binaural gains are observed, meaning that the ear with the lower SPL also influenced the loudness judgments (see Fig. 8).

When inspecting the laterality of binaural summation, the estimates for all listeners but subject IA are smaller for the sound source at $+90^\circ$ than for the source at -90° of azimuth. This suggests that the ipsilateral ear in the former and the contralateral ear in the latter case have a relatively high influence on binaural summation. It thus seems that for these four listeners the *left* ear is emphasized slightly more in their binaural loudness judgments. Note that the laterality was statistically significant for subjects PA and WE, when analysing their distal directional loudness matches (see Fig. 6), and these are the two participants, for whom the difference between the binaural-gain estimates for the two sides are larger 3 dB. For subjects RB and SC, showing smaller differences between the sides, the effect of side was statistically insignificant in the analysis of the distal matches.

The individually estimated binaural gains are in a reasonable agreement with the estimates by Sivo-

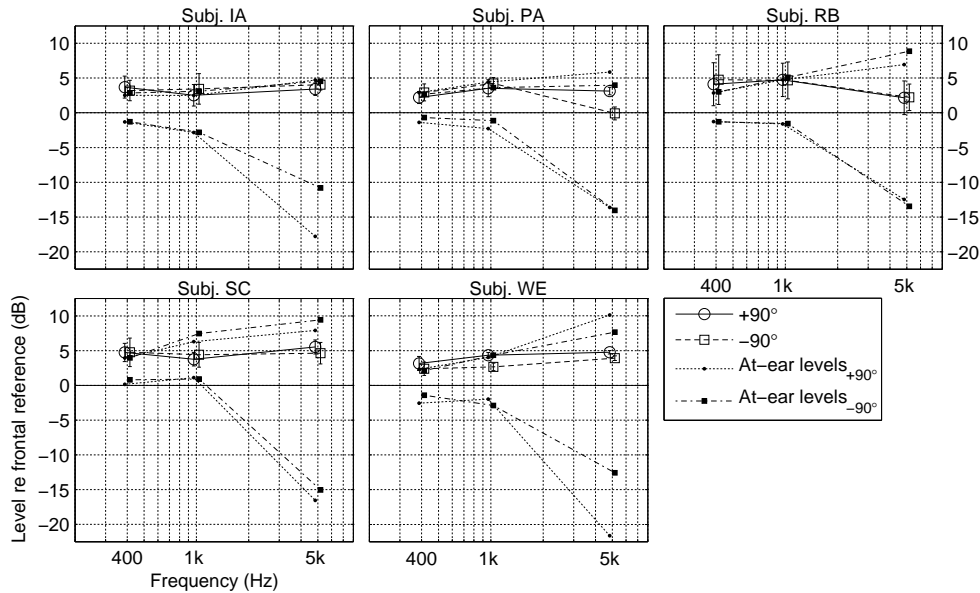


Figure 8: Directional loudness sensitivities with 95%-CI (dB) for $+90^\circ$ and -90° of azimuth (on the left and on the right side of the subjects) plotted along with the corresponding relative changes in at-ear levels, based on individual HRTFs.

Table II: Least-squares estimates for the amount of binaural loudness summation (in dB), aggregated over center frequencies, for the two lateral sound sources either at $+90^\circ$ or -90° of azimuth, i.e., on the left- and right-hand sides. The two bottom rows show the estimates when data are averaged across listeners, and finally aggregated across the sides. The estimates in the right-most column for sound sources at $+30^\circ$, $+60^\circ$, $+90^\circ$, $+135^\circ$ of azimuth in the left hemisphere of the horizontal plane are taken from Sivonen and Ellermeier (2006).

Subject	Azimuth		
	$+90^\circ$	-90°	$+30-135^\circ$ ^a
IA	0.1	0.1	0.1
PA (***)	1.3	4.8	2.1
RB	5.5	7.8	9.1
SC	2.8	5.0	3.8
WE (***)	0.1	4.9	2.8
Averaged data	2.3	4.2	2.6 ^b
Aggregated over sides		3.2	

^a Estimated in Sivonen and Ellermeier (2006).

^b The estimated average includes eight listeners.

nen and Ellermeier (2006), listed in the right-most column of Table II, and the order between the listeners is largely retained. Note that in Sivonen and Ellermeier (2006), the estimates were derived for sound sources located in the left hemisphere of the horizontal plane, with four different incidence angles at $+30^\circ$, $+60^\circ$, $+90^\circ$ and $+135^\circ$ of azimuth. The binaural summation estimates for the averaged data, and especially when aggregating over sides, come close to suggesting a 3-dB loudness summation between the two ears, a summation rule, which was also suggested in the mean data of Sivonen and Ellermeier (2006).

4. Summary

To sum up Experiment 2, the effect of side on the distal matches was significant for two out of five listeners, and only at some center frequencies. As in Experiment 1, the effect size was small compared to asymmetries observed at thresholds: the difference in the matches between the two sides was 3 dB at maximum.

When accounting for the differences in the physical transfer from the source to the listeners' ears, and modeling the underlying binaural summation, smaller binaural gains were obtained for the sound source on the left-hand side than for the sound source on the right-hand side. The smaller the

binaural gain, the more binaural loudness is determined by the ear receiving the higher exposure. This thus suggests that the influence of the left ear was relatively larger than that of the right ear on binaural loudness. Note that for the two listeners (subjects PA and WE), for whom this laterality was statistically significant in Experiment 2, an left-ear advantage was also observed in the monaural-to-binaural matches of Experiment 1.

C. Mean data

As in Sivonen and Ellermeier (2006), large differences between listeners were observed in the present investigation, both in the monaural-to-binaural loudness matches (see Fig. 4), and in the amount of binaural summation underlying the directional loudness matches (see Table II). Nevertheless, in order to assess the overall tendencies in the loudness judgments, the present data were averaged across listeners.

The mean monaural-to-binaural distal loudness matches of Experiment 1, relative to the frontal reference (the 0-dB horizontal), are plotted in panel 'a)' of Fig. 9. The mean left- and right-ear matches differ by less than 1 dB, and do not exhibit a clear ear difference, even though the right-ear matches are slightly higher at 1 and 5 kHz. Carrying out a repeated-measures ANOVA, similarly as for the individual data, but with the listener as a random factor, did not return any significant effects ($p > 0.17$), confirming the binaural loudness advantage, on average 3.4 dB, to be roughly constant across ears and center frequencies.

In panel 'b)' of Fig. 9, the mean directional loudness sensitivities obtained in Experiment 2 are plotted along with the corresponding changes in at-ear stimulation, and a 3-dB summation prediction based on the at-ear levels. The mean directional loudness sensitivities for both lateral sides are fairly constant, even though the at-ear levels vary with center frequency. The 3-dB summation, solely depending on the at-ear levels, thus fails to predict the subjective data at the two lowest center frequencies where the ILD is small. At 5 kHz, and with a considerably larger ILD, the 3-dB prediction is in good agreement with the data obtained at both lateral locations of $+90^\circ$ and -90° of azimuth. An ANOVA of the distal directional loudness matches did not detect any significant effects ($p > 0.39$), confirming that the side, on which the virtual source was created, did not have an effect

on loudness in the directional loudness paradigm when considering mean data.

The mean data from both Experiment 1 and 2 agree fairly well with the 3-dB binaural summation of loudness. They do not show unequivocal evidence for laterality: In Experiment 1, the monaural matches are approximately 3 dB higher than the binaural reference for *both* ears. In Experiment 2, even though for the mean data the estimated binaural gains differ between the two sides by almost 2 dB in Table II, the effect of side is insignificant. Furthermore, when data are aggregated across sides, the 3-dB binaural summation rule is corroborated.

IV. CONCLUSION

Laterality of loudness perception was investigated in the present study for spatial sound sources synthesized in individual binaural synthesis. In Experiment 1, the binaural loudness advantage was investigated acquiring loudness matches between monaural and binaural listening to a frontal sound source. In Experiment 2, directional loudness matches were obtained for sound sources either at $+90^\circ$ or -90° of azimuth, matched to the frontal sound source.

In both experiments, individual differences between listeners were observed, both in the listeners' loudness matches at various stimulus center frequencies, and occasionally in the laterality of the matches. In the data averaged across listeners, however, the effect of the side of stimulation on loudness matches was insignificant, the mean data supporting an earlier finding (Sivonen and Ellermeier, 2006) of a 3-dB binaural summation of loudness.

The present data, in addition to corroborating the 3-dB summation rule, thus suggest laterality only playing a minor role in binaural loudness perception of sound fields, it being negligible when considering mean data. The results can be used for modeling binaural loudness, and for the purpose of utilizing artificial-head measurements for loudness predictions.

Notes

¹Some listeners reported loudness matches of sounds, where one of the sounds was clearly localized close to the ear (monaural) and the other being further away (binaural), being difficult. This may have influenced the listeners' consistency in Experiment 1.

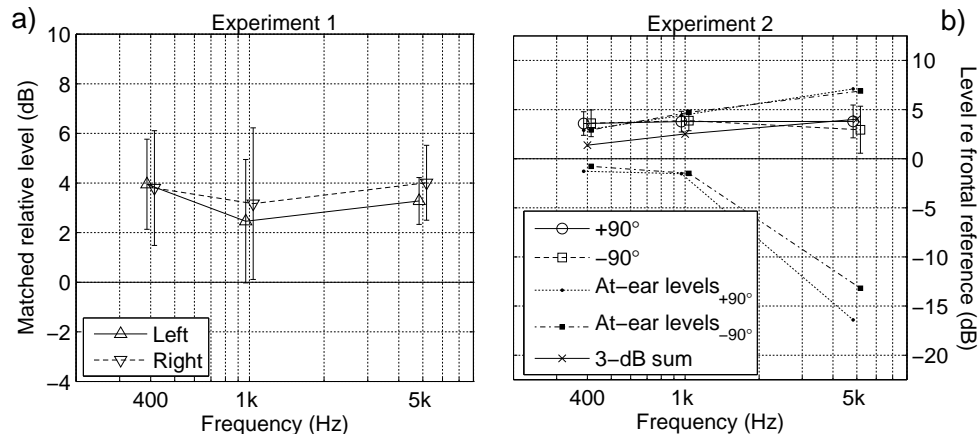


Figure 9: Average data: a) Monaural loudness matches (dB) for left and right ears and b) directional loudness sensitivities for $+90^\circ$ and -90° of azimuth in the horizontal plane relative to a binaural frontal reference with 95-% confidence intervals of the means. In panel 'b)', the corresponding relative changes in the at-ear SPLs and prediction based on a 3-dB summation of the at-ear levels are plotted along with the subjective data.

References

- Emmerich, D.S., Harris, J., Brown, W.S. and Springer, S.P. (1988). "The relationship between auditory sensitivity and ear asymmetry in a dichotic listening task," *Neuropsychologia*, **26**, 133–143.
- Hellman, R.P. and Zwislocki, J. (1963). "Monaural loudness function at 1000 cps and interaural summation," *J. Acoust. Soc. Am.*, **35**, 856–865.
- ISO 389-1 (1998). "Acoustics – Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," (International Organization for Standardization, Geneva, Switzerland).
- ISO 8253-1 (1991). "Acoustics – Audiometric test methods – Part 1: Basic pure tone air and bone conduction threshold audiometry," (International Organization for Standardization, Geneva, Switzerland).
- Jesteadt, W. (1980). "An adaptive procedure for subjective judgments," *Percept. Psychophys.*, **28**, 85–88.
- Marks, L.E. (1978). "Binaural summation of the loudness of pure tones," *J. Acoust. Soc. Am.*, **64**, 107–113.
- Moore, B.C.J. (2003). *An Introduction to the Psychology of Hearing* (Academic Press, London).
- Reynolds, G.S. and Stevens, S.S. (1960). "Binaural summation of loudness," *J. Acoust. Soc. Am.*, **32**, 1337–1344.
- Robinson, D.W. and Whittle, L.S. (1960). "The loudness of directional sound fields," *Acustica*, **10**, 74–80.
- Schneider, B.A. and Cohen, A.J. (1997). "Binaural additivity of loudness in children and adults," *Percept. Psychophys.*, **57**, 655–664.
- Sivonen, V.P. and Ellermeier, W. (2006). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation," *J. Acoust. Soc. Am.*, **119**, 2965–2980.
- Sivonen, V.P., Minnaar, P. and Ellermeier, W. (2005). "Effect of direction on loudness in individual binaural synthesis," in *Proceedings of the Audio Engineering Society 118th Int. Conv.*, Barcelona, Spain, May 28–31, preprint 6512.
- Zwicker, E. and Zwicker, U.T. (1991). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," *J. Acoust. Soc. Am.*, **89**, 756–764.

Binaural loudness summation for directional sounds*

Ville Pekka Sivonen^{1,2} and Wolfgang Ellermeier¹

¹*Sound Quality Research Unit, Dept. of Acoustics, Aalborg University, 9220 Aalborg, Denmark*

²*Brüel & Kjær Sound & Vibration Measurement A/S, 2850 Nærum, Denmark*

Abstract

In an earlier study (Sivonen and Ellermeier, 2006), directional loudness matches were obtained from eight listeners, and their individual head-related transfer functions (HRTFs) were utilized to model the binaural summation underlying the loudness matches. Even though the effects of HRTFs were taken into account, considerable individual differences in the binaural summation remained. In order to create conditions in which the directional changes in at-ear signals were identical for all participants, the present experiment employed 'generic' HRTFs measured for an artificial head to create sound sources positioned in space by means of binaural synthesis. In an adaptive forced-choice procedure, five listeners, all of whom had participated in the earlier study, matched the loudness of sounds coming from five different incidence angles in the horizontal plane to the same sound with frontal incidence. The directional loudness matches were obtained for narrow-band stimuli at two different center frequencies (1 and 5 kHz) and at an overall level corresponding to 65 dB SPL in the free field. When inspecting the results of the present experiment, however, large individual differences in the binaural summation were still evident. The generality of this finding was further corroborated by running an independent sample of 10 inexperienced participants, exclusively being exposed to the present generic HRTFs. Despite the individual differences, the average results suggest a relatively simple rule for combining the binaural at-ear signals when making acoustical measurements using an artificial head. That is particularly useful when predicting loudness for dichotic listening situations, as they often result from real sound sources positioned in space.

I. INTRODUCTION

Acoustical measurements of sound pressure increasingly rely on the use of artificial heads. It is unclear, however, how such measurements should be used to predict loudness. The standardized loudness model ISO 532 (1975) utilizes the sound signal measured with a *monophonic* microphone in the absence of a listener, while with an artificial head sound signals at *each ear* are measured, being obstructed by the physical dimensions of the head and torso.

In many applications this problem has been solved by the use of head-related transfer functions (HRTFs; Shaw, 1974; Wightman and Kistler, 1989; Møller et al., 1995; Blauert, 1997, chap. 5): these functions characterize the transfer of sound from an unobstructed free field to the listener's two

ears for a given sound incidence angle. In order to utilize artificial-head measurements for loudness predictions, the inverse of an HRTF can be used to transform an at-ear signal to the corresponding signal in the center of the head with the head absent. This technique of converting an at-ear exposure to the corresponding free- or diffuse-field exposure has recently been standardized for measurements on real ears (ISO 11904-1, 2002) or using an artificial head (ISO 11904-2, 2004).

In addition to correcting for the measurement point, the binaural nature of the stimulation has to be taken into account when making artificial-head measurements. When the measurement is diotic, i.e., the same signal arrives at the two ears, *either* of the at-ear signals can be converted to the corresponding free- or diffuse-field signal in the absence of the head, and subsequently be utilized for loudness computations¹. Free- and diffuse-field HRTFs are similar for the two ears due to symmetry, and

*Portions of the data presented at Euronoise 2006, May 31-June 1, 2006, Tampere, Finland.

thus, the conversion using either at-ear signal results essentially in the same outcome in the diotic case.

When the measurement is dichotic, the at-ear signals differ from one another. Since the free- and the diffuse-field HRTFs are similar for the two ears, a dichotic measurement yields two different signals in the absence of a listener when converting at-ear signals to the corresponding free- or diffuse-field signal. Applying these signals to compute loudness will result in different predictions, depending on the ear chosen. Dichotic at-ear signals are often obtained when making acoustical measurements in the field. This due to sound sources being located non-symmetrically with respect to the ears of an artificial head or reverberation in the sound field having non-symmetrical effects on the signals arriving at the ears.

It is thus important to investigate how, for a sound source located in space, the two at-ear signals are summed to yield a single percept of binaural loudness. Loudness matches of anechoic sounds coming from different directions, and HRTFs measured for the same incidence angles, can be utilized in investigating how changes in the at-ear signals influence perceived loudness. Since HRTFs describe the transfer from a free-field to the listeners' ears, they can be used in converting the matched sound pressure level (SPL) of a source measured in the absence of the head to equal-loudness at-ear SPLs. These at-ear SPLs can then be used to model binaural loudness summation of directional sounds arriving at the listeners' ears from various incidence angles.

Two investigations employed this strategy (Robinson and Whittle, 1960; Sivonen and Ellermeier, 2006), acquiring both *directional loudness matches* from various incidence angles to a reference direction and relating the matches to the changes in the at-ear stimulation as a function of incidence angle. Both described the directional loudness matches to agree with a relatively simple combination of the left-ear (L_{left}) and right-ear (L_{right}) sound pressure levels, expressed by equivalent monaural levels:

$$L_{mon} = g * \log_2(2^{L_{left}/g} + 2^{L_{right}/g}) \quad (1)$$

However, the ‘‘binaural gain’’ constants g were different, Robinson and Whittle (1960) reporting a ‘6-dB’ binaural-summation rule ($g = 6$) to fit the data best, while in Sivonen and Ellermeier (2006), a ‘3-dB’ gain ($g = 3$) was found to describe the

mean data much better.

In the latter study (Sivonen and Ellermeier, 2006), the modeling was also performed on an individual basis. Even though the effects of individual HRTFs were taken into account, large individual differences remained when fitting the parameter g . However, since the SPLs arriving at the listeners' ears owing to physical differences in the shapes of pinnae and heads were highly idiosyncratic as well, fairly large interindividual differences in stimulation resulted, questioning the validity of comparisons across individuals.

With non-individual human, or generic artificial-head HRTFs the *physical* differences between listeners can be ruled out: generating directional sounds with such HRTFs in *binaural synthesis*, the same directional effects on at-ear exposures can be delivered to each listener in obtaining directional loudness matches. Thus, the comparison of the raw directional loudness matches across individuals is more straightforward, since the effect of individual HRTF filtering does not have to be accounted for. Non-individual or generic HRTFs must be played back using binaural synthesis over headphones, which is justified in the directional loudness paradigm; individual binaural synthesis has been shown to result in essentially the same directional loudness matches as listening to real sound sources (Sivonen et al., 2005).

The goals of the present investigation are threefold: (1) To use generic, rather than individual HRTFs in order to investigate whether holding this source of interindividual variation constant will reduce the variability seen in directional loudness matches. (2) To run both a naive sample of listeners, and an expert group that had previously gone through numerous conditions of real-source, and binaurally synthesized directional loudness experiments, in order to explore the generality of the findings. (3) To sketch a model for predicting binaural loudness using artificial-head measurements based on the best account of the average data.

II. METHOD

A. Subjects

Five listeners (4 male, 1 female, age between 26 and 49 years; median 28 years) took part in the present study, and they all had participated in earlier experiments in a real sound field (Sivonen and

Ellermeier, 2006), or using binaural synthesis with individual HRTFs (Sivonen et al., 2005). These “expert” listeners were familiar with the task of comparing the loudness of sounds coming from different directions. An additional group of 10 listeners (7 male, 3 female, age between 22 and 31 years; median 26), considered “naïve” to the task, were recruited from a student population.

All listeners were screened for normal hearing, with the requirement that none of the hearing thresholds exceed 20 dB hearing level re ISO 389-1 (1998).

B. Apparatus

An artificial head (Brüel & Kjær Head and Torso Simulator Type 4100), which is commercially available and commonly used in the field, was used in acquiring the generic HRTFs. The HRTFs were measured with the same procedure as reported in Sivonen and Ellermeier (2006), with the exception that the sound pressures both in the absence of the artificial head (p_1 pressure, see Møller et al., 1995) and at the entrance to the ear canal (p_2 pressure, see Møller et al., 1995) were measured using the same type of microphones (the built-in microphones of the artificial head; Brüel & Kjær Type 4190). Despite using the same microphone type, a correction for the differences in the frequency responses of the dedicated left- and right-ear microphones was determined by measuring the same sound field simultaneously with both microphones placed approximately 3 mm apart. The subsequent signal processing for obtaining the HRTFs was as described in Sivonen and Ellermeier (2006).

In order to equalize the binaural playback, generic headphone transfer functions (PTFs) were measured for a pair of headphones (Beyerdynamic DT-990) using the artificial head, as typically measured in practice. The playback setup and the rest of the apparatus were as reported in Sivonen et al. (2005).

The listeners were seated inside an anechoic chamber in a chair with a headrest to restrict head movements. The listeners’ responses were collected with a two-button response box, and a model of the box having larger indicator lights was placed in front of the listeners, to avoid the listeners tilting their heads downwards to the response box in their hands. Loudspeakers mounted inside the anechoic chamber at the intended incidence angles were visible to the listeners in order to improve the plausibility of the binaural synthesis.

C. Stimuli

Third-octave noise bands centered at 1 and 5 kHz were used as stimuli. The length of each sound was 1 s, including 20-ms raised-cosine rise and decay ramps.

The two noise bands were convolved with the generic HRTFs for six incidence angles in the left hemisphere of the horizontal plane. The incidence angles were 0° , 30° , 60° , 90° , 135° and 180° of azimuth at an elevation of 0° , i.e. from ahead to behind the artificial head, at the height of the ears of the head.

The left- and right-ear magnitude spectra of the generic HRTFs are plotted in Fig. 1, in order to illustrate the effective stimulation for each incidence angle. As seen in Fig. 1, interaural level differences (ILDs) for the incidence angles ahead (0°) and behind (180°) the artificial head are small due to symmetry. By contrast, large ILDs of over 30 dB can be observed for the other incidence angles. Note that the same holds for the interaural time difference (ITD), where the ITD for the direction on the side (i.e., at $\pm 90^\circ$ of azimuth) reaches its maximum, while the ITDs for sound sources ahead and behind the head are close to zero.

In addition to convolving the two noise bands with the HRTFs, minimum-phase equalization filters for the headphones, as specified in Sivonen et al. (2005), were applied separately for the left- and right-ear signals in the binaural synthesis. Generic filters based on the PTF measurements using the artificial head were applied. Thus, the equalization was done for PTFs measured exactly at the same position in the ear canal as the HRTFs, but the individual coupling of the headphones to each listener’s ears was not accounted for. However, this coupling has no effect on the *directional* changes the HRTFs impose on the at-ear sound pressure levels. Furthermore, when inspecting the magnitude spectra of measured PTFs for each listener and the artificial head, the differences between individual and generic transfer functions were small (less than 4 dB, symmetrically affecting both ears) in the frequency range used in the present study.

In the listening test proper, the stimuli were played back at an overall level corresponding to 65 dB SPL in the free field. In order to derive the effective *at-ear* SPLs for the various incidence angles, the gains on the third-octave bands caused by the HRTFs are listed in Table I. The 65-dB overall level combined with the ‘Left’ and ‘Right’ gains yields the SPLs at the ears of the artificial head.

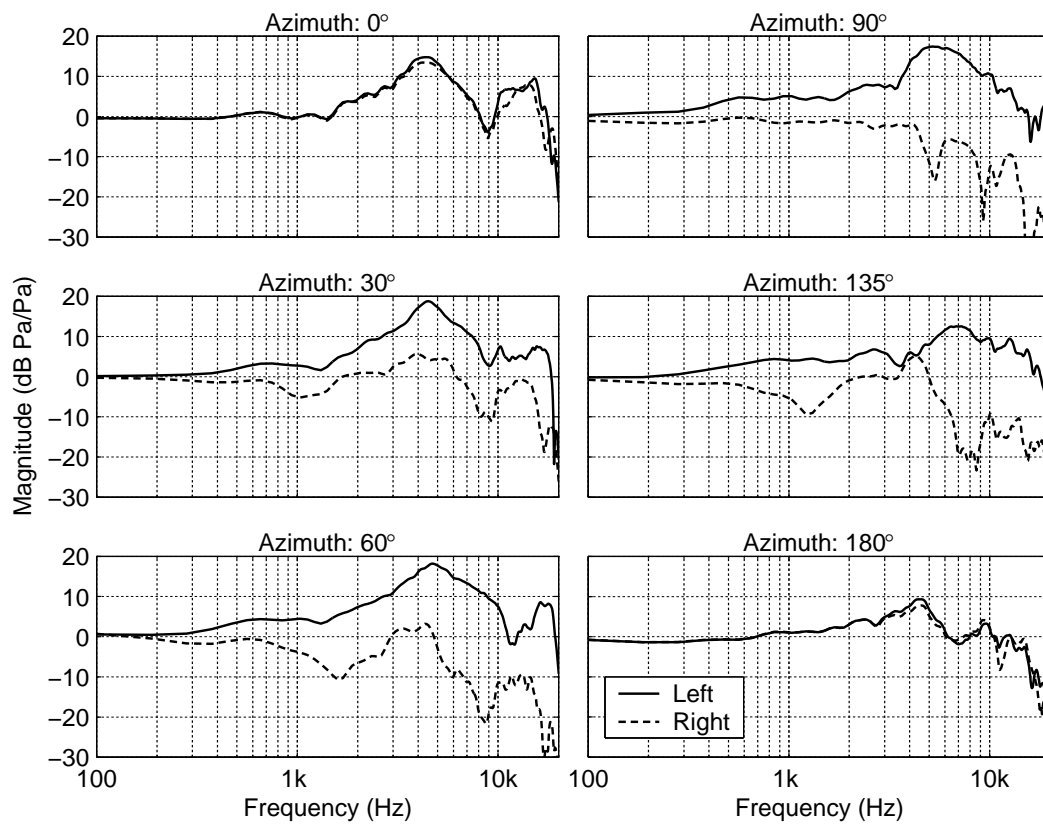


Figure 1: HRTF magnitude spectra for the artificial head, measured from six incidence angles in the left hemisphere of the horizontal plane (0° of azimuth: ahead, 180°: behind).

Table I: Third-octave-band levels (in dB Pa/Pa) for the generic HRTFs, and the ILD at the two stimulus center frequencies.

Az.	1 kHz			5 kHz		
	Left	Right	ILD	Left	Right	ILD
0°	-0.1	-0.2	0.1	13.1	12.2	0.9
30°	2.7	-4.8	7.5	17.1	4.2	12.9
60°	4.3	-3.8	8.1	17.5	-1.5	19.0
90°	4.9	-1.5	6.4	17.2	-9.8	27.0
135°	4.1	-5.6	9.7	8.5	0.7	7.8
180°	1.1	1.2	0.1	7.1	5.6	1.5

Furthermore, the ‘ILD’ column confirms the SPLs at the two ears being similar for the sound sources ahead and behind the head, while large ILDs are observed for the other incidences, especially at the higher center frequency.

D. Procedure

An adaptive, two-alternative, two-interval, forced-choice procedure was used in obtaining directional loudness matches between a sound of a given center frequency with synthesized frontal (reference) incidence and the same sound with one of the other five (comparison) incidence angles.

Loudness matches were obtained for 10 conditions (five comparison incidence angles \times two center frequencies), and eight replications were collected in each condition. In order to investigate the baseline variability of the loudness matches, the naïve group also had to match the loudness of the frontal reference to itself, just as the experienced listeners had done in Sivonen et al. (2005) and Sivonen and Ellermeier (2006).

The listener’s task in each trial was to judge which of the sounds in the pair was louder. The level of the frontal reference sound was fixed (corresponding to 65 dB SPL in the free field) and the level of the comparison sound was varied by the adaptive procedure. Half of the replications started +10 dB and the other half -10 dB from the level of the frontal reference. The initial step size of an adaptive track was 4 dB, and after two reversal it was decreased to 1 dB. A track was terminated after eight reversals, and the levels of the last six reversals were averaged for an estimate of the loudness match.

The adaptive tracks were interleaved in blocks of eight tracks and the order of the adaptive tracks was randomized within one set of replications.

One 1-hour listening session included either three or four blocks, each lasting approximately 10 mins, and rest breaks between the blocks. In total, the listening test lasted approximately three hours per listener, the expert listeners participating in 10 and the naïve listeners in 11 blocks.

III. RESULTS AND DISCUSSION

A. Directional loudness sensitivities

The listeners matched the loudness of a synthesized sound from the comparison incidence angles to the same sound with frontal incidence. These matches yielded sound pressure level adjustments for each incidence angle, denoting that if a sound emanating with a given angle had to be *attenuated* relative to the frontal reference, this incidence angle was perceived *louder*. Therefore, the relative level adjustments were inverted to denote *directional loudness sensitivities*, in agreement with the earlier studies (Sivonen et al., 2005; Sivonen and Ellermeier, 2006).

1. Expert listeners

The directional loudness sensitivities are plotted in Fig. 2 for the five expert listeners. For a given individual, the loudness-sensitivity curves vary as a function of incidence angle over a range of up to 4 dB and 10 dB at 1 and 5 kHz, respectively. At 1 kHz, the data points are generally *above* the 0-dB line, indicating that loudness is increased relative to the frontal reference. At 5 kHz, the data are both above and below the (0-dB) level of the frontal reference. A two-factor (incidence angle \times center frequency) repeated-measures analysis of variance (ANOVA) indicated that both the main effect of incidence angle [$F(4, 16) = 50.10; p < 0.001$], and its interaction with center frequency [$F(4, 16) = 39.48; p < 0.001$], were highly significant. This implies that loudness is not constant as a function of incidence angle, and that the directional loudness-sensitivity curves are different for the two center frequencies (left and right panel of Fig. 2).

Along with the subjective data, the changes in the at-ear SPLs are plotted in Fig. 2, normalizing the third-octave band levels of Table I by subtracting the levels of the frontal reference from them. Thus, the objective changes in the at-ear

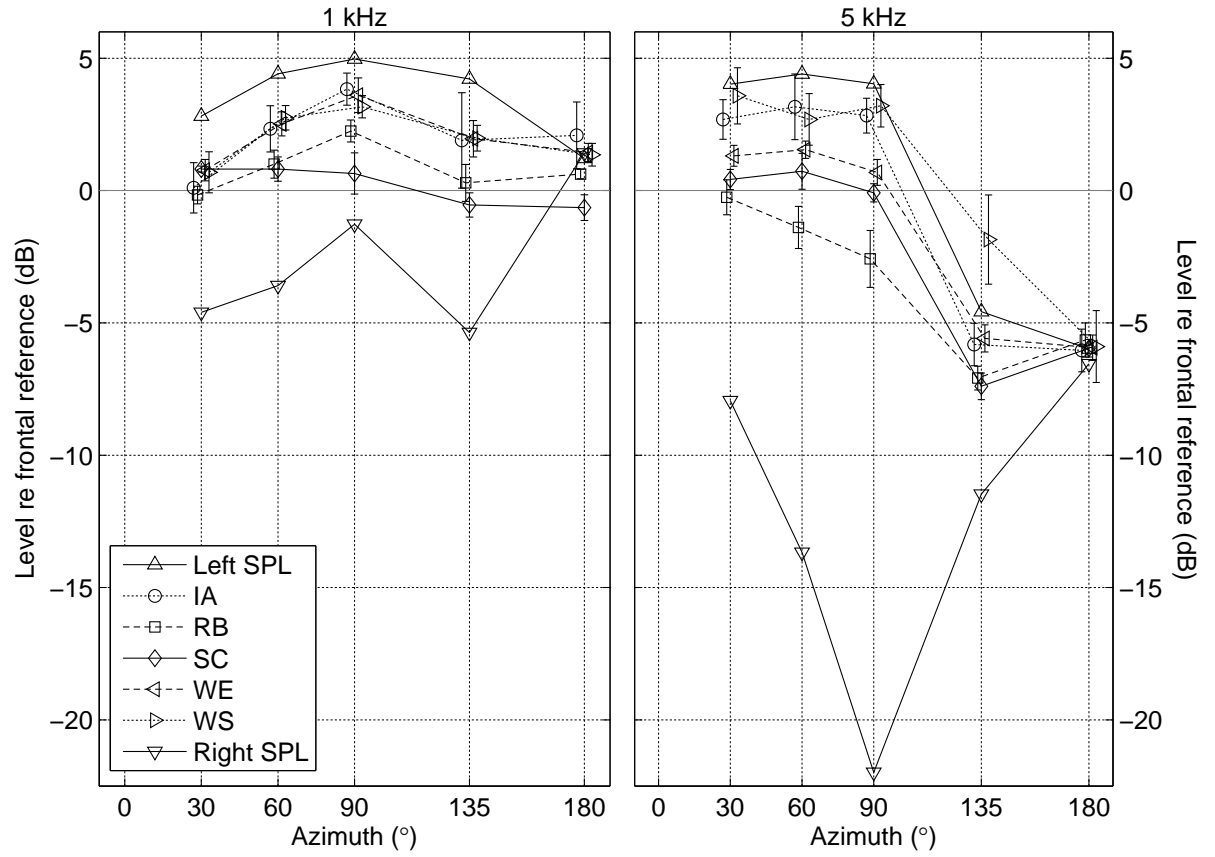


Figure 2: Directional loudness sensitivities $\pm 95\%$ confidence intervals for five expert listeners and the normalized at-ear SPLs derived from the generic HRTFs. See text for details.

stimulation can be compared with the directional loudness sensitivities. It is evident in Fig. 2 that for both center frequencies the directional loudness sensitivities largely follow the shape of the curve tracing the (left) ear receiving the higher SPL. However, individual differences can be observed, even though the changes in at-ear SPLs as a function of incidence angle are the same for all listeners. These individual differences reach up to 6 dB, see e.g., the 5-kHz data at an azimuth of 90° in Fig. 2. The statistical significance of the differences between listeners is confirmed by inspecting the interaction between incidence angle, center frequency and listener: $F(16, 350) = 8.28; p < 0.001$. As is evident in Fig. 2, the differences between the participants' curves are much larger than the confidence intervals defined by the replicability of a given experimental condition for a given listener.

2. Naïve listeners

The directional loudness sensitivities for the 10 naïve listeners are plotted in Fig. 3. As was the case for the expert listeners in the earlier studies (see e.g., Sivonen et al., 2005), all naïve listeners were also able to match the synthesized frontal reference to itself, see the data at 1 kHz at 0° azimuth, left-most data points in Fig. 3. For the remaining conditions, the range of variation in the individual directional loudness sensitivities at both center frequencies is similar to what is observed for the expert listeners. More importantly, the differences *between* the 10 individual listeners seem to be at least on the same order of magnitude as obtained for the group of experienced listeners, implying that the idiosyncracies in the data from the expert listeners were replicated for another independent sample of listeners.

Statistical analysis of the naïve listeners' data yielded the same significant effects (of incidence angles, and its interaction with center frequency) as had been obtained for the experts, including the significance of differences between listeners: $F(16, 700) = 2.59; p < 0.001$.

3. Mean data

In order to further explore the group differences between the expert and the naïve listeners, the means obtained from the two groups are plotted in Fig. 4. Even though the mean directional loudness sensitivities from the naïve group seem to be slightly below the corresponding data from the ex-

pert listeners, the 95-% confidence intervals are overlapping in all conditions suggesting the two curves to be statistically indistinguishable. This was confirmed by a three-factor (incidence angle \times center frequency \times group) mixed ANOVA, in which neither the group factor, nor any of its interactions with the other factors were significant (all $p > 0.25$).

Finally, Fig. 4 shows a prediction based on combining the at-ear levels according to a 3-dB summation rule (Eq. 1: $g = 3$), which is equivalent to a power sum of the at-ear pressures, and which appeared to fit the anechoic data reported by Sivonen and Ellermeier (2006). This 3-dB prediction seems to account for the mean data of the expert listeners quite well, even though it is based on mean data utilizing *individual* HRTFs and eight listeners (Sivonen and Ellermeier, 2006), and here, *generic* HRTFs of an artificial head for a subset of five listeners were used.

B. Binaural loudness summation

As seen in Figs. 2 and 3 for both groups of listeners, individual differences in the listeners' directional loudness sensitivities exist, even though the directional changes in the at-ear SPLs are the same for all listeners since generic HRTFs were used. In purely sensory terms, this appears to imply that the listeners sum the signals at their ears in an individual manner to yield a binaural loudness.

In Sivonen and Ellermeier (2006), binaural loudness summation was modeled based on individual directional loudness sensitivities and at-ear SPLs derived from individual HRTFs, utilizing Eq. 1 and estimating the binaural gain factor. Considerable differences between listeners were obtained, the amount of binaural summation ranging from close to 0 dB to far over 10 dB. This kind of modeling was also performed with the present data based on generic HRTFs. The third-octave-band SPLs listed in Table I were entered as left- and right-ear SPLs. Minimizing the sum-of-squares of the errors between the prediction (L_{mon}) and directional loudness sensitivity, the amount of binaural summation needed was estimated to best predict each expert listener's data, see Fig. 2. For details of the modeling, see Sivonen and Ellermeier (2006).

The listener-specific binaural loudness-summation values (g in Eq. 1) are listed in Table II. The individual (Ind.) values are taken from Sivonen and Ellermeier (2006), where the modeling was based on individual HRTFs, and the

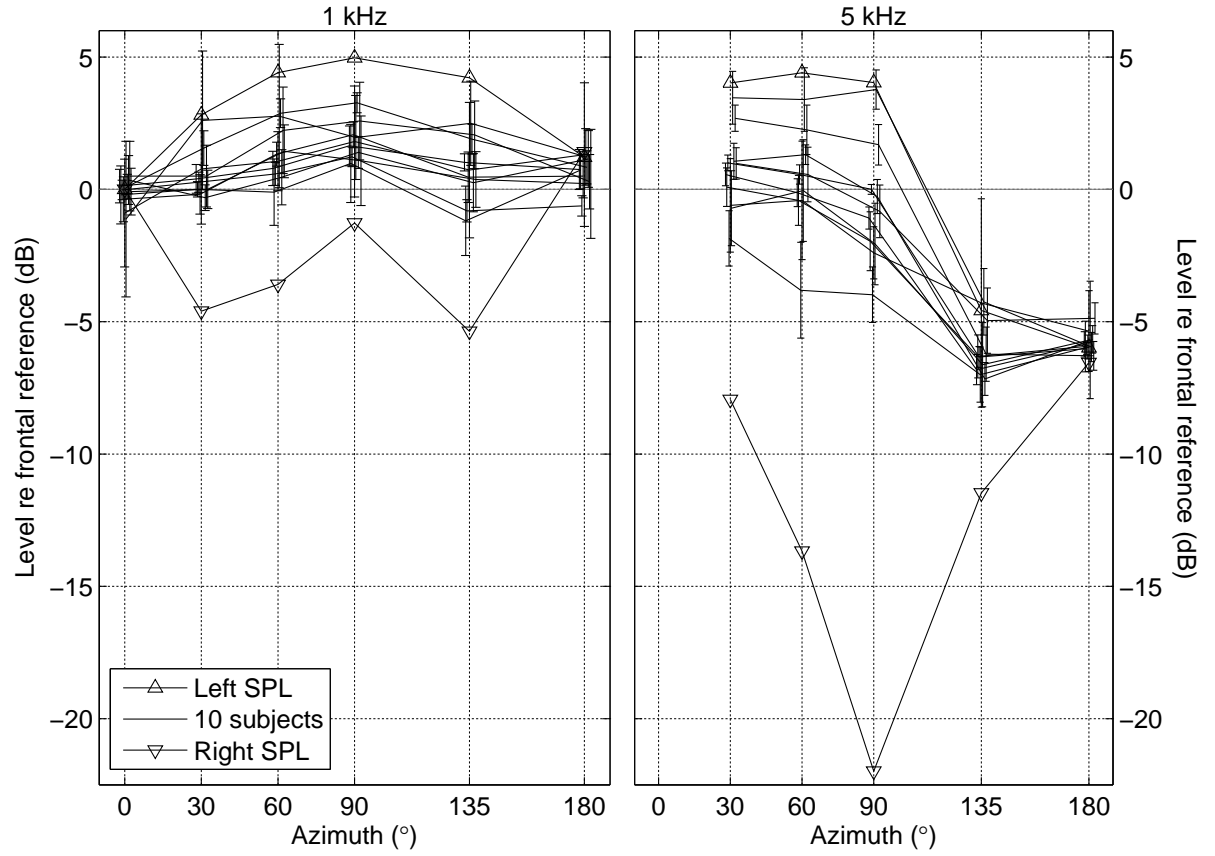


Figure 3: Directional loudness sensitivities $\pm 95\%$ confidence intervals for 10 naïve listeners and the normalized at-ear SPLs derived from the generic HRTFs. See text for details.

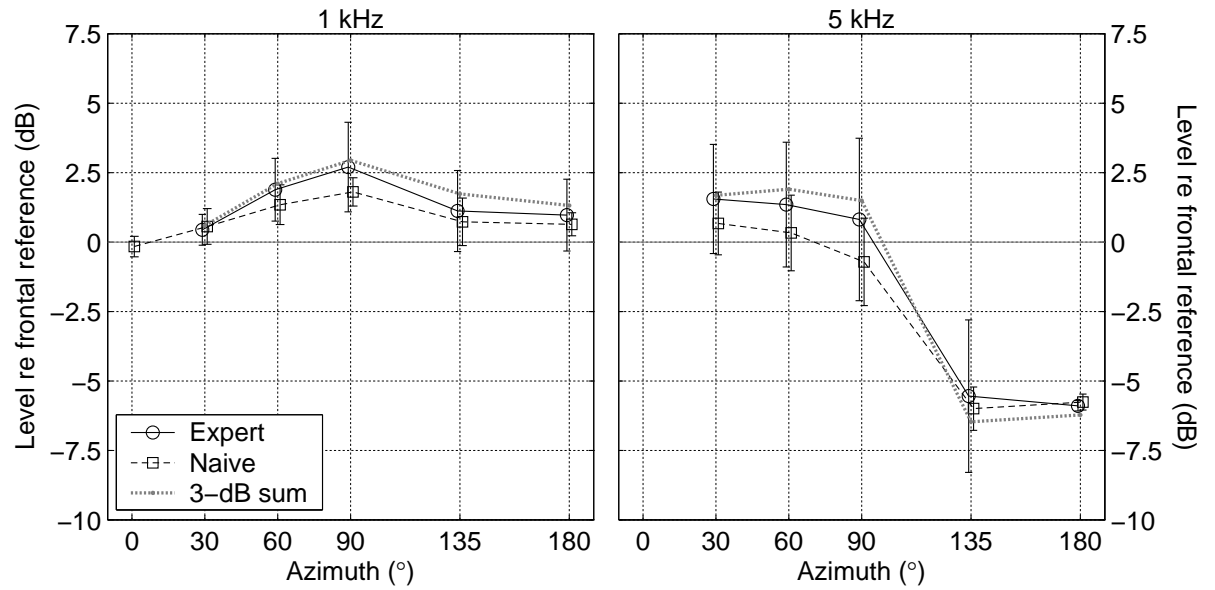


Figure 4: Mean directional loudness sensitivities $\pm 95\%$ confidence intervals for the expert ($n = 5$) and the naïve ($n = 10$) listeners, and a prediction based on a '3-dB' summation of the at-ear SPLs. See text for details.

generic (Gen.) values are based on the modeling of the present data. The modeling was also carried out aggregating the data across center frequencies, and finally, averaging the directional loudness sensitivities across listeners for both groups and aggregating over center frequencies (the right-most column and the two bottom rows in Table II, respectively).

As seen in Table II, there is considerable variation in the amount of binaural loudness summation between the listeners. This is true both for the data obtained using individual HRTFs (Ind.), and the present data utilizing generic HRTFs in binaural synthesis (Gen.). Thus, even though the same directional effects were played back in the latter case for all listeners, they still tend to sum their left- and right-ear SPLs differently.

Moreover, when comparing the summation values between the two studies in Table II for each listener and at each center frequency, or for the best fit across center frequencies, remarkable similarities may be observed. Even though the actual values may be numerically different, the rank order among listeners is largely preserved, see Table II. It thus seems that with a few exceptions, the individual binaural summation process of the at-ear SPLs is retained when generic HRTFs are used.

The two bottom rows in Table II show the amount of summation for the averaged directional loudness-sensitivity data, aggregated over the two center frequencies. The estimate for the expert listeners comes fairly close to a 3-dB summation for both studies. For the naïve listeners, for whom only for the generic HRTFs were investigated, the estimate is slightly larger, which can also be seen in Fig. 4 as a discrepancy between the data sets plotted². According to the statistical analysis, this small group difference of 0.6 dB on average in Fig. 4, and 1.7 dB in terms of the binaural gain in Table II, is most likely due to chance.

C. Binaural loudness for artificial-head measurements

Despite the apparently robust individual differences, the mean data may be utilized for developing a binaural loudness model. Note that mean data have typically been used for modeling loudness perception, even though the signals at the ears of a listener may be very individual due to the listener's pinnae, head and torso interacting with a real sound field or his ears coupling to the head-

phones.

The proposed binaural model is applicable to measurements carried out with an artificial head in any type of sound field, resulting in diotic *or* dichotic at-ear signals. Since the model is based on at-ear signals, no assumptions about the sound field need to be made, as opposed to measurements carried out with a monophonic microphone.

A block diagram of the binaural loudness model for artificial-head measurements is depicted in Fig. 5. Sound pressure signals are first measured at the left and right ears of the head [$p_{2L}(t)$ and $p_{2R}(t)$, respectively]. The magnitude spectra of these signals [$|P_{2L}(f)|$ and $|P_{2R}(f)|$, respectively] are then analyzed using a given frequency resolution, e.g., third-octave bands.

The 3-dB summation of sound pressure *levels* corresponds to a power summation of sound *pressures*. Thus, a power sum of the left- and right-ear magnitude spectra is subsequently computed. The loudness of this binaural power sum corresponds to the loudness of the same signal presented monotonically [$|P_{2M}(f)|$], i.e., to one ear only.

The standardized loudness model ISO 532 (1975) utilizes measurements carried out with a monophonic, omnidirectional microphone in the absence of a listener either in the free or diffuse field [$|P_1(f)|$ in Fig. 5]. The listening to the two types of sound fields, however, is always binaural (more specifically, the long-term spectra at the two ears are roughly the same due to symmetry). Therefore, the monotonic at-ear signal resulting from the power sum must be converted to the corresponding diotic signal. In case of the binaural power summation, the conversion for the linear magnitude spectra is simply a division by $\sqrt{2}$, corresponding to subtracting a binaural gain of 3 dB from the monotonic at-ear sound pressure level.

The diotic at-ear signal [$|P_{2D}(f)|$] is then converted to the corresponding signal in the absence of the head [$|P_1(f)|$] using the inverse of an HRTF. Here, the same incidence (free or diffuse) must be used both for the inverse HRTF and the sound field type in the loudness model.

In Sivonen and Ellermeier (2006), the performance of the present binaural model was contrasted with the most recent, and widely accepted loudness model by Moore et al. (1997). In this model, binaural dichotic loudness is handled by calculating monaural loudness values (in sones) separately for the left and the right at-ear signals assuming diotic stimulation, and computing bin-

Table II: The amount of binaural loudness summation (binaural gain in dB) estimated from data with individual (Ind.; Sivonen and Ellermeier, 2006) and generic (Gen.; the present data) HRTFs at an overall level of 65 dB SPL.

Subject	f_c				Best fit across f_c	
	1 kHz		5 kHz			
	Ind.	Gen.	Ind.	Gen.	Ind. ^a	Gen.
IA	0.1	2.7	0.7	1.7	0.1	2.0
RB	4.0	8.7	10.0	8.3	9.1	8.4
SC	13.1	20.4	3.1	4.8	3.8	5.2
WE	0.1	2.3	4.9	3.5	2.8	3.2
WS	1.7	2.4	0.1	0.3	0.7	1.3
Averaged data: Expert					2.6	3.5
Averaged data: Naïve						5.2

^aThe best fit included a third f_c at 0.4 kHz.

aural loudness as the arithmetic mean of the two monaural values: $N_{binaural} = (N_{left} + N_{right})/2$. Thus, it is assumed that monaural loudness is half of the corresponding diotic loudness, an assumption which is supported by earlier headphone studies on binaural loudness summation (see e.g., Hellman and Zwislöcki, 1963; Marks, 1978). It is worth noting, however, that headphone studies neglect the effect of HRTF filtering, producing auditory events localized inside the listeners head, and possibly generating ILDs that would never reach the listener’s ear in a real sound field.

In Sivonen and Ellermeier (2006, Fig. 8), the performance of the present binaural model (depicted in Fig. 5) was compared with predictions from the model by Moore et al. (1997). For externalized sound sources located in space, the present model performed considerably better than the model by Moore et al. (1997) which overestimated the effect of the ear receiving the lower sound pressure level. Since perfect summation of loudness in sones (as assumed in Moore et al., 1997) for the narrow-band stimuli of the present study fairly close corresponds to a 10-dB binaural gain, the analyses rendered in Table II, which generally show much smaller gains, also argue against this notion.

It thus seems that when investigating binaural loudness summation of directional sounds, the assumption of perfect summation in sones does not hold, while the 3-dB (power) summation³ gives a better prediction.

IV. CONCLUSION

The effect of sound incidence angle on loudness was investigated in the horizontal plane using *generic* HRTFs in binaural synthesis.

(1) Loudness matches to a frontal reference were by and large indistinguishable from earlier data obtained via *individual* binaural synthesis, as well as in a real sound field: they showed the same directional dependencies and interindividual variation.

(2) Naïve and expert listeners produced results that were indifferent, statistically, both with respect to mean values and the presence of reliable interindividual differences. This argues for the generality of the findings, and against experts developing peculiar, idiosyncratic listening strategies.

(3) The fact that derived binaural summation parameters largely preserved the same rank order across individuals, no matter whether the synthesis was individual (as in earlier studies) or generic (as in the present experiment), suggests that a large part of the interindividual differences is due to the way binaural inputs are weighted and combined, and not to peculiarities in HRTF filtering.

(4) The main feature of a model describing the mean results is a power summation of the sound pressures at the two ears. A proposal on how to implement this model when making artificial-head measurements in order to compute binaural loudness is outlined.

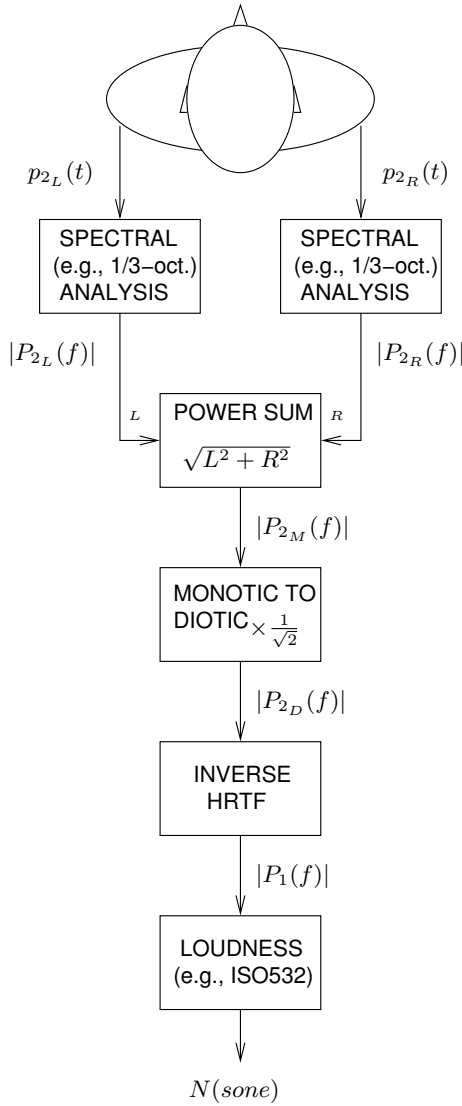


Figure 5: Binaural loudness model (pat.pend., Sivonen et al., 2004) for artificial-head measurements. Sound pressure signals are measured at the two ears, a power sum of the at-ear magnitude spectra is computed and converted to the corresponding diotic signal. An inverse HRTF is utilized to determine the sound signal in the absence of the head, which is subsequently used e.g., as an input to the standardized loudness model ISO 532 (1975). See text for details.

V. ACKNOWLEDGMENT

This research was carried out as part of the “Centercontract on Sound Quality” which establishes participation in and funding of the “Sound Quality Research Unit” (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP). Pauli Minnaar is thanked for his comments on the experimental details and on this manuscript. Finally, the participants in the listening test are thanked for their effort.

Notes

¹The inverse of the HRTF used for the conversion and the sound field in the loudness computations must be of the same type: either ‘free field’ or ‘diffuse field’.

²Note that the larger the amount of summation, the larger the effect of the ear with the lower SPL on binaural loudness. Thus, for the naïve sample of listeners, the (right) ear with the lower SPL input “pulls down” the binaural, directional loudness curve more than for the expert listeners.

³It is worth noting that a 3-dB binaural gain corresponds to a factor of approximately 1.2 in sones.

References

- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
- Hellman, R.P. and Zwislocki, J. (1963). “Monaural loudness function at 1000 cps and interaural summation,” *J. Acoust. Soc. Am.*, **35**, 856–865.
- ISO 11904-1 (2002). “Acoustics – Determination of sound immission from sound sources placed close to the ear – Part 1: Technique using a microphone in a real ear (MIRE technique),” (International Organization for Standardization, Geneva, Switzerland).
- ISO 11904-2 (2004). “Acoustics – Determination of sound immission from sound sources placed close to the ear – Part 2: Technique using a manikin (manikin technique),” (International Organization for Standardization, Geneva, Switzerland).

- ISO 389-1 (1998). “Acoustics – Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones,” (International Organization for Standardization, Geneva, Switzerland).
- ISO 532 (1975). “Acoustics – Method for calculating loudness level,” (International Organization for Standardization, Geneva, Switzerland).
- Marks, L.E. (1978). “Binaural summation of the loudness of pure tones,” *J. Acoust. Soc. Am.*, **64**, 107–113.
- Moore, B.C.J., Glasberg, B.R. and Baer, T. (1997). “A model for the prediction of thresholds, loudness, and partial loudness,” *J. Audio Eng. Soc.*, **45**, 224–239.
- Møller, H., Sørensen, M.F., Hammershøi, D. and Jensen, C.B. (1995). “Head-related transfer functions of human subjects,” *J. Audio Eng. Soc.*, **43**, 300–321.
- Minnaar, P., Olesen, S.K., Christensen, F. and Møller, H. (2001). “Localization with binaural recordings from artificial and human heads,” *J. Audio Eng. Soc.*, **49**, 323–336.
- Robinson, D.W. and Whittle, L.S. (1960). “The loudness of directional sound fields,” *Acustica*, **10**, 74–80.
- Shaw, E.A.G. (1974). “Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,” *J. Acoust. Soc. Am.*, **56**, 1848–1861.
- Sivonen, V.P., Ellermeier, W., Nielsen, F.K. and Bech, S. (2004). “A method of objectively determining subjective properties of a binaural sound signal,” DK2005/000287-PCT (international patent application), patent pending.
- Sivonen, V.P., Minnaar, P. and Ellermeier, W. (2005). “Effect of direction on loudness in individual binaural synthesis,” *Proceedings of the 118th Convention of the Audio Eng. Soc.*, Convention Paper No. 6512.
- Sivonen, V.P. and Ellermeier, W. (2006). “Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation,” *J. Acoust. Soc. Am.*, **119**, 2965–2980.
- Wightman, F.L. and Kistler, D.J. (1989). “Headphone simulation of free field listening. I: Stimulus synthesis,” *J. Acoust. Soc. Am.*, **85**, 858–867.

Directional loudness and the underlying binaural summation for wideband and reverberant sounds *

Ville Pekka Sivonen^{1,2}

¹*Sound Quality Research Unit, Dept. of Acoustics, Aalborg University, 9220 Aalborg, Denmark*

²*Brüel & Kjær Sound & Vibration Measurement A/S, 2850 Nærum, Denmark*

Abstract

The effect of incidence angle on loudness was investigated in Experiment 1 for anechoic, wideband and in Experiment 2 for reverberant, narrow-band sounds. Five listeners matched the loudness of a sound coming from five incidence angles in the horizontal plane to that of the same sound with frontal incidence. These directional loudness matches were obtained with an adaptive, two-alternative, two-interval, forced-choice procedure. The stimuli were presented to the listeners via binaural synthesis, utilizing individual head-related transfer functions in Experiment 1 and binaural room-impulse responses in Experiment 2. The results confirm that loudness depends on sound incidence angle in both experiments. The wideband and reverberant stimuli, however, exhibited distinct differences from the anechoic, narrow-band data earlier obtained from the same listeners. When modeling the binaural summation underlying the loudness matches based on at-ear exposure, a power summation of the at-ear signals yielded a good prediction for the obtained subjective data.

I. INTRODUCTION

Despite the vast amount of research on spatial hearing, little is known about how the location of a sound source affects its perceived loudness. The research on binaural and spatial hearing has largely focused on human sound localization (Blauert, 1997). In the classical research on loudness (Zwicker and Fastl, 1999), however, the emphasis has been on temporal and spectral aspects, while the spatial aspects have mostly been overlooked.

The spatial aspects have been explicitly addressed in a recent series of studies determining the effect of sound incidence angle on loudness (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a). In these investigations, anechoic, narrow-band stimuli were used for obtaining loudness matches between a frontal reference sound source and a comparison source located at various inci-

dence angles in the horizontal or the median plane. These *directional loudness matches* showed considerable dependence on the incidence angle, the center frequency of the stimuli, and the individual listener.

In Sivonen and Ellermeier (2006a), individual head-related transfer functions (HRTFs) were also measured for each incidence angle, in order to relate the changes in the actual at-ear exposure to the observed changes in loudness as a function of sound incidence. The individual at-ear exposures were then utilized in modeling the *binaural summation* underlying the directional loudness matches, i.e., it was determined how the signals at the two ears should be summed to result in a single binaural loudness percept. Anechoic, narrow-band stimuli were also used in an earlier study (Robinson and Whittle, 1960) for obtaining directional loudness matches and for modeling binaural summation, but, only reporting mean data as opposed to the individual analyses attempted in Sivonen and Ellermeier (2006a).

*Portions of the data presented at the 120th Meeting of the Audio Engineering Society, Paris, France, May 20-23, 2006.

A. Ecological validity of exposure

While the use of anechoic, narrow-band stimuli (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a; Robinson and Whittle, 1960) may be expected to maximize frequency-specific directional effects on loudness, they apparently lack ecological validity (e.g., Neuhoff, 2004). Note that most sounds in our environment are wideband and, in addition to the direct sound from a source to our ears, include reverberation. It is thus unclear whether the large variation (of up to 10 dB) in the directional loudness matches observed for the narrow-band stimuli (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a) is similarly quantifiable for stimuli resembling sounds our ears encounter in real life.

Based on measurements of HRTFs (see, Møller et al., 1995), the directional variation in anechoic at-ear exposures shows considerable frequency dependencies. At low frequencies, the human head and torso are relatively small compared to the wavelength. Thus, the directional differences in the HRTF magnitude spectra are relatively small. At high frequencies, the magnitude spectra are characterized by many closely spaced dips, the spectral location of the dips varying with the angle of incidence. Therefore, the directional differences relevant for loudness processing may to some extent be averaged out across a wider bandwidth.

In contrast to the direct sound from a source to our ears, reverberation in sound fields consists of sound waves which have been reflected a number of times from physical obstacles located in space. The incidence angles of the reflections are, by and large, different from that of the direct sound. In addition, the more time passing from the arrival of the direct sound, the more diffuse the reverberation becomes. In an ideal diffuse field, where the sound reaches the listener's ears with equal intensity from all directions, the orientation of the listener does not affect the at-ear exposure.

Therefore, stimulation which is closer to real-life sounds than the anechoic, narrow-band stimuli, should either be wider in bandwidth or include reverberation. This, however, may result in relatively small directional changes in at-ear exposure, and thus smaller directional effects on loudness may be expected than for the anechoic, narrow-band stimuli used earlier (Robinson and Whittle, 1960; Sivonen et al., 2005; Sivonen and Ellermeier, 2006a).

B. Earlier work

Anechoic, wideband sounds have been used for obtaining directional loudness matches by Remmers and Prante (1991) and Jørgensen (2002), where the matches varied by up to 3 dB as a function of incidence angle. In an investigation of multichannel audio reproduction by Bech (1998), a reverberant listening room was used, and a directional effect of approximately 1 dB for wideband stimuli was obtained, although investigated only over a limited number of angles.

In Bech (1998) and Jørgensen (2002), the loudness matches were reported only as relative level adjustments for the sound sources at various incidence angles. The results are thus not generalizable, since the actual at-ear exposure underlying the loudness matches was not considered.

In Remmers and Prante (1991), an attempt was made to predict the directional loudness matches from at-ear exposure using generic, artificial-head measurements. From the research on HRTFs, large individual differences in at-ear exposure for various incidences may be expected, and thus, relating mean loudness data to artificial-head predictions, as done in Remmers and Prante (1991), might wash out some of the detail in the results. Artificial-head measurements were also utilized in Tuomi and Zacharov (2000) for simulating the effect of incidence angle on loudness perception. However, subjective data, supporting the simulations, were not reported by the authors.

Relating at-ear exposure to directional loudness data is important, not only for the application of the level-alignment of multi-channel sound-reproduction systems, but also for predicting loudness from acoustical measurements made with an artificial head, which are more and more on the increase. Thus, in order to generalize the findings on binaural loudness summation for anechoic, narrow-band stimuli (Robinson and Whittle, 1960; Sivonen and Ellermeier, 2006a), a wider range of stimuli should be utilized in investigating the effect of incidence on loudness and the binaural summation underlying directional loudness matches.

C. Present investigation

The aim of the present study is to investigate the effect of incidence angle on loudness both for wideband and reverberant sounds. In addition to obtaining directional loudness matches, *individual*, listener-specific at-ear exposures are related to the

loudness matches, as was done in the earlier experiment (Sivonen and Ellermeier, 2006a) using anechoic, narrow-band stimuli.

The present study consists of two parts: Experiment 1, in which the effect of sound incidence angle on loudness is investigated for anechoic, *wide-band* sounds, and Experiment 2, where the stimuli are *reverberant* but narrow in bandwidth. In both experiments, the stimuli are calibrated so that if measured with a monaural microphone in the absence of a listener, the same sound pressure would be obtained for all incidence angles.

The present study makes use of individual binaural synthesis, as in Sivonen et al. (2005). In Experiment 1, directional sounds were synthesized from equidistant sound sources for various incidences using the listeners' individual HRTFs. In Experiment 2, the listeners' individual binaural room-impulse responses (BRIRs) were first measured for a fixed loudspeaker and listening position, while rotating the listeners in order to obtain various incidence angles. These BRIRs were then utilized in synthesizing the reverberant stimuli for the directional loudness experiment.

II. METHOD

A. Subjects

Five listeners (4 male, age between 26 and 48 years), all of whom had participated in previous experiments on directional loudness (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a), took part in the present study. The listeners' hearing thresholds were determined using pure-tone audiometry between 250 Hz and 8 kHz. None of the hearing thresholds revealed a hearing loss greater than 15 dB hearing level re ISO 389-1 (1998).

B. Apparatus

1. Listening tests

Listening-test stimuli were played back to the listeners over headphones (Beyerdynamic DT-990) using individual binaural synthesis. Since the binaural synthesis was static, the listeners were seated in a chair with a head rest to restrict head movements. For each listener, the chair was adjusted so that the center of the listener's head was aligned with the center position of the test setup.

The listeners' responses were collected using a two-button response box with light-emitting

diodes above each button indicating observation intervals. A model of this box (with larger lights) was placed in front of the listeners to avoid them tilting their heads towards the response box in their hands. The rest of the listening-test setup consisted of a computer (PC) with a high-quality sound card (RME DIGI96/8 PST), a custom-made programmable attenuator, and a power amplifier (Pioneer A-616) combined with a passive attenuator for unity gain, all placed in a control room (for details of the apparatus, see Sivonen et al., 2005).

In order to improve the plausibility of the binaural synthesis, loudspeakers at the intended incidence angles were visible to the listeners, even though the stimuli were actually played back over headphones. The distance from each loudspeaker to the center position of the setup was 2.0 m. Experiment 1 took place in an anechoic chamber, whereas a reverberant listening room, conforming to IEC 60268-13 (1998), was used in Experiment 2.

2. Measurements

In order to prepare stimuli for the individual binaural synthesis, and to relate at-ear exposures of the stimuli to the listening test results, a number of measurements were carried out. Individual (anechoic) HRTFs and headphone transfer functions (PTFs), measured in connection with an earlier investigation (Sivonen and Ellermeier, 2006a), were used to synthesize stimuli for Experiment 1.

For Experiment 2, individual BRIRs and PTFs were measured in a listening room, the dimensions of which were $2.8 \times 4.2 \times 7.8$ m (H \times W \times L). A loudspeaker (Vifa M10MD-39, mounted in a hard plastic ball) at a distance of 2.0 m from the measurement position was used as a sound source. The loudspeaker and the measurement position were at a height of 1.3 m and at a distance of 1.6 m from the closest wall, positioned non-symmetrically with the dimensions of the room. The BRIRs were measured with two miniature microphones (Sennheiser KE-4-211-2, one specifically for each ear), which were inserted in earplugs (E·A·R Classic, halved in length) and mounted flush with the ear-canal entrance. A pressure-field microphone (Brüel & Kjær Type 4133) was used as a reference microphone for measuring the loudspeaker's anechoic response and its response in the room, and determining the responses of the two miniature microphones.

From the measurements with the reference microphone, some room-acoustics parameters were

Table I: Room-acoustics parameters (ISO 3382, 1997) for the standard listening room for the specific source and receiver positions, measured on third-octave bands. The early decay time (EDT) is obtained from the initial 10 dB and reverberation time (RT) from the portion between -5 and -35 dB of the decay curve, both calculated as the time (in seconds) required for a decay of 60 dB. C_{50} and C_3 are the balance between early/late and direct/reverberant energy (as a ratio in dB), before and after 50 and 3 ms, respectively.

f_c (Hz)	EDT(s)	RT(s)	C_{50} (dB)	C_3 (dB)
1k	0.37	0.44	9.5	-10.5
5k	0.41	0.62	10.4	0.9

derived according to ISO 3382 (1997), see Table I. Values for third-octave bands at the two center frequencies used in the listening test are shown. Both the early decay time (EDT) and reverberation time (RT) show that the room is more reverberant at 5 kHz than at 1 kHz. The ratio between early and late energy (C_{50}) in decibels indicates that most of the energy arrives at the listening position within the first 50 ms, C_{50} being fairly similar at the two center frequencies. However, when looking at the ratio between direct and reverberant energy (in the present study, before and after 3 ms from the direct sound), at 1 kHz most of the energy is in the reverberation, whereas the ratio is close to unity at 5 kHz. This can be explained by the reflections in the room being stronger at 1 kHz and the loudspeaker¹ being more directive at 5 kHz.

For the BRIR measurements, the listeners were seated in the same chair, which was used in the listening test. The listener's position was aligned with a fixed laser, pointing to the center of the setup from 90° incidence. The orientation of each listener was calibrated so that for frontal incidence, the chair was rotated until time-aligned impulse responses were obtained at the two ears. Then, a polar-coordinate graph was attached on the listener's head, denoting 0° incidence on the monitor of a video camera above the center position of the setup. Reference points for given listener orientations were marked on the video monitor, and the chair was rotated to obtain the desired incidence angle.

In addition to the PC, the sound card, and the power amplifier of the listening test setup, the equipment for the BRIR measurements included

an audio interface for AD/DA-conversion (RME ADI DS-8) and a custom-made pre-amplifier for the miniature microphones. The measurements were made using a maximum-length sequence with an order of 15 and a sampling rate of 48 kHz, resulting the length of an impulse response of 0.68 s. All microphone signals were bandpass filtered between 22.5 Hz and 22.5 kHz and appropriately gained by measurement amplifiers (Brüel & Kjær Type 2607, Type 2236 or Type 2690 Nexus).

C. Stimuli

Before processing stimuli for the binaural synthesis, all sound signals were 1 s long, including 20-ms raised-cosine rise and fall ramps in the beginning and at the end of each signal. Stimuli were generated for six different incidence angles: 0°, 30°, 60°, 90°, 135° and 180° of azimuth, in the left hemisphere of the horizontal plane with 0° elevation.

1. Experiment 1: Wideband noise

A pink-noise signal, bandpass filtered between 100 Hz and 10 kHz was used as a wideband stimulus. It was played back at an overall level corresponding to 55 dB sound pressure level (SPL) measured in the free field. At this overall level, the loudness of the pink noise was roughly the same as that of the 65-dB-SPL narrow-band noises used in the earlier studies (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a), based on predictions made from the loudness model by Moore et al. (1997).

For the binaural synthesis, the pink noise was convolved with individual HRTFs for the incidence angles. Furthermore, individual equalization filters for the headphones were applied (for details concerning the equalization, see Sivonen et al., 2005).

2. Experiment 2: Reverberant stimuli

Third-octave noise bands, centered at 1 and 5 kHz, were used as reverberant stimuli. These narrow-band stimuli were synthesized at an overall level corresponding to 65 dB SPL measured in the free field, i.e., without the effect of reverberation. The listening room and each listener's physique had idiosyncratic effects on the at-ear exposure levels. This is captured by the individual BRIRs. The narrow-band stimuli were convolved with the individual BRIRs, and equalization filters for the

headphones were applied, based on individual PTFs measured in connection with the BRIR measurements.

Binaural room impulse responses A single source at a fixed position in the room was used. Thus, the effect of the room at the measurement position was constant. BRIRs for all listeners and for the six incidence angles were then obtained by rotating the listener in the chair, carrying out measurements at each incidence. In this manner, changes in the direct sound (comparable to HRTFs) were obtained, while the effect of the reverberant energy was constant as a function of incidence angle². Thus, in the experiment proper, the virtual acoustics of the room were “rotated” around the listener. This went largely unnoticed, because static sound sources were synthesized and the listeners were not blindfolded.

The anechoic, on-axis response of the loudspeaker, measured at the same distance as the BRIRs in the absence of a listener, was equalized for in the BRIR computations. Thus, the direct sound of the BRIRs could be compared to the individual HRTFs, obtained in Sivonen and Ellermeier (2006a).

Examples of the measured BRIRs are plotted in Fig. 1 for subject IA. Panel ‘a)’ shows the direct sound of the BRIRs to be fairly similar to the head-related impulse responses (HRIRs) for the same incidence angle and the same listener in the time domain.

In panel ‘b)’ of Fig. 1, the time decay for the contralateral (right) ear at 90° is plotted for the 1-kHz center frequency, normalized to the maximum level of the BRIR. For this incidence angle, the direct sound is attenuated the most, due to the obstruction by the listener. Despite the low acoustic input, a signal-to-noise ratio (SNR) of approximately 50 dB is obtained, as seen in the level of flat (noise) part in the measured BRIR after 400 ms. Even though the SNR is reasonable, the noise may still become audible in the binaural synthesis, when the BRIR is convolved with a stimulus, since the natural decay of the sound is limited by the noise floor of the measurement. Thus, the reverberation tail was suppressed from 400 ms onwards, using a determined average BRIR decay rate between 100 and 300 ms, separately at both center frequencies. The BRIRs with the suppressed tails continue to decrease in level below -90 dB from the direct sound. The decay rate

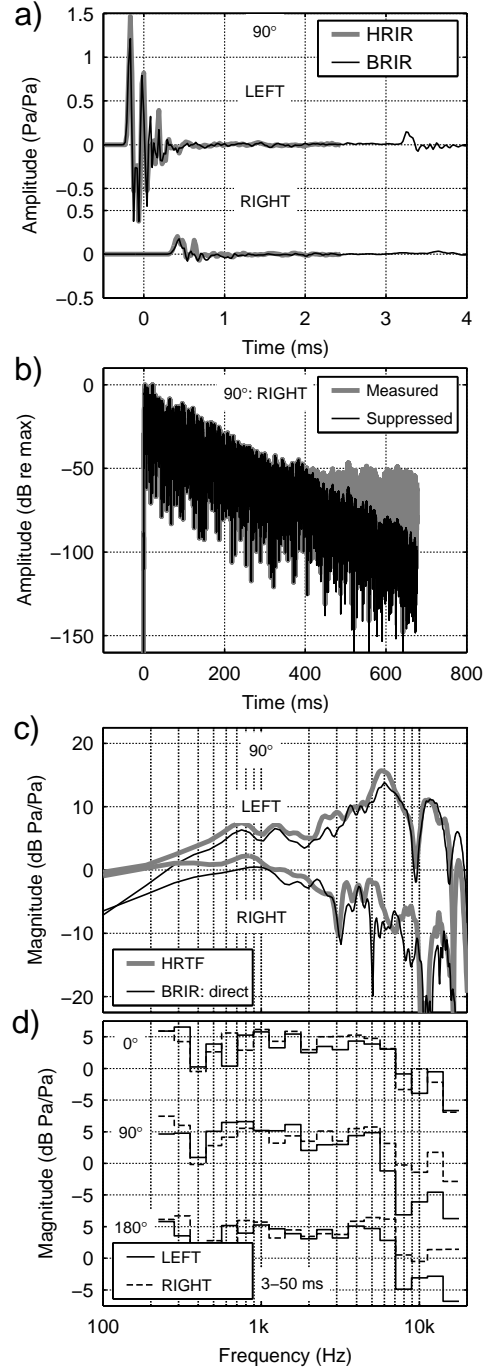


Figure 1: Subject IA, measured BRIRs: a) 4 ms of the BRIRs plotted along with HRIRs, b) measured and suppressed time decays of the BRIRs at 1 kHz for the contralateral ear (90° of azimuth), c) the direct sound (< 3 ms) from the BRIRs and the HRTFs in the frequency domain, and d) third-octave-band levels for the early part (from 3 to 50 ms) of the BRIR decay for three incidences (0°, 90° and 180° of azimuth).

(see the time at -60 dB from the maximum level in Fig. 1b) is in good agreement with the estimated reverberation time at 1 kHz in Table I.

In panel 'c)' of Fig. 1, the frequency responses of the direct part of the BRIRs are contrasted with the HRTFs in the frequency domain, similarly as for the time domain in panel 'a)'. The agreement between the BRIRs and the HRTFs is good, even though some deviations at low frequencies can be seen. Note that in HRTF computations the DC value is set to unity, whereas for BRIRs, uncertainties, caused by a poor SNR at low frequencies, remain. As with HRTFs, large interaural differences are obtained in the direct part of the BRIRs.

In panel 'd)' of Fig. 1, the early reflections (from 3 to 50 ms after the direct sound) are plotted on third-octave bands for three incidence angles, where the ears of the listener were in different positions in the room. This panel shows that the early reflections in the room give an increase of roughly 5 dB to the at-ear signals, and more importantly, confirms the reflections to be fairly *independent* of the sound incidence angle. This is also the case for sounds arriving later (after 50 ms), since the sound field becomes more and more diffuse with time. Thus, the aim of having fairly constant reverberation, while still obtaining changes in the direct sound comparable to the HRTFs, seems to have been achieved in the present stimulation.

D. Procedure

In an attempt to compare the present results with earlier studies using the same listeners, the listening-test procedure was largely as reported in Sivonen et al. (2005) and Sivonen and Ellermeier (2006a).

The listeners' task was to compare the loudness of a sound with a given frequency content (Experiment 1: pink noise, Experiment 2: 1 or 5 kHz narrow-band noise) synthesized with one of the comparison incidence angles (30° , 60° , 90° , 135° or 180° of azimuth) to the same sound with frontal incidence, i.e., 0° of azimuth. The loudness matches with the frontal reference were obtained in an adaptive, two-interval, two-alternative, forced-choice procedure.

In each trial, the listeners were played back a pair of sounds, the frontal reference and a comparison in a randomized order, and they had to judge which of the two sounds was louder. If the comparison was judged louder its level was decreased by a given amount, whereas if the reference was

judged louder the level of the comparison was increased by the same amount. The initial step size of such an adaptive track was 4 dB before the second reversal in the listener's judgments and 1 dB for the rest of the track. A schematic of the setup for obtaining the loudness matches is plotted in Fig. 2.

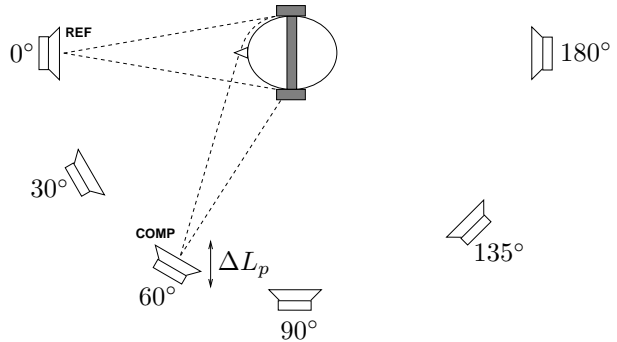


Figure 2: A schematic of the experimental setup for obtaining directional loudness matches. 0° is the frontal loudness reference (REF), while the other incidence angles serve as comparison (COMP) directions, the sound pressure levels (ΔL_p) of which are adjusted in the listening-test procedure.

In one adaptive track, eight reversals were obtained, of which the last six were used in estimating the loudness match. Eight replications of each factorial combination of stimulus frequency and incidence angle were collected for each listener. Half of the replications started $+10$ dB and the other half -10 dB from the level of the frontal reference. In Experiment 2, the listeners had to match also the frontal reference to itself, as an indication of the baseline variability of the loudness matches.

The order of the adaptive tracks (stimulus frequency, incidence angle and starting level) was randomized within one set of replications. The tracks were divided in blocks of eight and interleaved within a block, one 1-hour listening session consisting of maximum of four blocks, each lasting approximately 10 mins. All listeners participated in Experiment 1 first and then in Experiment 2. The total duration was 6 h, including the BRIR- and PTF-measurement sessions.

III. RESULTS

In both Experiments 1 and 2, loudness matches with the frontal reference were obtained for five incidence angles in the horizontal plane. Thus,

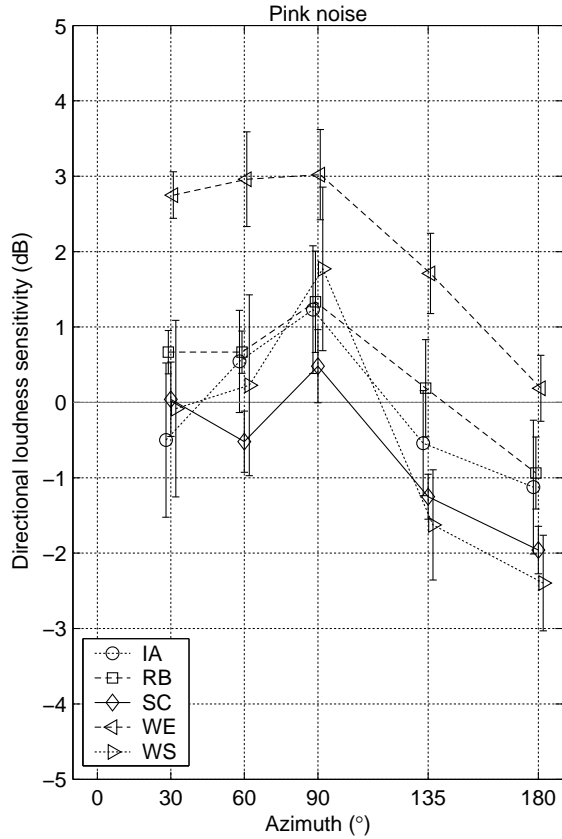


Figure 3: Results of Experiment 1: Individual directional loudness sensitivities (in dB) $\pm 95\%$ confidence intervals in the left hemisphere of the horizontal plane for anechoic pink-noise stimuli, obtained from five listeners.

the raw data are the relative sound pressure levels each synthesized sound source would have to be set to, in order to be perceived equally loud as the reference. The inverses of these gain adjustments were termed *directional loudness sensitivities*, in agreement with earlier studies, a positive value denoting a direction which is perceived louder than the frontal reference.

A. Experiment 1: Wideband noise

Individual directional loudness sensitivities for the wideband, pink-noise stimulus are plotted in Fig. 3 for all five listeners. The data show that loudness is not constant as a function of incidence angle: for 16 out of 25 data points in Fig. 3, do the 95-% confidence intervals not cover the 0-dB line denoting the level of the frontal reference for the directional loudness matches.

The directional loudness sensitivities also show individual differences. For subject WE, all data points are above the 0-dB line, the maximum value being around +3 dB for the 90° incidence, i.e., a sound source at that angle is perceived louder and has to be attenuated by the same amount to produce a loudness match with the frontal reference. For subject SC, however, most of the data points are either close to or below the 0-dB line, the latter implying that the level of the sound source has to be increased for a match with the reference source in front.

In order to inspect the significance of the observed directional loudness sensitivities, a one-factor (namely, incidence angle) repeated-measures analysis of variance (ANOVA), with the listener as a random factor, was performed. The ANOVA returned a significant effect of incidence angle [$F(4, 16) = 26.63; p < 0.001$], confirming, in addition to what was observed in Fig. 3, that loudness is dependent on the angle of incidence for the wideband stimulus. Also, the interaction between incidence angle and listener was significant [$F(16, 175) = 2.59; p = 0.001$], revealing statistically significant individual differences in the listeners' directional loudness matches.

B. Experiment 2: Reverberant stimuli

Individual directional loudness sensitivities for the reverberant narrow-band stimuli at the two center frequencies are plotted in Fig. 4. The directional effect at 1 kHz is smaller than what was observed in Experiment 1: Most of the data vary within ± 1 dB, from the 0-dB line, and only for 9 out of 30 data points do the 95-% confidence intervals not cover the 0-dB line. Due to the smaller spread of the individual data, the comparison between listeners is more difficult than in Experiment 1 (see Fig. 3), although, again the data for subject WE display the highest directional loudness sensitivities for the comparison incidence angles. All listeners were able to match the frontal source to itself, as seen in the 0-dB (i.e., identity) outcome for the left-most data point in Fig. 4.

At 5 kHz, however, the data vary over a larger range of approximately 6 dB, and in all but one case the 95-% confidence intervals do not cover the 0-dB line. The individual differences are also larger in magnitude, even though for all listeners except subject WE, the shapes of the curves as a

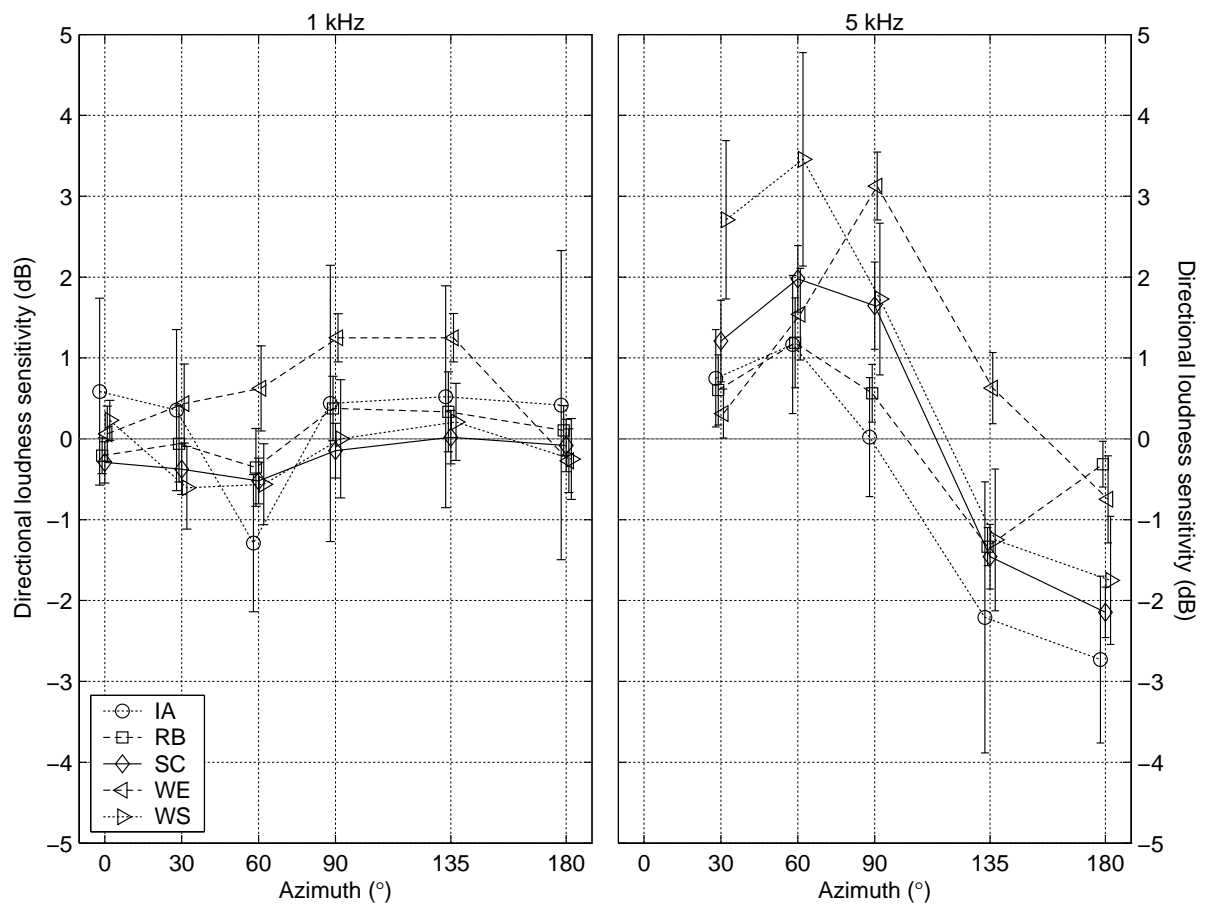


Figure 4: Results of Experiment 2: Individual directional loudness sensitivities (in dB) $\pm 95\%$ confidence intervals in the left hemisphere of the horizontal plane for reverberant narrow-band stimuli, obtained from five listeners.

function of incidence angle are quite similar.

Performing a two-factor (incidence angle \times center frequency) ANOVA on the data of Experiment 2 returned a significant two-way interaction between incidence angle and center frequency [$F(4, 16) = 15.78; p < 0.001$]. This implies that the shapes of the directional loudness-sensitivity curves are different at the two center frequencies, see Fig. 4. Furthermore, the interaction between incidence angle, center frequency and listener [$F(16, 350) = 4.95; p < 0.001$], revealed statistically significant individual differences for the matches of the reverberant stimuli, similarly as for the wideband stimuli of Experiment 1.

IV. DISCUSSION

Sound incidence angle had an effect on loudness both for the anechoic, wideband noise of Experiment 1 and the reverberant, narrow-band stimuli of Experiment 2. The directional loudness sensitivities derived from the loudness matches also exhibited considerable dependency on the individual listener.

The sound sources in the present study were synthesized with a flat frequency response, and calibrated so that the same sound pressure would be measured in the center position of the listening-test setup for all incidence angles (p_1 , in the absence of a listener, see Møller et al., 1995). The listeners then adjusted the relative levels of the comparison incidence angles, matched for equal loudness with the frontal reference. Due to the calibration of the p_1 pressure to be constant over incidence angles, a monophonic measurement would fail to predict the observed directional loudness matches, and reasons for the matches must be inspected from the binaural at-ear exposure.

A straightforward approach is thus to analyze the directional changes in the at-ear exposures and relate them to the directional loudness sensitivities, as had been done in the earlier directional-loudness investigation (Sivonen and Ellermeier, 2006a). In the present study, the HRTFs and BRIRs, as measured at the blocked entrance to ear canal (p_2 , see Møller et al., 1995), were used in the stimulation. These measurements of individual HRTFs and BRIRs were utilized in determining the exposures at each listeners ears for Experiment 1 with anechoic wideband stimuli and for Experiment 2 with reverberant narrow-band stimuli, respectively.

A. Experiment 1

1. Considerations for stimuli

The spectral bandwidth of a stimulus has been shown to affect perceived loudness, a phenomenon which is known in the literature as the spectral summation of loudness (Zwicker and Fastl, 1999). Loudness models (e.g. Moore et al., 1997) have been developed to account for this summation of loudness across critical bands. Since wideband stimuli were used in Experiment 1, the spectral summation of loudness could not be omitted in the analysis. Therefore, the at-ear exposures must be utilized as loudness values in sones, rather than acoustical measures of sound pressure level, as had been done in Sivonen and Ellermeier (2006a) for the narrow-band stimuli.

Furthermore, for each subjective data point in Figs. 3 and 4, two at-ear exposure values, one for each ear, are determined. Thus the binaural loudness summation of the left- and right-ear exposure to a single binaural percept must also be considered when predicting the subjective data.

2. Modeling

The loudness model by Moore et al. (1997) was utilized in the present analysis, with two different predictions of binaural loudness summation between the two ears:

(1) The first prediction is a power-summation³ model, favored in the analysis of Sivonen and Ellermeier (2006a) in a directional sound field. In this prediction, a binaural power sum of the linear left- and right-ear magnitude HRTF spectra with a given frequency resolution is first determined. Then the power sum is converted to the corresponding diotic (i.e., with the same signal at both ears) stimulation by halving the summed power. Finally, this (diotic) signal is utilized for loudness computations. For details of the power-summation model, see Sivonen and Ellermeier (2006b).

(2) The second prediction is a loudness-summation model, where loudness is computed separately for the left- and right-ear signals and binaural loudness determined as an arithmetic mean of the left- and right-ear sone values. This manner of summation is supported by some headphone studies on binaural loudness (e.g., Marks, 1978), and is implemented in Moore et al. (1997) and Tuomi and Zacharov (2000).

In order to predict loudness from at-ear exposures, the loudness model by Moore et al. (1997)

allows substitution of the HRTF-filtering stage of the model by a measured at-ear exposure, utilizing ear-drum pressures (p_4) for loudness computations. Since the individual HRTFs (and the BRIRs) of the present study were measured at the blocked entrance to the ear canal (p_2), a direction-independent transfer from the blocked entrance to the ear-drum (mean p_4/p_2) was adopted from Fig. 13 in Hammershøi and Møller (1996). For the two predictions of binaural loudness summation (power and loudness sum), individual loudness matches were then obtained by varying the p_1 pressure so that the difference in sones between the frontal reference and each comparison incidence was minimized, based on the individual HRTF effects on the left- and right-ear p_4 pressures.

3. Outcome

In Fig. 5, the two predictions are plotted along with the obtained directional loudness sensitivities (of Fig. 3) in individual panels for each listener and a mean across listeners. As was the case for the subjective data, the predicted loudness matches were converted to directional loudness sensitivities by inverting them. Even though the predictions deviate from one another by a maximum of 2 dB, the order is the same in all panels of Fig. 5: The power sum is always above the loudness sum prediction, due to the fact that in a power sum the ear with the higher SPL (see the plotted HRTF magnitude spectra in Fig. 1b) is emphasized more, while in a loudness sum the ear with the lower SPL has a relatively large effect on binaural loudness. A more detailed discussion concerning different binaural summation models can be found in Sivonen and Ellermeier (2006a).

The power sum yields a considerably better prediction of the subjective data. For subjects IA and RB, and partly for subject WS, the agreement between the actual data and the power-sum prediction is fair. For subjects SC and WE, neither prediction aligns with the subjective data, although for subject SC, the loudness sum is somewhat closer. However, when inspecting the mean data obtained by averaging third-octave-band HRTFs across listeners for the two predictions and the mean directional loudness sensitivities, the power-sum prediction is clearly favored for the anechoic wideband stimuli (see the lower-right panel in Fig. 5). Note that this was also the case for anechoic narrow-band stimuli in Sivonen and Ellermeier (2006a).

B. Experiment 2

1. Considerations for stimuli

For Experiment 2, reverberant, narrow-band stimuli were used. The mean directional loudness sensitivities, averaged across the five listeners, are plotted for comparison in Fig. 6, for anechoic (Sivonen et al., 2005) and reverberant (present study) stimuli. In both studies, the stimuli were presented via individual binaural synthesis, the only difference being that for the present data, the sounds included fairly constant reverberation ($RT \approx 0.5$ s) as a function of incidence angle.

The directional effect is smaller for the reverberant stimuli, the mean loudness data being close to 0-dB at 1 kHz, while at 5 kHz, the effect size is roughly halved compared to the anechoic case, see Fig. 6. This discrepancy between center frequencies can be explained by the measured room-acoustics parameters in Table I: The balance between direct and reverberant energy (C_3) reveals the sound field at 1 kHz to be dominated by reverberation, while at 5 kHz the direct and reverberant energy are fairly equal. Thus, the changes in the direct sound, as a function of the listener's orientation relative to the sound source, were more likely to affect loudness at the higher center frequency. Furthermore, as reported in Sivonen and Ellermeier (2006a), the directional effect on loudness is larger at higher center frequencies, which can be explained by the respective changes in HRTFs. Finally, the error bars in Fig. 6 denote smaller variation between listeners for the present data, due to the room reverberation being the same for all listeners, whereas in Sivonen et al. (2005) with (pure, anechoic) HRTFs the individual differences were more pronounced.

2. Modeling

In order to inspect the effect of the BRIRs on the obtained directional loudness sensitivities on an *individual* basis, third-octave-band, left- and right-ear sound pressure levels were computed from the stimuli convolved with the BRIRs for each listener, incidence angle and center frequency. The at-ear SPLs were then normalized to those of the frontal reference, separately for the left and the right ears.

Then, a power-sum prediction was computed from the left- and right-ear SPLs, also normalized to the frontal reference. Note that this prediction is based on a summation of at-ear SPLs, which was fairly successful for the an anechoic, narrow-band

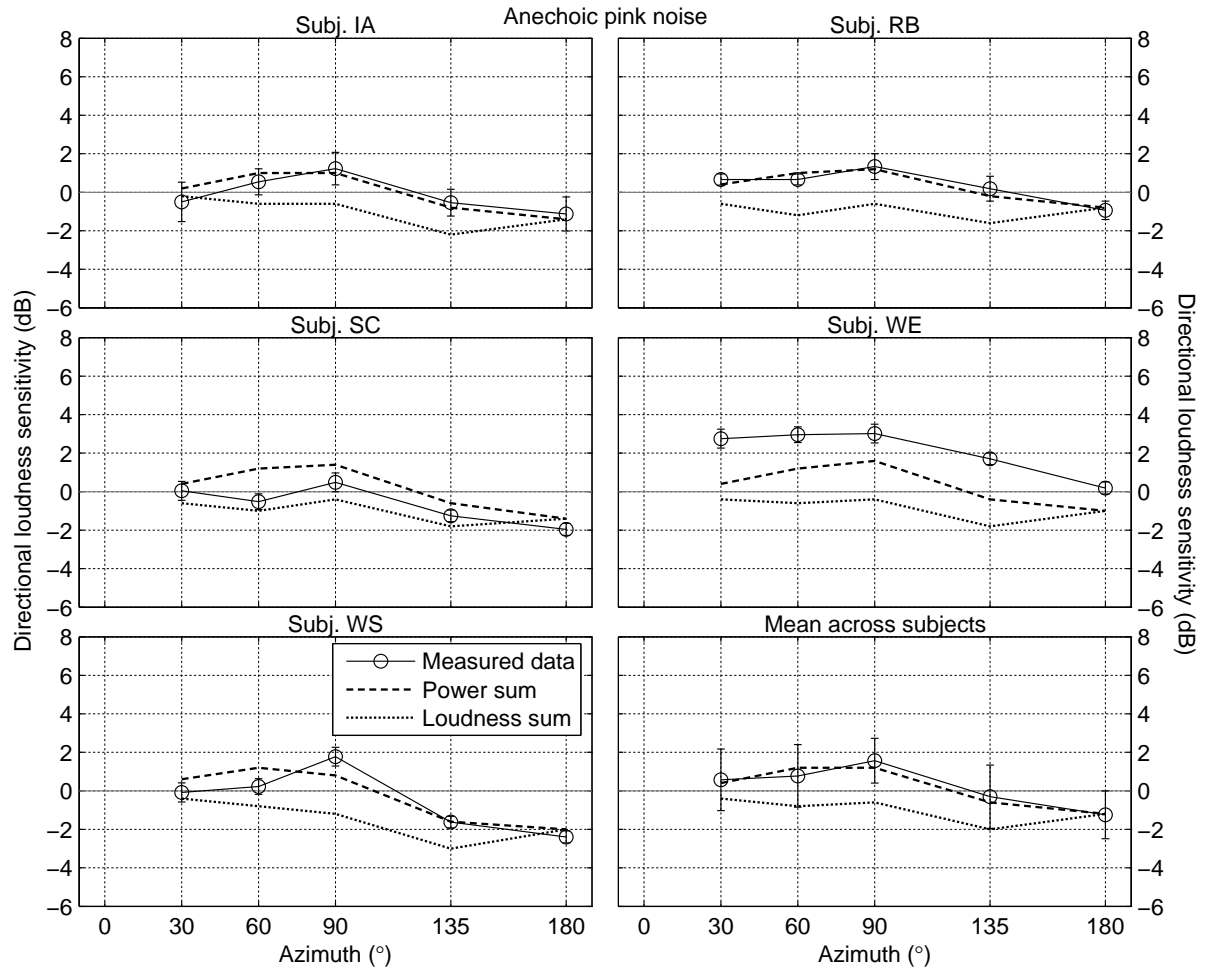


Figure 5: Modeling Experiment 1: Directional loudness sensitivities (in dB) $\pm 95\%$ confidence intervals for the anechoic wideband (pink-noise) stimuli, plotted along with predictions based on a binaural power sum and loudness sum for the five individual listeners and a mean across listeners.

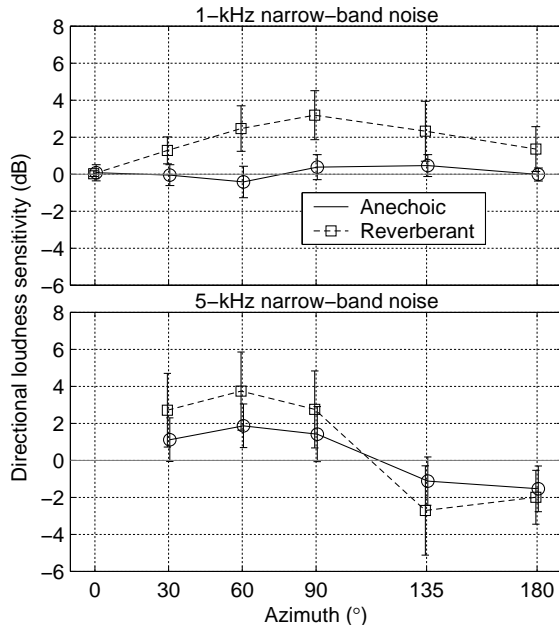


Figure 6: Effect of reverberation on directional loudness: Mean directional loudness sensitivities $\pm 95\%$ confidence intervals of the means, obtained in individual binaural synthesis from five listeners for anechoic (Sivonen et al., 2005) and reverberant (Experiment 2) narrow-band stimuli.

stimuli of Sivonen and Ellermeier (2006a). Unlike in the analysis of the results for the wideband stimuli, a loudness model is not utilized here.

3. Outcome

The individual directional loudness sensitivities, the relative directional changes in the at-ear SPLs and the power-sum prediction are plotted in Figs. 7 and 8 for the 1- and 5-kHz center frequencies, respectively. Even though interaural level differences (ILDs) are present at both center frequencies, the effect of the room on the at-ear exposure is strong: In an anechoic sound field, for the same listeners and incidence angles, ILDs of over 10, and up to 30 dB were obtained in Sivonen and Ellermeier (2006a), while in the present study the corresponding ranges are 4 and 10 dB at 1 and 5 kHz, respectively. Despite the smaller variation in ILDs, the subjective directional loudness sensitivities (with a few exceptions) lie within the bounds defined by at-ear SPLs, exhibiting a good agreement, similar to that reported for the anechoic stimuli of Sivonen and Ellermeier (2006a).

When predicting the directional loudness data

from the at-ear SPLs, differences between individuals can be observed, i.e., for some listeners (e.g., subjects IA, SC and WS) the power sum yields a fair prediction, while for others (e.g., subject WE), the prediction is at odds with the obtained subjective data, see Figs. 7 and 8. These peculiarities, which remain even after accounting for individual differences in at-ear exposure, imply that there are differences in the underlying binaural loudness summation of the left- and right-ear signals. These individual differences have been speculated and elaborated in more detail in Sivonen and Ellermeier (2006a). The mean data in both Figs. 7 and 8, however, is in good agreement with the power-sum prediction.

C. Summary

Despite the (small) individual differences between the directional loudness sensitivities and predictions based on at-ear exposures, the mean data of both for the wideband and the reverberant stimuli agree well with a binaural power summation, see Figs. 5, 7 and 8. Based on the present results, it seems that for spatial sounds reaching the listener's ears with various incidences, adjustments to the modeling of the underlying binaural summation of loudness are called for.

A perfect binaural summation of loudness in sones does not seem to be warranted for spatial, wideband sounds. Inspecting the mean data of Fig. 5 shows that the prediction based on a loudness summation is below the measured data, and also the power-summation prediction. In Sivonen and Ellermeier (2006a) this was reported to be due to the loudness summation overestimating the effect of the (ipsilateral) ear having the lower exposure (in dB SPL) when combining the exposures at the two ears for binaural loudness. Here, the same reasoning can be used, although plotting the at-ear exposure in dB SPL along with the subjective data is not meaningful, due to the wideband nature of the stimuli. A binaural power sum, however, gives more emphasis on the ear having the higher exposure, and thus arrives at predictions above those of the loudness summation.

To sum up, computing a power sum of the HRTF magnitude spectra, converting that sum to the corresponding diotic spectrum, and only then utilizing this at-ear spectrum for loudness computations (according to the binaural power-summation model, see Sivonen and Ellermeier,

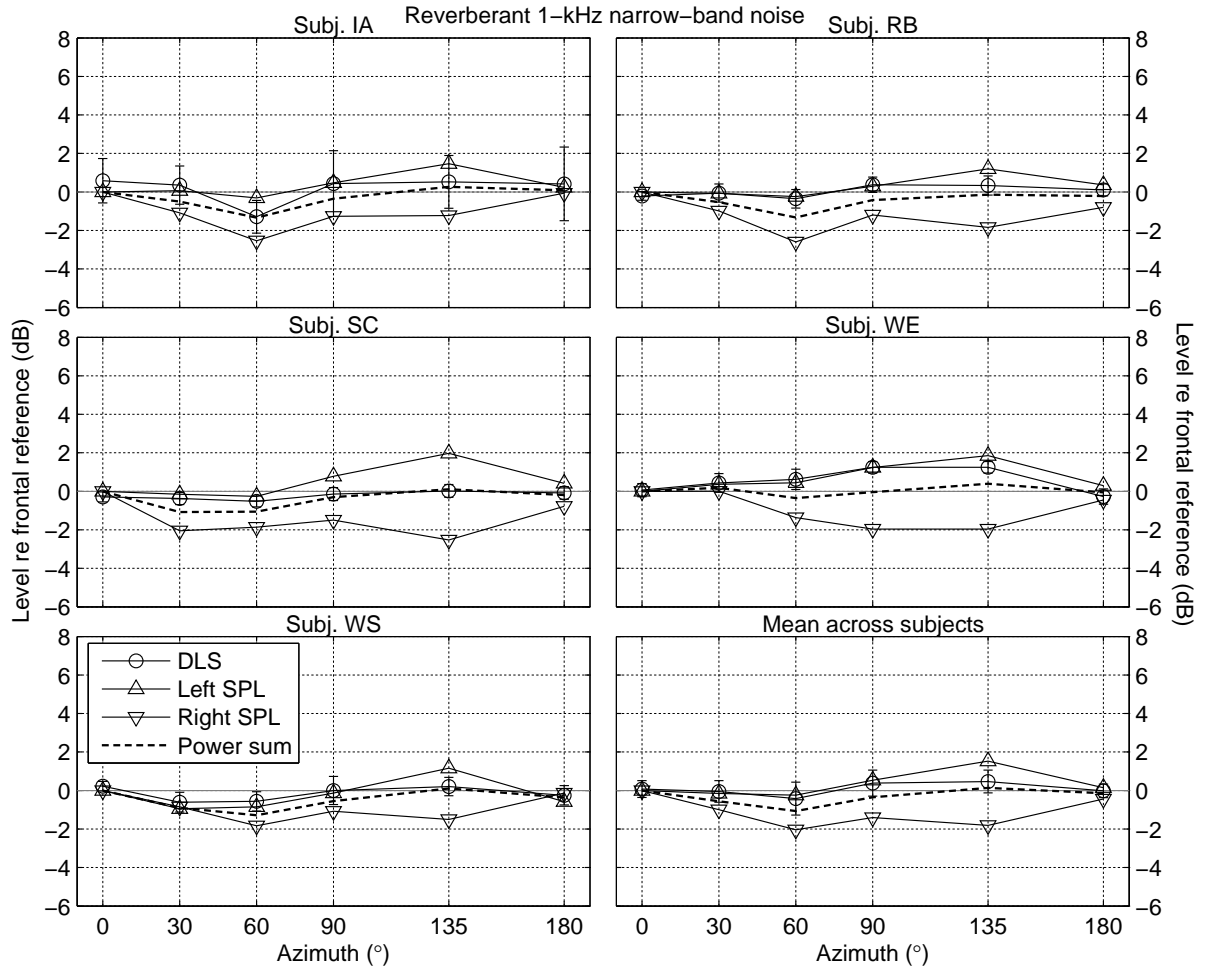


Figure 7: Modeling Experiment 2, 1 kHz: Individual directional loudness sensitivities (DLS) $\pm 95\%$ confidence intervals for five listeners, and a mean across listeners, plotted along with the effective changes in the left- and right-ear SPLs, and a prediction based on a binaural power sum.

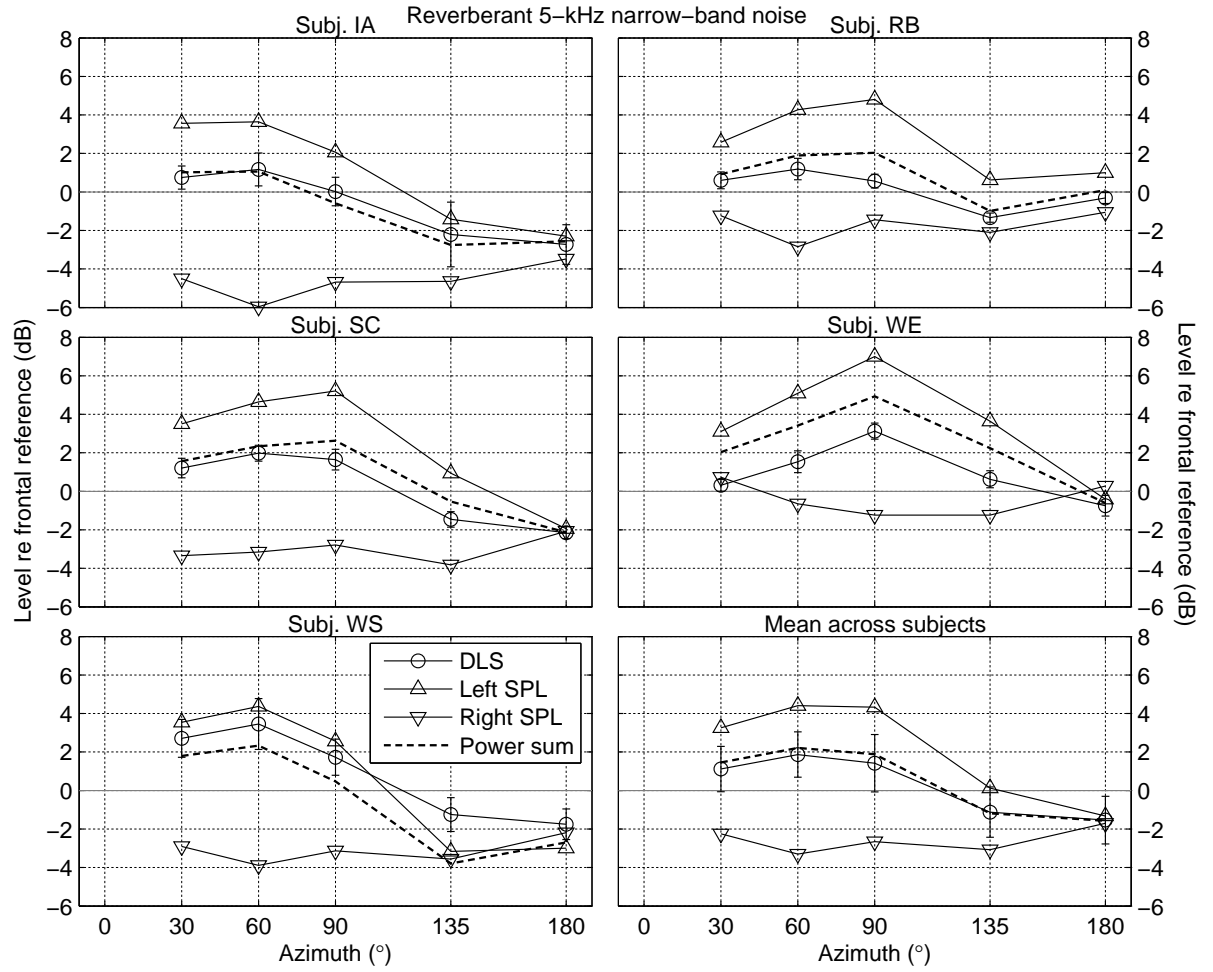


Figure 8: Modeling Experiment 2, 5 kHz: Individual directional loudness sensitivities (DLS) $\pm 95\%$ confidence intervals for five listeners, plotted along with the effective changes in the left- and right-ear SPLs, and a prediction based on a binaural power sum.

2006b) gives a good prediction of the directional loudness matches for the wideband stimuli.

When the sound field is reverberant, the exposure differences between directions, and between the two ears, are smaller than those of the anechoic case, due to the diffuseness of the reverberation. Despite the reverberation in the stimuli, a quantifiable, statistically significant effect of direction on loudness was observed in the present study. Therefore, these results do not support the phenomenon of loudness constancy. In Zahorik and Wightman (2001) (the source) loudness was reported to be constant for various sound-source distances in a reverberant room, irrespective of profound changes in at-ear exposure. This observation of constancy could be due to instructing listeners to focus on the source loudness, whereas in the present series of experiments (Sivonen et al., 2005; Sivonen and Ellermeier, 2006a,b, present study), no such instructions, possibly biasing listeners, were given.

The small exposure differences between the two ears for the reverberant stimuli are not optimal for the purpose of modeling binaural loudness summation. The closer to diotic the stimulation is, the less the summation method plays a role for binaural loudness matches, since the same binaural advantage is present both for the reference and the comparison (a more detailed discussion is found in Sivonen and Ellermeier, 2006a). The present results for reverberant sounds, however, agree well with the binaural power summation, a model which is based on anechoic stimuli with larger exposure differences between the two ears (Sivonen and Ellermeier, 2006a).

All in all, the present results, obtained using stimuli which resemble real-life sounds more than the anechoic, narrow-band stimuli used in Sivonen and Ellermeier (2006a), corroborate the validity of binaural power summation for spatial, directional sounds.

V. CONCLUSION

The effect of sound incidence angle on loudness was investigated for wideband and reverberant sounds. The results for both types of sounds showed that loudness is not constant over incidence angles, and the directional loudness matches exhibit considerable dependence on the listener. These idiosyncrasies could, however, largely be accounted for by determined individual at-ear exposures. A binaural power-summation model yielded

a fair prediction of the obtained directional loudness data, and was clearly favored over a perfect summation of loudness in sones between the two ears, as commonly used. The present results thus suggest adjustments to the modeling of binaural loudness summation for spatial sounds.

VI. ACKNOWLEDGMENT

This research was carried out as part of the “Centercontract on Sound Quality” which establishes participation in and funding of the “Sound Quality Research Unit” (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP). Wolfgang Ellermeier and Pauli Minnaar are thanked for their comments on the experimental design and the acoustical measurements, respectively, and their comments on the manuscript. Finally, the listeners are thanked for their participation in the experiment.

Notes

¹An omni-directional sound source should be used for room-acoustics measurements. Even though this is not the case for the loudspeakers, its on-axis frequency response has roughly the same magnitude at 1 and 5 kHz.

²Note that when rotating the listener, the measurement positions at the ears move along the perimeter of the head. Assuming that the sound field is uniform in the at-ear plane and the reverberant sound is diffuse, the rotation will not affect the reverberation at the listener’s ears.

³Note that a power summation of the sound pressures corresponds to a 3-dB summation rule of sound pressure levels, as proposed in Sivonen and Ellermeier (2006a).

References

- Bech, S. (1998). “Calibration of relative level differences of a domestic multichannel sound reproduction systems,” *J. Audio Eng. Soc.*, **46**, 304–313.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

- Hammershøi, D. and Møller, H. (1996). "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.*, **100**, 408–427.
- IEC 60268-13 (1998). "Sound system equipment – Part 13: Listening tests on loudspeakers," (International Electrotechnical Commission, Geneva, Switzerland).
- ISO 3382 (1997). "Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters," (International Organization for Standardization, Geneva, Switzerland).
- ISO 389-1 (1998). "Acoustics – Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," (International Organization for Standardization, Geneva, Switzerland).
- Jørgensen, S.L. (2002). *Matlab model for loudness perception - directional loudness*. Unpublished masters thesis, Technical University of Denmark (DTU), Copenhagen.
- Marks, L.E. (1978). "Binaural summation of the loudness of pure tones," *J. Acoust. Soc. Am.*, **64**, 107–113.
- Moore, B.C.J., Glasberg, B.R. and Baer, T. (1997). "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, **45**, 224–239.
- Møller, H., Sørensen, M.F., Hammershøi, D. and Jensen, C.B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, **43**, 300–321.
- Neuhoff, J.G. (2004). *Ecological Psychoacoustics* (Academic Press, London).
- Remmers, H. and Prante, H. (1991). "Untersuchung zur Richtungsabhängigkeit der Lautstärkeempfindung von breitbandigen Schallen" ("Investigation of the directional dependence of the loudness of broadband sounds"), *Fortschritte der Akustik - Deutsche Arbeitsgemeinschaft für Akustik (DAGA)*, 537–540.
- Robinson, D.W. and Whittle, L.S. (1960). "The loudness of directional sound fields," *Acustica*, **10**, 74–80.
- Sivonen, V.P., Minnaar, P. and Ellermeier, W. (2005). "Effect of direction on loudness in individual binaural synthesis," *Proceedings of the 118th Convention of the Audio Eng. Soc.*, Convention Paper No. 6512.
- Sivonen, V.P. and Ellermeier, W. (2006a). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation," *J. Acoust. Soc. Am.*, **119**, 2965–2980.
- Sivonen, V.P. and Ellermeier, W. (2006b). "Binaural loudness summation for directional sounds," In preparation.
- Tuomi, O. and Zacharov, N. (2000). "A real-time binaural loudness meter," *Presented at the 139th meeting of the Acoustical Society of America*, Atlanta, USA.
- Zahorik, P. and Wightman, F.L. (2001). "Loudness constancy with varying sound source distance," *Nat. Neurosci.*, **4**, 78–83.
- Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and Models* (Springer, New York).

Appendix: Extended discussion of individual differences

This appendix investigates whether the dispersion in the individual directional loudness sensitivities of [M1] and [M4] could be accounted for by differences in the listeners' loudness functions. First, *individual loudness functions* as typically found in the literature are presented. Then, these individual functions are used in *predicting* directional loudness sensitivities and finally, *interindividual standard deviations* of the predicted sensitivities are compared with data obtained from listeners.

Individual loudness functions

The loudness function of a binaurally presented 1-kHz pure tone, as perceived by an *average* listener, can be approximated by (Scharf, 1978):

$$N_b = k_b(p - p_{0b})^{\beta_b}, \quad (1)$$

where N_b is loudness, k_b is a constant, p and p_{0b} are the (linear) sound pressure and the hearing threshold in μPa , respectively, and β_b the exponent determining the *slope* of the function. Subscript b stands for binaural values.

Data averaged across listeners have suggested a value of 0.6 for β (for a review, see Scharf, 1978). However, sizeable individual differences in the slope of the function have been reported, e.g., by Stevens and Guirao (1964) using loudness scaling, and by Collins and Gescheider (1989) using scaling and in addition, correcting for the individual slopes by cross-modality matching (CMM). These individual differences, obtained for a 1-kHz pure tone, are listed in Table 1, denoted as listener-specific exponents β .

The range of individual variation in the exponents obtained by scaling (see columns 'ME/MP' and 'AME' in Table 1) is almost threefold for the data reported by Stevens and Guirao (1964), and almost fivefold for these reported by Collins and Gescheider (1989). Since loudness scaling is susceptible to various numerical response biases (for a discussion, see section 1.2.1), non-auditory factors may be confounded with the loudness sensation in these exponents. The variation in the exponents is markedly reduced to a range from 0.49 to 0.74, when corrected for via cross-modality matching between loudness and apparent line length (see column 'CMM corr.' in Table 1). Thus corrected for, the exponents supposedly describe more accurately the "true" growth of loudness and rule out individual differences concerning the usage of scales and numbers. Note that the mean data are close to the value of 0.6.

The effect of changing the exponent β in the 'standard' loudness function of Eq. 1 is illustrated in Fig. 1a, using the range reported by Stevens and Guirao (1964) and the CMM-corrected range by Collins and Gescheider (1989) (the thin solid and the thick dashed lines in Fig. 1a, respectively). Sizable differences in the slopes of the functions can be observed, even for the reduced range of the CMM-corrected exponents.

Given the considerable range of individual exponents in Table 1, implying vastly different loudness-vs-SPL functions, could this dispersion be responsible for the interindividual variation of the directional loudness matches, obtained in [M1] and [M4] for anechoic, narrow-band noises at 1.0 and 5.0 kHz?

Table 1: Exponents (β) of individual loudness functions, obtained in two investigations via (1) a 'hybrid' method of magnitude estimation and production (ME/MP), and (2) the method of absolute magnitude estimation (AME), as well as correcting for numerical biases by cross-modality matching (CMM). The minimum and maximum values of each column are in boldface.

Stevens and Guirao (1964)		Collins and Gescheider (1989)*		
Subj.	ME/MP	Subj.	AME	CMM corr.
PK	1.1	A	1.5	0.55
BS	0.99	B	0.70	0.54
HR	0.98	C	0.69	0.60
SS	0.92	D	0.54	0.58
TI	0.85	E	0.49	0.53
MW	0.80	F	0.44	0.51
LD	0.73	G	0.62	0.74
LM	0.73	H	0.49	0.58
DD	0.56	I	0.46	0.63
LF	0.44	J	0.42	0.64
EG	0.40	K	0.31	0.49
		L	0.37	0.72
Mean	0.77		0.58	0.59

* Values originally reported for intensity ($I \propto p^2$), hence, multiplied here by 2.

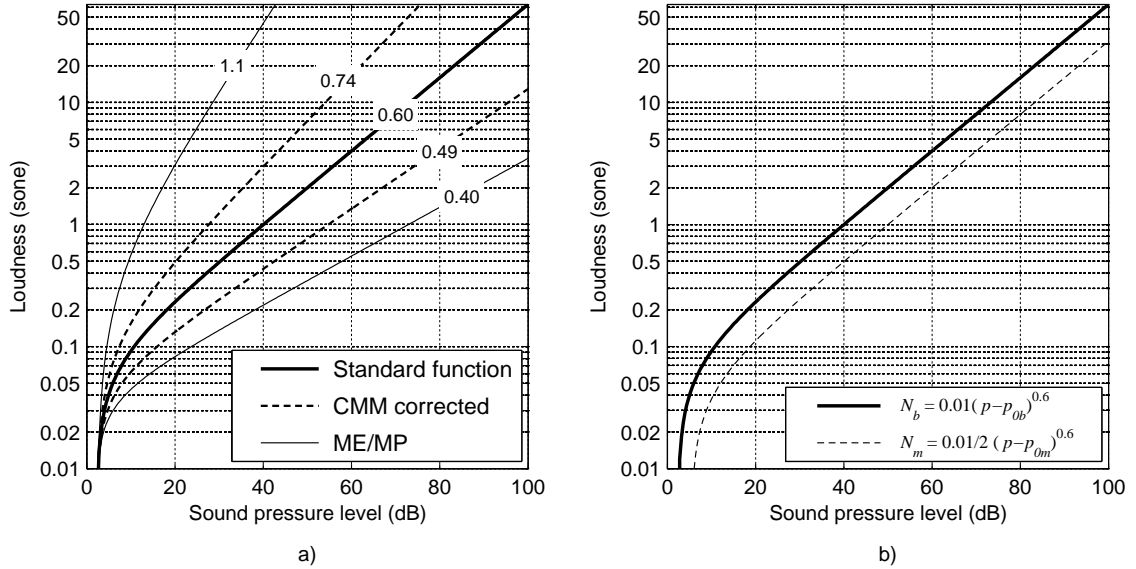


Figure 1: a) Loudness functions with different exponents β . (b) Monaural and binaural loudness functions (N_m and N_b , respectively). See text for details.

Table 2: Third-octave-band levels (L_p in dB SPL) for the generic HRTFs, for six incidence angles (azimuths ϕ) in the horizontal plane. These SPLs were converted to linear sound pressure (p in μPa) in the computations.

	1 kHz		5 kHz	
ϕ	$L_{p_{left}}$	$L_{p_{right}}$	$L_{p_{left}}$	$L_{p_{right}}$
0°	64.9	64.8	78.1	77.2
30°	67.7	60.2	82.1	69.2
60°	69.3	61.2	82.5	63.5
90°	69.9	63.5	82.2	55.2
135°	69.1	59.4	73.5	65.7
180°	66.1	66.2	72.1	70.6

Predicting directional loudness sensitivities

To explore this possibility, the third-octave-band levels for the *generic* HRTFs of [M4] were utilized as input sound pressures at an overall level of 65 dB SPL, creating at-ear levels as listed in Table 2. Thus, any individual differences in at-ear sound pressures were disregarded, the emphasis being on the value of β , and its effect on the directional loudness sensitivities. In addition, since the growth of loudness with SPL is approximately equal at 1.0 and 5.0 kHz (see the equal-loudness-level contours in section 1.3), a frequency-independent exponent β was used.

In [M1] and [M4] for a given incidence angle, the sound pressures at the two ears typically differed from one another ($p_{left} \neq p_{right}$). Thus, in order to compute loudness separately for the two ears, a monaural loudness function was defined as:

$$N_m = k_m(p - p_{0_m})^{\beta_m}, \quad (2)$$

where N_m is monaural loudness, k_m a constant, p and p_{0_m} the (linear) sound pressure and the monaural hearing threshold in μPa , respectively, and β_m the exponent of the function. A monaural loudness function is illustrated in Fig. 1b, together with the binaural function.

Assuming that binaural loudness is a simple summation of the monaural loudness components (Marks, 1978), Eq. 1 for binaural loudness can be written as follows:

$$N_b = N_{left} + N_{right} = k_{left}(p_{left} - p_{0_{left}})^{\beta_{left}} + k_{right}(p_{right} - p_{0_{right}})^{\beta_{right}}. \quad (3)$$

Due to supra-threshold at-ear levels (see the levels of Table 2 and the shape of the functions in Fig. 1b), the left- and right-ear loudness functions can be assumed to be simple power functions. Thus, the sloping of the functions near the threshold can be omitted, i.e., the effect of p_0 is negligible in Eq. 3. Furthermore, assuming that the two ears have the same "transducer sensitivity" ($k_{left} = k_{right}$; which is inspected explicitly in [M3]), and that loudness grows with SPL at the same rate for each ear, as well as for monaural and binaural listening ($\beta_b = \beta_{left} = \beta_{right}$), Eq. 3 can be written with a single constant k_m and exponent β (for a discussion of these assumptions, see Marks. 1978):

$$N_b = k_m p_{left}^\beta + k_m p_{right}^\beta = k_m(p_{left}^\beta + p_{right}^\beta) \quad (4)$$

Eq. 4 was then utilized for computing binaural loudness and *predicting* directional loudness sensitivities (*DLS*) for various exponents β . This was achieved by finding the relative levels (L , in 0.1-dB resolution) the comparison directions had to be set to, to minimize the absolute difference between the loudness prediction for the frontal reference (*ref*) direction and a comparison (*comp*) direction at a given azimuth ϕ . These predictions were made using Eq. 5, separately at the two center frequencies (f_c):

$$DLS(\beta, \phi, f_c) = \min_L |N_{b_{comp}}(\beta, \phi, f_c, L) - N_{b_{ref}}(\beta, f_c)| \quad (5)$$

Note, that this approach was similar to the one taken in [M1] and [M5], when utilizing the loudness model by Moore et al. (1997) for predicting directional loudness.

Interindividual standard deviations

To get a fair picture of the variability to be expected on the basis of different psychophysical loudness functions, the entire samples of exponents found by Stevens and Guirao (1964) and Collins and Gescheider (1989), respectively, were used in making predictions. Directional loudness sensitivities were predicted using Eq. 5 for each of the individual values of β obtained (1) by mere scaling (Stevens and Guirao, 1964, $n = 11$) and (2) by cross-modality matching (Collins and Gescheider, 1989, $n = 12$), see the values in the left- and right-most columns of Table 1 and the slopes plotted in Fig. 1a. The *interindividual* means and standard deviations of the predictions were then computed, and contrasted with the corresponding loudness matching data obtained from the listeners participating in the present set of experiments [M1 & M4]. While there were differences between the means of the data sets, this discussion will focus on the *interindividual standard deviations* only.

The interindividual standard deviations are listed in Table 3, for the two center frequencies (1 and 5 kHz) and five incidence angles in the horizontal plane, each row being assigned to a data set: The three top rows are the interindividual standard deviations for the subjective data of [M1] for eight listeners in a real sound field, and of [M4] for five expert and ten naïve listeners using binaural synthesis based on generic HRTFs. The two bottom rows are the predictions made using vastly different loudness functions. For all data sets, the deviations are larger at the higher center frequency, and they reach their maxima at an azimuth of 90° , where the interaural level difference also is the greatest (see Table 2).

The interindividual standard deviations obtained by the predictions for various exponents are generally small compared to the subjective data, see the two bottom rows in Table 3. The predicted deviations are the largest for the 90° azimuth at 5 kHz. However, when corrected for biases (Collins and Gescheider, 1989), the predicted standard deviation is less than 25% of the deviation found in the actual data. It thus seems, that the variation in the exponent of individual loudness functions, especially when corrected for biases (see the dashed lines of Fig. 1), can not explain the distribution of directional loudness sensitivities as obtained from the listeners' responses.

In obtaining the subjective data of [M4], the directional, *physical* differences between listeners were minimized by the use of generic HRTFs. When comparing with [M1], where an addi-

Table 3: Interindividual standard deviations (in dB) of directional loudness sensitivities, obtained via loudness matching from a number of listeners, and predictions made with different loudness exponents. See text for details.

	n	1 kHz					5 kHz				
		30°	60°	90°	135°	180°	30°	60°	90°	135°	180°
Intramodal matching											
[M1] Real sound field	8	0.7	0.9	1.0	0.9	1.0	1.4	1.4	2.7	2.5	2.3
[M4] Gen. HRTF, exp.	5	0.4	0.9	1.3	1.2	1.0	1.6	1.8	2.4	2.2	0.1
[M4] Gen. HRTF, naï.	10	0.9	1.0	0.7	1.2	0.6	1.6	1.9	2.2	1.1	0.4
Predictions from scaling											
Individual exponents	11	0.2	0.2	0.1	0.3	0.0	0.5	0.9	1.4	0.2	0.0
Corrected for biases	12	0.1	0.1	0.1	0.1	0.0	0.2	0.3	0.5	0.1	0.0

tional interindividual variation was caused by the fact that the at-ear signals were due to the listeners' individual *physical* obstruction in the sound field, the use of generic HRTFs does not seem to have reduced the interindividual variation, see Table 3. Possibly, additional *cognitive* variations may have been involved when using non-individual HRTFs, creating directional cues which the listener may not be used to.

It is worth noting that the interindividual standard deviations in the subjective data of Table 3 are small compared to other loudness matching studies, where values on the order of 4 dB, or larger, have been reported (Robinson and Dadson, 1956; Buus et al., 1997; Verhey and Kollmeier, 2002). However, a direct comparison between studies may not be justified, as the loudness is being matched by controlling different physical dimensions (spatial location vs. spectrum or duration), possibly having very different effects on the variability of loudness matches.

References

- Buus, S., Florentine, M. and Poulsen, T. (1997). "Temporal intergration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am., **101**, 669–680.
- Collins, A.A. and Gescheider, G.A. (1989). "The measurement of loudness in individual children and adults by absolute magnitude estimation and cross-modality matching," J. Acoust. Soc. Am., **85**, 2012–2021.
- Marks, L.E. (1978). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am., **64**, 107–113.
- Moore, B.C.J., Glasberg, B.R. and Baer, T. (1997). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc., **45**, 224–240.
- Robinson, D.W. and Dadson, R.S. (1960). "A re-determination of the equal-loudness relations for pure tones," Brit. J. Appl. Phys., **7**, 166–181.
- Scharf, B. (1978). "Loudness," In Carterette, E.C. and Friedman, M.P. (Eds.), *Handbook of Perception, Vol. IV: Hearing* (Academic Press, New York), pp. 187–242.

Stevens, J.C. and Guirao, M. (1964). "Individual Loudness Functions," J. Acoust. Soc. Am., **36**, 2210–2213.

Verhey, J.L. and Kollmeier, B. (2002). "Spectral loudness summation as a function of duration," J. Acoust. Soc. Am., **111**, 1349–1358.