Published in final edited form as:
Hear Res. 2010 February ; 260(1-2): 30. doi:10.1016/j.heares.2009.11.001.

On the ability of human listeners to distinguish between front and back

Peter Xinya Zhang[a,b] and **William M. Hartmann**[a]
Peter Xinya Zhang: pzhang@colum.edu; William M. Hartmann: hartmann@pa.msu.edu
[a] Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA
[b] Department of Audio Arts and Acoustics, Columbia College Chicago, Chicago, IL 60605, USA

Abstract

In order to determine whether a sound source is in front or in back, listeners can use location-dependent spectral cues caused by diffraction from their anatomy. This capability was studied using a precise virtual-reality technique (VRX) based on a transaural technology. Presented with a virtual baseline simulation accurate up to 16 kHz, listeners could not distinguish between the simulation and a real source. Experiments requiring listeners to discriminate between front and back locations were performed using controlled modifications of the baseline simulation to test hypotheses about the important spectral cues. The experiments concluded: (1) Front/back cues were not confined to any particular 1/3rd or 2/3rd octave frequency region. Often adequate cues were available in any of several disjoint frequency regions. (2) Spectral dips were more important than spectral peaks. (3) Neither monaural cues nor interaural spectral level difference cues were adequate. (4) Replacing baseline spectra by sharpened spectra had minimal effect on discrimination performance. (5) When presented with an interaural time difference less than 200 μs, which pulled the image to the side, listeners still successfully discriminated between front and back, suggesting that front/back discrimination is independent of azimuthal localization within certain limits.

Keywords

localization; front/back; human; simulation; transaural; localization bands

1. Introduction

The human auditory system localizes sound sources using different stimulus cues, such as interaural level difference cues, interaural time difference cues, and spectral cues. For localization in the median sagittal plane, e.g. for locations in front and in back, interaural cues are minimally informative (Oldfield and Parker, 1982; Middlebrooks, 1992; Wightman and Kistler, 1997; Langendijk and Bronkhorst, 2002). Instead, the spectral cues arising from unsymmetrical anatomical filtering are dominant (Musicant and Butler, 1984).

The roles of diverse localization cues can usefully be studied with virtual reality experiments (e.g. Wightman and Kistler, 1989a, b). Probe-microphones inside a listener's ear-canals are used to measure the head-related transfer functions (HRTFs) from real sound sources in an

anechoic environment. When these transfer functions are simulated through headphones, the listener perceives locations correctly for these virtual signals. Then by modifying the simulations in different ways one can test ideas about which physical attributes of the signal cues are most important in determining a listener's perception of location.

Wightman and Kistler (1989b) measured the ability of listeners to determine azimuth and elevation from virtual signals in comparison with results for real signals from the actual loudspeakers. It was found that the listeners localized the virtual signals well in the azimuthal plane, but much less well in sagittal planes. There were more front/back confusions with virtual sources. The relatively poor front/back localization performance with virtual signals might be attributed to the diffculty of accurately simulating the spectral cues, especially the high-frequency cues caused by the asymmetry of the pinna.

Kulkarni and Colburn (2000) showed that different fittings of headphones on a KEMAR (Knowles Electronics Manikin for Acoustic Research) led to different signals at the eardrums. The discrepancies were so pronounced above 8 kHz that a simulation of HRTFs using headphones became inadequate.

On the other hand, Asano *et al.* (1990) applied filters of different orders to smooth the microscopic structures at high frequencies and found that front/back discrimination did not depend on fine details at high frequencies – only macroscopic patterns seemed to be important.

The experiments of the present article also used a virtual reality technique called "extreme virtual reality (VRX)." The technique led to extreme accuracy in the amplitude and phase information presented to listeners for components with frequencies as high as 16 kHz. It also permitted the experimenters to be extremely confident about the simulation of real sources and carefully controlled modifications of them. Loudspeakers were used to present real sources, and other loudspeakers (synthesis speakers) were used to present baseline and modified simulations of the real sources. The loudspeakers gave the listener the opportunity to use his or her own anatomical filtering to discriminate the sources. Because headphones were not used, there was no need to compensate for a headphone response. As described in the method section below, the experimental technique was demanding and it proved possible only to study the ability to discriminate between two locations – directly in front and directly in back – a study that has been resistant to previous virtual reality experiments.

The goal of the experiments was to determine which cues are important to front/back discrimination. The strategy was to modify the amplitude and phase spectra of the simulated sources, to discover which modifications caused errors in discrimination. The VRX technique began by measuring the spectra of front and back real sources using probe tubes in the ear canals. Then signals were synthesized and delivered by the synthesis speakers such that the real-source spectra were precisely reproduced in the ear canals. This was the baseline synthesis. Because the spectra sent to the synthesis speakers were known exactly, only the assumption of linearity in the audio chain was required to generate a modified synthesis such that the spectra in the ear canals took on any desired values. This was the modified synthesis.

2. Materials and methods

The experiments studied front/back discrimination in free-field conditions. Real sources, virtual sources, and modified virtual sources could be presented in any desired order within an experimental run. This flexibility made it possible to verify the validity of baseline stimuli with a real/virtual discrimination task. Probe microphones in the listener's ear canals throughout the entire experiment ensured that the stimuli were well controlled. The above features were the same as in the azimuthal plane study by Hartmann and Wittenberg (1996), but the implementation was so greatly improved by the VRX technique that it was possible to simulate

real-source spectra up to 16 kHz and to present well-controlled modifications of real-source signals to the listener's ears.

2.1. Spatial setup

The experiments were performed in an anechoic room, IAC 107840, with a volume of 37 cubic meters. As shown in Fig. 1, there were four loudspeakers, all RadioShack Minimus 3.5 single-driver loudspeakers with a diameter of 6.5 cm. The front and back speakers (called "source speakers" below) were selected to have similar frequency responses. The left and right loudspeakers were "synthesis speakers," called α and β, with no requirements on matched frequency response. All loudspeakers were at the ear level of a listener. The listener was seated at the center of the room, facing the front source. The distance from the source speakers to the listener's ears was always 1.5 meters, and each synthesis speaker was 37 cm from the near ear. A vacuum fluorescent display on top of the front speaker displayed messages to the listener during the experiments. Two response buttons were held in the listener's left and right hands. Using the hand-held buttons instead of a response box was found to reduce head-motion. During the experiments, the listener made responses by pushing either button or both.

2.2. Alignment

In order to minimize head motion, the position of the listener's jaw was fixed using a bite bar, a rod attached rigidly to the listener's chair. In order to minimize binaural differences, the source speakers were positioned equidistant from the ends of the bite bar. The bite bar was 53 cm long, and at each end there was a 1/4-inch electret microphone for alignment. A source speaker was positioned by playing a sine tone through the speaker and modifying the speaker location so that the two microphone signals were in phase, as observed on an oscilloscope, while the 1.5-m distance to the center of the bite bar was maintained. The alignment procedure began at a low frequency and proceeded to higher frequencies, up to 17 kHz, making modifications as needed at each stage. Ultimately the procedure ensured that each source speaker was equidistant from each end of the bar; intermicrophone delays were within 10 μs, equivalent to 3.4 mm. Therefore, to a good approximation, a line drawn between the two source speakers was the perpendicular bisector of the bite bar.

At the beginning of an experimental run, the listener used a hand mirror to set his top incisors on either side of a pencil line drawn at the center of the bite bar. If the listener's anatomy is left-right symmetrical, this approach put the two ears equally distant from the front source speaker and equally distant from the back source speaker. A listener maintained this contact with the bite bar during the entire run.

2.3. Stimuli and listeners

The stimulus used in the experiments was a complex pseudo-tone with a fundamental frequency of 65.6 Hz and with 248 components that were pseudo-harmonics, beginning with the third harmonic (about 197 Hz). Pseudo-harmonic frequencies were chosen by starting with harmonic frequencies (harmonics of 65.6 Hz) and then randomly offsetting them according to a rectangular distribution with a width of ±15 Hz. The reason for the pseudo-tone is described in Appendix A.

Component amplitudes were chosen by starting with equal amplitudes and then applying a broadband valley to avoid the large emphasis of the external ear resonances. At various times during the years of experimenting, different amplitude spectra were used, as shown in Fig. 2 [Footnote 1]. Component phases were chosen to be Schroeder phases (Schroeder, 1970). The procedures for amplitudes and phases attempted to maximize the dynamic range for each component within the six-octave bandwidth. The Schroeder-minus phase condition was used because when added to the phase shifts caused by cochlear delays, these phase shifts tend

towards a uniform distribution of power throughout a cycle of the stimulus (Smith *et al.*, 1986). The highest frequency of the pseudo-tone was 16.4 kHz. A frequency of 16 kHz was identified by Hebrank and Wright (1974b) as the upper limit of useful median plane cues.

There were 11 listeners (B, D, E, F, G, L, M, P, R, V, and Z), five female and six male, who participated in some or all of the experiments. Listeners were all between the ages of 20 and 26 except for listener Z, the first author, who was 31. Listeners all had normal hearing, defined as thresholds within 15 dB of nominal from 250 to 16,000 Hz, as measured by Békésy audiometry using headphones. Because of the importance of high-frequency hearing to sagittal plane localization, listeners were also tested in the anechoic room using the front source loudspeaker. Again the test was an eight-minute pure-tone Békésy track from 250 to 16,000 Hz. Each ear was tested individually by plugging the other ear. It was found that listener thresholds were below the level of the pseudo-tone components for all components up to 16000 Hz. There were three exceptions to that result: Thresholds for listeners F and Z exceeded the pseudo-tone levels above 14 kHz, and listener G was not tested for thresholds using the loudspeaker.

2.4. Signal generating and recording

Signals were generated by the digital-to-analog converters on the DD1 module of a Tucker Davis System II, with a sampling rate of 50 ksps and a buffer length of 32768 words. After low-pass filtering at 20 kHz with a roll-off rate of −143 dB/octave, the signals were sent to a two-channel power amplifier and then to individual loudspeakers in the anechoic room by way of computer-controlled relays. Tones were 1.3 s in duration, turned on and off with 100-ms raised cosine edges, and were presented at a level of 80 dB SPL as measured with an A-weighted sound level meter at the position of the listener's head.

For recording, Etymotic ER-7C probe-microphones were placed in the listener's ear-canals. Each probe microphone was connected to its own preamplifier with frequency-dependent gain (about 25 dB) compensating the frequency response of the probe tube. The outputs, were then passed to a second preamplifier adding 42 dB of gain, before the signals left the anechoic room. The output signals from the preamplifiers were lowpass filtered at 18 kHz with a roll-off rate of −143 dB/octave, and then sent to the analog-to-digital converters on the DD1 module, with a sampling rate of 50 ksps and a buffer length of 32768 words.

Capturing the probe-microphone signals in the computer will be called "recording" in the text that follows. Once a signal was recorded, it was analyzed. Because the frequency, f, of each of the 248 components was known exactly, it was possible to extract 248 amplitudes and 248 phases for each ear. The complex phasor array with two elements (left-ear and right-ear) for all frequencies will be called the "analyzed signal," given the symbol $Y(f)$ or $W(f)$ below.

2.5. VRX procedure

The VRX technique was based on a transaural synthesis known as "cross-talk cancellation" (Schroeder and Atal, 1963; Morimoto and Ando, 1980). As defined by Cooper and Bauck (1989), a transaural method has the goal of generating an appropriate signal at each of the listener's ears. The idea of "cross-talk cancellation" is that no part of the signal intended for the left ear should appear in the right ear canal and vice versa. The technique simulates a

[1]Pseudo-tone spectra were changed several times in an attempt to improve the dynamic range of the synthesis procedure, given some dramatic individual differences in head-related transfer functions. Listeners in early experiments showed dips in ear canal pressure in the 8–11 kHz region. This was compensated by the rectangular spectral boost (EQ 2) in Fig. 2, and later by the smoother boost (EQ 3). When other listeners failed to show such pressure dips, the boost in this spectral region was abandoned for all listeners and EQ1 was used. Whenever a change in equalization was made, the *before* and *after* conditions were used in front/back discrimination experiments to try to detect changes. No changes in localization performance were ever found that could be attributed to the change in stimulus equalization.

real source having an arbitrary location by means of two synthesis loudspeakers which produce signals identical to the real source in a listener's ear canals. For every frequency component there are four unknowns, the amplitudes and phases for the two synthesis speakers. Knowing the desired amplitudes and phases in the ear canals for the real source, in addition to knowing the transfer functions between the two synthesis speakers and the two ear canals leads to four equations which can be solved for the four unknowns.

The VRX calibration steps were described mathematically in the thesis by the first author (Zhang, 2006, page 160–177). An abbreviated version follows:

1. The pseudo-tone stimulus, with complex components $X(f)$, was played through the front source speaker (F) and recorded and analyzed as left (L) and right (R) ear-canal signals $Y_{F,L}(f)$ and $Y_{F,R}(f)$.
2. The pseudo-tone was played through the synthesis speaker α and recorded and analyzed as left and right ear-canal signals $W_{\alpha,L}(f)$ and $W_{\alpha,R}(f)$.
3. The pseudo-tone was played through the synthesis speaker β and recorded and analyzed as left and right ear-canal signals $W_{\beta,L}(f)$ and $W_{\beta,R}(f)$.

   These three steps provide enough information to determine the signals $S_\alpha(f)$ and $S_\beta(f)$ which can be sent to the synthesis speakers in order to reproduce the recordings of the front source, $Y_{F,L}(f)$ and $Y_{F,R}(f)$. The mathematical key to the synthesis technique is to regard the four values of $W(f)$ as a two-by-two matrix, and then use its inverse to multiply array $Y_F(f)$.

   In principle, signal S is an adequate synthesis signal. However, we realized that the pseudo-tone, X, sent to the synthesis speakers in calibration steps 2 and 3 would be very different from the synthesis signal S. If the speakers and recording chain are perfectly linear then the difference is of no consequence, but if there is nonlinear distortion, it is possible that a large difference between the calibration signal and the computed synthesis signal might lead to errors in the simulation. Therefore, the calibration was iterated with the following steps.
4. Signal $S_\alpha(f)$ was played through the synthesis speaker α and recorded and analyzed as left and right ear-canal signals $W'_{\alpha,L}(f)$ and $W'_{\alpha,R}(f)$.
5. Signal $S_\beta(f)$ was played through the synthesis speaker β and recorded and analyzed as left and right ear-canal signals $W'_{\beta,L}(f)$ and $W'_{\beta,R}(f)$.

   Inverting the two-by-two matrix $W'(f)$ then led to alternative synthesis signals $S'_\alpha(f)$ and $S'_\beta(f)$. It was expected that S' would be less affected by nonlinear distortion than S. However, for some values of frequency f, the α or β part of the $S(f)$ was quite small, and that led to a noisy estimate for $W'(f)$ and consequently for $S'(f)$.
6. Therefore, the next step was to record and analyze the signals in the ear canals when synthesis S and synthesis S' were presented to determine, for each frequency, which synthesis led to closer agreement with the target signals $Y_{F,L}(f)$ and $Y_{F,R}(f)$. In this way, the trade off between distortion and noise was optimized component-by-component.
7. Sometimes neither S nor S' led to an acceptable amplitude in both left and right ear canals. In the final signal generation step the error measurement from step (6) was used to accept or eliminate each frequency component. If a component deviated from the target $Y_{F,L}(f)$ or $Y_{F,R}(f)$ by more than 50% in amplitude, i.e. an error outside the range −6 to +3.5 dB, then the component was eliminated from the synthesis. Normally there were only a few eliminated components, and their number was limited by the

protocol. If a component was eliminated in the calibration of the front source, it was also eliminated from the synthesis for the back source. If more than 20 out of the 248 components were eliminated, the calibration was considered to be a failure and the procedure started over from step (1). Otherwise the synthesis was tentatively accepted and called the "baseline simulation" for the front source.

8. After the tentative baseline simulation was determined, the VRX protocol included a confirmation test to discover whether the listener could learn to distinguish between real (front source) and virtual (baseline simulation) signals. The confirmation test began with a training sequence of four intervals, known by the listener to be *real-virtual-real-virtual*. The listener could hear the sequence as many times as desired. When the listener was satisfied with the training, or gave it up as hopeless, the test phase followed. The test phase contained 20 single-interval trials (10 real and 10 virtual in a random order). In each trial, the listener tried to decide whether the sound was real or virtual and then reported the decision using the push buttons. If the percentage of correct responses was between 25% and 75%, it was concluded that the listener could not distinguish between the real and virtual signals, and the experiment continued; otherwise the calibration sequence started again from the very beginning.

9. –(16) If the front baseline simulation passed the confirmation test, the eight-step calibration sequence was repeated for the back source. As for the front source, components were optimized (S vs S') and possibly eliminated.

The total number of components eliminated by the front and back calibrations was limited to 20. The spectrum of eliminated components was displayed to the experimenter during the calibration procedure. In addition to the limit on the number of eliminated components, the experimenter was wary of blocks of adjacent eliminated components possibly leading to spectral gaps. No study was made of the distribution of eliminated components. Instead, the runs for any given experiment were not all done successively, a procedural element that was intended to randomize the distribution of eliminated components. If the back-source simulation was unsuccessful at some stage, the experiment re-started from the very beginning with step (1).

Figure 3(a) shows an example of two recordings in the right ear canal for the back source. The open symbols show the recording of the real source, and the dots show the recording of the virtual signal, i.e. the baseline simulation. The agreement between real and virtual recordings is typical of VRX calibrations. Two points above 15 kHz are plotted off the graph, below the horizontal axis. They were eliminated from the baseline simulation for both front and back because they did not meet the ±50% amplitude error criterion.

Figure 3(b) shows the corresponding phase information. It shows the difference between the virtual phase and the real phase. The figure shows that all components had an absolute phase error less than 15 degrees, and only two components had an absolute phase error greater than 10 degrees.

The duration for the calibration sequences and the confirmation tests was approximately 2.5 minutes. As we gained experience with the VRX protocol, we discovered that the confirmation tests, such as step (8) above, could normally be omitted because whenever a simulation met the objective standard – fewer than 20 components eliminated and no long blocks of continuous eliminated components – then the listeners could not discriminate real and virtual signals. Therefore, to make the runs shorter we relied on the objective standard for most of the runs and employed confirmation runs occasionally, approximately on every tenth run, and especially after a new fitting of the probe microphones.

2.6. VRX experiments

If both front and back sources were adequately simulated in the baseline synthesis, as indicated by the component level measurements and by the optional confirmation test, the experimental run continued with modifications to the baseline. All the modifications were focused on a frequency domain representation of the stimulus – eliminating or distorting spectral features with the goal of discovering critical spectral features. As in previous virtual reality experiments, the goal was to control the spectral features as they appear in the listener's ear canal. Therefore, the spectra described in the sections to follow are spectra measured in the ear canals. The advantages of the VRX technique over other virtual reality techniques are that it does not use headphones and it enjoys a self-compensating feature, as described in Appendix B. Spectral modifications were selected, often tailored to the individual listener, to test hypotheses about front/back localization.

The methods used in the experiments were approved by the Institutional Review Board of Michigan State University.

3. Experiment 1: flattening above and below

Experiments 1 and 2 tried to determine whether the cues to front/back discrimination were in a single frequency band or in multiple frequency bands, and which band or bands were involved. By flattening the amplitude spectra within a frequency band, the detailed front/back spectral cues within the band were eliminated because the flattening process made them the same for front and back sources. Then the listener had to use cues outside the band to discriminate front from back. Performance of each listener was examined with various flattened frequency bands.

Changing spectra to determine relevant spectral regions for localization is not new. He-brank and Wright (1974b) used high-pass, low-pass, band-pass, and band-reject filtered stimuli in their localization experiments. These filtered stimuli removed power from selected spectral regions. By contrast, our flattened spectra left the average power unchanged in broad frequency regions. Therefore,

- **i.** No extra spectral gradient was introduced, which might itself be a localization cue (Macpherson and Middlebrooks, 1999, 2003).
- **ii.** Listeners could not immediately distinguish flattened spectra from baseline spectra. By contrast, if the signals are filtered to remove energy from a spectral region, listeners know that they are being given less information.
- **iii.** The spectrum level and overall level were unchanged. For filtering experiments, as available information is reduced, either the spectrum level or the overall level must change, which might affect performance.

In headphone experiments with goals similar to ours, Langendijk and Bronkhorst (2002) flattened directional transfer functions (DTF) in various frequency bands. They flattened a DTF by taking the average of the amplitude spectrum for each source independently. Similarly, the experiments by Asano *et al.* (1990) simplified the HRTFs for one source location at a time. In our flattening experiments the average was taken over front and back sources together. Thus, experiments by Langendijk and Bronkhorst and experiments by Asano *et al.* only removed or simplified the local spectral structure within certain bands, whereas the flattened bands in our experiments also eliminated the spectral differences between the two sources that the listeners had to distinguish.

3.1 Necessary and/or sufficient bands

The flattening experiments were motivated by the idea that the relevant spectral cues for front/back discrimination might lie in a single frequency band, narrower than the 16-kHz bandwidth of our stimuli. One can imagine a *necessary band*, defined by upper and lower edge frequencies, every part of which is essential to discriminate front from back. Alternatively one can imagine a single *adequate band*, i.e. *sufficient band*. The spectral information in an adequate band is, by itself, sufficient for the listener to successfully discriminate.

The alternative to single-band models is multiple-band models. A listener may compare the spectral structure in one frequency region with the structure in a remote region. With this strategy, both frequency regions are necessary and neither is sufficient. Alternatively a listener may have a flexible strategy. If deprived of information in one frequency region, the listener can use the information in another. For such a listener there are multiple adequate bands. Experiments 1 and 2 below were designed to look for single or multiple necessary or adequate bands.

Concepts of “necessary” and “sufficient” bands have previously appeared in studies of spectral cues for front/back discrimination. Experiments by Asano *et al*. (1990) found that macroscopic patterns at high frequencies were necessary for front/back judgement. They also found that if energy was present in the band below 2 kHz then it was necessary that precise microscopic spectral cues be available, though these cues alone were not sufficient.

3.2. Experiment 1A: flattening above

Experiment 1A examined the role of high-frequency spectral cues. In Experiment 1A, the amplitudes of high-frequency components were all caused to be equal (flattened) as measured in the ear canals. All amplitudes for frequencies above and including a *boundary frequency* (f_b) in the baseline spectra were replaced by the root-mean-square average, where the average was computed over both front and back sources, at each ear independently. The components below f_b in the baseline spectra were unchanged. The phase spectra of the modified signals were identical to baseline.

By applying the transaural matrix equations, the modified syntheses with flattened spectra were computed for presentation through the synthesis speakers. The choice of matrices was made for each frequency, depending on which of the baseline synthesis signals, S or S', was better.

The modified spectra, as measured in the ear canals, were compared with the desired modified spectra. The frequency components that deviated from the desired spectra by more than 50% (corresponding to an error larger than −6 dB or +3.5 dB) were eliminated. Overall, there were few eliminated components, and never more than 20. If more than 20 components failed the comparison test (including those eliminated in the calibration sequence), the simulation was considered a failure, and the entire calibration sequence was repeated. Figure 4 shows a modified spectrum for the back source, flattened above the boundary, f_b = 10 kHz, together with the baseline, as measured in the right ear of a listener. The overall power in any broad spectral region is the same for modified and baseline signals, but the information above the boundary frequency is eliminated in the modified version.

In each run of this experiment, the modified syntheses for the front and back sources were presented to the listener in a random order for 20 trials (10 for the front source and 10 for the back source). The listener’s task was to respond whether sound came from front or back, by pressing the corresponding buttons. There was no feedback. Besides these 20 trials, eight trials of baseline simulation (four for the front source and four for the back source) were added randomly, to make sure that the listener could still do the discrimination task. Therefore, each run included 28 trials. If the listener failed to discriminate the baseline simulation more than

once in the eight baseline trials, it meant that either the synthesis was failing or that the listener had temporarily lost the ability to discriminate. The data from that run were discarded. The procedure described in this paragraph was practiced in all of the following experiments.

Eight listeners (B, D, E, F, L, M, R, and Z) participated in Experiment 1. The testing range of boundary frequencies was chosen for each listener so that the performance decreased from almost perfect (100%) to close to the 50%-limit. [Footnote 2]. The filled circles in Fig. 5 show the results of Experiment 1A in the form of percent correct on front/back judgement as a function of boundary frequency. Each listener did four runs for each condition. Hence each data point on the figure is a mean of four runs, and the error-bar is the standard deviation over the four runs.

The filled circles show decreasing performance with decreasing boundary frequency. For example, the data of listener R show that she could successfully discriminate front and back sources having all the information below 14 kHz, but she failed the task with only information below 10 kHz. Figure 5 shows large individual differences among the listeners. The scores for listeners E, F, L, R and Z dropped sharply, within a frequency span of 4 kHz, as f_b decreased below a value that ranged from 6 to 12 kHz. The scores for listeners B and D decreased very slowly over a much broader frequency range.

Listeners B, D, L, and M scored greater than 80% correct when presented with information only below 4 kHz. An ability to use low-frequency information like this was suggested by Blauert, who found significant cues for front/back localization around 500 and 1000 Hz (Blauert, 1983, p. 109). Both Experiment 1A and Blauert's experiment show that it is not necessary to have cues above 4 kHz to successfully discriminate front from back. Moreover, Asano *et al.* (1990) found that listeners' front/back judgements were successful with smoothed spectra that eliminated the detailed structure above 3 kHz, though listeners failed the task with smoothing below 2 kHz. This suggests that it is not always necessary to have the information above 3 kHz for successful front/back judgement.

In their lowpass experiments, Hebrank and Wright (1974b) found that information above 11 kHz was required for localization in the median sagittal plane, which clearly disagrees with the results of all listeners in Experiment 1A except for listeners B and R. However, their loudspeaker did not pass energy below 2.5 kHz, whereas low frequencies were included in Experiment 1A. This could explain the difference between the results in Experiment 1A and the results from Hebrank and Wright.

3.3. Experiment 1B: flattening below

Experiment 1B examined the importance of spectral cues at low frequencies. It was similar to Experiment 1A, except that it was the frequency components *below* the boundary frequency f_b for which amplitudes were flattened, and the frequency components above f_b were unchanged.

The eight listeners from Experiment 1A also participated in Experiment 1B. Their success rates are shown by the open circles in Fig. 5. The open circles in Fig. 5 show decreasing performance in Experiment 1B as the boundary frequency increased, which is reasonable because useful

[2]A score of 50% correct can arise in different ways. Sometimes, listeners heard sound images that were either diffuse or in the center of head. Sometimes, they found that they could hear the sound images from both directions. For these two conditions, the 50%-limit corresponds to random guessing. Alternatively, listeners sometimes perceived that all the sound images were in only one direction, clearly in front or clearly in back, and they made their responses accordingly. For this condition, a score of 50% arises because sources in front and in back were presented the same number of times. For all of these conditions with scores close to 50%, listeners could not find an effective localization cue to discriminate front from back. Thus this article does not distinguish among these conditions, and simply notes them as near the "50%-limit".

front/back cues were eliminated below that increasing boundary. For example, the data of listener R show that having all the information above 6 kHz (f_b = 6 kHz) was adequate for her to discriminate front and back, but having only the information above 8 kHz (f_b = 8 kHz) was inadequate.

Apart from this general decreasing tendency, listeners demonstrated large individual differences. The drop in performance occurred at different boundary frequencies for different listeners, and the frequency spans of the drop were also different. As the boundary frequency increased, performance of listeners B, L, R and Z dropped sharply within a span of only 2 kHz. The performance of listeners D, E, F and M decreased over a much wider span.

3.4. Discussion of Experiments 1A and 1B

The heavy horizontal lines in Fig. 5 are our best estimates of the adequate bands based on the results of flattening high and low frequency regions. If a listener is presented with all the detailed spectral information in an adequate band, front/back discrimination will be good; scores will be greater than 85 percent correct, equivalent to the score required for baseline synthesis. By definition, it follows that no part of a necessary band can lie outside an adequate band.

Classifying listener types—Listeners were classified according to the shape of their performance functions in Fig. 5. Listeners D, E L, and maybe M, were classified as “A-shape” listeners because the shape looked like a letter “A.” Listeners B, F, and Z were classified as “X-shape” because of the crossing of the plots near the 85% correct point. Listener R was called “V-shape” and not “X-shape” because her performance dropped so rapidly that the plots only crossed near the 50-percent limit.

The heavy lines in Fig. 5 show that for A-shape listeners there is no single necessary band. For these listeners, there is a low-frequency adequate band and a high-frequency adequate band, and these bands do not overlap. Deprived of low-frequency information these listeners can use high-frequency information and vice versa. For all the other listeners there may be a necessary band somewhere in the frequency region where the heavy lines overlap.

4. Experiment 2: flattening inside and outside

Experiment 2 was designed for X-shape and V-shape listeners with the goal of determining whether there is a necessary band for them. Following the logic above, the experiment focussed on the region of overlap between high- and low-frequency adequate bands. This region was called the “central band.” It was hypothesized that this central band includes a necessary band.

4.1. Experiment 2A: flattening inside

Experiment 2A was similar to Experiment 1 except that the frequency components within the central band were flattened. The upper and lower boundary frequencies for each listener were determined from Experiments 1A and 1B, so that the central band included the necessary band, if it exists.

Five of the eight listeners in Experiments 1A and 1B participated in Experiment 2A. Three of the five listeners (B, R and Z) were V-shape or X-shape listeners. Listeners E and L, who were A-shape listeners, also participated though the experiment was not designed for them. Central bands were chosen as follows: For V-shape listener R, 6–13 kHz. For X-shape listeners B and Z, 8–14 kHz and 6–9 kHz, respectively. For A-shape listeners E and L, 4–9 kHz and 6–10 kHz respectively. The results are shown by open squares in Fig. 6.

For the three V- and X-shape listeners (B, R and Z), for whom this experiment was designed, the necessary band hypothesis predicts that performance should be poor because information in the necessary band was removed. However, the open squares in Fig. 6 show that only listener R did poorly. Listener R was the only V-shape listener in this experiment, and poor results were especially expected for her. A V-shape listener requires information over a wider frequency band compared to X-shape or A-shape. Contrary to the hypothesis, listener Z achieved a nearly perfect score and listener B's score was very close to the 85% criterion. The good performance by listeners Z and B clearly disagreed with the necessary-band hypothesis.

The two A-shape listeners, E and L, achieved perfect scores, which was not surprising. According to Experiment 1 these listeners could discriminate between front and back sources with even less information than actually provided in Experiment 2A.

4.2. Experiment 2B: flattening inside with wider central band

Listeners L and Z had fairly narrow central bands in Experiment 2A, and flattening within those bands eliminated very little front/back information. Both listeners did very well in Experiment 2A. The purpose of Experiment 2B was to test whether listeners L and Z could still succeed in a task with wider flattened central bands. For listener L, the central band of flattened amplitudes was increased to 3–12 kHz. For listener Z, the central band was increased to 4–11 kHz.

The results of Experiment 2B are shown as solid squares in Fig. 6. Clearly both listeners performed well above the 85%-criterion. Most impressive was the performance by listener L, who received even less spectral detail than in Experiment 1 and yet managed a perfect score. Possibly this listener benefited from having both extremely high and extremely low frequency information available simultaneously.

4.3. Experiment 2C: flattening outside

Experiment 2C was simply the reverse of Experiment 2A. In Experiment 2C, the spectrum outside the central band was flattened, and the frequency components within the band were unchanged, i.e. they were identical to baseline. For V-shape and X-shape listeners, the central band is part of both the low-frequency adequate band and the high-frequency adequate band. This experiment determined whether the central band is adequate by itself. The five listeners from Experiment 2A participated in this experiment. Their results are shown by open circles in Fig. 6.

None of the listeners did well in this experiment. Their scores were close to the 50%-limit. The scores for listeners E, L, R and Z, were exactly 50% with no error-bar because these listeners heard all the modified synthesis coming from only one direction, either front or back. The poor performance indicates that the central band is not an adequate band. Because any single necessary band was included in the central band, it can be further said that if a necessary band exists, it is not an adequate band.

5. Summary of Experiments 1 and 2

In Experiments 1 and 2, spectral patterns in various frequency bands, bearing information for front/back discrimination, were eliminated by flattening the amplitude spectrum. One goal of the experiments was to discover whether there is a necessary band that is essential for a given listener to successfully discriminate front from back. Another goal was to find an adequate band or bands.

For four of the eight listeners in the experiment (called A-shape listeners), the concept of the necessary band was immediately rejected because they exhibited low-frequency and high-frequency adequate bands that did not overlap.

The remaining listeners, except for one, were X-shape listeners. Experiment 1 hinted strongly at a single necessary band somewhere in the region of overlap between the low-and high-frequency adequate bands for these listeners. For both flattening-above (1A) and flattening-below (1B) experiments, as the boundary moved through this region, the discrimination performance rate changed from near 100% to near 50%. Thus, Experiment 1 suggested that this frequency region contained critical information. That observation motivated the hypothesis that this region (the central band) contained a necessary band for the X-shape listeners. That hypothesis drove Experiment 2.

Neither of the two X-shape listeners in Experiment 2A supported the necessary band hypothesis. The central region did not prove to be necessary for correct discrimination. On the contrary, Experiment 2C showed that what was necessary for all the listeners in Experiment 2 was spectral detail *outside* the central region. This result is difficult enough to understand as to demand some speculation as to how it might occur. The case of listener Z will serve as an example.

As shown in Fig. 5, Listener Z has a central band from 6 to 9 kHz. One can conjecture that this listener discriminates front from back by making comparisons in three critical frequency regions, one near 2 kHz, another near 7 kHz, and yet another near 10 kHz. In Experiment 1A, as everything was flattened above 6 kHz, information in the two higher-frequency bands was eliminated and no comparison could be made. In Experiment 1B, as everything was flattened below 8 kHz, information in the two lower-frequency bands was eliminated, again permitting no comparison. In Experiment 2A only the band near 7 kHz was affected and the listener could compare structure in the highest and lowest bands in order to make successful decisions.

In summary, for six out of the eight listeners there was no single necessary band. Instead, these listeners appear to be capable of making comparisons across a variety of frequency regions. For the other two listeners, only one, listener R, participated in Experiment 2. For Listener R, the only V-shape listener, there was evidence of a necessary band from 6 to 13 kHz, a very broad band.

The abilities of listeners to use information in different frequency bands as measured in the flattened-band paradigms of Experiments 1 and 2, particularly the classification as A-, X-, and V-shape listeners, have implications for capabilities in other circumstances. These implications were tested in an entirely different kind of experiment in Section 11. The conclusions of that section are that the abilities measured in Experiments 1 and 2 continue to apply outside the narrow context of a flattening experiment.

6. Experiment 3: peaks and dips

Experiments with sine tones or with one-third-octave noises (Blauert, 1969/70), or with one-twelfth-octave noises (Mellert, 1971), or with one-sixth-octave noises (Middlebrooks, 1992) show elevation cues that correspond to peaks in the spectrum. Blauert (1983) refers to them as boosted bands, serving as directional bands. However, other research, based on stimuli with broader bands, has pointed to notches, i.e. dips in the spectrum (Bloom, 1977a, b; Hebrank and Wright, 1974b). Experiment 3 was performed to determine whether peaks or dips were dominant in the ability to distinguish front from back.

6.1. Method and results

The modifications in Experiment 3 were all applied above a chosen boundary frequency. The components below the boundary frequency were identical to baseline. The boundary frequency was different for different listeners and was taken from the *flattening-above* portion of Experiment 1, where the listener's performance dropped to 60%. By choosing the boundary frequency in this way we were sure that critical information was affected by the modifications. Four listeners (B, L, R, and Z) participated in Experiment 3. Their individual boundary frequencies are shown in Fig. 7.

In Experiment 3A, dips in the baseline spectra were removed and only peaks were left. To remove the dips, the RMS amplitude was first calculated from the baseline spectra above the boundary frequency, averaging over both front and back sources. The final modification was achieved by finding those components having frequencies above the boundary frequency and amplitudes less than the RMS amplitude, then setting the amplitudes of those components equal to the RMS amplitude. Flattened amplitudes were the same for front and back modified signals. The open circles in Fig. 7 show the results of Experiment 3A. The scores of all of the four listeners were somewhere between perfect (100%) and the 50%-limit.

Experiment 3B was the reverse of Experiment 3A in that peaks in the baseline spectra were removed and dips were preserved. As for Experiment 3A, the altered components were above the boundary frequency, and the flattened amplitudes were given by the RMS values, the same for front and back.

The solid circles in Fig. 7 show the results of Experiment 3B. Three out of the four listeners achieved nearly perfect scores (100%). Compared to the scores achieved with peaks only, the scores with dips only were better for all four listeners. A one-tailed t-test showed that the difference was significant for three of them (for listeners B and Z, significant at the 0.05-level; for listener L, significant at 0.1-level).

The small diamonds in Fig. 7 indicate the performance on the *flattening-above* experiment at the same boundary frequency. The diamonds serve as a reference. One would not expect performance on either *Peaks only* or *Dips only* to be below the diamonds and they are not.

6.2. Discussion

Mellert (1971) and Blauert (1972) hypothesized that both peaks and dips in the spectra are important for localization in sagittal planes. Hebrank and Wright (1974b) argued that a dip is a particularly important cue for the forward direction. The results of Experiment 3 suggest that dips are the more important cues for front/back localization. Obviously, the validity of this conclusion depends on the definition of peak and dip as a deviation from the RMS value as well as the restriction to a critical high-frequency region. Neurons have been found in the dorsal cochlear nucleus of cat (Nelken and Young, 1994) and of gerbil (Parsons *et al*., 2001) that show sharp tuning for notches in noise. It was conjectured that these units mediate localization in sagittal planes. Experiment 3 provides support for the importance of notches for front/back discrimination.

7. Experiment 4: monaural information

Because spectral cues are thought to be the basis for front/back discrimination, one might expect that a listener could discriminate front from back using the spectral information in only one ear. An obvious way to test this idea is to make the listener effectively monaural by completely plugging one ear. However, plugging one ear causes the sound image to move to the extreme opposite side, and therefore the front/back discrimination experiment requires listeners to rely on percepts other than localization (Blauert, 1983). Our informal listening tests

confirmed that listeners with one ear plugged found the front/back task to be unnatural in the sense that all the images were on one side and there was no front/back impression.

As an alternative to plugging one ear, Gardner (1973) and Morimoto (2001) partially filled the pinna cavities of one ear but left an open channel to avoid extreme lateralization of the image. This technique severely modified the pinna cues, but it retained other features of directional filtering, e.g. the diffraction due to head, neck and torso. Experiment 4 removed the spectral details from the signal to one ear, thereby removing the cues to front/back localization, while retaining the spectral power in that ear to avoid extreme lateralization.

7.1. Method and results

Experiment 4 tested monaural front/back discrimination by flattening the right-ear spectrum while leaving the left-ear spectrum identical to baseline. (The modified phase spectra in both ears were identical to baseline.) Flattening the spectrum in one ear while leaving the power the same does not lead to an extremely lateralized image. Completely flattening the spectrum in one ear, as in Experiment 4, eliminates *all* the directional filtering cues, in a controlled way, regardless of their anatomical origin.

Seven listeners (B, E, F, L, M, R, and Z) participated in Experiment 4, and their results are shown as circles in Fig. 8. Except for listener E, the listeners performed poorly (below 75%) on this experiment, suggesting that monaural cues are not adequate for most listeners for successful front/back judgement. Listener Z's score was exactly 50% correct, with no error-bar, because he localized all modified signals in the back.

Squares in Fig. 8 indicate performances on runs with baseline stimuli for comparison (open squares). Three listeners did not do complete baseline runs, and their baseline scores were calculated from the 80 baseline trials in the first ten continuous runs (filled squares). Ideally, results with baseline should be perfect, and Fig. 8 shows that they usually were. It is clear from the figure, and was confirmed by one-tailed t-tests at the 0.05-level, that performance with flattened spectra in the right ear was significantly worse than baseline performance for all listeners.

7.2. Discussion

When listeners hear unusual spectra, such as the flattened right-ear spectrum of Experiment 4, they typically find the image to be located diffusely, or inside the head, or both. Some listeners in Experiment 4 perceived diffuse images with no preference for one side, while others reported split images with different spatial locations. Other listeners reported hearing a spatially compact image, but they were usually unable to identify the image as front or back.

There is contradictory evidence from previous research indicating that listeners *do* gain useful front/back information monaurally. Tests with modified pinna cues on one side by Morimoto (2001) studied the role of different azimuths in vertical plane localization. The results showed that the far ear stopped contributing to elevation judgements when the azimuth exceeded 60° measured from the midline. An abstract from Wightman and Kistler (1990) reported that flattening the spectrum in the ear contralateral to the source azimuth led to rather little degradation of the ability to localize. Martin *et al*. (2004) found good elevation perception and front/back discrimination from monaural high-frequency cues when low-frequency binaural cues to azimuth were available. All these studies provide evidence that monaural cues can be effective in vertical plane localization. Further, it is known that listeners can be trained to distinguish front from back when the HRTFs are changed (Hofman *et al.*, 1998; Zahorik *et al.*, 2006). Hebrank and Wright (1974a) even concluded that with modest training listeners could successfully *localize* sources in the vertical plane with one ear completely occluded.

A way to square the results from previous research with the results from Experiment 4 is to conclude that listeners may be able to localize successfully using only monaural cues, but listeners are unable to ignore the useless cues from the other ear when those useless cues are as intense as the useful cues across the entire spectrum. This interpretation is consistent with the observations made by Hofman and Van Opstal (2003) on binaural weighting of elevation cues. It would not seem to apply to the abstract by Wightman and Kistler (1990). A second point to note is that the listeners in Experiment 4 were not specially trained in monaural discrimination.

8. Experiment 5: interaural differences

There is an intrinsic problem in using spectral cues for front/back localization: How can a listener know that the peaks and dips at certain characteristic frequencies are due to directional filtering and not properties of the original sound? One way to solve this problem is to use interaural spectral level differences (ISLD) as front/back cues. ISLD is defined as the interaural level difference between left and right ears at each frequency. The logical advantage of the ISLD is that peaks and dips in the ISLD do not depend on the spectrum of the original source, and therefore they unambiguously encode information on directional filtering (Duda, 1997; Algazi *et al.*, 2001).

8.1. Method and results

Experiment 5 was designed to discover whether ISLD cues alone were adequate for front/back discrimination. In Experiment 5, the modified spectra in the right ear were flattened over all frequencies for both front and back sources. The modified amplitudes in the left ear were chosen to produce the baseline ISLD for front and back sources independently. Thereby, the ISLD was maintained perfectly while the spectra in the left and right ears were greatly changed.

Eight listeners (B, D, E, F, L, M, R, and Z) participated in Experiment 5, and their results are shown as circles in Fig. 9. The results show that, except for listener M, all the decisions were less than 75% correct. One-tailed t-tests show that scores for listeners B, D, L, R and Z were significantly below the 75%-threshold at the 0.05-level; for E, the difference was significant at the 0.1-level; and for listener F, the difference was not significant. Listeners B and Z always heard stimuli in one direction, either front or back, which led to a score of 50% with no standard deviation.

Squares in Fig. 9 show scores with baseline stimuli. Open squares are from baseline runs. For the three listeners who did not do complete baseline runs, baseline scores, plotted as solid squares on the figure, are based on the 80 baseline trials from the first ten continuous runs. When compared with baselines, the scores for all listeners in Experiment 5 were significantly worse (one-tailed t-tests at the 0.05-level), indicating that ISLD is not an adequate cue for front/back discrimination.

8.2. Discussion

Experiments 4 and 5 show that the monaural cues and the ISLD cues are not adequate for front/back discrimination. In headphone experiments that included front, back, and other locations in the median sagittal plane, Jin *et al.* (2004) similarly found that these cues are not sufficient. The headphone experiments by Jin *et al.* used broadband noise filtered by directional transfer functions (DTFs) derived from head-related transfer functions. The VRX experiments, with the advantages and limitations of transaural technology, do not lead to different conclusions. In general, Experiments 4 and 5 support the conclusions of Morimoto (2001), who found that both ears contribute to localization in the median sagittal plane. The idea that ISLD alone is not a sufficient cue also emerged from the azimuthal plane experiments of Hartmann and

Wittenberg (1996), who concluded that ISLD is not an adequate cue to provide externalization of sound images.

9. Experiment 6: sharpening

It is believed that the frequencies of peaks and dips in the amplitude spectrum give a listener the sensation of front or back locations (Shaw and Teranishi, 1968; Blauert, 1969/70; Hebrank and Wright, 1974b). It is less clear whether the heights and depths of these peaks and dips are important. For example, if a peak at a particular frequency happens to be a useful cue then increasing the height of that peak by some artifice could have two opposite effects. It might improve localization performance because the useful peak is now more prominent. Alternatively, it might degrade performance because the peak now has the wrong height.

Experiment 6 was designed to test these ideas. The modified spectra were obtained by starting with the baseline spectra and convolving them in the frequency domain with a normalized contrast enhancement function having the shape of a Mexican-hat (Table I). Equation (1) shows the formula for calculating all the modified spectra – both ears independently, both front and back sources. This algorithm sharpened the baseline spectra by increasing the level difference between the peaks and the dips (with negative elements in the enhancement function), and it smoothed the curves between adjacent components as well (with positive elements in the enhancement function). After modification, the level became L^+, given by

$$L^+(f_n)=\frac{1}{G_n}\left(\sum_{i=\mathrm{Max}(3,n-4)}^{\mathrm{Min}(250,n+4)} S_{i-n}L(f_i)\right) \quad (1)$$

where $L(f_i)$ is the level, in decibels, of the baseline stimulus component with frequency f_i, and the discrete function S_j is given in Table I. Also, S_{i-n} was set to zero if component i was eliminated in the calibration. The normalizing function G_n is simply the sum of the S_j values Therefore, G_n was always equal to 1.0 except when a component near n was eliminated. The width of the enhancement function S_j, namely 8 × 65.6 = 525 Hz, was chosen based on our observations of typical spacing of peaks and valleys in the baseline spectra and was intended to emphasize local structure. Figure 10 shows the baseline and modified syntheses for the front source as measured in the left ear canal of one listener for illustration. The figure shows that peaks and dips in the modified spectra occurred at the same frequencies as in the original spectra, but the level differences between the peaks and dips were magnified.

9.1. Experiment 6A - pseudo-tone

Experiment 6A used the pseudo-tone stimulus, as in the other experiments of this article. Eight listeners (B, D, G, L, P, R, V, and Z) were in Experiment 6A. The scores are given in Fig. 11 (a), which simply shows that performance was almost perfect for baseline stimuli, and that performance was not made worse by sharpening the contrasts in peaks and valleys.

It may be worth mentioning that an older listener (male, age 67) with sloping bilateral hearing loss above 4 kHz also participated in this experiment. His performance for baseline stimuli was much worse than all the other listeners, (75.0 ± 13.5) % correct. However, with the sharpened stimuli, his performance increased to (95.0±4.1) % correct. Nothing is proved by this experience with one listener, but the idea that sharpening the spectra can improve deficient front/back localization is anyhow an intriguing conjecture. For other listeners, a ceiling effect was evident in Experiment 6A. That motivated Experiment 6B using the Schroeder-phase periodic tone.

9.2. Experiment 6B - Schroeder-phase tone

Because of the ceiling effects observed in Experiment 6A, Experiment 6B used a periodic complex tone with Schroeder phases. As noted in Appendix A, some listeners localized unsuccessfully with that tone, and it was thought that sharpening the spectra might improve the performance.

The results of Experiment 6B are shown in Fig. 11(b). That figure shows essentially no effect of sharpening though some error bars are rather large. Listeners who were successful in the baseline experiments remained successful after sharpening. Listeners who were less successful remained that way as well.

The results of Experiments 6A and 6B are consistent with the idea that only the frequencies of the peaks and dips are important and the magnitudes are not important. This result agrees with Sabin *et al.* (2005), who found that increasing the contrast of the magnitude of DTF up to 4 times did not impair performance.

Zakarauskas and Cynader (1993) suggested that localization might be better predicted by an algorithm based on the first or second derivatives of level spectrum with respect to frequency, especially the second derivative. They constructed a computational algorithm and tested it against stimuli modified by HRTFs for different locations in the median sagittal plane. The connection with Experiment 6 is that higher-order derivatives show larger effects when the spectral differences are enhanced. Because the role of the magnitude of the spectral structure in their computational algorithm is unclear, it cannot be said that Experiment 6 necessarily argues against an algorithm based on derivatives. It does argue against a model in which the *size* of the derivatives plays a role.

10. Experiment 7: interaural time difference

It is widely believed that interaural time difference (ITD) cues are most important for localization in the azimuthal plane in an anechoic environment when low frequencies are present (Wightman and Kistler, 1992; Hartmann and Wittenberg, 1996; Macpherson and Middlebrooks, 2002). In the sagittal plane, the spectral cues are most important. Experiment 7 examined whether a stimulus ITD would affect front/back discrimination, i.e. whether the spectral cues for front and back sources are orthogonal to the ITD cues. Bloom (1977a, b) and Watkins (1978) claimed such a high degree of independence that spectral cues for elevation maintain their effectiveness even when the sound image is far off to one side due to monaural presentation.

10.1 Method and results

In Experiment 7, the modified spectra were achieved by advancing the right-ear baseline signal by a certain time shift. The advance (negative-delay) was accomplished by subtracting an extra phase that increased linearly with increasing frequency. The slope of the linear function determined the shift. Five values of interaural advance were used: 200, 400, 600, 800 and 1000 μs. The modified amplitude spectra were identical to baseline.

Five listeners (D, E, M, Z, and R) participated in Experiment 7. Listener R also did runs with 50 and 100 μs. Listeners perceived the sound images to be displaced to the right side. Their task was to discriminate front and back sources.

Results of Experiment 7 are shown in Fig. 12. Baseline results are shown on the figure at an ITD value of 0 μs. As seen in Fig. 12 there were large individual differences. Scores for listeners E and Z dropped below 75% at around 600 μs, which is close to the physiological range of the human head. However listener R's score dropped below 75% at 200 μs. By contrast, listeners

D and M scored above 75% even at 1000 μs, and listener D responded almost perfectly up to 1000 μs.

There were several common tendencies: All listeners successfully discriminated front from back with an ITD less than 200 μs, and performance tended to decrease as ITD increased. These tendencies suggest that spectral cues for front/back localization and ITD cues for horizontal localization are approximately independent for small ITD (less than about 500 μs) but not normally for larger ITD.

It is known that the spectral cues for elevation from the HRTFs are different for different azimuths (Algazi *et al.*, 2001). Listeners can be expected to apply their experience with these differing sets of cues depending on their knowledge of azimuth. Consequently, the stimuli of Experiment 7 presented conflicting cues in that spectral cues corresponding to zero azimuth were accompanied by ITD cues pointing to azimuths ranging over the entire space to the right of midline. Conflicting cues often lead to a diffuse image instead of a compact image and a lower externalization score. Therefore, it was of interest to measure the perceived externalization of the stimuli in Experiment 7.

10.2 Externalization

Externalization scores between 0 (inside head) and 3 (perfectly externalized) were recorded for listeners D, R and Z. Listener D always reported 3 for all conditions. Listener R reported scores above 2.8 for all conditions except for an ITD of 200 μs, where the score was about 2, i.e. she perceived a less externalized sound image for the smallest finite ITD, which is surprising. Listener Z reported scores above 2 for all conditions, except for front sources with ITDs of 200 and 400 μs. Similar to listener R, listener Z gave the surprising report that images were less externalized for small ITDs (200 and 400 μs) than for larger ITDs. Listeners D, R and Z all gave the perfect externalization score of 3 for the baseline stimulus, with zero ITD. In general, the externalization was good even with ITD of 1000 μs. The fact that inconsistency between ITD and spectral information did not lead to markedly reduced externalization may be further evidence of orthogonality between binaural and spectral cues.

10.3 Discussion

The ITD image displacement in Experiment 7 resembled one of the conditions in the headphone experiments of Macpherson and Sabin (2007): an ITD of 300 μs with spectral elevation cues taken from zero-azimuth sources. To measure ear dominance, the elevation cues were either identical in both ears or different. As in previous experiments (Humanski and Butler, 1988; Morimoto, 2001) Macpherson and Sabin found that significant displacement of an image to one side causes the spectrum in the ipsilateral ear to dominate the vertical plane localization almost entirely. However, the 300-μs displacement resulted in a large number of front/back confusions. In the experiment most similar to ours, where the two ears received the same amplitude cues and all sources were in front of the listener, the confusion rate was 34%. This high confusion rate agrees with the results of Experiment 7 in concluding that although some measures of vertical plane localization may be orthogonal to horizontal plane localization, this orthogonality holds less well with respect to front/back discrimination. The individual differences in Experiment 7 may reflect different interpretations of the conflicting cues. Listeners who make errors for large ITDs may attempt to track front/back location differences as they vanish geometrically when the perceived azimuth approaches 90 degrees. Listeners who do not make errors may use perceived timbral differences without geometrical interpretation.

11. Experiment 8: competing-sources

Experiment 1 in this article measured the ability of listeners to discriminate front from back using only high-frequency information or only low-frequency information. The experiment showed that different listeners required different amounts of information to discriminate. The listeners were categorized into V-, X- and A-shape listeners. V-shape listeners require the most information; A-shape listeners require the least. Experiment 8 was a test to determine whether the individual differences and our interpretation of these differences would continue to hold good under very different experimental circumstances.

11.1. Methods

Experiment 8 used only two loudspeakers, the front and the back. It did not use the VRX technique at all except that the stimulus was the pseudo-tone. On every experimental trial both loudspeakers sounded simultaneously. On a Type-1 trial, the front speaker played the low-frequency components and the back speaker played the high. A Type-2 trial was the reverse. The boundary between high- and low-frequency regions, called the “boundary frequency,” varied in 2-kHz steps from 2 to 16 kHz. The listener’s task was to decide whether the sound came from the front or the back.

There were 20 trials in each run, 10 of each type, presented in random order. Six listeners (R, Z, F, B, L, and D) participated in this experiment, and their results are shown as solid circles in Fig. 13. Each listener did four runs, and hence each data point is an average over 80 trials. The vertical axis on the figure is the percentage of trials on which the listener’s response agreed with the source assigned to the *low* frequencies.

11.2. Results and discussion

The critical question in Experiment 8 is whether the results can be interpreted in terms of Experiment 1. Therefore, the results from Experiment 1 are also plotted in Fig. 13. In Experiment 1A, high-frequency cues were flattened, and hence only low-frequency cues existed. Performance in Experiment 1A was a measure of a listener’s use of low-frequency information, and the percentage of correct responses in Experiment 1A, as shown in Fig. 5, was copied directly to Fig. 13 and plotted as squares. In Experiment 1B, low-frequency cues were flattened, and only high-frequency cues existed. Thus, performance in Experiment 1B as shown in Fig. 5 was a measure of a listener’s use of high-frequency information, which is opposite to the measure plotted in Fig. 13. Therefore the percentage scores in Experiment 1B were subtracted from 100%, and that difference was plotted as triangles in Fig. 13.

In Experiment 8, every listener’s low-frequency tracking score increased from 0% to 100% as the boundary frequency increased over the complete frequency range. This common feature is expected because as boundary frequency increased, more front/back information was played through the low-frequency speaker.

The comparison with Experiment 1 is easiest for Listener R, a V-shape listener who required a great deal of information to discriminate front from back. Therefore, as the boundary frequency varied in Experiment 8, only one speaker ever presented adequate information for this listener, and sometimes no speaker presented adequate information. There was never any contradiction. Therefore, it is expected that the listener’s choice in Experiment 8 should agree with the listener’s capabilities, as measured in Experiment 1. That agreement was actually observed, as shown in Fig. 13.

Listener Z was an X-shape listener, but nearly V-shape. The agreement between Experiment 8 and Experiment 1 resembles the agreement seen for Listener R as expected. Listener F was

another X-shape listener, and the choices made in Experiment 8 also resemble the average of the capabilities measured in Experiment 1.

Listener B was an X-shape listener whose performance in Experiment 1A showed that she could make some use of low-frequency information, but she was never very good at it. Therefore, it is not surprising to find that her choices in Experiment 8 favored the high-frequency source even for rather high boundary frequencies.

Listeners L and D were A-shape listeners. They required very little information to discriminate front from back. In Experiment 8, boundary frequencies in the middle of the range caused both speakers to present adequate front/back cues. Thus, unlike a V-shape listener, these A-shape listeners received contradictory information. It is hard to predict what choices these listeners might make in Experiment 8, and Fig. 13 shows that they responded to the ambiguity in very different ways. The choices for Listener L resemble an average of high- and low-frequency capabilities, to the extent that these can be determined, but the error bars are huge. This listener was obviously very confused by the superabundance of contradictory information. Listener D apparently decided to track the high-frequency source on all the trials until the boundary frequency became so high that there was hardly any power in the high-frequency speaker.

In summary, for V-shape and some X-shape listeners, the choices made in Experiment 8 agree with expectations based on Experiment 1, because there is minimal competition among adequate cues. For A-shape listeners and other X-shape listeners, both low-frequency and high-frequency bands include adequate cues, resulting in competition. It is hard to predict how these listeners will respond in Experiment 8, and the value of the comparison can only be judged by the plausibility of post-hoc explanations. The agreements and reasonable explanations in this section suggest that listeners retain their individual front/back decision strategies and capabilities under quite different stimulus conditions.

12. Conclusion

The extreme virtual reality experiments (VRX) reported this article simulated external complex sound sources using a transaural synthesis in an anechoic room with two synthesis loudspeakers. Simulation was good up to 16 kHz by objective measures. Subjectively, sound images for baseline stimuli were perceived to be well externalized, and listeners could not discriminate between real and virtual signals.

The VRX technique allows an experimenter to present a signal with a fixed spectrum to a listeners' eardrums, while giving the listener an opportunity to use personal pinna cues. Given the stimulus uncertainties in virtual reality experiments that use headphones, as exposed by Kulkarni and Colburn (2000), the VRX technique is expected to be more accurate than any method using headphones and head-related transfer functions. Extreme accuracy is especially important in connection with the delicate matter of front/back discrimination because of the potential importance of short-wavelength signal components.

The VRX method is less flexible than virtual reality methods that use HRTFs. In principle, if the HRTFs are known, an experimenter can present a virtual image of any sound, speech or music for instance, by convolving the head-related impulse response with the sound waveform. Additionally, one can apply head-tracking methods and interpolate among impulse responses as the listener moves his head. By contrast, the VRX method specializes its calculations to a particular stimulus.

Seven VRX front/back discrimination experiments were performed to discover the importance of various front/back cues. There were large individual differences among listeners, suggesting that different listeners have developed quite different strategies for localizing front and back

sound sources. The large individual differences in performance may be due to large individual differences in the directional transfer functions, which were found to be highly correlated with geometric properties of listeners' ears and heads (Middlebrooks, 1999).

Experiment 1 flattened the spectra in high or low frequency ranges in an attempt to find subbands that are either necessary or adequate for front/back discrimination. After interesting bands were found for each listener, Experiment 2 flattened the spectra inside and outside those bands. The conclusion from these two experiments was that most listeners can use a variety of strategies, employing comparisons among multiple bands to distinguish front from back.

It is important to distinguish the flattening experiments, which eliminate the spectral structure in selected bands, from filtering experiments, which eliminate the energy in selected bands. The flattening experiments avoid some of the artifacts of the filtering experiments, but the flattening experiments preserve the power averaged over the affected range. Therefore, if a listener's localization strategy requires a comparison involving the level over a wide band, the same for front and back sources, then that strategy could be available in a flattening experiment but not in a filtering experiment. Flattening experiments do not make filtering experiments irrelevant. The two kinds of experiments give different information.

Experiment 3 compared the roles of peaks and dips in the spectrum and found that dips were more important than peaks for accurate front/back localization. Experiments 4 and 5 showed that monaural cues alone and interaural spectral level difference cues alone were not sufficient for correct front/back judgement for most of our listeners.

Experiment 6 presented sharpened spectra, with enhanced contrasts between peaks and dips. The experiments showed that sharpening had no effect on the ability to discriminate front from back for normal hearing listeners. This result is consistent with a directional band concept, such as Blauert's (1983), in which characteristic frequencies of peaks and dips point to locations in the vertical plane. According to this concept the magnitudes of the peaks and valleys play no particular role, though logically it is clear that the magnitudes must exceed some threshold in order to be effective. It is interesting to conjecture that enhancing the contrast between peaks and valleys might enhance the localization information contained in spectral features for some listeners, as suggested by an informal experiment with one listener with a modest sloping hearing loss.

According to the results of Experiment 7, applying an interaural time difference up to 200 μs did not ruin listeners' front/back judgement, although the sound image was displaced to one side. As the ITD increased beyond 200 μs, front/back discrimination decreased for most of the listeners though there were important individual differences. Most listeners could not follow front/back cues with an ITD greater than 800 μs.

The spectral manipulations in the seven VRX experiments led to large degradations in the ability of listeners to distinguish front from back. There were enormous individual differences, especially in the ability to make successful use of information in selected frequency bands. Among normal hearing listeners who were perfectly able to distinguish front from back given the baseline stimulus, some required a wide bandwidth while others required a much smaller bandwidth. For some, either high-frequency information alone or low-frequency information alone was sufficient, but others always required high-frequency information.

The ability to distinguish front from back based on information in different frequency bands was tested in Experiment 8. This experiment was not a VRX experiment but used real sources, one in front and one in back, both sounding simultaneously. Low frequencies were produced by the front source and high frequencies were produced by the back source, or vice versa. The boundary between low and high was an experimental variable.

The value of Experiment 8 was that it enabled the conclusions about specific listeners drawn from Experiments 1 and 2 to be tested in very different stimulus conditions. Those listeners who required a great deal of information to discriminate front from back were not subjected to any contradiction in Experiment 8, though they sometimes received inadequate information. Their decisions, as a function of the low-high boundary, were predictable from the VRX results in Experiment 1. Those listeners who required little information to discriminate, were subjected to contradictions because they were presented with adequate information from both front and back sources simultaneously. Those listeners made unpredictable and highly individualistic decisions in Experiment 8. The results for those listeners could be given plausible interpretations after the fact. The comparison of the decisions made in Experiment 8 with individual listener abilities as determined by flattening cues in Experiment 1, suggests that the information learned from the VRX flattening procedure holds good in other contexts.

Acknowledgments

Dr. Brad Rakerd provided important technical help in early stages of this work and suggested Experiment 8. We are also grateful to Dr. John Middlebrooks, and Dr. Ewan Macpherson for helpful discussions. This work was supported by the NIDCD grant DC-00181.

LIST OF ABBREVIATIONS

DTF	directional transfer function
HRTF	head-related transfer function
ISLD	interaural spectral level difference
ITD	interaural time difference
KEMAR	Knowles Electronics Manikin for Acoustic Research
RMS	root mean square
SPL	sound pressure level
VRX	extreme virtual reality

References

Algazi VR, Avendano C, Duda RO. Elevation localization and head-related transfer function analysis at low frequencies. J Acoust Soc Am 2001;109:1110–1122. [PubMed: 11303925]

Asano F, Suzuki Y, Sone T. Role of spectral cues in median plane localization. J Acoust Soc Am 1990;88:159–168. [PubMed: 2380444]

Blauert J. Sound localization in the median plane. Acustica 196970;22:205–213.

Blauert, J. Spatial hearing: the psychophysics of human sound localization. MIT Press; Cambridge MA: 1983.

Bloom PJ. Determination of monaural sensitivity changes due to the pinna by use of minimum-audible field measurements in the lateral vertical plane. J Acoust Soc Am 1977a;61:820–828. [PubMed: 853154]

Bloom PJ. Creating source elevation illusions by spectral manipulation. J Audio Engr Soc 1977b;25:560–565.

Cooper DH, Bauck JL. Prospects for transaural recording. J Audio Engr Soc 1989;37:3–19.

Hartmann WM, Rakerd B. Auditory spectral discrimination and the localization of clicks in the sagittal plane. J Acoust Soc Am 1993;94:2083–2092. [PubMed: 8227750]

Hartmann WM, Wittenberg A. On the externalization of sound images. J Acoust Soc Am 1996;99:3678–3688. [PubMed: 8655799]

Hartmann WM, Rakerd B, Koller A. Binaural coherence in rooms. Acta Acustica United with Acustica 2005;91:451–462.

Hebrank J, Wright D. Are two ears necessary for localization of sound sources on the median plane? J Acoust Soc Am 1974a;56:935–938. [PubMed: 4424970]

Hebrank J, Wright D. Spectral cues used in the localization of sound sources on the median plane. J Acoust Soc Am 1974b;56:1829–1834. [PubMed: 4443482]

Hofman PM, Van Opstal AJ. Binaural weighting of pinna cues in human sound localization. Exp Brain Res 2003;148:458–470. [PubMed: 12582829]

Hofman PM, Van Riswick JGA, Van Opstal AJ. Relearning sound localization with new ears. Nature Neuroscience 1998;1:417–421.

Humanski RA, Butler RA. The contribution of the near and far ear towards localization of sound in the median plane. J Acoust Soc Am 1988;83:2300–2310. [PubMed: 3411022]

Jin C, Corderoy A, Carlile S, Schaik A. Contrasting monaural and interaural spectral cues for human sound localization. J Acoust Soc Am 2004;115:3124–3141. [PubMed: 15237837]

Kulkarni A, Isabelle SK, Colburn HS. Sensitivity of human subjects to head-related transfer-function phase spectra. J Acoust Soc Am 1999;105:2821–2840. [PubMed: 10335633]

Kulkarni A, Colburn HS. Variability in the characterization of the headphone transfer-function. J Acoust Soc Am 2000;107:1071–1074. [PubMed: 10687721]

Langendijk EHA, Bronkhorst AW. Contribution of spectral cues to human sound localization. J Acoust Soc Am 2002;112:1583–1596. [PubMed: 12398464]

Macpherson EA, Middlebrooks JC. Sound localization illusions produced by source spectrum discontinuities. Assoc Res Otolaryngology Abstracts. 1999

Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. J Acoust Soc Am 2002;111:2219–2236. [PubMed: 12051442]

Macpherson EA, Middlebrooks JC. Vertical-plane sound localization probed with ripple-spectrum noise. J Acoust Soc 2003;114:430–445.

Macpherson EA, Sabin AT. Binaural weighting of monaural spectral cues for sound localization. J Acoust Soc Am 2007;121:3677–3688. [PubMed: 17552719]

Martin RL, Paterson M, McAnally KI. Utility of monaural spectral cues is enhanced in the presence of cues to sound-source lateral angle. J Assn Res Otolarygol 2004;5:80–89.

Mellert, V. op cit. 1971. Directional hearing in the median plane and diffraction of sound around the head, thesis, Gottingen, cited by Blauert (1983).

Middlebrooks JC. Narrow-band sound localization related to external ear acoustics. J Acoust Soc Am 1992;92:2607–2624. [PubMed: 1479124]

Middlebrooks JC. Individual differences in external-ear transfer functions reduced by scaling in frequency. J Acoust Soc Am 1999;106:1480–1492. [PubMed: 10489705]

Middlebrooks JC, Green DM. Directional dependence of interaural envelope delays. J Acoust Soc Am 1990;87:2149–2162. [PubMed: 2348020]

Morimoto M. The contribution of two ears to the perception of vertical angle in sagittal planes. J Acoust Soc Am 2001;109:1596–1603. [PubMed: 11325130]

Morimoto M, Ando Y. On the simulation of sound localization. J Acoust Soc Japan (E) 1980;1:167–174.

Musicant AD, Butler RA. The influence of pinnae-based spectral cues on sound localization. J Acoust Soc Am 1984;75:1195–1200. [PubMed: 6725769]

Nelken I, Young ED. Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrow-band and wide-band stimuli. J Neurophysiol 1994;71:2446–2462. [PubMed: 7931527]

Oldfield SR, Parker SPA. Acuity of sound localization: a topography of auditory space II. Pinna cues absent Perception 1984;13:601–619.

Parsons JE, Lim E, Voigt HF. Type III units in the gerbil dorsal cochlear nucleus may be spectral notch detectors. Annals of Biomedical Engineering 2001;29:887–896. [PubMed: 11764319]

Sabin AT, Macpherson EA, Middlebrooks JC. Vertical-plane localization of sounds with distorted spectral cues. Assoc Res Otolaryngology Abstracts. 2005

Schroeder MR. Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. IEEE Trans on Information Theory 1970;IT-16:85–89.

Schroeder MR, Atal BS. Computer simulation of sound transmission in rooms. IEEE Intl Conv Rec 1963;11:150–155.

Shaw EAG, Teranishi R. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. J Acoust Soc Am 1968;44:240–249. [PubMed: 5659838]

Smith BK, Sieben UK, Kohlrausch A, Schroeder MR. Phase effects in masking related to dispersion in the inner ear. J Acoust Soc Am 1986;80:1631–1637. [PubMed: 3794068]

Vliegen J, Van Opstal AJ. The influence of duration and level on human sound localization. J Acoust Soc Am 2004;115:1705–1713. [PubMed: 15101649]

Watkins AJ. Psychoacoustical aspects of synthesized vertical locale cues. J Acoust Soc Am 1978;63:1152–1165. [PubMed: 649874]

Wightman FL, Kistler DJ. Headphone simulation of free-field listening. I: stimulus synthesis. J Acoust Soc Am 1989a;85:858–867. [PubMed: 2926000]

Wightman FL, Kistler DJ. Headphone simulation of free-field listening. II: psychophysical validation. J Acoust Soc Am 1989b;85:868–878. [PubMed: 2926001]

Wightman FL, Kistler DJ. Sound localization with unilaterally degraded spectral cues. J Acoust Soc Am (abst) 1990;105:1162.

Wightman FL, Kistler DJ. The dominant role of low-frequency interaural time differences in sound localization. J Acoust Soc Am 1992;91:1648–1661. [PubMed: 1564201]

Wightman, FL.; Kistler, DJ. Factors affecting the relative salience of sound localization cues. In: Gilkey, RH.; Anderson, T., editors. Binaural and Spatial Hearing in Real and Virtual Environments. Erlbaum; Mahwah, NJ: 1997. p. 1-24.

Zahorik P, Bangayan P, Sundareswaran V, Wang K, Tam C. Perceptual recalibration in human sound localization: learning to remediate front-back reversals. J Acoust Soc Am 2006;120:343–359. [PubMed: 16875231]

Zakarauskas P, Cynader MS. A computational theory of spectral cue localization. J Acoust Soc Am 1993;94:1323–1331.

Zhang, PX. Detection and localization of virtual tones and virtual images. Michigan State University; East Lansing, MI: 2006. Ph.D. thesis

Appendix A. Preliminary experiment: noise, tones, and pseudo-tones

To be sure that listeners in VRX experiments could discriminate between real front and back sound sources, a preliminary discrimination experiment was performed for each listener. The setup was similar to Fig. 1, but with only front and back source speakers. The first stimulus tested was a complex, periodic tone having a fundamental frequency of 65.6 Hz and all harmonics from the 3rd to the 250th in Schroeder-minus phase relationship. The level was 70 dB SPL at the listener's ears. Each experimental run contained 80 trials, 40 trials from the front source speaker and 40 trials from the back source speaker in a random order. The listener's task was to press buttons indicating whether the sound came from front or back. In addition to ten of the listeners from the VRX experiments (listeners D, E, F, G, L, M, P, R, V, and Z) there were eight other listeners, females (C, H, J, S, and Y) and males (K, N, and Q), who did not continue with the VRX experiments. These other listeners were in their 20s except, H (65), K (54), and Q (37). Each listener did two runs. Figure 14 (squares) shows the results, expressed as the percentage of correct responses averaged over the two runs. The error-bars are two standard deviations in overall width. We expected that this would be a very easy task. Surprisingly, most listeners found that it was not easy to do, and some listeners even felt it was rather difficult, which was confirmed by the low percent correct in Fig. 14 with a mean score of 78.0% across listeners.

The second stimulus was white noise. Everything else about the experiment was the same. With white noise, most listeners found the task to be very easy, and the percentages of correct

responses, shown as circles in Fig. 14, were also much higher with a mean score of 96.5% across listeners. White noise would have been an excellent stimulus for VRX experiments, but the large number of components would have imposed a considerable computational load on the real-time simulations.

Finally, a new signal, the "pseudo-tone," was generated by offsetting the frequency of each harmonic of the Schroeder-phase periodic tone by a random value within a range of ±15 Hz. The pseudo-tone has the same number of components as the Schroeder-phase periodic tone, and has the same overall frequency distribution. However, the random frequency displacement of each component makes the pseudo-tone rather noisy. Thus, the pseudo-tone is subjectively somewhere between the Schroeder-phase periodic tone and noise. The scores on the front/back discrimination task also fell somewhere between scores for the periodic tone and the noise. The results (Fig. 14 triangles) show that most listeners succeeded in this task (91.3% correct averaged across listeners).

Subjectively, most listeners said that localization was more difficult for the Schroeder-phase periodic tone compared to the other two signals. Listeners E, L, P, and Z agreed that the Schroeder-phase periodic tone was much more difficult even though their scores for that tone were also high. There was one exception: Listener K found the pseudo-tone more difficult to localize than the periodic tone.

The diffculty with the Schroeder-phase periodic tone might have been anticipated given the experience of Hofman and Van Opstal (2003). The frequency of 65.6 Hz is in the range where they found strikingly deficient elevation perception. According to their interpretation, elevation perception requires that excitation be present in many frequency channels simultaneously (5 ms or less temporal disparity) but the excitation from a low-frequency periodic tone with Schroeder phases is only present in these channels sequentially. Because the excitation in any channel is impulsive one might expect that Schroeder phase tones will show a negative level effect (Hartmann and Rakerd, 1993; Macpherson and Middlebrooks, 2000; Vliegen and Van Opstal, 2004).

Because most listeners performed better with the pseudo-tone in the preliminary experiment, the pseudo-tone was used instead of the periodic tone. The frequency set of the pseudo-tone was frozen. It was generated once and was used throughout the experiments.

Appendix B. Accuracy of simulation at the ear-drums

The transaural technique controlled the signal being played through the synthesis loudspeakers so that the recorded spectra at the probe-microphone tips in the listener's ear-canals were the same as the recorded spectra for the real source. Of course, the goal of the technique is that the synthesized signal and the real-source signal should be the same at the ear-drums – not at the probe-tips. Further, in a real ear-canal the incident sound wave and the sound wave reflected by the ear-drum establish standing waves, which may make the probe-microphone recording very sensitive to the position of the probe-tips. However, the VRX technique tends to be self-compensating against such changes in position.

Self-compensation was demonstrated by an experiment on the artificial ear-canals of a KEMAR head, terminated by Etymotic ER-11 microphones. The ER-11 microphones will be called "KEMAR microphones" in the following text to distinguish them from the probe-microphones. During the test, the KEMAR was placed in the anechoic room, and the probe-microphones were inserted, as in the VRX experiments with human subjects. The probe-microphones were inserted to within 1 mm of the KEMAR microphone. A complete calibration sequence was performed, and recordings were made through both the probe-microphones and the KEMAR microphones. Then the probes were pulled out in several increments of about 1.5

mm, and spectra were recorded for the new probe-tip locations, with new calibrations at each stage. The recorded spectra in the right ear are shown in Fig. 15. Part (a) shows that the three recordings at the probe-tip were quite different from one another. However, when presented with the simulation that had been calculated by the VRX method based on those three different recordings, the KEMAR microphone (analogous to human ear-drum) recorded very similar spectra, as shown in part (b). For example, the figure shows that at 14 kHz, the levels recorded at different tip-positions could differ by as much as 9 dB, while the levels recorded at the KEMAR microphone differed by only 1 dB. Thus, the self-compensation of the VRX technique led to 8 dB of compensation, and the technique provided a stable simulation at the listeners' ear-drums.

Although the VRX technique is rather insensitive to variations in probe-microphone location, nevertheless, in any one probe-tip position, the tip may coincide with a node for a particular frequency. Then the recorded amplitude is low for that frequency, and that can lead to a large error in the simulation. The VRX technique selectively eliminated such frequency components by setting their amplitudes to zero. Apart from nodes, the simulation is stable against changes in probe-tip positions because the VRX calibration sequence compensates for different spectra at the probe-tips. This is an advantage of a transaural technique over the use of head-related transfer functions.

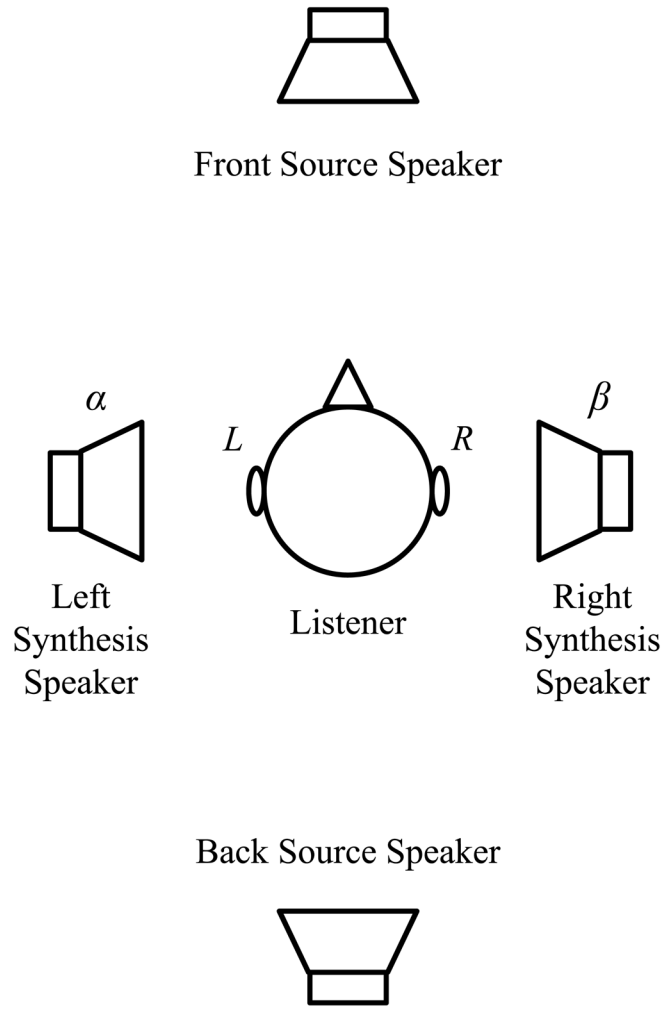

Figure 1.
Setup of loudspeakers in the anechoic room with real sources 150 cm from the listener in front (F) and in back (B) and synthesis speakers α, and β to the sides, each 37 cm from the near ear.

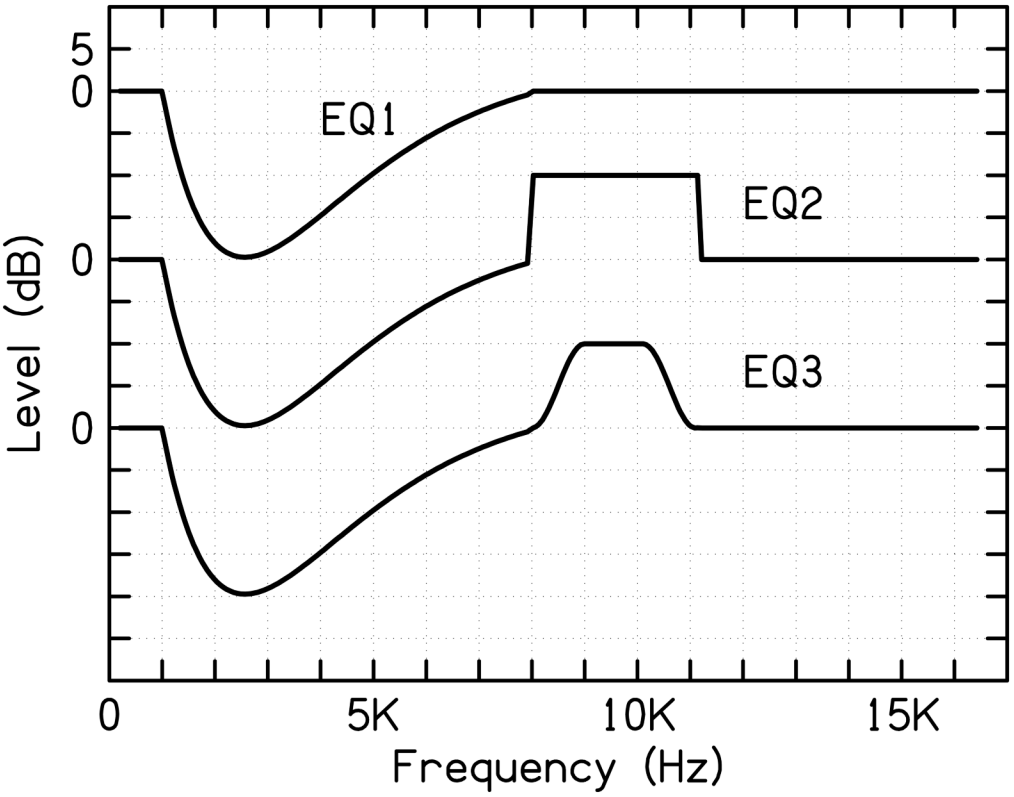

Figure 2.
Three equalizations, used for the signal sent to the real-source loudspeakers, optimizing the crest factor for various listeners and conditions [Footnote 1].

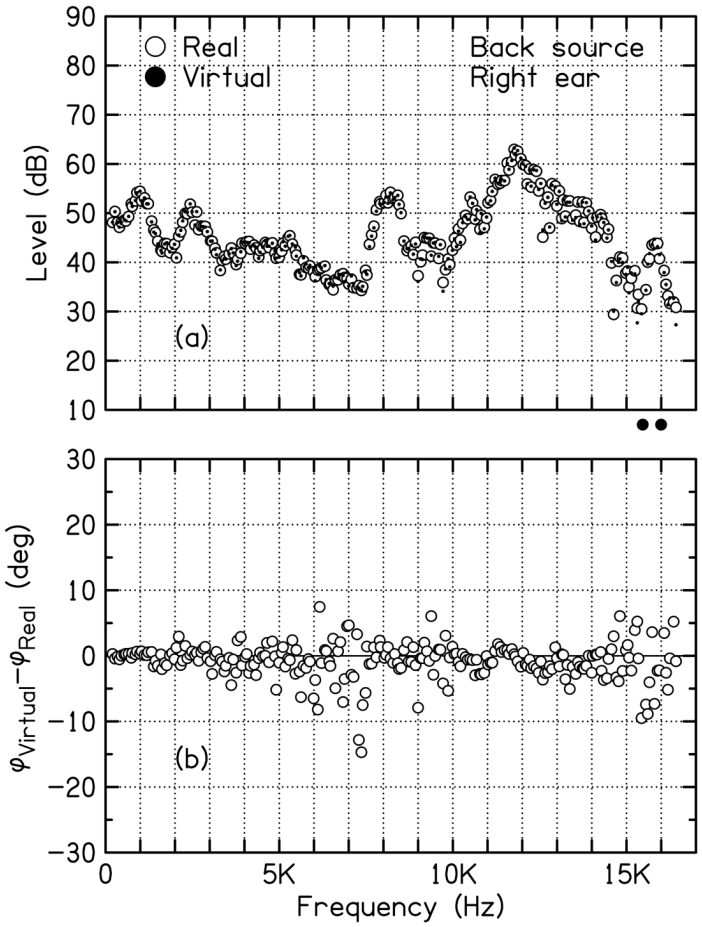

Figure 3.

Typical baseline simulation for the back source as measured in the right ear. (a) The amplitude spectrum for the real source is compared with the virtual source (simulation). Two points below the plot show components that did not meet the ±50% criterion and were eliminated. (b) The spectrum of phase differences between recordings of real and virtual signals.

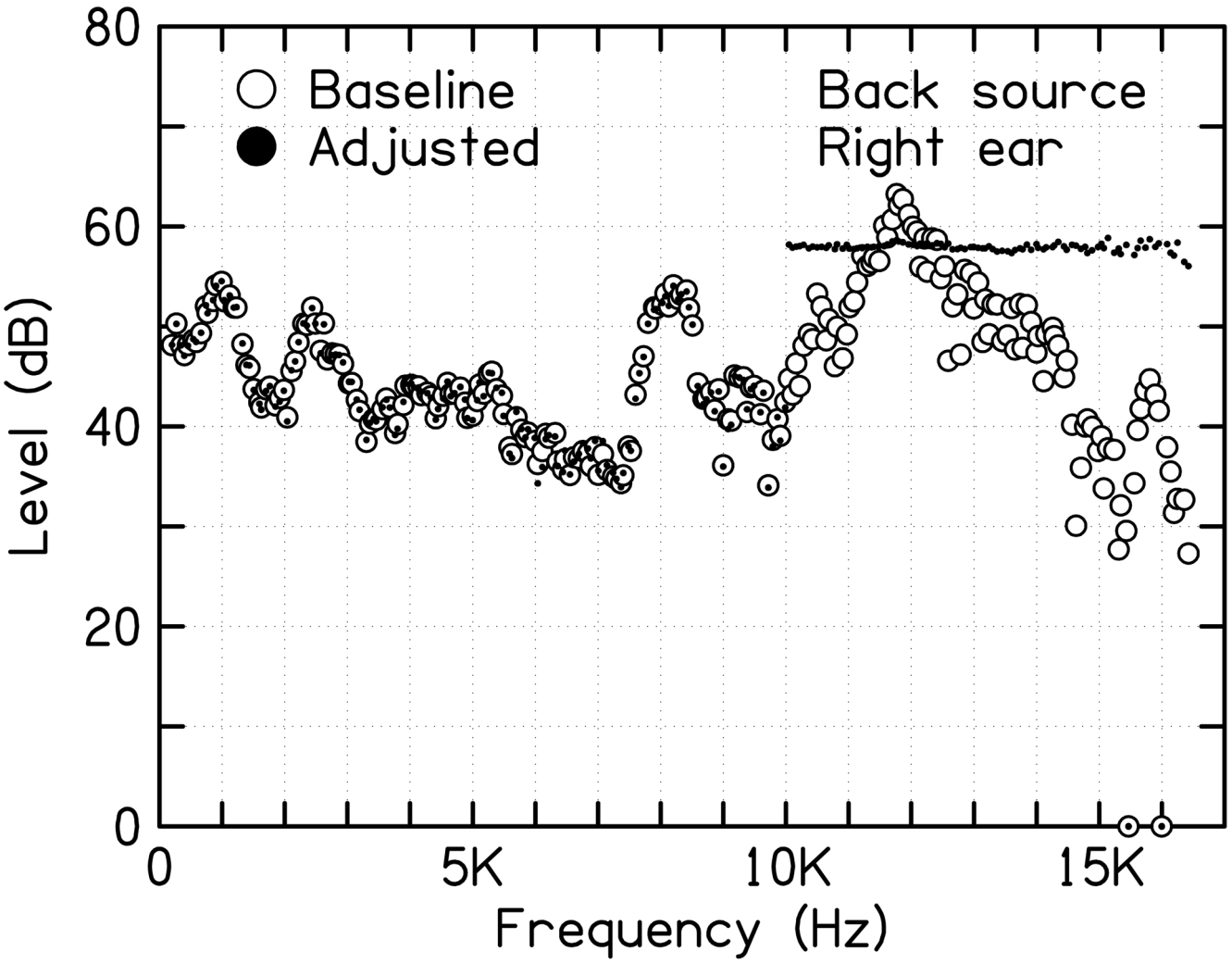

Figure 4.

Experiment 1: Typical simulation for the back source as measured in the right ear. Open circles show the baseline amplitude spectrum. Filled symbols show the modified amplitude spectrum, flattened above 10 kHz. Two components, shown by points on the horizontal axis, were eliminated in the calibration process for the front or the back source.

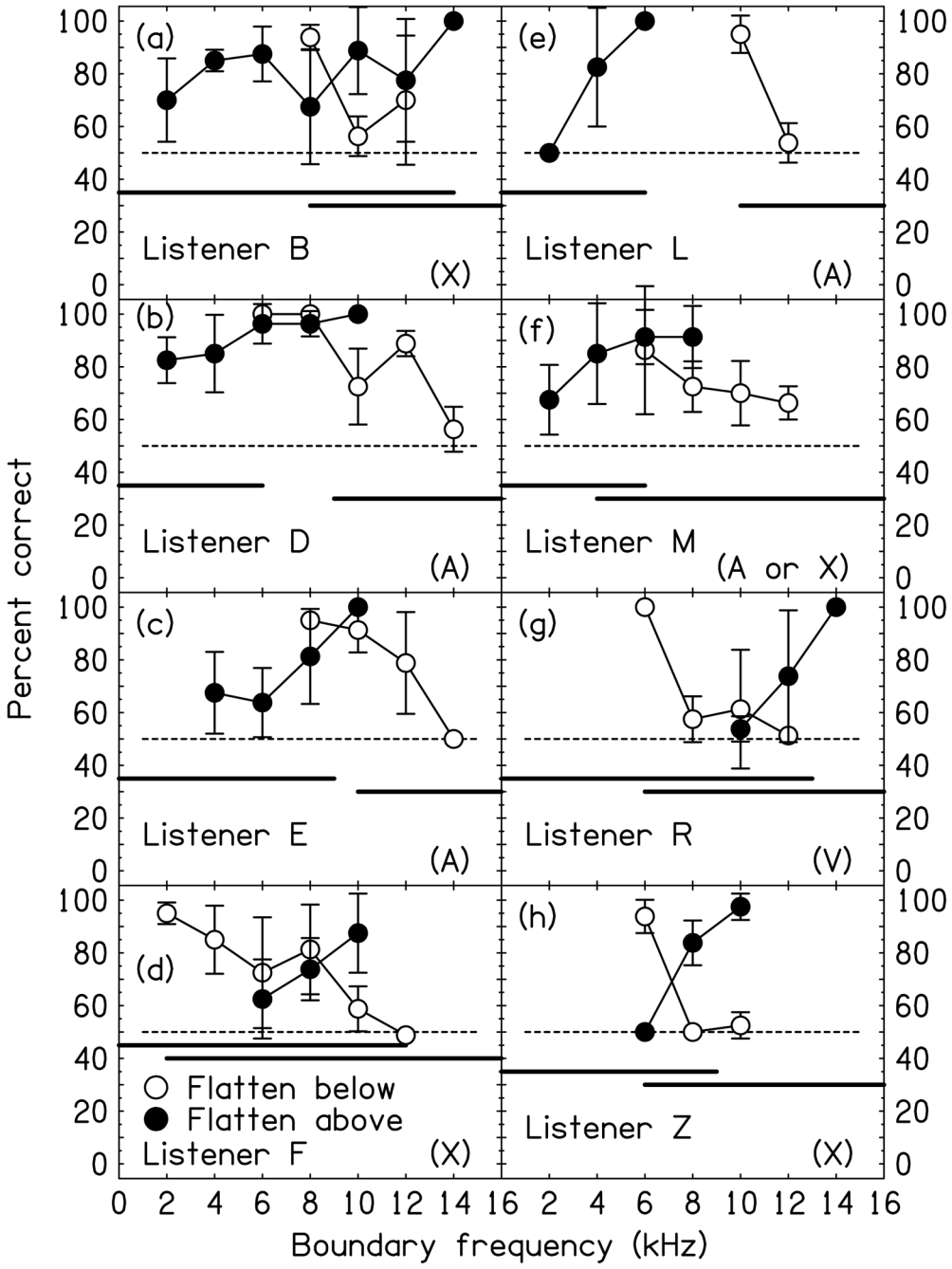

Figure 5.

Experiment 1: Percentage of correct responses for eight listeners with flattened amplitude spectra above (solid symbols) and below (open symbols) the boundary frequency. Listener response types are characterized as A, V, or X. Heavy horizontal lines indicate bands that are adequate for good discrimination. The dashed horizontal line at 50% correct is the random-guessing limit. Error bars are two standard deviations in overall length.

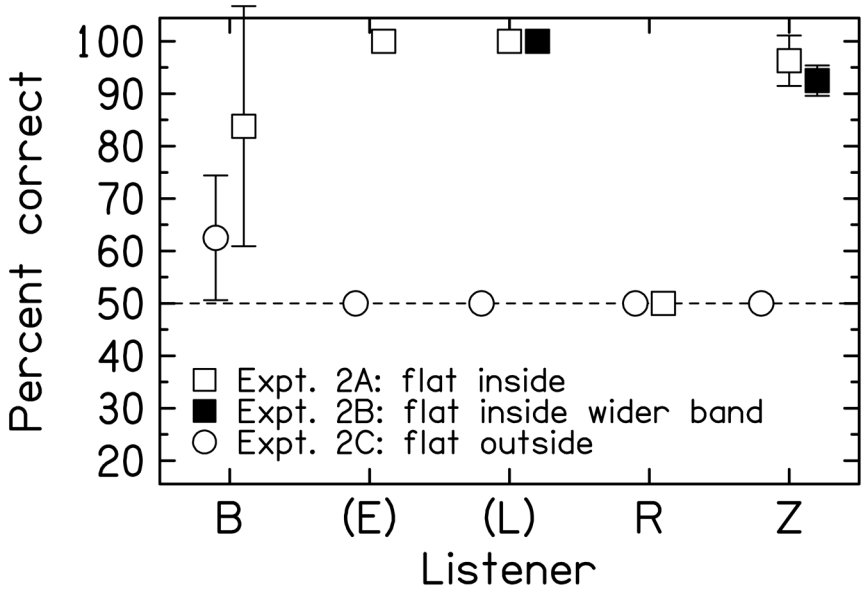

Figure 6.
Experiment 2: Percentage of correct responses for five listeners with flattened amplitude spectra inside or outside a central frequency region. Parentheses for listeners E and L indicate that the experiment was not designed for them. Error bars are two standard deviations in overall length.

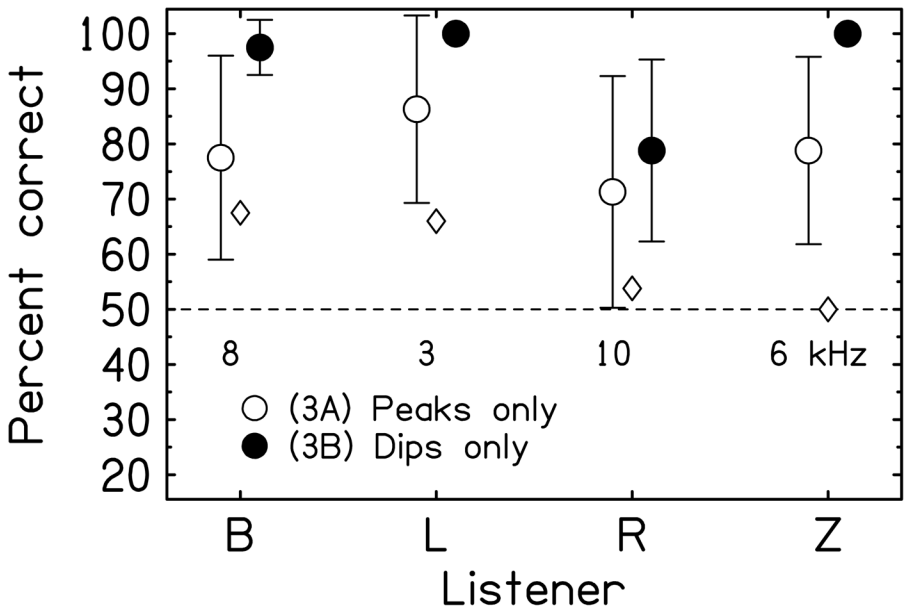

Figure 7.
Experiment 3: Evaluating the importance of peaks and dips. Above a boundary frequency, shown below the dashed line, the spectral dips were flattened in Experiment 3A and the peaks were flattened in Experiment 3B. Diamonds indicate performance in the *flattening-above* experiment (Experiment 1A) for the same boundary frequency. Error bars are two standard deviations in overall length.

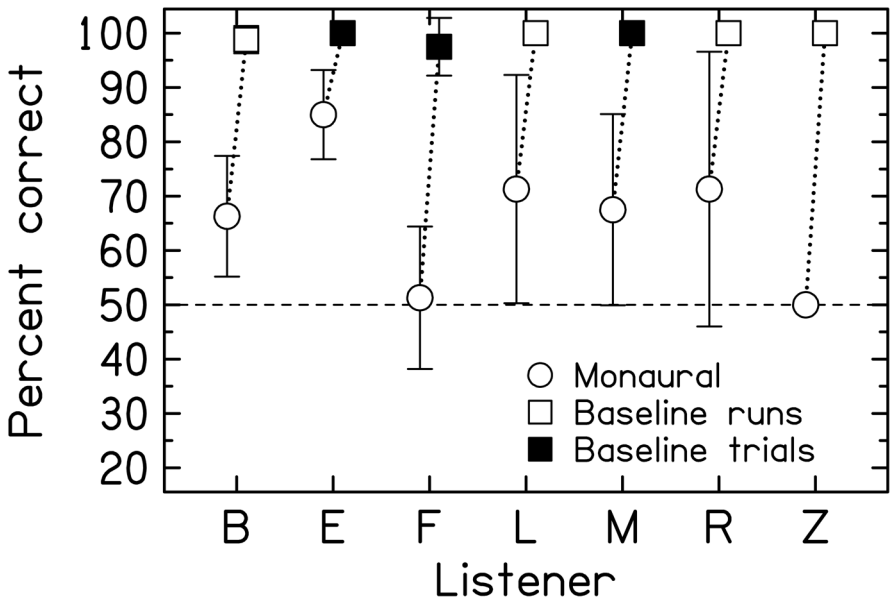

Figure 8.

Experiment 4: Monaural information. Circles show the performance for seven listeners when the amplitude spectrum in the right ear was flattened by setting all amplitudes to the RMS value, averaged over all frequencies. Squares show baseline performance when both ears obtained accurate information. Baseline performance is expected to be perfect. Error bars are two standard deviations in overall length.

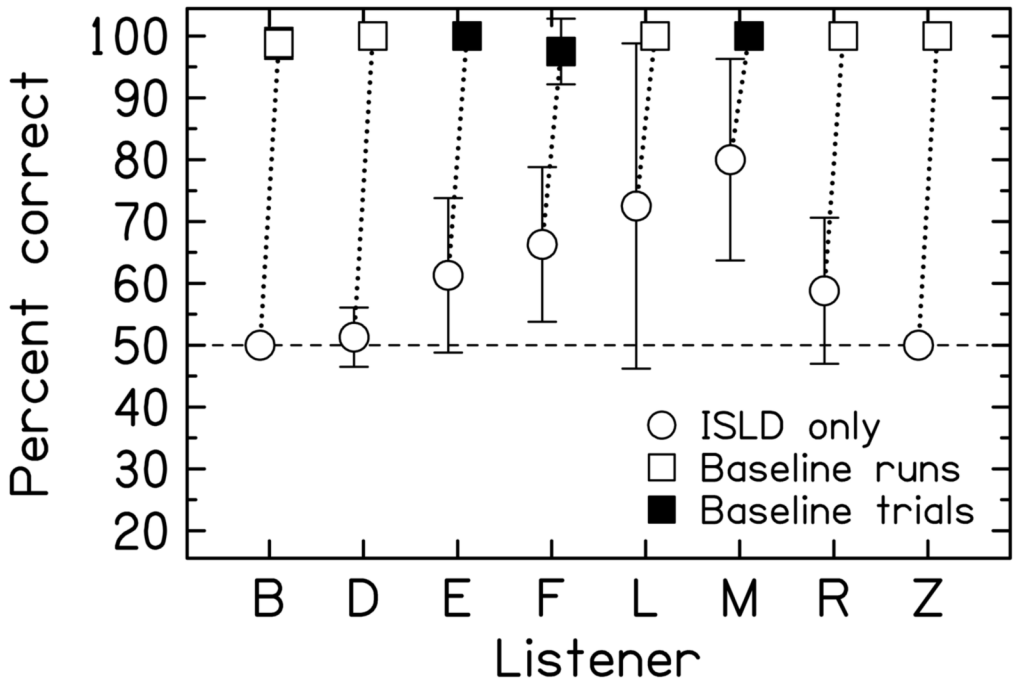

Figure 9.

Experiment 5: Interaural spectral level difference only. Circles show the performance for eight listeners when the amplitude spectrum in the right ear was flattened and the amplitudes in the left ear were modified so as to perfectly maintain the interaural level difference at each frequency. Squares show baseline performance when both ears obtained accurate information. Error bars are two standard deviations in overall length.

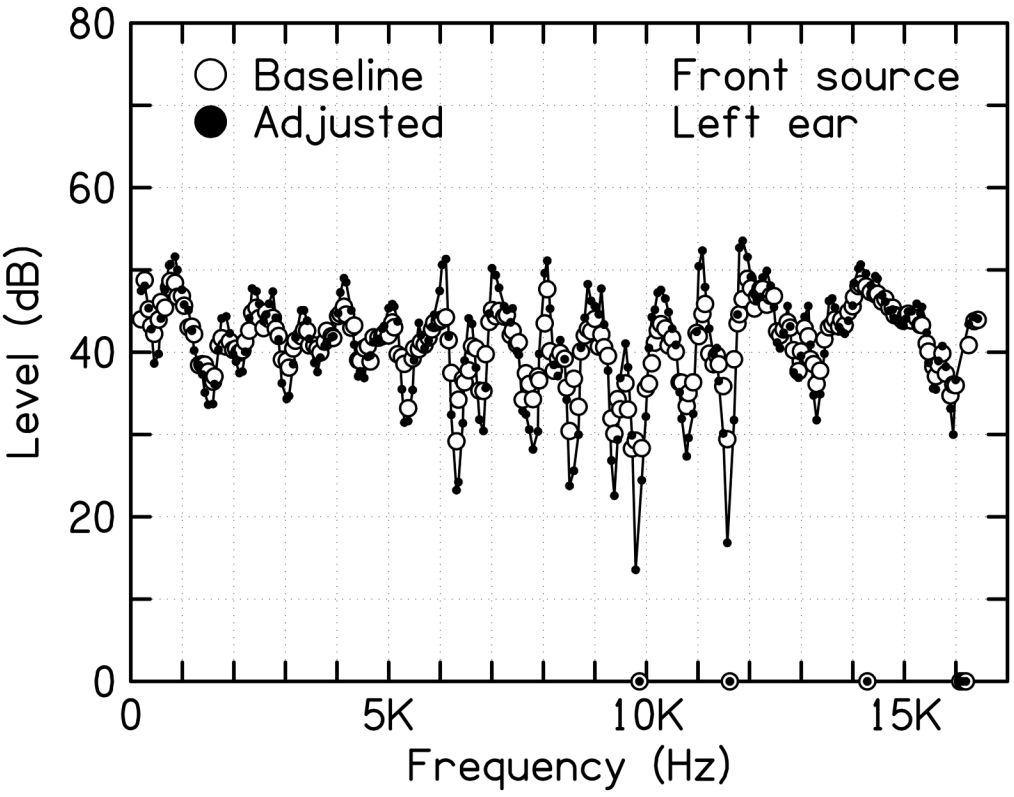

Figure 10.
Experiment 6: Example of baseline and modified spectra for the sharpening experiment 6A, where spectral contrasts are enhanced.

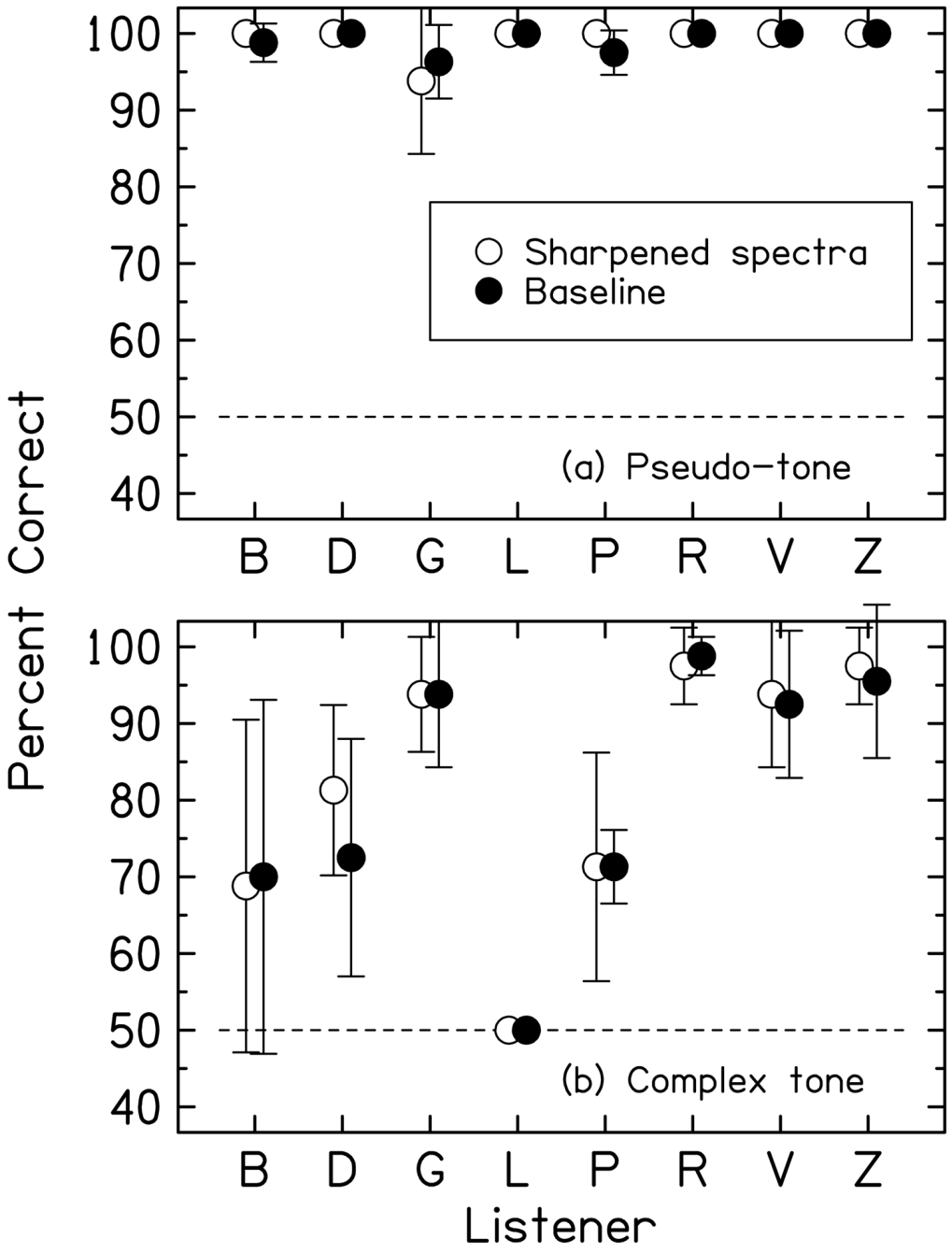

Figure 11.
Experiment 6: Performance by eight listeners in the sharpening experiment for (a) the pseudo-tone and (b) the Schroeder-phase periodic tone. Error bars are two standard deviations in overall length.

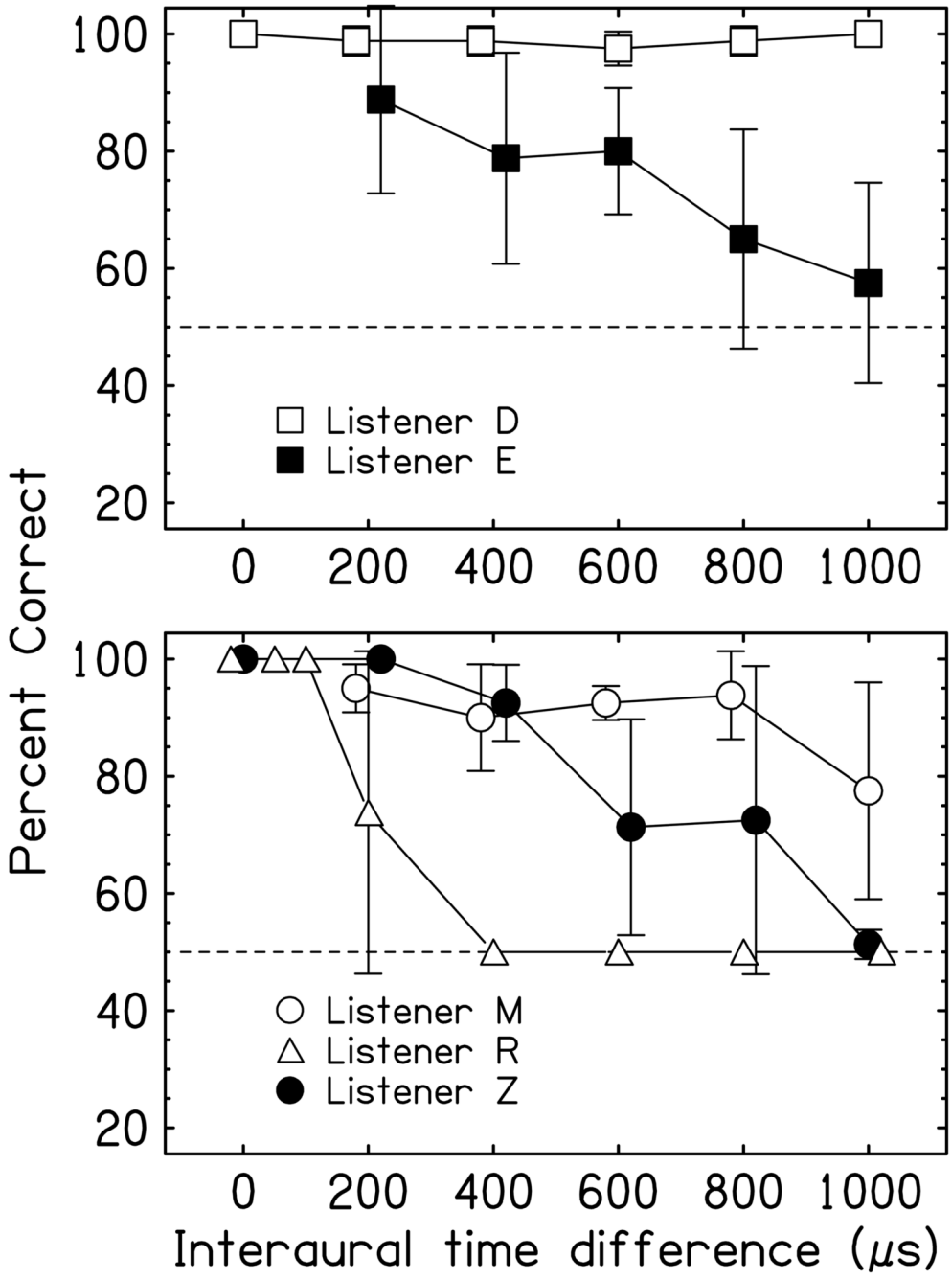

Figure 12.
Experiment 7: Discrimination between front and back by five listeners when an interaural time difference was applied. Amplitude spectra were baseline values. Error bars are two standard deviations in overall length.

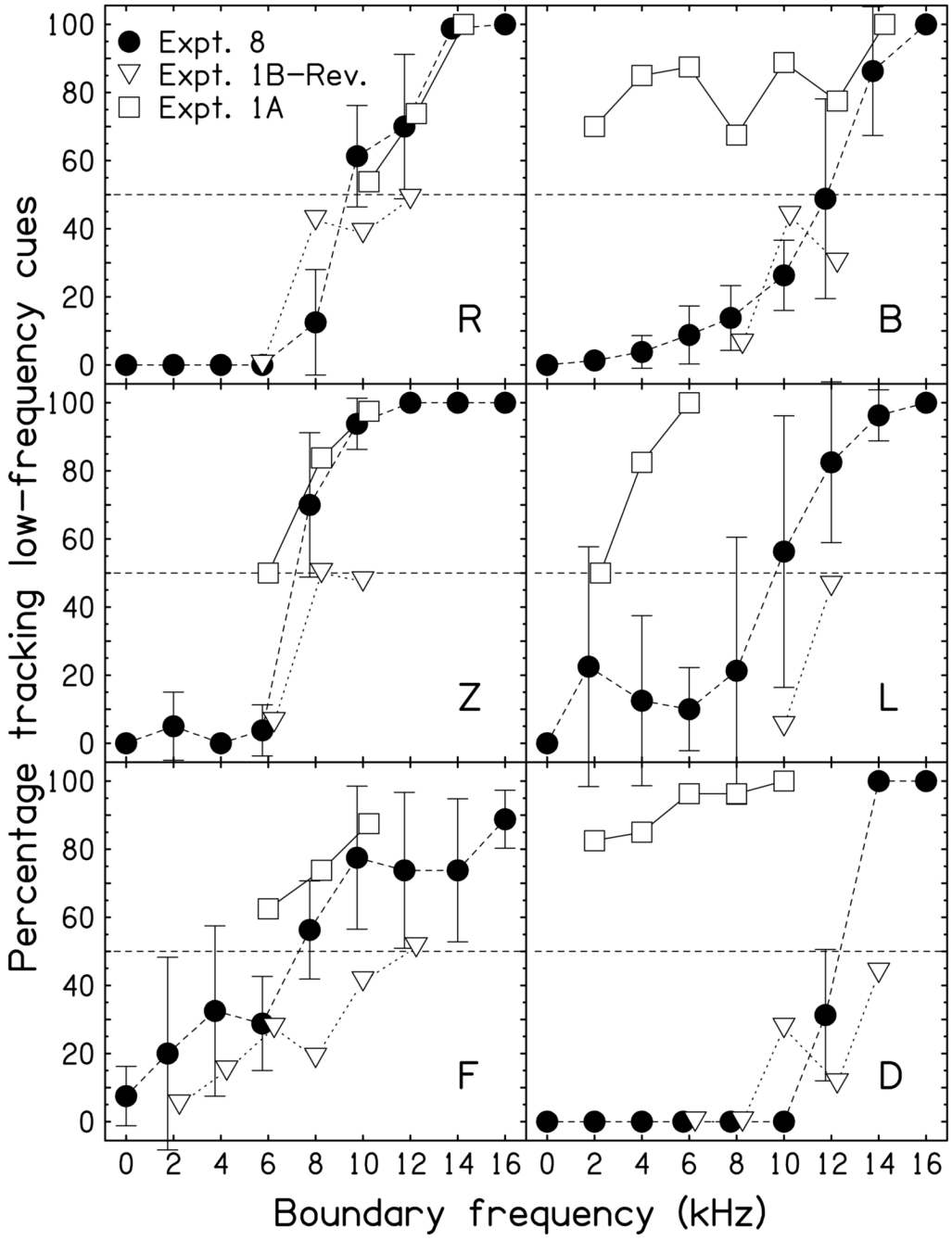

Figure 13.

Experiment 8: Filled symbols show the percentage of trials on which the listener's choice followed the low-frequency source in the competition experiment from Appendix C. Open squares show percent correct from Experiment 1A, *flattening above*. Open triangles show (100% correct) from Experiment 1B, *flattening below*. Error bars are two standard deviations in overall length. The dashed line shows the random-guessing limit.

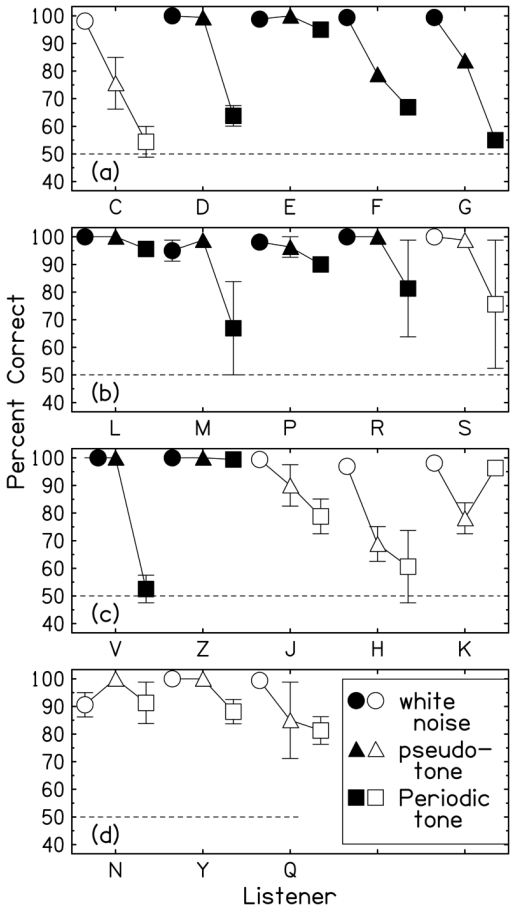

Figure 14.

Preliminary experiment: Percentage of correct front/back judgements by 18 listeners given three different broad-band stimuli. Filled symbols are for listeners who participated in other experiments in this article. Open symbols are for supplementary listeners. Error bars are two standard deviations in overall length.

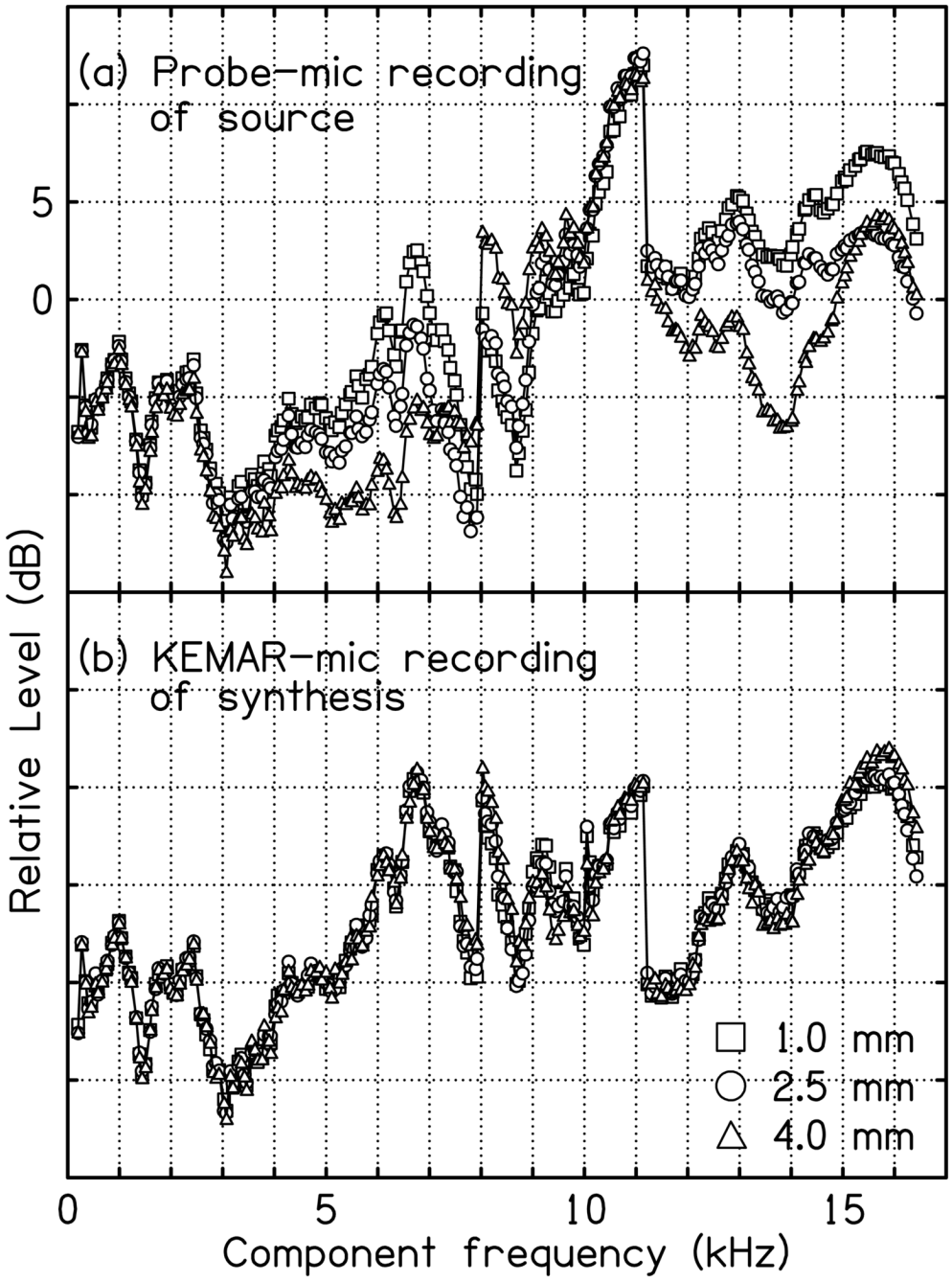

Figure 15.

Self-compensation in VRX. Part (a) shows the recorded power spectrum for three different positions of the probe-microphone tip in the ear canal of the KEMAR, 1.0, 2.5, and 4.0 mm. The recorded spectra were used to calibrate the system, leading to the recordings of part (b), measured at the KEMAR microphone (ear drum). Each vertical division is 5 dB. The discontinuities near 8 and 11 kHz reflect equalization EQ2 shown in Fig. 2.

Table 1

Values of the Mexican-hat function S_j

j	±4	±3	±2	±1	0
S_j	−0.2	−0.5	0.2	0.5	1