CLINICAL APPLICATIONS OF 3-D AUDITORY TESTS

Janet Koehnke, Ph.D., and Joan Besing, Ph.D.

ABSTRACT—One of the primary goals of any audiological evaluation is to assess the impact of the hearing impairment on the communication ability of the individual in everyday listening situations. However, there are presently no clinical tests available that measure these abilities in realistic environments. To fill this gap, two three-dimensional (3-D) auditory tests measuring localization and speech intelligibility in noise are being developed for clinical use. These tests eliminate many of the problems inherent in free-field testing while providing a means for measuring performance in typical listening situations. The procedures used to create the stimuli needed in these tests and the methods for administering these tests are described. Results of these 3-D tests for adults with normal hearing are shown to illustrate typical results obtained with these tests. Some preliminary data obtained in an ongoing research project with children with a history of recurrent otitis media are also shown for comparison. These 3-D tests of localization and speech intelligibility in noise are easy to administer and score. Further testing is being conducted to prepare these tests for clinical use.

INTRODUCTION

The goal of the audiological evaluation has traditionally been to assess the auditory sensitivity of an individual by determining both the degree of hearing loss and the type of hearing loss. In addition, in order to determine strategies for remediation and assess complex auditory perception, it is important to examine the effects of the hearing loss on communication. This is usually done by measuring performance on tests of speech intelligibility, such as word identification tests and sentence identification tests. These tests are typically administered in quiet, under earphones, to each ear individually or via a speaker in a sound-treated test room. Since the complaint most commonly voiced by individuals with impaired hearing is difficulty understanding speech in noise, clinicians sometimes examine this problem by testing speech intelligibility with a speech and a noise presented through the same earphone or with the speech presented from one speaker and the noise presented from the same speaker or from a different speaker in a sound-treated room.

While such an evaluation does yield important information, it does not assess the impact of the hearing impairment on the in-
individual in typical, everyday listening situations. Not only do these individuals have difficulty understanding speech in noise, they also frequently report difficulty localizing sound sources. The tests and test facilities currently available for audiological evaluation, however, do not provide a means for assessing sound source localization or speech intelligibility in noise in realistic environments. There are presently, no clinical tests available to assess an individual's ability to localize sound sources in quiet or in the presence of background noise, no tests to assess the effects of speech and noise source separation on speech intelligibility in noise, and no tests to determine the effects of reverberation on either of these tasks. With the advent of virtual reality techniques using digital signal-processing technology, we have been developing realistic, 3-D auditory tests to measure sound source localization and speech intelligibility in noise. The stimuli in these tests are presented via earphones, but they are externalized and are perceived to be coming from various spatial locations outside the head. In this article the procedures used to create the tests and some results obtained using these tests are described.

In order to successfully localize sound sources (at least in the horizontal plane) and understand speech in noise, it is necessary for the listener to use binaural processing. It has been shown repeatedly that on measures of basic binaural processing such as masking-level differences, interaural time discrimination, and interaural intensity discrimination, performance of listeners with compromised auditory systems\(^1\) is usually poorer than performance of listeners with normal hearing (e.g., Colburn & Trahtiotis, 1992; Durlach, Thompson, & Colburn, 1981; Gabriel, Koehnke, & Colburn, 1992; Hall & Grose, 1993; Hauser, Colburn, & Marr, 1983; Koehnke, Culotta, Hawley, & Colburn, 1995). A number of studies have also quantified the difficulty with localization reported by listeners with compromised auditory systems (e.g., Abel, Birt, & McLean, 1978; Wilmington, Gray, & Jahrsdoerfer, 1994) and the difficulty with speech intelligibility in noise experienced by many of these individuals (e.g., Bronkhorst & Plomp, 1989; vanRooij, Plomp, & Orlebeke, 1989), but there are no clinical tests presently available to assess the ability of listeners to use binaural information to localize sound sources and understand speech in noise.

There is clearly a need for clinical tests to assess the ability to use binaural processing for these tasks in realistic listening situations, but the advantages of developing tests using complex, 3-D auditory processing techniques may not be as obvious. There are, however, a number of reasons virtual reality techniques are preferable to testing in a real sound field for tests of localization and speech intelligibility in noise. First, 3-D auditory tests eliminate problems with calibration of the room and the speakers. Such calibration is time-consuming and costly when conducted as often as would be necessary if the tests were administered regularly. Another problem with sound field testing is replicating speaker and listener placement in the room, which is important for test consistency. Most clinics do not have rooms dedicated to sound field testing where an array of speakers can remain set up and not be disturbed. With 3-D auditory tests, the need to replicate the set-up of the sources and the listener in the sound field is eliminated. A third issue associated with sound field testing is head movements. Allowing head movements might well be important to answer certain research questions. But, for purposes of clinical test--retest reliability and intersubject comparison, it is desirable to eliminate head movements by using 3-D auditory tests with stimuli presented via earphones.

The final advantage gained by using 3-D auditory tests is the ability to control reverberation. It is known that reverberation occurring in typical listening situations diminishes binaural cues used in understand-
SIGNAL PROCESSING FOR 3-D AUDITORY TESTS

The techniques used to process the stimuli for the 3-D auditory tests are designed to preserve, as much as possible, the binaural information present in signals occurring in real rooms. The first step in the creation of the 3-D auditory test signals is to record head-related transfer functions (HRTFs) in the environments or rooms of interest. For the tests described here, both an anechoic and a reverberant room have been used. The reverberant room is a typical conference room with a reverberation time ($RT_{60}$) of 200 to 400 msec. This is comparable to the reverberation in a living room; it is two to three times less reverberant than the average classroom, and about twice as reverberant as the sound-treated rooms used for most audiological evaluations. If desired,

![Figure 1](image)

Figure 1. Experimental set-up for measurement of HRTFs in the free field. KEMAR is seated 3 feet from the signal source in the desired room. Signals reaching each ear are picked up by the microphones in KEMAR’s ears, and these HRTFs are then stored on a personal computer.
other more reverberant rooms could also be used for this test.

Figure 1 shows the experimental set-up used to measure the HRTFs. The HRTFs were recorded in each room using the KEMAR manikin fit with Zwislocki occluded ear simulators (Knowles Electronics). Each of KEMAR’s ears was also fit with a microphone connected to a preamplifier (Etymotic Research). KEMAR was “seated” in the room, 3 feet from the sound source. The same speaker was used as the source in all of the HRTF measurements. For each location of the speaker in the sound field, the HRTFs for each ear were measured in each room in response to a wideband noise stimulus. HRTFs were obtained at nine locations from −90° (near the left ear) to +90° (near the right ear) in 22.5° steps. Each of these HRTFs is unique to the room, the ear, and the source location and preserves the temporal and spectral characteristics of the signal. As indicated in Figure 1, the HRTFs are stored on the computer to be used in the next phase of signal processing.

Figure 2 shows the procedure used to process the speech and noise signals for the 3-D localization and speech intelligibility in noise tests. The speech stimuli (described later) for the localization and speech intelligibility in noise tests were digitized at a 20 kHz sampling rate and stored on a 386 personal computer. Speech-spectrum noise was generated by creating a white noise and shaping it using the long-term speech spectrum. This noise was also stored on the personal computer. In addition, both the speech and noise were filtered using the individual HRTFs obtained from KEMAR for each location for each room in each ear. For example, to create a 3-D auditory stimulus perceived to be coming from a source near the right ear in a reverberant room, the right and left ear HRTFs that were recorded in the reverberant room with the speaker located at +90° would be used to filter the speech signal.

![Figure 2](image)

**Figure 2.** Block diagram indicating the signal processing used in the 3-D localization and speech intelligibility in noise tests. The speech signals for the right and left ears and the noise for the right and left ears are processed separately according to the source-to-eardrum transfer functions (HRTFs) for the desired locations in the desired room and then presented to the listeners via earphones.
PROCEDURES

3-D LOCALIZATION TEST

The target stimuli used for the 3-D localization test are three-word phrases, such as "mark the spot," that are digitized and stored on the computer. Then, the stimuli are processed for each of the nine source locations for each ear in the anechoic and reverberant environments, as described in the preceding section, using the HRTFs obtained using the KEMAR manikin. Both anechoic and reverberant environments are included because the natural reverberation occurring in everyday listening situations tends to diminish the binaural cues used in localizing sound sources. We are also testing 3-D localization in noise. Twenty independent samples of speech-spectrum noise were generated by filtering wideband noise signals generated at a 20 kHz sampling rate. These speech-spectrum noises were then processed using the same technique described earlier. However, the noise samples were processed for only three source locations, 0°, +90°, and −90°. The presentation level of the target stimulus is 70 dB sound pressure level (SPL) for a source located at 0° in the horizontal plane. So, as the source moves away from 0° toward the right or the left ear, the intensity increases at the ear nearer the sound source and decreases at the ear further away from the sound source due to the head-shadow effect. All stimuli are presented via phased and matched TDH earphones. A more complete description of the virtual localization test procedures can be found in Besing and Koehnke (1995a).

The listener's task is to choose the location of the speech from among nine locations. A visual display indicating the nine possible source locations is presented on a computer monitor. Each of the nine locations is represented by either a number or a picture. The listener indicates the perceived source location by touching the appropriate visual image. After the listener responds, feedback is provided by showing the listener the correct source location. The number of trials in a run varies with the subject population. For the adults, there are usually 45 trials per run, including five trials at each of the nine source locations. It takes approximately 4 minutes to complete a run. For children there are usually 18 or 27 trials per run, including two or three trials at each location. The time to complete a run varies from 2 to 4 minutes depending upon the individual child. Data are stored in 9 × 9 confusion matrices, and the percent correct and root mean square (RMS) error are calculated for each experimental run. The RMS error indicates, on the average, how the listener's perception of the source location differs from the actual source location.

3-D SPEECH INTELLIGIBILITY IN NOISE TEST

Phonetically balanced, monosyllabic words from Egan's (1948) 1000-word list are used as the target stimuli for the 3-D speech intelligibility in noise test. Speech-spectrum noise is used as the interference. The monosyllabic words are processed using KEMAR's HRTFs at 0° for the anechoic and reverberant rooms as shown in Figure 2. The speech spectrum noise is processed for the 0°, +90°,

---

For the adult listeners, numbers are typically used to indicate the nine possible locations of the speech source. For the children, other graphic displays are often used to indicate the source locations, including smiling faces, rockets, and firecrackers. When these other graphic displays are used, the correct source location is indicated to the listener after they respond by animating the graphic display (e.g., the rocket at the source location takes off or the firecracker explodes.

RMS error = \[ \sqrt{\frac{\sum (S_k - R_k)^2}{n}} \] where \( n \) = number of trials in a run, \( S_k \) = stimulus location, and \( R_k \) = response location. In other words, it is the square root of the sum of the difference between the perceived location of the stimulus and the actual location of the stimulus on each trial squared and divided by the number of trials in a run.
and −90° source locations in both environments using KEMAR's HRTFs. So, the speech always comes from a location straight ahead of the listener, and the noise comes either from the right, the left, or straight ahead. These speech and noise locations were chosen since they are the configurations most commonly encountered in everyday listening situations. As for the localization task, both anechoic and reverberant environments are included because natural reverberation occurring in everyday listening situations tends to diminish the binaural cues important for understanding speech in noise.

The speech is presented at a normal conversational level of 70 dB SPL, and the noise level is varied to find the 50% correct intelligibility threshold using a single interval, adaptive procedure. The listeners’ task is to repeat the word they hear. A complete description of the virtual speech intelligibility in noise test can be found in Koehnke and Besing (1996).

The 3-D localization test is easy to administer and score, and listeners consider the task to be fairly interesting. Results of the virtual localization test for a group of five normal-hearing adults and a group of five children with a history of otitis media are shown in Figure 3. This bar graph shows the average RMS localization error in degrees for adults and children as well as the standard errors in both the anechoic and reverberant listening environments. The ordinate is incremented in 22.5° steps since the source locations were separated by 22.5°. As indicated by the light bars in Figure 3, localization of virtual targets by adults with normal hearing is quite accurate in both listening environments, and reverberation does not appear to degrade their performance. On the average these adult listeners have errors of less than one location in both conditions. Standard errors of the adult listeners show that the variability among listeners is quite small.

APPLICATION OF THE 3-D AUDITORY TESTS

As described earlier, there are many advantages to using 3-D auditory tests to assess the ability of listeners to localize sound sources and understand speech in noise. We have been using these tests in the laboratory to investigate the ability of individuals with compromised auditory systems to localize sound sources and understand speech in noise. To examine the utility of these tests and to provide a baseline for comparison, performance is also being measured for adults with normal hearing. To illustrate the results of these tests, some selected data are presented showing the performance of a group of normal hearing adults. In addition, some preliminary results obtained for a few of the children we have been testing with a history of chronic, recurrent otitis media are presented and compared to the results of the normal hearing adults.

![Figure 3. 3-D sound source localization for a group of adults with normal hearing (light bars) and a group of children with a history of recurrent otitis media (dark bars). The RMS localization error in degrees is indicated by the height of the bar versus the type of listening environment. Data for the anechoic environment are on the left, and data for the reverberant environment are on the right. Chance performance is indicated by the solid line at 82°, which is the RMS error that would result if a subject responded without listening to the stimuli. Standard errors are indicated for each condition for both groups of subjects.](image-url)
Figure 3 shows that 3-D localization by children with a history of chronic, recurrent otitis media tends to be much poorer than that by normal-hearing adults. The dark bars indicate the average results for five children ages 8 to 11 with a history of otitis media. It should be noted that, although these children have a history of recurrent otitis media, their auditory acuity was normal when the 3-D localization test was administered, so signal audibility is not an issue. As indicated by the standard error bars for the children, there is considerably more variability in the performance of these listeners than there is in the normal-hearing adult listeners. While these data represent only a small subset of the complete group of listeners tested, it is also clear that performance of this small group of children with a history of otitis media is much poorer than performance of the adults. These children also localize more poorly than their peers with no history of otitis media whose localization is comparable to that of the adults (Besing & Koehnke, 1995a). The children’s RMS localization errors are more than twice as large as those of the adults. But, like the adults, there is no effect of listening environment on performance. A more complete discussion of these results is presented in Besing and Koehnke (1995a).

This test is also being used to evaluate the effectiveness of monaural and binaural amplification for listeners with sensorineural hearing loss. This 3-D localization test is effective in determining the benefits of monaural and binaural amplification for complex binaural processing tasks (Koehnke & Besing, 1997). These results, as well as those shown earlier, illustrate the utility of the 3-D localization test for identifying individuals with impaired binaural processing ability.

The 3-D speech intelligibility in noise test is, like the 3-D localization test, easy to administer and score and provides a means for testing speech intelligibility in typical listening situations. This test makes it possible to evaluate both the effects of reverberation and the effects of speech and noise source separation on speech intelligibility. Figure 4 shows the results of the 3-D speech intelligibility in noise test for a group of seven adults with normal hearing and a group of four children with a history of otitis media. The average intelligibility gain in dB is plotted for each group of subjects in each of the listening environments. The intelligibility gain is the difference (in dB) between the signal-to-noise ratio (SNR) for 50% correct intelligibility with the speech and noise source both at 0°, and the SNR for 50% correct intelligibility with the speech at 0° and the noise at −90° or +90° [i.e., \((SNR @0°) - (SNR @−90°)\) or \((SNR @0°) - (SNR @+90°)\)]. The average intelligibility gains for the −90° and +90° conditions are plotted in both listening environments.

Results for these normal-hearing adults indicated by the light bars show a large intelligibility gain in the anechoic environment and a smaller, but nonetheless consistent, gain in the reverberant environment. The gains obtained by individual subjects range from about 10 to 17 dB in the anechoic environment and from about 2.5 to 5.8 dB in the reverberant environment. Clearly, these listeners obtain a consistent intelligibility advantage by separating the speech and noise sources in both listening
environments, but the reverberation decreases the binaural information available to the listeners, resulting in smaller intelligibility gains in the reverberant environment.

Figure 4 also shows the 3-D speech intelligibility in noise results for four children with a history of chronic, recurrent otitis media indicated by the dark bars on the right. Although these data for the children are preliminary, the results for these children with a history of otitis media are clearly different from the results of the adults with normal hearing. Again, as for the children whose localization data were shown earlier, these children had normal auditory thresholds at the time the 3-D speech intelligibility in noise test was administered. Like the adults, the children obtain a much larger intelligibility gain in the anechoic environment than in the reverberant environment. In fact, in the reverberant environment, these children obtain only minimal gain when the speech and noise sources are separated. The intelligibility gains are consistently smaller for this group of children than for the normal-hearing adults. On the average, the gains are about 3 dB smaller in both listening environments for these children than for the adults. The children’s gains range from about 9 to 13 dB in the anechoic environment and from about −0.75 to 3.0 dB in the reverberant environment.

CONCLUDING REMARKS

The 3-D localization and speech intelligibility in noise tests described here are easy to administer and provide a means for testing performance in typical listening situations without the confounding effects involved in free-field testing. These virtual auditory tests eliminate calibration problems and problems replicating source and listener locations across test sessions. They also remove any questions about the characteristics of the signals reaching the ears of the listeners. No training is required to learn to perform these tasks, and the instructions provided to the listeners are simple and straightforward.

In addition to measuring performance on these tests in 8- to 12-year-old children with a history of chronic otitis media, we are measuring performance for children in the same age group with no history of otitis media (Besing & Koehnke, 1995a, b; Besing, Koehnke, & Goulet, 1993). To compliment this work, the 3-D localization in quiet and in noise and 3-D speech intelligibility in noise tests are being administered to children ages 5 to 7 with and without a history of otitis media. As part of another project, these tests are being used in our laboratory to compare the effectiveness of different amplification configurations in adults with varying degrees and configurations of hearing impairment (Koehnke & Besing, 1997).

Results obtained thus far indicate that these 3-D auditory tests are useful for measuring binaural performance in children and adults. Overall, the data indicate that individuals with compromised auditory systems perform more poorly on 3-D tests of localization and speech intelligibility in noise than individuals with no evidence of auditory pathology. The apparent ability of these 3-D auditory tests to identify individuals likely to experience binaural processing problems indicates that they have the potential to be good clinical tools. Data collection with individuals with normal hearing and individuals with compromised auditory systems is continuing so that tests designed for clinical use can be available in the near future.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NIDCD #DC00428), the Deafness Research Foundation, and the Baton Rouge Sertoma group. We would like to thank Pat Zurek for his much needed assistance measuring the EMAR impulse responses and processing the speech stimuli for the 3-D localization and speech intelligibility in noise tests. Thanks also to Marla Allard, Sandie Bass-Ringdahl, and Christine Goulet for their assistance with data collection.
REFERENCES


ARTICLE THREE

SELF-ASSESSMENT QUESTIONS

1. The advantages of 3-D auditory tests over free-field testing include:
   (a) elimination of calibration problems
   (b) control over reverberation
   (c) less time required
   (d) larger test room required
   (e) source locations must be replicated

2. An HRTF is:
   (a) a human research test function
   (b) used to process the signals used in the 3-D auditory tests
   (c) a head related transfer function
   (d) b and c
   (e) all of the above

3. For the 3-D localization test, the sim-
   uli:
   353
(a) are presented from locations in the horizontal and vertical planes
(b) are processed in an anechoic and reverberant environment
(c) are presented through speakers in a sound-treated room
(d) are presented at the same intensity from each source location
(e) consist of continuous discourse with or without noise

4. For the adult listeners with normal hearing described here,
(a) 3-D localization ability is not affected by reverberation
(b) RMS localization errors are about 18°
(c) Speech intelligibility in noise is better when the speech and noise sources are separated than when they are at the same location
(d) The intelligibility gain is larger in the anechoic environment than in the reverberant environment
(e) All of the above

5. The results shown here for the children with a history of chronic, recurrent otitis media
(a) indicate better localization in the anechoic environment than in the reverberant environment
(b) reveal smaller intelligibility gains than the adults with normal hearing
(c) are the same as the results of the adults on these two 3-D tests
(d) indicate larger RMS localization errors in the anechoic environment than in the reverberant environment
(e) reveal smaller intelligibility gains in the anechoic environment than in the reverberant environment