ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 91 (2005) 451 – 462

# Binaural Coherence in Rooms

William M. Hartmann[1], Brad Rakerd[2], Aaron Koller[1]

[1]: Department of Physics and Astronomy, Michigan State University, East Lansing, MI, 48824, USA.
hartmann@pa.msu.edu

[2]: Department of Audiology and Speech Sciences, Michigan State University, East Lansing, MI, 48824, USA

**Summary**
In a study of the binaural properties of rooms, binaural cross-correlation (CC) functions were measured in 19 one-third-octave bands, and broad band, using an artificial head and torso (KEMAR). For a fixed source-receiver distance, twenty CC functions were collected as the source and receiver were moved to different locations in the room. The peaks of the CC functions were taken as measures of binaural coherence. Plots were made of binaural coherence as a function of frequency for five rooms and several source-receiver distances (3 m, 6 m, and 12 m). Attempts were made to interpret the coherence plots in terms of geometrical, material, and acoustical properties of the different rooms in order to develop some intuition about the statistical behavior of the coherence function in rooms. Correctly interpreted, in coordination with headphone lateralization studies, the measured values of binaural coherence are expected to indicate the utility of the steady-state interaural time difference cue in sound localization.

**PACS no. 43.66.Cb, 43.66.Dc, 43.66.Lj**

## 1. Introduction

At the beginning of the 20th century it was discovered that human listeners can make use of the interaural time difference (ITD) to localize sounds [1, 2]. The ITD arises physically because a sound wave coming from a direction to the listener's right takes longer to reach the left ear than to reach the right ear – hence, an interaural time difference. The interaural time difference may be as large as 800 microseconds for a source at 90 degrees azimuth with respect to the listener's forward direction, but the human binaural system is capable of localizing a sound based on an ITD of only 10 $\mu$s, leading to a sensitivity of about 1 degree of arc. There is evidence that when a listener is required to localize a broadband, continuous sound in a free field, the ITD is the most important acoustical cue available [3]. Of course, if the sound is periodic, there is the danger that a binaural comparison will make a mistake of a full period because every cycle is the same as every other. A binaural comparison finds the correct ITD only if the ITD is less than half a period. This kind of problem does not occur if the sound is noise, which partly accounts for the appeal of noise as a sound source in localization studies, as in the present article.

The binaural ITD calculation, which works so well in a free field, can encounter troubles if sound is heard in a room. Reflections from the surfaces of rooms lead to complex sound fields wherein the signal in one ear does not exactly resemble a delayed version of the signal in the other. If the source of sound is close by, the sound in the listener's ears is mainly "direct sound," arriving by a straight-line path from the source, and the disturbing effect of reflections is not an important consideration. But if the source is remote, in a large room, it is common for the power in the reflected sound to be much greater than the power in the direct sound. Then the binaural system faces logical difficulties in matching waveform features in the left ear with features in the right ear.

A mathematical measure of the similarity of waveforms in the two ears is the binaural cross-correlation function (CC function). According to the standard model of azimuthal sound localization [4] the mammalian binaural system registers the interaural time difference (ITD) in a way that is akin to a neural computation of cross-correlation. Such a computation requires a process by which neural spikes from one ear can be delayed with respect to spikes from the other, followed by coincidence cells which respond only if spikes from the two ear-related channels arrive almost simultaneously. Such a system has been found in the medial superior olive of dogs [5] and cats [6], and human models have been developed [7].

The CC function, $\rho$, is given as a function of the lag $\tau$ by

$$\rho(\tau) = \int \mathrm{d}t\, x_L(t)\, x_R(t+\tau) / \sqrt{E_L E_R},$$

Received 31 January 2004, Revised 18 October 2004, accepted 26 November 2004.

where $x_L(t)$ is the pressure in the left ear as a function of time, $E_L$ is the energy in that ear,

$$E_L = \int \mathrm{d}t\, x_L^2(t),$$

and $E_R$ is similarly defined for the right ear. All three integrals cover the same span of time, much longer than $\tau$.

The CC function measures the average similarity of signals in the left and right ears. If the signal in the right ear is the same as the signal in the left except for a delay $T_d$ then the CC function will have its maximum possible value, namely 1.0, when the lag equals the delay, i.e. $\rho(T_d) = 1.0$. If $x_L$ and $x_R$ are almost independent, as can occur in a reverberant room, then $\rho(\tau)$ is nearly zero because the positive and negative parts of the product $x_L\, x_R$ cancel on the average.

The maximum value of the CC function, occurring over the range of lags from $-\tau_m$ to $+\tau_m$, is defined as the *binaural coherence* $\gamma$, i.e.

$$\gamma = \mathrm{Max}_{-\tau_m<\tau<\tau_m}\left[\rho(\tau)\right].$$

*A priori* one expects that the ITD may be useful in localizing a steady state sound if the binaural coherence is not small, i.e. not much less than 1.0.

The limits $\pm\tau_m$ represent the fact that a high value of crosscorrelation is of use to the binaural system only if it occurs for a value of lag that is meaningful in terms of sound localization. For example, a CC function that is large only for $\tau$ greater than 10 seconds would be of no practical value. The definition of the quantity IACC (interaural cross-correlation), as used in architectural acoustics, limits $\tau$ to the range -1 to +1 ms [8]. Apart from sound localization, architectural acousticians find that the binaural coherence determines the sensation of listener envelopment when music is heard in rooms and is also related to the apparent width of a sound source [9, 10, 11].

Binaural coherence can be studied in headphone experiments that provide idealizations, or simplifications, of the signals and effects that occur in real rooms. For example, the masking level difference paradigm in its classic N0S$\pi$ configuration explores a binaural detection advantage that has been attributed to a change in binaural coherence [12].

In connection with sound localization, the role of binaural coherence can be studied with headphones by beginning with diotic noise, adding an interaural delay to cause the noise image to be lateralized, and then adding an independent noise to one ear to create a controlled amount of binaural incoherence. Experiments of this kind were done by Constan [13] to determine the amount of binaural coherence needed for a listener to distinguish between left and right lateralizations. Such a stimulus is clearly an idealization. The binaural coherence is constructed to be frequency independent, and the statistics of the temporal factors depend only on the noise bandwidth. For such a stimulus, the ensemble average cross-correlation is a uniform function of lag away from the main peak without systematic secondary peaks characteristic of special room reflections.

Ultimately, an important goal of research on sound localization in rooms is to understand the relative roles of the many different signal cues that listeners can use to localize. Some cues are available in the steady-state portions of sounds, others come from transients. The measurements of binaural coherence in rooms in the present article were done to learn about the strength of steady-state ITD cues that are available to listeners in different frequency bands in different rooms. Together with experiments such as those by Constan, it is expected that the room measurements will enable us to predict the conditions under which the steady-state ITD cue is useful in ordinary life.

## 2. Broadband cross-correlation function

The degree to which binaural coherence can be reduced by reflections in a room can be illustrated dramatically by coherence measurements in a reverberation room. Because the reflected sound in this environment is such a large fraction of the total sound power, the room is able to reduce the coherence to a small value as shown below.

### 2.1. Measurements in a reverberation room

The CC function was measured in a reverberation room with dimensions 7.67 × 6.35 × 3.58 meters and a reverberation time from 2 to 3 seconds at speech frequencies. The receiver was a KEMAR manikin with two large ears placed near the center of the room. The sound sources were 24 loudspeakers (Minimus 3.5) placed in an arc 3 meters from the manikin. The loudspeakers were separated by 2 degrees so that they extended from 23 degrees to the left of midline to 23 degrees to the right. The loudspeakers were at an elevation of 1.17 m, the same height as the KEMAR ears. The loudspeakers reproduced a five-second noise burst that was flat from 200 to 17,000 Hz.

The two ears of the KEMAR were fitted with Etymotic ER-11 microphones, and the microphone outputs were amplified by ER-11 preamps (battery powered) followed by a two-channel dBX 760x preamp. The signals were lowpass filtered at 18 kHz and converted to digitized form by a 16-bit two-channel ADC (TDT DD1) at a sample rate of 50 ksps per channel.

The CC functions, based on a digital recording of the last second of the five-second noise burst, are shown in Figure 1 for each of the 24 loudspeakers. Cross-correlation is shown as a function of lag, ranging from $-1000\,\mu$s to $1000\,\mu$s, ie. $\tau_m = 1000\,\mu$s. Each function oscillates about a cross-correlation of 0, and the magnitude of the function is given by the tic marks on the vertical axis, which are spaced by 0.1 unit of correlation.

Figure 1 shows that the CC functions have many peaks and valleys, but the tallest peak turns out to be an excellent indicator of location because that peak occurs at a lag that agrees with the expected ITD for the source. For loudspeakers 12 and 13, immediately to the left and right of the midline, the tallest peak occurs at zero, or very close to zero. For loudspeakers 1 through 11 the tallest peak occurs at negative values of the lag, and for loudspeakers 13-24 it

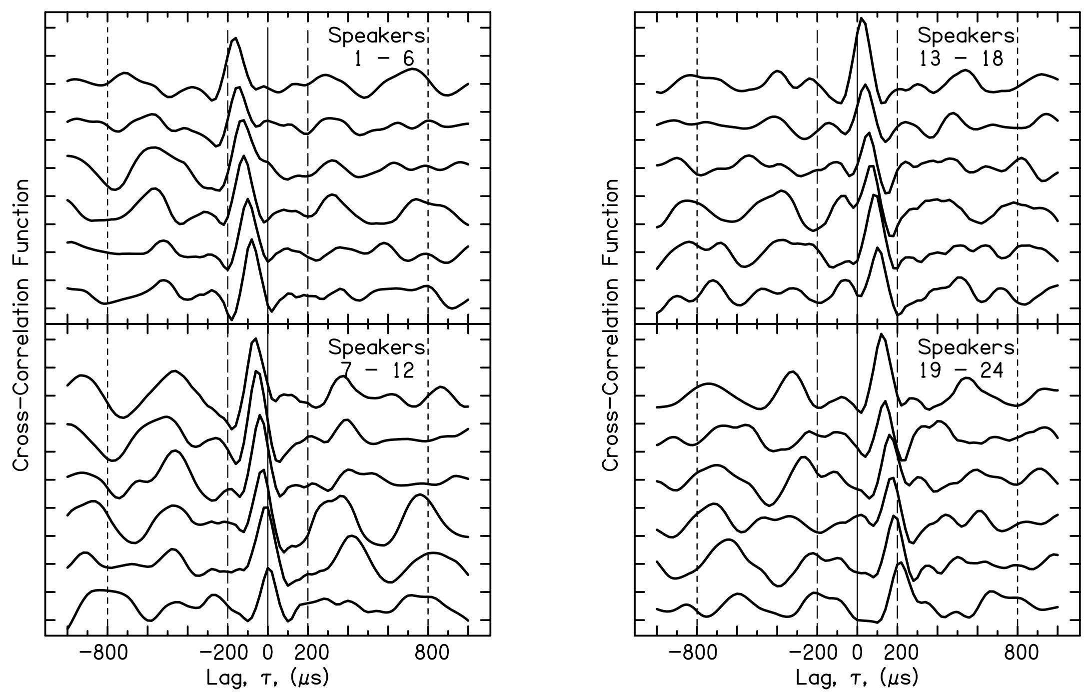


Figure 1. Binaural cross-correlation function for broadband noise from 24 loudspeakers in a reverberation room, 3 meters from a KEMAR manikin. Part (a) shows loudspeakers 1-12, all to the left of midline. Part (b) shows loudspeakers 13-24, all to the right of midline. Loudspeaker 1 is at -23 degrees azimuth; loudspeaker 24 is at +23 degrees. Intervening loudspeakers are in numerical order from left to right, separated by 2 degrees. The horizontal axis shows the lag variable of the cross-correlation function from -1000 to +1000 microseconds. Dashed lines at $\pm 800\mu$s show the range of ITDs encountered by a human listener in free field. Dashed lines at $\pm 200\ \mu$s show the expected ITDs for $\pm 23$ degrees. Cross-correlation functions all oscillate about a correlation of zero. The scale of the vertical axis is given by the tic marks, spaced by 0.1 unit.

occurs at positive values of the lag. The tallest peak occurs at lag values that increase monotonically as the azimuth of the source increases from $-23$ degrees (loudspeaker 1) to +23 degrees (loudspeaker 24).

For loudspeakers 1 and 24 the peak occurs at $-160\,\mu$s and +220 $\mu$s. These lags are in reasonable agreement with the expected ITDs, calculated as follows. In the low-frequency limit, the diffraction formula for a sphere with antipodal ears leads to an ITD given by $\mathrm{ITD} = (3r/v)\sin(\theta)$, where $r$ is the head radius (8.75 cm), and $v$ is the speed of sound, 34400 cm/s. For $\theta = 23$ degrees the predicted ITD is $300\,\mu$s, which is higher than observed. However, in the high-frequency limit, the formula leads to $\mathrm{ITD} = (2r/v)\sin(\theta)$ [14, 15], and the predicted ITD is $200\,\mu$s. Because the direct sound becomes a significant fraction of the total sound in the reverberation room only at high frequencies, the high-frequency limit is the better choice. The agreement between the observed peaks and the high-frequency limit is essentially perfect if it is assumed that the dummy head was misaligned by about 2 degrees.

### 2.2. Application to sound localization

The physical analysis above shows that excellent information about source location is present in the measured broadband CC functions. Given the task of localizing a source, a computer with the information in Figure 1 could achieve excellent source identification performance based only on the lag of the cross-correlation peak. However, there are several difficulties in applying this idea to actual sound localization by human listeners. First, the peak heights (i.e. coherence) are rather small. They average 0.18, and none of them is taller than 0.24.

The amount of coherence actually necessary to localize a broadband noise was studied by Constan [13] in headphone experiments measuring ITD sensitivity. Those experiments generated partially coherent noise by mixing independent noise sources, thought to be a reasonable model for noise measured in the center of a reverberation room. The results of Constan's experiment indicated that when the coherence is no greater than 0.2, listeners cannot begin to discriminate sources that are only 2 degrees apart. In contrast to the orderly march of narrow peaks across two-degree intervals shown in Figure 1, Constan's perceptual measurements showed that if listeners had been exposed to the cross-correlation functions measured in the reverberation room they would only have been able to tell the difference between loudspeakers on the extreme left and those on the extreme right.

Second, there is no good evidence that listeners can make use of broadband information in the way that the measured cross-correlation does. Instead, the standard model of the binaural system is tuned so that the cross-correlation is computed within individual critical bands. The tuning in the model is supported physiologically in that broadband noise experiments by Yin and Chan [6] led to MSO recordings that showed an evident bandpass character, presumably because of MSO input tuning.

Consequently, the remainder of this report describes measurements of CC functions in one-third-octave bands as estimates of channels for binaural auditory processing. The cross-correlations are represented here only by the coherence because without adequate coherence it is impossible to use ITD to localize sound.

## 3. Measurements in rooms - Technique

The binaural coherence was measured in a variety of rooms, ranging from an anechoic room to a reverberation room, using techniques described in this section.

### 3.1. Definitions

Individual one-third-octave bands are labeled with index $m$. Therefore, the coherence is a function of band index,

$$\gamma_m = \mathrm{Max}[\rho_m(\tau)],$$

where

$$\rho_m(\tau) = \int \mathrm{d}t\, x_{L,m}(t)\, x_{R,m}(t+\tau) / \sqrt{E_{L,m} E_{R,m}}.$$

The meaning of subscript $m$ on physical variables is that signals are rectangularly filtered into band $m$ prior to integration. Rectangular filtering allows the broadband CC function to be expressed in terms of the individual band functions,

$$\rho(\tau) = \sum_m \sqrt{\frac{E_{L,m}}{E_L} \cdot \frac{E_{R,m}}{E_R}}\, \rho_m(\tau).$$

### 3.2. Sources

CC functions were measured using a loudspeaker source and the KEMAR manikin as described above. Sources were one-third octave bands of equal-amplitude random-phase noise as given in Table I. Measurements were also made in a broad band (BB).

Noises were created in the frequency domain and hence had sharp spectral edges. They were reproduced by a TDT DD1 DAC with a sample rate of 50 ksps, attenuated to optimize the level for each band by a PA4 attenuator, and lowpass filtered at 20 kHz. Noises were amplified and reproduced by a B&W Model DM302 two-way loudspeaker[1]. Five-second noise bursts were presented with the 20 frequency bands shown in Table I in succession to make a "run."

[1] The B&W DM302 loudspeaker is a two-way 4th-order vented box with 5-inch and 1-inch drivers and a crossover frequency of 3 kHz. The response within a horizontal angle of 40 degrees and within a vertical angle of 10 degrees is within 2 dB of the response on axis at all the frequencies of interest in this article.

Table I. The table shows the limits and center frequencies of the rectangular noise bands used in the experiments. Band labels include a number indicating octaves, e.g. band 2A is an octave above band 1A.

| 1/3 Octave band | Bottom (Hz) | Center (Hz) | Top (Hz) |
|---|---|---|---|
| 1A | 125 | 142 | 160 |
| 1B | 160 | 180 | 200 |
| 1C | 200 | 225 | 250 |
| 2A | 250 | 285 | 315 |
| 2B | 315 | 358 | 400 |
| 2C | 400 | 450 | 500 |
| 3A | 500 | 565 | 630 |
| 3B | 630 | 715 | 800 |
| 3C | 800 | 900 | 1000 |
| 4A | 1000 | 1125 | 1250 |
| 4B | 1250 | 1425 | 1600 |
| 4C | 1600 | 1800 | 2000 |
| 5A | 2000 | 2250 | 2500 |
| 5B | 2500 | 2850 | 3200 |
| 5C | 3200 | 3600 | 4000 |
| 6A | 4000 | 4500 | 5000 |
| 6B | 5000 | 5625 | 6250 |
| 6C | 6250 | 7125 | 8000 |
| 7A | 8000 | 9000 | 10000 |
| BB | 125 | 5000 | 10000 |

### 3.3. Recording

Recordings were made using the KEMAR system as described in section 2.1. The separation between the KEMAR ears and the loudspeaker was 3 m, 6 m, or (for the largest room only) 12 m. The KEMAR nose always pointed directly at the loudspeaker. Therefore, the azimuth was 0 degrees. Two kinds of recording were made – same location and different location.

In a *same-location* set 20 measurements were made with the loudspeaker and the KEMAR in fixed positions throughout. The variation in coherence measured in such same-location recordings indicates the variation caused by different noise samples plus whatever variation can be attributed to a non-stationary environment (e.g. air currents in the room).

In a *different-location* set 20 measurements were made with the positions of the loudspeaker and/or the KEMAR haphazardly moved about the room between each measurement. To the extent possible, the 20 different locations were near the center of the room, avoiding walls for both the loudspeaker and the KEMAR. The advantage of central locations is that the measurements should reflect the properties of the room as a whole. By contrast, measurements made with a wall in a nearby position risk domination by a strong reflection from that surface. However, in smaller rooms it was impossible to completely avoid proximity to walls when the source to KEMAR distance became large. In all the different locations the KEMAR nose continued to point towards the loudspeaker (0 degrees azimuth). The average coherence measured in the different-location recordings was intended to provide a room average so that the measurement reflected the room itself without the idiosyncrasies of special locations within the room. In comparison with the same-location measurements, the variation in the different-location measurements was intended to indicate the additional variation attributable to different room locations.

### 3.4. Calculation

The cross-correlation integral was done in the time domain with lags varying from $-800\,\mu$s to $800\,\mu$s, the physiological range for human listeners. Because the maximum lag was less than 0.1 percent of the recorded duration, the normalization of the CC function was taken to be lag independent. The sample rate of 50 ksps quantized lags in $20\mu$s intervals. The peak of the CC function, at whatever lag, was taken as a coherence datum. Therefore, a single run accumulated 20 coherence data points, one for each band. At the end of 20 runs the coherence data for each band were averaged to find a coherence for the band.

### 3.5. The rooms

Six different rooms were chosen for study. One was an anechoic room, another was a reverberation room. Between those extremes rooms were chosen for their availability and variety.

- Anechoic: An anechoic room (IAC 107840, 32 cubic meters). Because there should be negligible reflections in the room the coherence is expected to be 1.0 for all bands. With room effects absent, the primary value of the anechoic room measurements is to test the measurement system.
- Room 16: A classroom, 9 by 8 meters with a 3-meter ceiling height. The room has a light-carpet floor covering, acoustical tile suspended ceiling, and sound absorbing panels on 36 percent of the wall area. This room was chosen because it is especially dry, with reverberation times less than 0.5 s.
- Room 25: A classroom, 8 by 7 meters with a 3-meter ceiling. The floor is vinyl tile and the ceiling is acoustical tile. About 38 percent of the wall surface area is covered by sound-absorbing panels. Room 25 is typical of university classrooms built in the United States in the early 1980s. It was chosen because it resembles Room 16, but is somewhat more reverberant.
- Room 147: An auditorium (3800 cubic meters) with 271 cushioned and raked seats on concrete or light carpet. Room 147 is diamond shaped, 16 m from front to back and 17.5 m in width. The ceiling height varies from 17 m in front to 11 m in back. About 41 percent of the brick wall surface is acoustically absorbing because the bricks (20 × 9 cm × 9 cm deep) form an open lattice with absorbing material behind them. For the measurements reported here, the source was on the centerline, about 2 meters from the front wall, where a lecturer would be. The KEMAR was near the geometrical center of the room for the 6-meter spacing. It was within a few meters of the back wall or a side wall on two different measurements with a 12-meter spacing. Room 147 was chosen to study the effect of room scale on coherence.
- Room 10B: A lab, 6.5 by 7.5 meters with a 4.5-meter ceiling. The floor is vinyl tile and the ceiling is concrete. This room was chosen because it is almost identical to Room 16 in volume, but its surface treatment is much less absorbent.
- Reverb room: The same as the 174-cubic-meter room used for the broadband measurements described in section 2.

Table II. The table gives room volumes and reverberation times in seconds for seven octave bands in the experimental rooms.

| Room | | 16 | 25 | 147 | 10B | RR |
|---|---|---|---|---|---|---|
| Volume | ($m^3$) | 216 | 168 | 3800 | 219 | 174 |
| 1/3 Octave band | freq (kHz) | (s) | (s) | (s) | (s) | (s) |
| 2 | 0.25 | 0.7 | 0.8 | 1.0 | 0.9 | 1.2 |
| 3 | 0.5 | 0.5 | 0.6 | 0.9 | 0.8 | 2.0 |
| 4 | 1. | 0.4 | 0.5 | 0.9 | 0.8 | 2.3 |
| 5 | 2. | 0.3 | 0.4 | 0.8 | 0.9 | 2.5 |
| 6 | 4. | 0.4 | 0.4 | 0.8 | 0.8 | 1.7 |
| 7 | 8. | 0.4 | 0.4 | 0.7 | 0.7 | 1.1 |
| 8 | 16. | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 |

#### 3.5.1. Reverberation times

The (60-dB) reverberation times of the rooms were measured in seven one-third octave bands, separated by an octave, using a Larson Davis sound level meter, model 800B. The results are shown in Table II. The measured reverberation times show the large absorption of the surfaces in Room 16. Although Room 16 is actually somewhat larger in volume (216 $m^3$) than Room 25 (168 $m^3$), reverberation times are smaller in Room 16. Although Room 10B has a volume comparable to Room 16, its reverberation times resemble those of Room 147, which has a volume more than 15 times larger. Room 10B has hard surfaces with minimal absorption except for the highest frequency bands.

## 4. Measured coherence

### 4.1. Anechoic room

Coherence measured in the anechoic room is shown in Figure 2. The coherence should be 1.0 in all bands, but the figure shows that the coherence differs from 1.0, by increasing amounts as frequencies increase beyond 3 kHz. The discrepancy may be caused by the room – perhaps by reflections from a patch panel. Alternatively the discrepancy may be caused by the artificial head recording system. Head dispersion may be responsible in part [16]. The fact that the two KEMAR ears are not mirror images may be a factor in the highest three bands. Even if the non-ideal high-frequency behavior of the system is the result of an

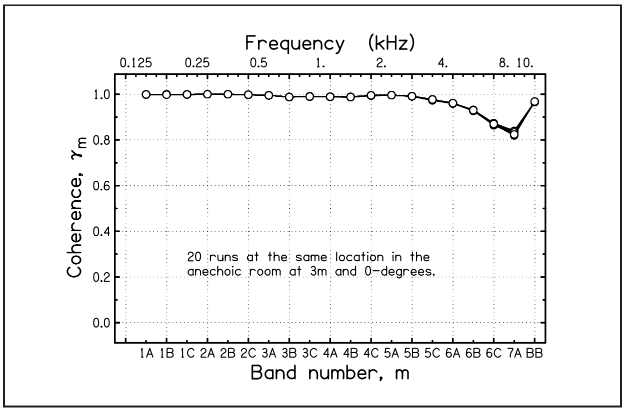


Figure 2. Binaural coherence in an anechoic room with the source (B&W loudspeaker) and receiver (KEMAR) separated by 3 meters and the receiver facing the loudspeaker, 0-degree azimuth. The coherence is shown for each of the 20 runs, but the points overlap almost completely.

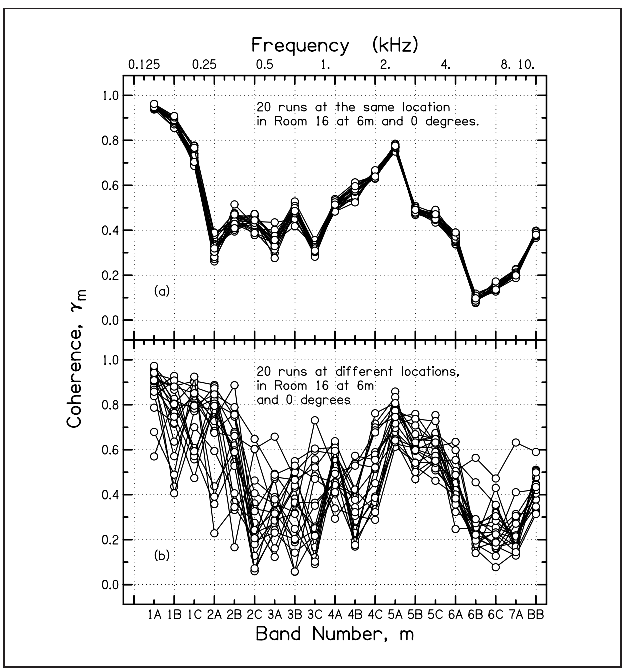


Figure 3. Binaural coherence in Room 16 (classroom). The coherence is shown for every run. (a) Source and receiver were in the same locations, separated by 6 meters, for all 20 runs. The variability among runs is due to the different phases in the noise. (b) Source and receiver were in different locations, again separated by 6 meters, on each of the 20 runs. The variability among runs is partly due to the different phases in the noise, as in part (a), but mostly due to the different locations for source and receiver.

imperfect recording system, the imperfections are unlikely to have an important effect on the measurements in other rooms because the coherence is so much less in the other environments.

### 4.2. Room 16

Coherence measured in Room 16 at a 6-meter distance is shown in Figure 3, part (a) for same-location recordings

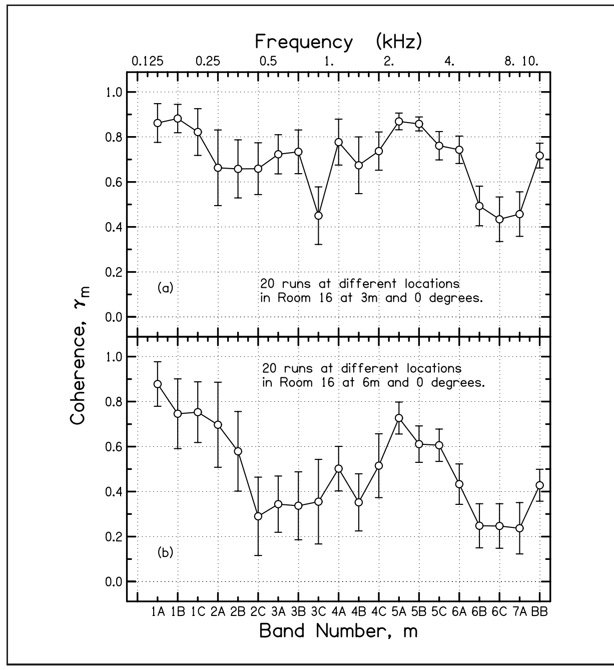


Figure 4. Binaural coherence in Room 16 (classroom) with source and receiver in 20 different locations. The mean coherence is indicated by a circle, and the error bars are two standard deviations in overall length. (a) Source-receiver separation of 3 meters. (b) Source-receiver separation of 6 meters. Data in 4b are the mean and SD transformation of the data in Figure 3b.

and part (b) for different-location recordings. Coherence values in one-third-octave bands are shown by open circles, and the measurements of a given run are connected by straight lines. Comparing parts (a) and (b) shows that the variation was greatly increased by randomizing the locations in the room. Figure 3 is unique in that it presents every measurement made in the experiment. The other figures in this report give only the mean and standard deviation. The data in Figure 3b are replotted according to this standard in Figure 4b. By comparing Figures 3b and 4b, the reader may develop the intuition needed to translate a plot of mean and standard deviation into an image of the actual data.

Figure 4 shows the coherence measured in Room 16 for two source-receiver distances, 3 meters in part (a) and 6 meters in part (b). Coherence is almost always less at the larger distance, reflecting the fact that the direct sound, which leads to a coherent image, is a smaller part of the total sound power when the distance is larger.

Figures 3 and 4 show that the coherence is large at low frequencies. At low frequencies the wavelength is long and signals received by the two ears are nearly identical. Therefore, the cross-correlation tends to be large leading to a large value of the coherence. The long-wavelength region extends out to about 500 Hz, where the coherence tends towards zero in a uniform randomized field [17]. The frequency range below 500 Hz will be referred to as the *low-frequency range*. Figure 4a shows that coherence at 500 Hz is well above zero, indicating that the sound field is far from being randomized by the environment.

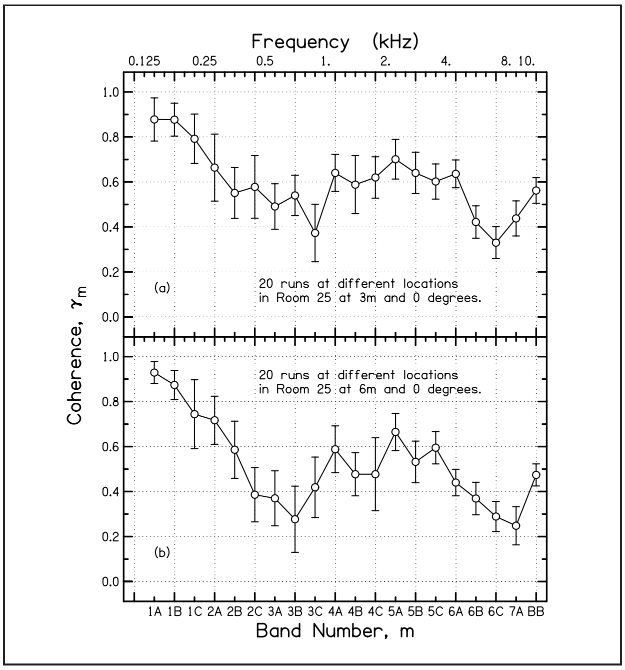


Figure 5. Binaural coherence in Room 25 (classroom) with source and receiver in 20 different locations. The mean coherence is indicated by a circle, and the error bars are two standard deviations in overall length. (a) Source-receiver separation of 3 meters. (b) Source-receiver separation of 6 meters.

### 4.3. Room 25

Figure 5 shows the coherence for Room 25 at two distances, part (a) for 3 meters and part (b) for 6 meters. Outside the low-frequency range, coherence is always less at the larger distance with the exception of band 3C. Figure 5 for Room 25 is in the same format as Figure 4 for Room 16. At the 3-meter distance the coherence is always less in Room 25. At the 6-meter distance the coherence is comparable in the two rooms.

Rooms 16 and 25 are similar. The most important difference is that the floor is carpeted in Room 16 and is vinyl tile in Room 25. A reasonable interpretation of the observations begins by noting that at 3 meters the experiment is near the center of the room, well away from the walls, and important reflections come from the floor. If the floor were perfectly flat, then floor reflections would have the same azimuth as the direct sound and would enhance the direct sound. However, in rooms 16 and 25 the floor was covered with student desks. Consequently the floor was highly irregular in both rooms. Then, because the floor is more reflective in Room 25 the direct sound is a smaller fraction of the total sound and that would explain why the coherence is smaller in Room 25. The importance of the floor is partly attributable to the fact that the ears are low, equal to the ear height for a seated listener. At the 6-meter spacing reflections from the walls play a larger role in the total sound compared to the floor. Because the walls are comparable in the two rooms the values of coherence in Rooms 16 and 25 are similar when measured at the 6-meter distance.

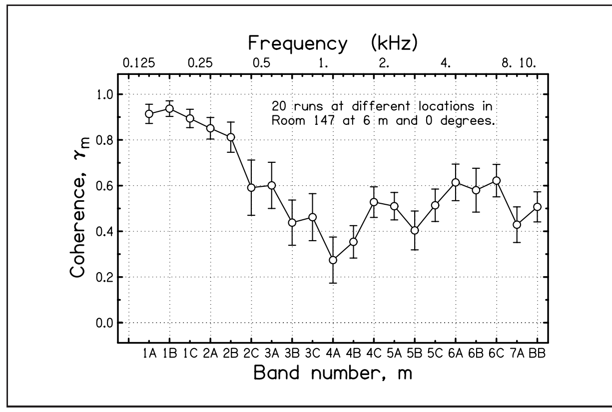


Figure 6. Binaural coherence in Room 147 (lecture hall) with source and receiver in 20 different locations. The mean coherence is indicated by a circle, and the error bars are two standard deviations in overall length. The source-receiver separation is 6 meters.

### 4.4. Room 147

Figure 6 shows the coherence for the large room, Room 147, at 6 meters. Therefore, this figure can be compared with Figures 4b and 5b for the smaller rooms at 6 meters. Comparison shows that coherence is larger in Room 147 in both the low-frequency region and in the high-frequency region. By contrast, in the intermediate frequency region between 1 and 3 kHz the coherence is smaller in Room 147 compared to the smaller rooms. A large value of coherence is favored for Room 147 in this configuration because measurements tended to be concentrated along the midline of the hall, well away from the side walls. The room is wider than it is long, and the diamond shape makes first lateral reflection paths especially long for source and receiver near the midline. It is possible that the unusually high coherence in the 6-10 kHz region (wavelength near 4 cm) is caused by an absorption resonance of the lattice of holes in the brick wall. Figure 7 shows the coherence for Room 147 at 12 meters in two conditions, (a) with the KEMAR on the room centerline within a few meters of the back wall and (b) with the KEMAR near a side wall, with its right ear sometimes less than 1 meter from the wall.

For the 12-meter center location (Figure 7a) the coherence at low frequency resembles that observed for the 6-meter distance, but above the mid-frequency dip the coherence is always smaller at the larger distance. An interpretation of this observation is that the center location is approximately symmetrical with respect to the two ears, and when the wavelength is long the detailed lack of complete symmetry is not apparent. Therefore, coherence at low frequency tends to be high. Evidence in favor of this interpretation is that the variation of coherence with position, as shown by the error bars, tends to be smaller than usual in Figure 7a. At higher frequencies the detailed asymmetry of the environment leads to significant interaural differences in the reflected sound and the direct sound is smaller than at 6 meters because of the longer path length.

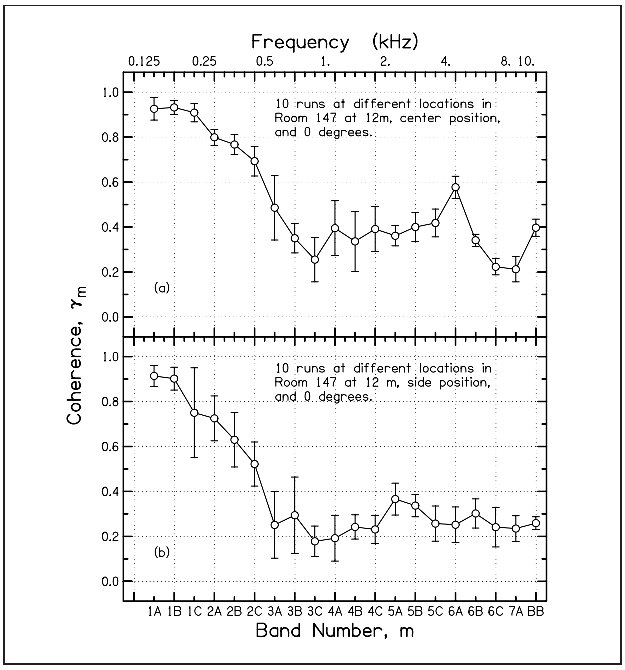


Figure 7. Binaural coherence in Room 147 (lecture hall) with source and receiver in 10 different locations, separated by 12 meters. The mean coherence is indicated by a circle, and the error bars are two standard deviations in overall length. (a) The receiver is in the center, equally distant from the two side walls and in the back row of seats. (b) The receiver's right ear is near a side wall, near the back of the auditorium.

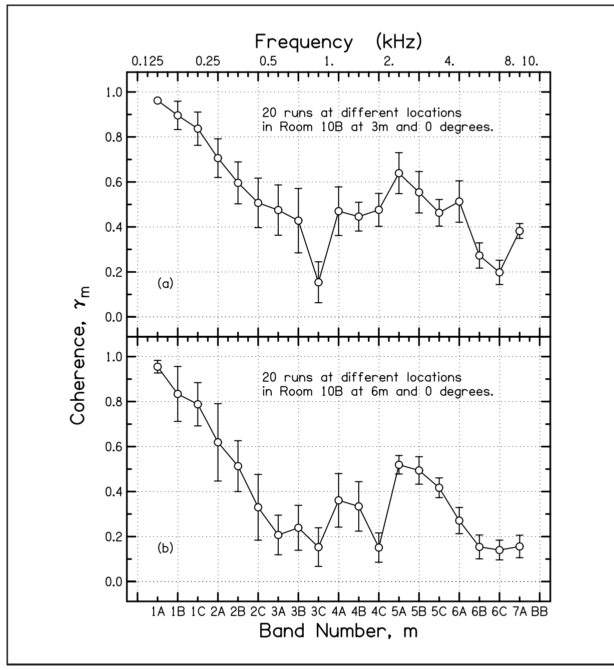


Figure 8. Binaural coherence in Room 10B (lab) with source and receiver in 20 different locations. The mean coherence is indicated by a circle, and the error bars are two standard deviations in overall length. (a) Source-receiver separation of 3 meters. (b) Source-receiver separation of 6 meters.

For the 12-meter side location (Figure 7b) the coherence overall is the smallest seen in any of our conditions except for the reverberation room. As might have been expected, the gross asymmetry introduced by placing the KEMAR near a wall led to large interaural differences and a small binaural coherence. In the low-frequency region the coherence is comparable to that measured in Rooms 16 and 25 at a 6-meter distance - similar situations in that side walls were nearby. Side walls appear to have a perturbing effect on the familiar form of the coherence in the low-frequency region. At higher frequencies the coherence in 147 is considerably smaller.

### 4.5. Lab 10B

The coherence for Room 10B is shown in Figure 8, part (a) for a distance of 3 meters and part (b) for a distance of 6 meters. With the exception of band 3C, the coherence is always smaller at the larger distance as expected. Room 10B has almost the same volume as Room 16, but the coherence is smaller in 10B, presumably because none of the surfaces in 10B are deliberately absorbing. In the low-frequency range, the coherence in 10B is similar to that in Room 25, both at 3 and 6 meters. Both rooms have vinyl floors. However, at higher frequencies the coherence is smaller in Room 10B.

Figure 8a shows a sharp dip in band 3C, centered on 900 Hz. This dip is a somewhat enlarged version of dips in this band seen in Figures 4a and 5a, all for 3-meter distances. A plausible interpretation of the dip in this band is that it occurs when lateral reflections are particularly important and isotropic. Like Rooms 16 and 25, Room 10B is of a size such that lateral reflections can arrive about equally from all directions when the distance is 3 meters, but tend to arrive preferentially from a nearby wall when the distance is 6 meters. Therefore, the effect of isotropic reflections is best seen in coherence measured at 3 meters, and special reflections may dominate at 6 meters. Unlike Rooms 16 and 25, Room 10B has little absorption on its lateral surfaces. Surfaces in 10B are hard, though they tend to be irregular due to a lot of large laboratory equipment along the walls. Because lateral reflections are more intense in Room 10B the above interpretation would predict a deeper dip in band 3C for 10B compared to Rooms 16 and 25, as observed.

### 4.6. Reverberation room

The coherence measured in the reverberation room, and shown in Figure 9, is the smallest ever seen in our study. In the low-frequency region, both 3- and 6-meter coherence measurements are comparable to that measured in Room 147 with the head at 12 meters and near the side wall. At higher frequencies the coherence in the reverberation room is smaller than anywhere else. This result is not unexpected because the direct to reverberant ratio is smaller in the reverberation room than in any other room.

Outside the low-frequency region, the coherence in the reverberation room is smaller for the 6-meter distance than at the 3-meter distance, but it is not much smaller. The broadband data serve as a summary. The BB data in Fig-

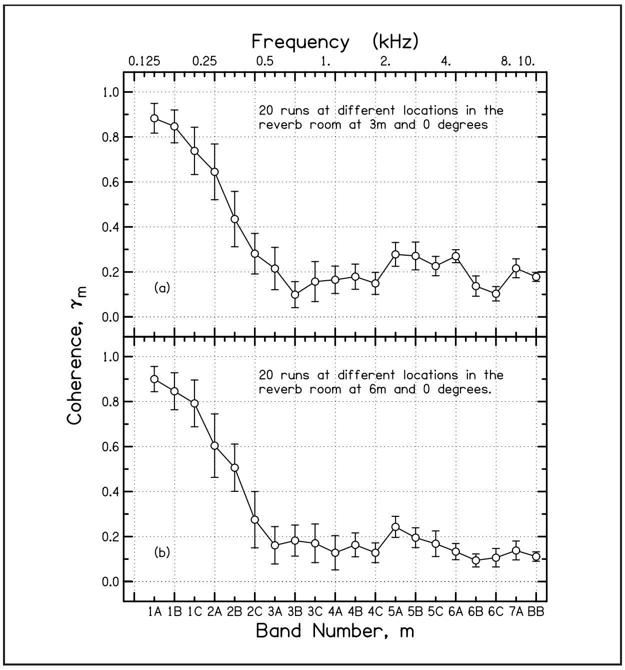


Figure 9. Binaural coherence in the Reverberation Room with source and receiver in 20 different locations. The mean coherence is indicated by a circle, and the error bars are two standard deviations in overall length. (a) Source-receiver separation of 3 meters. (b) Source-receiver separation of 6 meters.

ure 9 lead to a 6 m/3 m ratio of 0.62. Returning to the original measurements made in this room shown in Figure 1, the averages for the 24 loudspeakers lead to a ratio of 0.51. These two ratios for the coherence at 6 m compared to 3 m are similar and they are both higher than expected on the basis of the simplest room model.

In the simplest room model, the reverberant sound is completely incoherent so that the cross-correlation for the reverberant part of the sound is zero. The only finite contribution comes from the direct sound. In the reverberation room, the total sound intensity (direct plus reverberant) is mostly reverberant and a sound level meter shows that it is the same at 3 meters and 6 meters. By contrast, the direct sound intensity is four times smaller at 6 meters than at 3. Therefore, one expects that the cross-correlation at 6 meters should be one fourth that at 3 meters, much smaller than the observed ratios of 0.61 or 0.51. It is apparent that the sound field in the reverberant room was not ideally diffuse. The two-point correlation measurements of Cook *et al.* [18] showed that an adequately diffuse field cannot be obtained without moving vanes, which our room did not have, though Cook et al. did not randomize the position as we did. It seems possible that for the 6-meter distance the KEMAR was too often close to the back wall, which tended to lead to reflections that were equivalent in the two ears and thereby increased the coherence.

## 5. Discussion

Given the goal of measuring coherence in rooms to gain insight into mechanisms of sound localization, this discussion considers the applicability of the measurements to the usefulness of the ITD localization cue. Although the focus of the present article is on coherence, this discussion gained perspective from an inspection of the measured ITDs. Both the coherence and the ITDs originate in the peaks of the cross-correlation function. The coherence is the height of a peak, the ITD is the lag at which the peak occurs.

### 5.1. Variation in measured coherence

Except for Figure 3a, the figures in this report show the results of experiments where the locations of source and receiver were randomized. However, the experimental situations were also studied with the source and receiver in fixed positions. Twenty runs were done for both randomized and fixed conditions. The object of the exercise was to discover how much of the variability could be assigned to the difference in locations. Figure 3 for Room 16 gives an example of the considerable difference between results at the fixed and randomized locations – on the average a factor of six in variation. Comparisons between fixed and randomized locations made in other rooms lead to similar impressions – most of the variation in the coherence measured with randomized locations is the result of the different locations. However, the comparison in other rooms is not as dramatic as in room 16. Generally, the variation in coherence for fixed-location conditions − as evidenced by the standard deviation, or half the length of the error bars in the figures − is about 0.02. For dry rooms like 16 and 25, the variation for the randomized-location conditions is about 0.1. In more reverberant environments, like the reverberation room, the variation caused by different locations is less – about 0.07. These place-to-place variations in coherence are large enough that they could have an idiosyncratic effect on sound localization via ITD. The fact that listeners are not aware of such variations is reminiscent of the observation by deVries *et al.* [11] who found dramatic location dependence of the IACC in the Concertgebouw of which listeners are said to be unaware.

Initially, it was not obvious why the same-location experiments (e.g. Figure 3a) exhibit as much variability as they do. Possibly the variation is caused by thermal currents in the room. Possibly it is caused by the sample-to-sample variation in the stimulus noise due to the randomization of component phases across runs. To distinguish between these explanations, we performed a frozen-noise experiment in which the stimulus phases were not changed over the course of 20 runs. This experiment was performed in the reverberation room at the 6-meter distance with a timing identical to all the other same-location runs. The results of the the frozen noise experiment showed very little variation, an order of magnitude less than that obtained in the corresponding experiment wherein the locations were fixed but the phases were randomized. Therefore, it became evident that essentially all of the variation seen in the same-location experiments was caused by stimulus variability and almost none was caused by temporal variations in the acoustics of the room itself.

### 5.2. The low-frequency limit

Inevitably, the coherence is high for low frequencies. In the low-frequency limit, the wavelength is long, the signals at the two ears are nearly identical, and the coherence is high. However, the fact that the coherence is high does not necessarily mean that listeners can use the ITD for localization because the actual ITD may itself be totally misleading. Although the waveform phases are similar in the two ears, the ITD, defined as a phase delay, is equal to the phase difference divided by the frequency. Therefore, there may still be a significant ITD because the frequency is low. The ITD may contradict the ILD and be discounted. According to our view of sound localization in rooms, this is exactly what happens in all but the driest rooms and this is responsible for the Franssen effect and other steady-state location confusions [19].

The long-wavelength region extends out to about 500 Hz, where the coherence shows a local minimum [17]. The reverb-room Figure 9 shows this effect best. If the coherence near and above 500 Hz is appreciably greater than the reverb-room value (about 0.2), then the room probably supports consistent and useful ITDs because high coherence at such frequencies is not a long-wavelength effect. Then, according to Constan's headphone experiments, the waveform ITD cue should be influential. Room 16 (Figure 4) provides an interesting example where the waveform ITD cue is likely to be influential at 3 meters but not at 6 meters.

Within the long-wavelength region below 500 Hz the span of ITDs measured in our environments was often large. The spans were found to correlate with the coherence. For instance, coherence was high at low frequencies in Room 16 at the 3m distance (Figure 4a). The low-frequency spread in ITD was correspondingly small - about 400$\mu$s. By contrast at the 6m distance (Figure 4b), the coherence dropped rapidly in the 300–500 Hz region. The span of ITD values was 800 $\mu$s there, rising to more than 1200 $\mu$s in bands 2C and 3A, where the coherence function was especially low.

But although the variation in ITD was smaller when the coherence was higher, the span of ITDs in the region below 500 Hz was never tiny. It was always 200$\mu$s or more, corresponding to at least 15 degrees of azimuth. This variation is large compared to broadband psychophysical angular difference limens of several degrees. Therefore, although the low-frequency peaks may have coherence that is high enough to make them influential, these peaks are not necessarily useful.

### 5.3. The special nature of the region near 500 Hz

A survey of the binaural literature shows that experimenters have often focused on the 500-Hz region for listening experiments. That is because the binaural system, especially the ITD-sensitive part, works best in that region [20]. The special role of 500 $(\pm 250)$ Hz in binaural neural processing may not be entirely fortuitous. Because of the size of the human head, this frequency range corresponds to a minimum in coherence when the sound field is isotropic. If a listener is required to localize a source in the presence of an interfering reverberant field that is approximately isotropic, then any peak that occurs in this frequency region is likely to come from the direct sound from the source and not from the environment. Consequently it is to the listener's advantage to pay special attention to the 500-Hz frequency region. To our knowledge this observation has not been made previously.

### 5.4. Mid frequencies and high frequencies

As the frequency of a sound increases beyond 1000 Hz, there is a substantial degradation in the ability of the binaural system to make use of ITD in the waveform. Timing in the fine structure of a tone or noise ceases to be of value. Instead, listeners are able to make use of ITD in the envelope of sounds. If there is no structure in the envelope, as for a continuous sine tone, then listeners cannot localize. For noise, like the third-octave noises used in our experiments, the ITD in the temporal fluctuations can be used.

Given the significance of envelope ITDs at mid and high frequencies, it would seem that the waveform cross-correlation and waveform coherence, as measured in the experiments reported here, are less interesting than the cross-correlation and coherence of the envelope. However, the coherence of the waveform and of the envelope are statistically related. Because the envelope always has a finite average value two different forms of the envelope need to be considered.

In the first form of envelope coherence, the average value is subtracted off before the cross-correlation is computed. This calculation leads to the cross covariance or Pearson product moment. Then Monte Carlo calculations find that the envelope coherence is very close to being equal to the square of the waveform coherence. For more precision, the following formula may be used,

$$\gamma_E = \gamma^2 - 0.25\gamma^3(1-\gamma),$$

where $\gamma_E$ is envelope coherence and $\gamma$ is the waveform coherence as measured in our experiments.

In the second form of envelope coherence the average value is retained in each envelope before the envelopes are cross-correlated. This is the form of envelope coherence recommended by Bernstein and Trahiotis [21] to model the ability of listeners to detect incoherence. Corresponding Monte Carlo calculations for the envelope (or amplitude) coherence agree with a formula found by Bernstein [22],

$$\gamma_A = 0.78 + 0.22\gamma^{1.853}.$$

Because the average value is retained, the amplitude coherence is never less than 0.78 regardless of the coherence of the waveforms.

These Monte Carlo calculations are based on mixtures of independent noise samples chosen to target a range of $\gamma$ values. The formulas above are independent of bandwidth, though deviations from the formulas for individual noise samples increase with decreasing bandwidth-duration products. The calculations appear in a report on

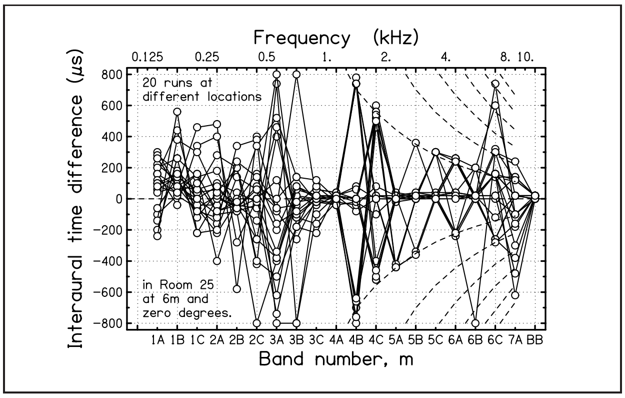


Figure 10. Interaural time difference (ITD) in Room 25 for 20 different locations with separation of 6 m, as in Figure 5b. The ITD is determined by the lag of the largest peak in the CC function. Dashed lines at $\pm n/f_c$ as a function of the band center frequency $f_c$ show loci of expected alias points.

the mathematical properties of the cross-correlation function [23]. As a consequence of these relationships between envelope coherence and waveform coherence, the waveform coherence, as measured in the experiments of this article, contain information that is relevant at high frequencies where the time structure in the envelope is more important perceptually than the time structure in the waveform.

The coherence of the waveform is useful also in that Constan's headphone experiments were reported in terms of waveform coherence. Constan [13] studied the ability of listeners to make use of ITD cues in lateralizing bands of running noise as a function of the noise coherence. His results showed that in order for listeners to distinguish between ITDs of +200 $\mu$s and $-200\mu$s, coherence $\gamma$ had to be greater than 0.6 when the band was above 1500 Hz. This critical frequency range 2-4 kHz corresponds to the mid-frequency bump in the coherence shown in the figures of this report. As noted previously, the range of $\pm 200\,\mu$s corresponds to about $\pm 15$ degrees of azimuth. This comparison allows us to conclude that unless the mid- or high-frequency coherence breaks through the $\gamma = 0.6$ barrier, the ITD envelope cue is of little use to a listener. For resolution better than $\pm 15$ degrees the coherence should be still higher. Based on this criterion, we think it likely that listeners can benefit from the ITD in the envelope of running noise in Rooms 16 and 25, but perhaps not in the other environments described in this report.

### 5.4.1. Interaural time differences

Binaural coherence, as highlighted in this report, is determined by the height of the tallest peak in the cross-correlation function. A second feature of a peak is the value of the lag for which it occurs. This value corresponds to the interaural time difference (ITD), which serves as a potential cue for azimuthal localization. Such ITD values were measured along with the peak heights in our experiments. The purpose of the present section is to describe the trends that were observed. A representative plot appears in Figure 10, showing the measured ITD values for 20 different runs at different 6-meter locations in Room 25. This plot is based on the same peaks as Figure 5b.

The measured ITD values are most easily described by beginning at the right hand edges of the plots and progressing toward lower frequencies. First, for the broadband (BB) peaks, ITD values were always near zero, within 20 or 40 $\mu$s, as expected for a source at zero azimuth. This result held good for all environments and distances in our experiments. Deviations from zero can all be attributed to small errors in alignment of the head. Therefore, like the data shown in Figure 1, the BB measurements in the different rooms demonstrated that if the human binaural system were somehow able to make use of a broadband cross-correlation, the ITD would be a powerful sound localization cue even in a reverberation room. A robot with access to broadband information could make better use of ITD cues than a human listener.

In some respects, the ITD values in one-third-octave high-frequency bands, 5A and above, resembled those found in the broad band. However, in the high-frequency bands CC peaks often occurred at lag values that are near integer multiples of the band center period (reciprocal of the band center frequency), as shown by points near the dashed lines in Figure 10. Such aliasing never occurs for a broad band. This aliasing effect causes the high-frequency ITD data to appear rather scattered, but also highly systematic because the aliases fall near distinct frequency-ITD hyperbolas (dashed lines). Aliases become more common as the coherence decreases.

It is easy to prove that the CC peaks ought to appear at zero ITD and not at the alias values [24]. For a one-third octave band, the $n$-th alias ($n = \pm 1, \pm 2, \ldots$) should be down by $\sin(0.23\,\pi n)/(0.23\,\pi n)$. However, the randomness that leads to small coherence values in the first place can apparently cause an alias peak to be taller than the main peak ($n = 0$).

As defined in the present article, coherence is based on peaks that fall within the $\pm 800\,\mu$s range of lags. An attractive alternative definition would restrict the range of lags to exclude alias peaks. That choice would have reduced the value of the coherence reported in Figures 3-9 because all tall alias peaks that dominated the peak at zero lag would have been excluded. It is unknown how large an effect that would be.

The conclusion from the high-frequency CC functions is that highly precise ITD information would be available to a listener, assuming that the listener could ignore the aliases and could cope with small peak height. Very often the span of ITD values (modulo a band center period) was less than $100\,\mu$s. For example band 5C (centered on 3600 Hz) filled this condition for every room and distance tested. However, from Constan's experiments it is known that the peaks must be tall, with coherence greater than 0.6, for listeners to make use of high-frequency ITD cues, and the coherence values measured in the rooms of this article are often not high enough.

At mid frequencies, between 500 and 2000 Hz, the span of ITD values tends to correlate with the value of coherence. For bands with a wide span of ITDs the coherence tends to be small, but the correspondence is not rigorous because incoherence can arise from interaural amplitude variations as well as interaural phase variations. Dramatic dips in coherence, as for band 3C in Figure 8a, generally correspond to large spans in ITD *circa* 800 $\mu$s.

At low frequencies, 500 Hz and below, the span of ITDs is normally larger than 200 $\mu$s, illustrating the point made previously that the ITD may be quite uncertain even if the coherence is large. An exception occurred for the center of Room 147 (Figure 6) where the low-frequency coherence is unusually high and the span of ITDs is not more than 200 $\mu$s.

## 6. Conclusion

The binaural cross-correlation function was measured in five different rooms at two different distances between the source and receiver. The binaural cross-correlation function is interesting because of the possibility that something like it is used by human listeners to localize sounds using the interaural time difference (ITD) of steady-state complex sounds. The measurements were particularly concerned with the coherence, which is the height of peaks in the cross-correlation function, but they also determined the ITD value, which is the lag at which the peaks occur. The measurements suggest that listeners should have difficulty using the steady-state ITD to localize. In low-frequency bands, the coherence is adequate but the ITDs show a large span of values that should lead to confusion. In high-frequency bands the span of ITD values is much smaller, but the coherence is inadequate for many of our room configurations. The division between high- and low-frequency regions depends on the room.

### Acknowledgement

Millicent Ow, Wendy Thorpe, and Tim Vander Velde contributed to some of the measurements in this report. We are grateful to Barbara Shinn-Cunningham and Adelbert Bronkhorst for useful comments on a earlier version of this manuscript. The work was supported in part by the United States NIH NIDCD.

### References

[1] J. W. Strutt: On our perception of sound direction. Phil Mag. **13** (1907) 214–232.

[2] J. W. Strutt: On our perception of the direction of sound. Proc. Roy Soc. **83** (1909) 61–64.

[3] F. L. Wightman, D. J. Kistler: The dominant role of low-frequency interaural time differences in sound localization. J. Acoust. Soc. Am. **91** (1992) 1648–1661.

[4] L. A. Jeffress: A place theory of sound localization. J. Comp. Physiol. Psychol. **41** (1948) 35–39.

[5] J. M. Goldberg, P. B. Brown: Response of binaural neurons of dog SOC to dichotic tonal stimuli: Some physiological mechanisms of sound localization. J. Neurophysiol. **32** (1969) 613–632.

[6] T. C. T. Yin, J. C. K. Chan: Interaural time sensitivity in medial superior olive of cat. J. Neurophysiol. **64** (1990) 465–488.

[7] H. S. Colburn, Y. Han, C. Culotta: Coincidence model of MSO responses. Hear. Res. **49** (1990) 335–346.

[8] L. Beranek: Concert halls and opera houses. Springer, New York, 2002. p. 615.

[9] J. Blauert, W. Lindemann: Auditory spaciousness: Some further psychoacoustic analyses. J. Acoust. Soc. Am. **80** (1986) 533–542.

[10] Y. Ando: Architectural acoustics - Blending sound sources, sound fields, and listeners. Springer, New York, 1998.

[11] D. de Vries, E. M. Hulsebos, J. Baan: Spatial fluctuations in measures for spaciousness. J. Acoust. Soc. Am. **110** (2001) 947–954.

[12] J. Koehnke, H. S. Colburn, N. I. Durlach: Performance in several binaural interaction experiments. J. Acoust. Soc. Am. **79** (1986) 1558–1562.

[13] Z. A. Constan: All things coherence. Thesis, Michigan State University (unpublished), 2002.

[14] G. F. Kuhn: Model for the interaural time differences in the azimuthal plane. J. Acoust. Soc. Am. **62** (1977) 157–167.

[15] G. F. Kuhn: The pressure transformation from a diffuse sound field to the external ear and to the body and head surface. J. Acoust. Soc. Am. **65** (1979) 991–1000.

[16] Z. A. Constan, W. M. Hartmann: On the detection of dispersion in the head-related transfer function. J. Acoust. Soc. Am. **114** (2003) 998–1008.

[17] I. M. Lindevald, A. H. Benade: Two-ear correlations in the statistical sound fields of rooms. J. Acoust. Soc. Am. **80** (1986) 661–664.

[18] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, J. Thompson, M. C.: Measurement of correlation-coefficients in reverberant sound fields. J. Acoust. Soc. Am. **27** (1955) 1072–1077.

[19] W. M. Hartmann, B. Rakerd: Localization of sound in rooms IV - The Franssen effect. J. Acoust. Soc. Am. **86** (1989) 1366–1373.

[20] J. Raatgever: On the binaural processing of stimuli with different interaural phase relations. Thesis, Dutch Efficiency Bureau - Pijnacker, Delft, The Netherlands, 1980.

[21] L. R. Bernstein, C. Trahiotis: The normalized correlation: Accounting for binaural detection across center frequency. J. Acoust. Soc. Am. **100** (1996) 3774–3784.

[22] L. R. Bernstein. personal communication, 2004.

[23] W. M. Hartmann: Mathematical properties of the cross-correlation function. MSU Psychoacoustics Report 103, 2004.

[24] W. M. Hartmann: Signals, sound, and sensation. Springer, New York, 1997. p. 343 ff.