A STUDY OF BINAURAL INTERACTION*

- The Relationship between Directional Hearing and Binaural Synthesis -

Ichiro Kirikae, Kenji Nakamura, Tsunemasa Sato and Tetsuya Shitara

The directional hearing and binaural synthesis tests are used in differentiating central deafness from other types of deafness. When there is a lesion somewhere in the retrocochlear auditory pathway, both of these tests reveal invariably poor performance. The purpose of our study is to clarify whether this apparent correlation is merely a coincidence or whether it represents any inherent relation between the two auditory processes in the underlying mechanism.

In the directional hearing tests in this study, the same signal was presented separately to both ears with a lag of time between the two ears. In the binaural synthesis tests, on the other hand, two different signals were presented separately to each of the ears. In spite of the apparent difference between the two tests, the results show clear correspondence in the following aspects.

First, the formation of a sound image as a sensation of localization of the subjective sound source is inherently associated with a sensation of a merged and integral sound quality of the dichotic stimuli. Second, for two different signals simultaneously presented to the two ears to be perceptually synthesized, the spectral characteristics may be vastly different but the repetitive occurrences of waveform envelopes in the time domain must be approximately synchronous.

We aimed at investigating the conditions of sound image formation and binaural synthesis in respect to different criteria including vowel identification. Our experiments were performed in the following manner. Our test material consisted of vowel-like stimuli obtained by feeding periodic pulses through a set of band-pass filters and by amplitude- and frequency-modulation of sinusoidals (Fig. 1).

---

* Oral paper presented at the Xth International Congress of Audiology, held at Dallas, U. S. A., on 12-15, October, 1970.

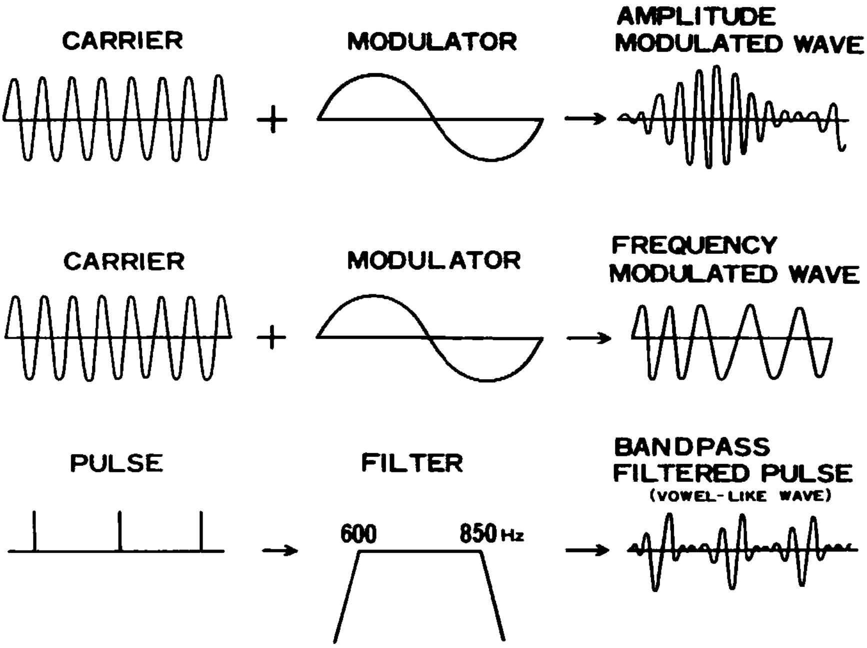


Figure 1. Wave forms of test sounds.

Under various conditions, in terms of coherence of two different signals as periodic time functions and their spectral envelopes, the perceptual feasibility of sound-image formation was investigated. For the binaural synthesis of tones, we employed synthetic Japanese vowels separated into two spectral components each representing one of the two lowest formant regions. The results of these experiments were as follows.

Our first experiment was designed to see whether it was possible to create a sound image using two complex sounds presented separately to both ears. When the occurrences in time of the repeated envelopes of the waveforms between the two signals slightly disagreed, binaural beats were perceived. Using these binaural beats as the criterion, the conditions for the sound image formation were examined.

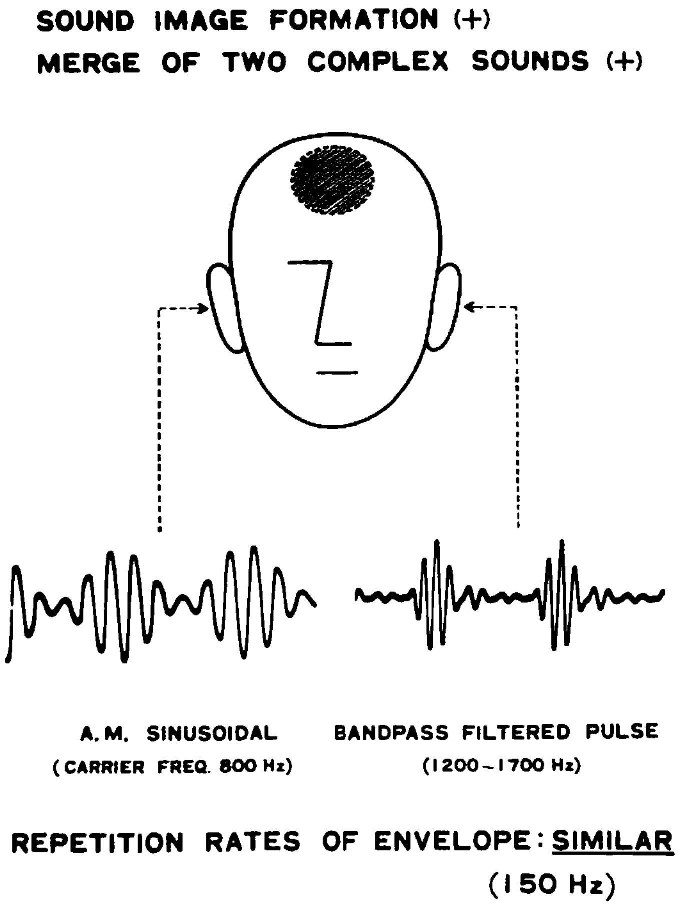


Figure 2. Sound image formation for two complex sounds presented separately to both ears.

The sound image of two complex sounds is readily produced when approximate coherence between the two waveforms is observed. For the complex sounds to be compared, we used periodic pulses, amplitude-modulated and frequency-modulated sinusoidals (Fig. 2).

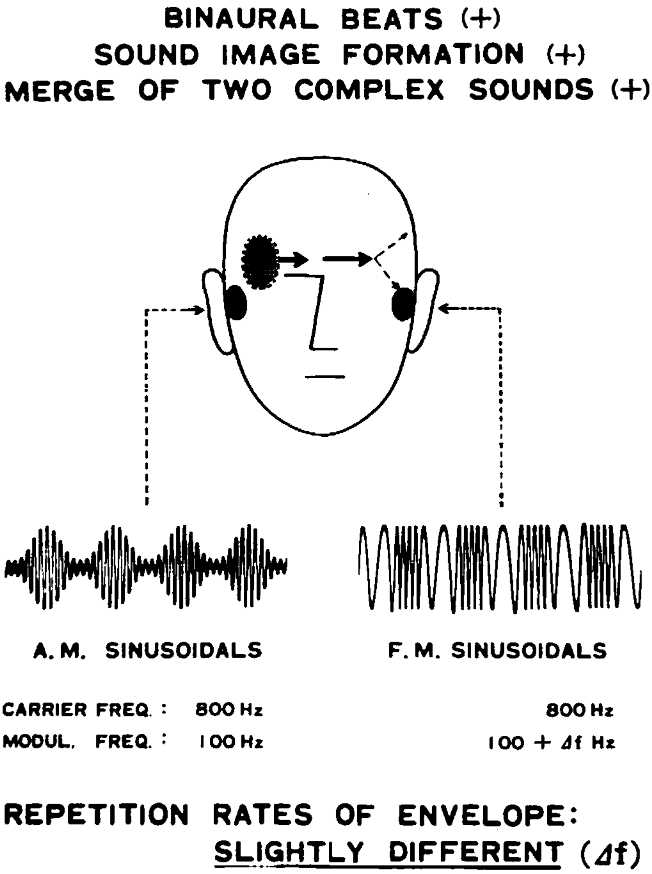


Figure 3. Sound image shift by the difference in repetition rates of envelope between the dichotic signals.

These periodic complex sounds have spectra with quasi-harmonic components concentrated in limited frequency bands. When the coherence was imperfect and the rate of repetition of the waveform envelopes between the dichotic signals was slightly different, the sound image moved and lateralized periodically. The periodicity of this lateralization agreed with the beat frequency, viz. the difference of the two repetition rates. In these cases, it may be mentioned that the spectral structures for the two signals may be substantially different: i. e. there may be no apparent correspondence in terms of individual pure-tone components. The sound image moved from the side closer to the ear with the lower repetition rate to the other, until it split into two unrelated sounds, and this sequence of sound image formation was repeated. When the sound image is produced, the two complex tones may be said to be perceptually synthesized into a single tone (Fig. 3).

Our second experiment is concerned with the determination of the maximum permissible synchronized dichotic signals for formation of a single sound image. Figure 4 is a block diagram of the experimental setup.

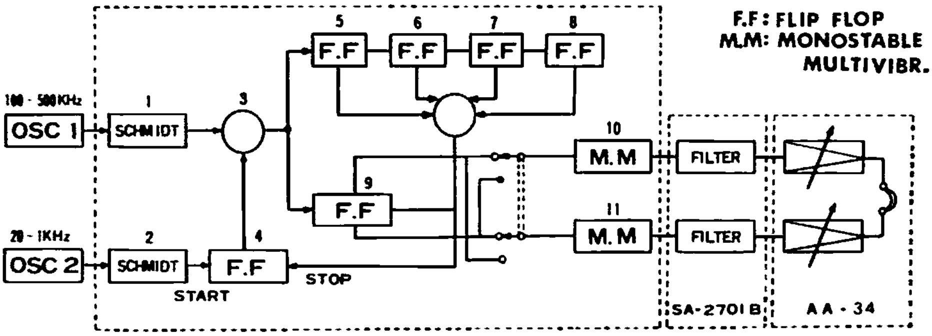


Figure 4. Blockdiagram of the setup for the second experiment.

This system, which was originally designed for evaluating the directional hearing in pathological cases, was used in this experiment for investigating the sound image formation of normal subjects. This system is a combination of a pulse train generator and two half-octave band pass filters. One of the outputs from the filters can be given adjustable time delay in respect to the other.

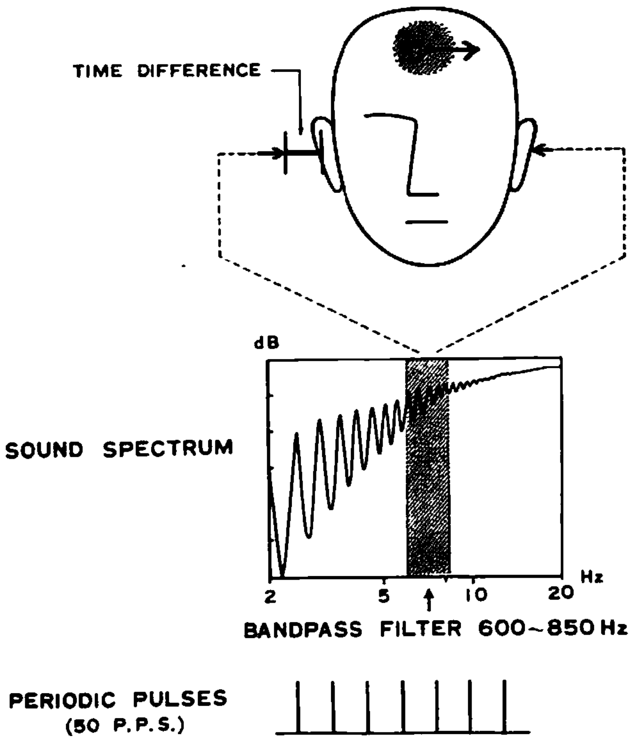


Figure 5. Presentation of bandpass filtered pulses with interaural time difference.

Figure 5 shows the central frequencies of the bandpass filters. In this experiment, the two filtering conditions for the two ears were identical, and the only difference between the dichotic signals was in the amount of delay. Figure 6 shows threshold values in terms of the delay for shift of sound lateralization and also for sound image splitting. It is easier to detect a shift in sound lateralization at a lower frequency band--600 to 850 Hz rather than at a higher frequency band--1200 to 1700 Hz. Thus the threshold for discriminating sound lateralizations is lower at lower frequency bands. The splitting of the sound image, on the other hand, more easily occurs at higher frequency bands, namely the threshold value for splitting is smaller at higher frequency bands (Fig. 7).

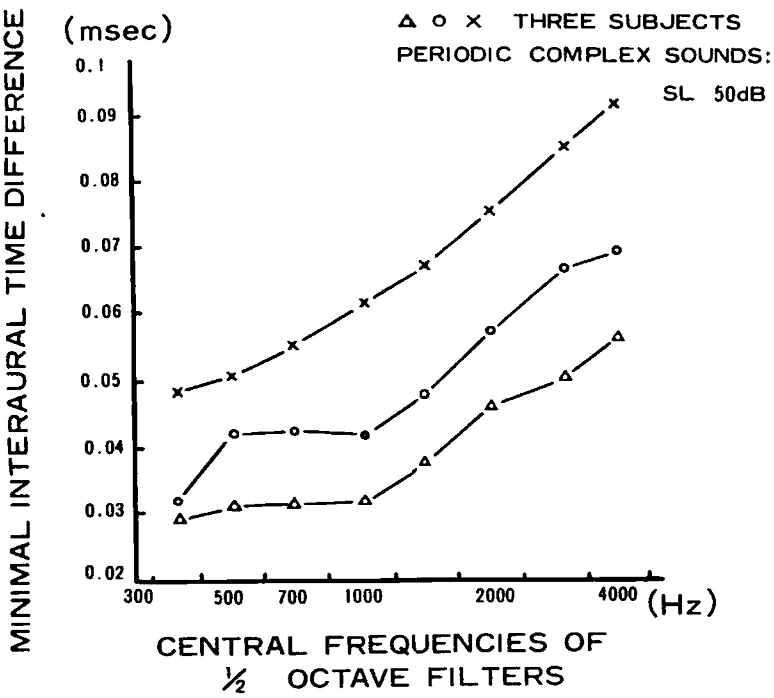


Figure 6. Thresholds of sound lateralization by interaural time difference.

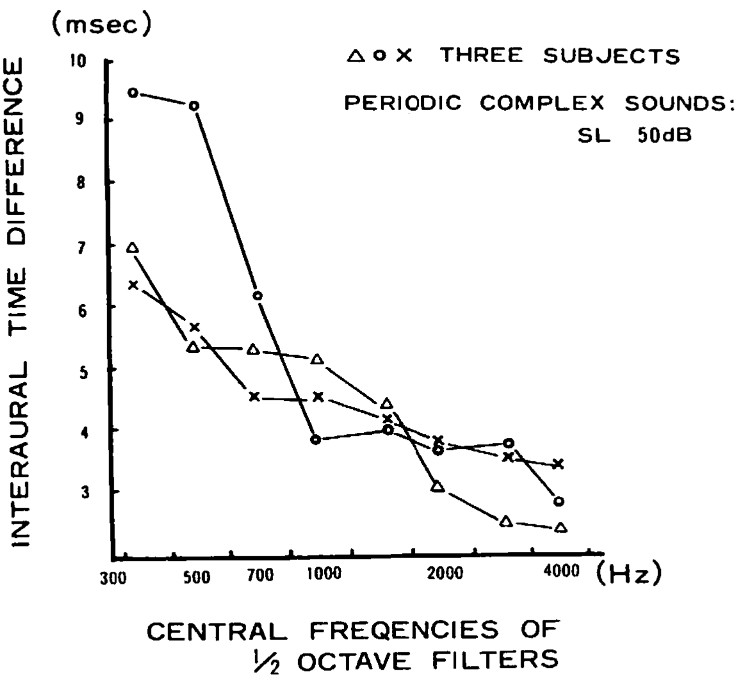


Figure 7. Maximal limits of time difference for sound image formation: slipping points of sound image.

We assume that this was due to the fact that in lower frequency bands, not only the burst of the nerve impulses, but also all single impulses are synchronized to the low frequency stimulus waveform.

In another series of test sessions of the experiment, the two filters were set for different pass bands and thus the spectral envelopes of the dichotic signals were different. In this case, the sound image either lateralized to the side of the ear that received higher frequencies or splitted into two unrelated images. The splitting was observed when the two frequency bands were widely separated.

In the former case, in order to produce a sound image at the mid-occiput, the lower frequency signal had to precede by a certain time difference the higher frequency signal. This phenomenon can be accounted for by the difference in the traveling time within the inner ear for higher and lower frequencies to be perceived. The delay time was varied and the permissible range for sound image formation was evaluated to be about 2 msec. Under these conditions, not only the localization of the sound image is perceived, but also typically, a fused timbre is perceived as different from either of the two separate tones.

Our third experiment was on the monaural synthesis of two complex sounds of different frequency bands. This was performed by feeding the two filter outputs through a mixer and presenting them to the ear. When the time difference between the sounds was within the range of sound formation in binaural synthesis, the two sounds merged into a single tone and assumed a vowel-like quality and when the time difference was out of range, the merging did not take place.

Our last experiment was on mono- and binaural tonal synthesis using synthetic Japanese vowels (Fig. 8). As in our second and third experiments, we presented two complex sounds either mono- or binaurally in order to investigate tonal synthesis. This time, we used synthetic Japanese vowels as our test material and performed vowel identification tests.

Synthetic vowels were produced by a series-type terminal analog speech synthesizer. The first and second formants were used as variable parameters, while the third and fourth formants were held constant at 2, 700 Hz and 3, 500 Hz, respectively. The fundamental frequency was set at 140 Hz. A set of 64 synthetic stimuli was chosen by combining different

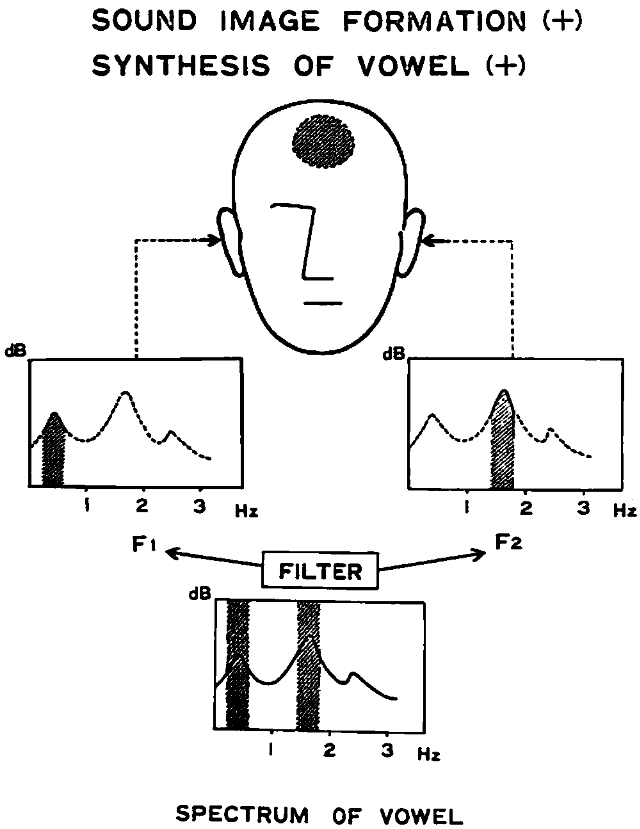
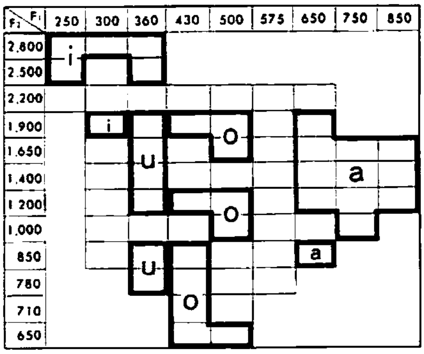
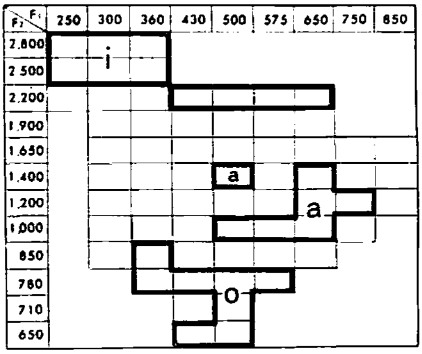


Figure 8. Separate presentation of F1 and F2 of a synthesized vowel.

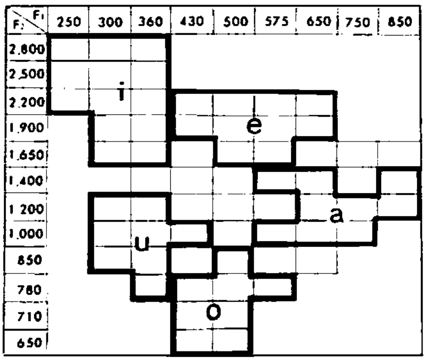
values of Fig. 1 and Fig. 2. The stimuli were tape-recorded in random order. The synthetic vowel signal was passed through a band-pass filter set at 425 - 600 Hz, 600 - 850 Hz,or 850 - 1, 200 Hz in accordance with the first formant frequency of the stimulus, and the output was recorded on one channel of a sterephonic recorder. The output of the other band pass filter set at 600 - 850 Hz, 850 - 1, 200 Hz, 1, 200 - 1, 780 Hz, 1, 700 - 2, 400 Hz, or 2, 400 - 3, 400 Hz, was recorded on the second channel.



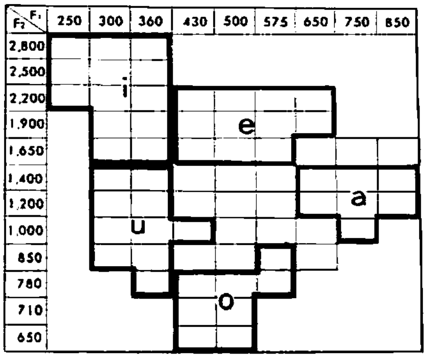
9a F1 ONLY (MONAURAL)



9b F2 ONLY (MONAURAL)



9c F1 AND F2 MIXED (BINAURAL)



9d F1 AND F2 SEPARATE (DICHOTIC)

Figure 9. Identification of five Japanese vowels.

Figure 9a shows the result when only the first channel, i. e. the F1 band was presented to the ear. Figure 9b, when only F2 was presented.

Ten young adults with normal hearing participated in these tests. The areas within the heavy lines indicate the condition under which more than 80% of the responses gave consistent vowel identification.

Figure 9d shows the result obtained when F1 was fed to one ear and F2 to the other, simultaneously. In contrast to the monaural hearing, all the five Japanese vowels were correctly identified. Vowel /e/, which was not identified monaurally, was now clearly identified. Incidentally, it can be shown by using single formant stimuli for vowel identification that this particular vowel cannot be perceived unless the two formant regions simultaneously exist.

Figure 9c shows that the result of binaural tests with the same stimuli of mixed F1 and F2 is similar to the dichotic test with F1 to one ear and F2 to another ear.

We may summarize the results of our various experiments as follows:

1. Sound image is formed when a (single) complex sound is presented simultaneously to both ears. An image is also formed when two complex sounds of different bands of frequency distribution are fed to the ears, provided that the repetition of the waveform envelopes are synchronized. The mechanisms of these two methods of sound image formation are concluded to be basically mutually common.

2. In binaural experiments, sound image formation and tonal synthesis are made possible under similar conditions. When two complex sounds have the same rate of envelope repetition, the fusion takes place under the condition that the discrepancy between the envelope peaks is within a certain limit of time.

3. The mechanisms of monaural and binaural tonal syntheses are similar.

4. Binaural synthetic tests of vowel stimuli involve sound image formation as well as speech discrimination.