Ann. Bull. RILP (1983) No. 17, 123–128

LOCALIZATION OF PURE TONES, WHITE NOISE, SINGLE RESONANCE TONES, AND SYNTHETIC VOWELS*

*Sumi Shigeno*** and *Naobumi Sakamoto***

Introduction

It is widely known that the perception of speech sounds is different from that of non-speech sounds in various kinds of psychological phenomena such as categorical perception, context effects, and recognition masking. To explain these results, many types of models have been suggested.

On the other hand, many studies have been conducted on our ability to localize non-speech sounds. The results of these studies indicate that the perception of a sound's direction rests upon three factors: interaural differences in the phase, time, and intensity of the tone (Mills, 1958). It has also been indicated that the detection of a change in the spatial location for noise is more accurate than for musical tones, and even more so than for pure tones (Hornbostel, 1926). However, few studies have looked at the localization of speech sounds. In monkeys, Brown, Beecher, Moody & Stebbins (1978) found that acuity of localization was a function of the magnitude of the change in pitch (frequency modulation) of different clear calls.

The purpose of this investigation is to compare the localization of speech and non-speech sounds under the same experimental conditions and to discuss the effects of both the physical characteristics of the sounds and the subjective factors involved in the localization of sounds.

Method

Subjects

Four adult subjects with normal hearing (tested with an audiometer) participated in this experiment. None of them had a past history of speech or hearing problems. Prior to testing, they were familiarized with synthetic sounds.

Apparatus

The experiment was conducted in a soundproof room. Fig. 1 shows the layout of the two loudspeakers, one of which was the reference and the other was the comparison source. They were placed within a radius of three meters from the subject, whose face was held in a fixed position in the sound field by a chin-and-head rest. The reference speaker was placed in front of the subject, and the comparison

---

* A part of this paper is a version of the paper submitted to Japanese Psychological Research.

** Department of Psychology, Faculty of Letters.

Stimuli

Four kinds of stimuli were used; synthetic vowels as speech stimuli, and as non-speech stimuli pure tones, white noise, and single resonance tones. A PDP 11 computer (DEC) was used to produce the stimuli; the synthetic vowels were generated by computer simulation of a terminal-analog speech synthesizer, while the single resonance tones were generated with one formant characteristic of the synthesizer, which was excited by a repetitive impulse voltage source having a fundamental frequency of 130 Hz. All of the stimuli were read out at a sampling rate of 10 kHz with an accuracy of 10 bits, converted into analog waveform, and recorded through a low-pass filter with a cutoff frequency of 4.5 kHz.

The synthetic vowels were typical sounds for /a/ (F_1 = 750 Hz, F_2 = 1,200 Hz) and typical sounds for /o/ (F_1 = 450 Hz, F_2 = 750 Hz). The other formant frequencies (F_3, F_4 and F_5) were held constant at 2,600, 3,500 and 4,500 Hz, respectively. The bandwidths of these formants were also held constant at 60, 100, 120, 175 and 280 Hz, for F_1 to F_5 respectively. The resonance frequencies (RF) of the single resonance tones and the frequencies (F) of the pure tones had the same value as the first or the second formant frequencies of the synthetic vowels.

Procedure

The experiment was divided into two sessions of from 30 min. to 40 min. On any one day each subject served for two sessions with a 10 min. inter-session rest period.

The duration of each stimulus was 200 ms, and nine kinds of stimulus sounds were used—pure tones at three different frequencies (450 Hz, 750 Hz and 1,200 Hz): white noise: single resonance tones at three different resonance frequencies (450 Hz, 750 Hz and 1,200 Hz) and two synthetic vowels (typical sounds for /a/ and typical sounds for /o/).

The interval between the first tone and the second tone was 1 second, and the inter-trial interval was 3 seconds. The first stimulus was presented constantly at 54 dB(c) through the reference speaker, while the second stimulus was presented through the reference speaker or the comparison speaker with a probability of 0.5 for each speaker. Its amplitude was varied randomly among five levels from 48 dB(c) to 60 dB(c) in order to minimize the possibility of the subjects using slight differences in signal amplitude between the speakers as a cue.

All of the stimuli [(9 kinds of stimuli) × (5 intensity levels)]were completely randomized within each session. Ten trials per each intensity for each stimulus were given to each subject.

For each stimulus, the subject was asked to indicate whether the direction of the stimulus sound was the same as for the reference sound or had shifted to the left. He/she was further asked to report whether the stimulus was a speech or a non-speech sound.

speaker was 2.5° or 5° to the left of the reference. These two azimuths were selected on the basis of the results of a preliminary experiment from which the minimum audible angle of pure tones for the subjects was obtained (2° to 7° in the frequency range of 450 to 1,200 Hz), as shown in Fig. 2. The subject's head and the centers of the speakers were positioned 1.15 m above the floor of the room.

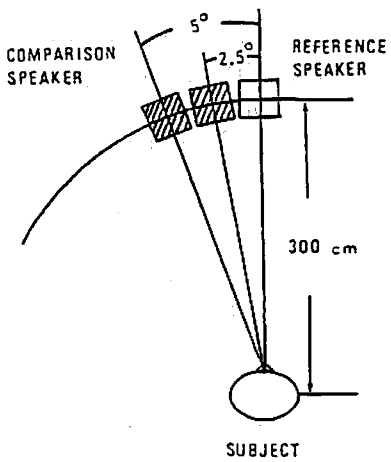

Fig. 1 *Loudspeaker array for localization tests in the horizontal plane.*

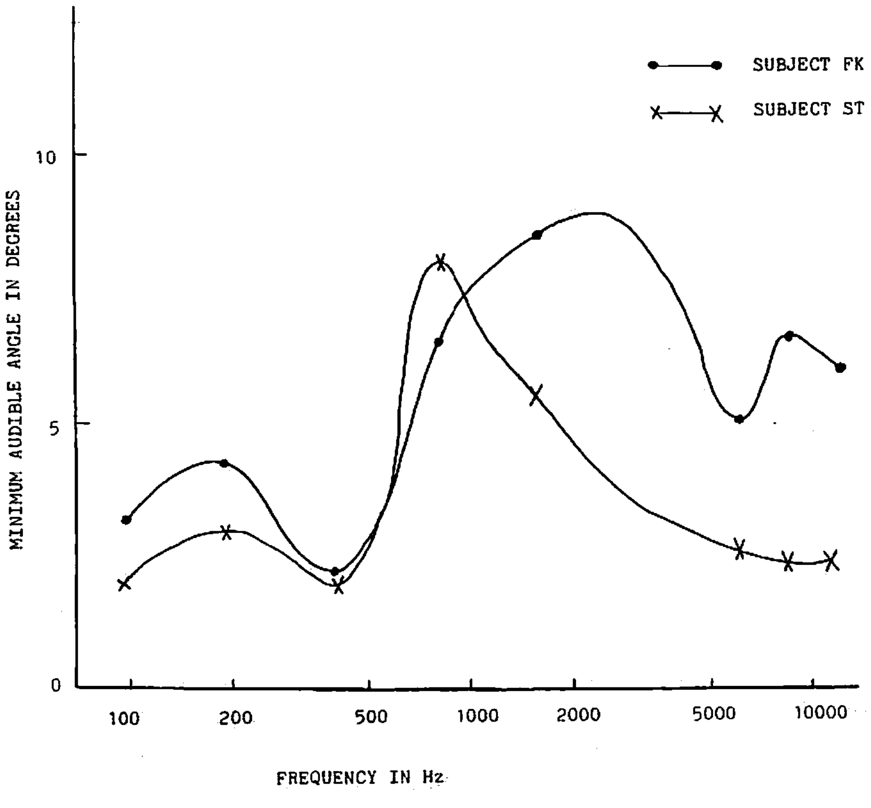

Fig. 2 *The minimum audible angle about the median plane. The ordinate is the minimum audible angle of degrees, and the abscissa is the signal frequency.*

Results

The experimental results were very similar among the five different intensities, and consequently the fifty judgments were pooled across this variable for each stimulus condition and for each subject.

Table 1 lists the results of the experiment pooled across the different frequencies. When the azimuth was 5°, the percentage of correct responses for the speech stimuli and that for the non-speech stimuli (except for the pure tones), reached almost 100% Accordingly, it is obvious that there are great differences between the pure tones and the other stimulus sounds ($\chi^2 = 112.50, p < 0.001$, for the synthetic vowels; $\chi^2 = 125.23$, $p < 0.001$, for the single resonance tones; $\chi^2 = 173.4$, $p < 0.001$, for the white noise). However, when the azimuth was 2.5°, some systematic differences among the stimulus sounds were obtained. The percentage of correct responses for the white noise was greater than that for both the synthetic vowels and the single resonance tones ($\chi^2 = 18.55, p < 0.001$; $\chi^2 = 32.53, p < 0.001$), and the percentage of correct responses for the pure tones was the lowest ($\chi^2 = 32.49$, $p < 0.001$, for the synthetic vowels; $\chi^2 = 29.74$, $p < 0.001$, for the single resonance tones; $\chi^2 = 119.03, p < 0.001$, for the white noise).

Table 1 *Percentage of correct responses for speech and non-speech stimuli*

Azimuth	Stimulus sounds	Subject			
		SS	YA	TI	NS
2.5°	Synthetic vowels	82	68	71	80
	Single resonance tones	77	71	60	83
	White noise	95	82	75	93
	Pure tones	64	53	59	55
5°	Synthetic vowels	95	97	97	97
	Single resonance tones	93	91	95	98
	White noise	95	97	100	99
	Pure tones	68	67	72	69

Fig. 3 shows the performance as a function of the different frequencies or the first formant (or resonance) frequencies when the azimuth was 2.5°. It also presents the percentage of correct responses only for the stimuli judged as speech sounds in each frequency. The results for the four subjects were pooled in this figure. In the case of 450 Hz, there was little difference in the percentage of correct responses between the synthetic vowels and the single resonance tones ($\chi^2 = 0.63$, $p > 0.05$). On the other hand, the percentage of correct responses for only the stimuli judged as speech sounds among the synthetic vowels was no greater than that for the single resonance tones. In general, the percentage of correct responses for the pure tones was smaller than that for the other stimuli.

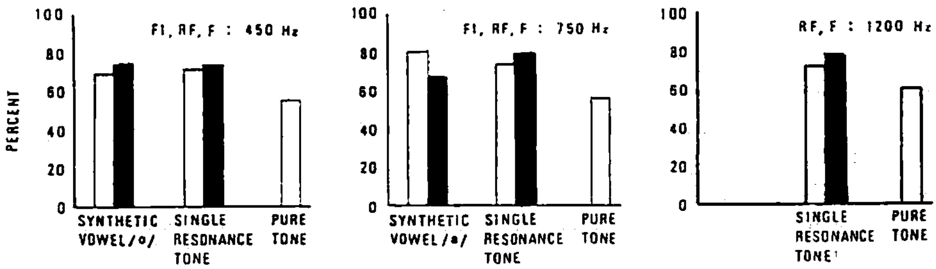

Fig. 3 *Percentage of correct responses and percentage of judgments as speech sounds for the synthetic vowels, the single resonance tones, and the pure tones. The abscissa represents the three kinds of stimuli, and the ordinate represents the percentage of correct responses (blank bars) or the percentage of correct responses only for the stimuli judged as speech sounds (solid bars). The azimuth was 2.5°.*

Discussion

The present experimental results show that complex tones such as single resonance tones and synthetic vowels, or non-cyclic tones such as white noise, can be localized more accurately than pure tones. The results also show that the localization of acoustical stimuli is influenced by the physical characteristics of the stimuli rather than by subjective factors.

The present data have several implications for the localizations of speech and non-speech sounds. Though the percentage of judgments of the synthetic vowels as speech sounds was much greater than that for the single resonance tones (which have only one formant characteristic of the synthetic vowels), the percentage of correct responses in localization for the synthetic vowels was only slightly greater than that for the single resonance tones. Furthermore, the percentage of correct responses for the white noise was greater than that for the synthetic vowels.

It might be said that the critical factor for localization is neither the recognition of the stimulus sounds as speech or non-speech, nor the identification of their content, but rather their acoustic complexity. Though complex sounds contain high frequency components, the data obtained by Mills (1958) rule out this idea: the performance on higher frequency tones was poorer than that for the 450–1,200 Hz tones used in this experiment.

Acknowledgment

The authors wish to thank Professor Tadasu Oyama of the Department of Psychology, University of Tokyo for his encouragement and helpful suggestions during the course of this work.

References

Brown, C.H., M.D. Beecher, D.B. Moody and W.C. Stebbins (1978); Localization of primate calls by old world monkeys. Science, 201, 753-754.

Hornbostel, E.M. (1926): Das räumliche Hören. In A. Bethe, G. v. Bergmann, G. Embden, and A. Ellinger (Eds.). Handbuch der normalen und pathologischen physiologie. Berlin: Verlag Von Julius Springer. pp. 602-618.

Mills, A.W. (1958): On the minimum audible angle. Journal of the Acoustical Society of America, 30, 237-246.