# PERCEIVING COMPLEX SPEECH AND NON-SPEECH SOUNDS

J. C. WEBSTER*—A. CARPENTER**—M. WOODHEAD**

The purpose of this paper is to see if naive listeners can learn supposedly meaningless sounds quicker if some are indeed sustained vowels, some musical instrument tones and some truly meaningless. Nine sounds were chosen; three were identifiably different but meaningless buzzes; three were two-second portions taken from the middle of the sustained vowels a, i and u; and three were mid-portions of sustained tones on the clarinet, the cello and the tromba- stop on an experimental electronic organ. The subjects were not told that there were three different classes of sounds although they were told "... some may sound like voice sounds, some... musical... (or) perhaps they will (all) sound like nothing you've heard before".

## *STIMULUS MATERIALS*

A 54-item sound identification (*SI*) test was recorded on magnetic tape. Each item consisted of four tonal complexes. The duration of each tone, which could be any one of nine sounds, was two seconds. Each interval between tones was also two seconds. A five second interval was left between items to give the listener time to write a response. The vowels were intoned for 15 seconds at a fundamental frequency of 110 Hz by a native British speaker; the two second samples were chosen from the middle of the 15-second utterance. The musical tones were originally recorded at 220 Hz for 15 seconds but played back at half speed for the two-second mid-portion samples used in the recorded test. The meaningless sounds were buzzes with a fundamental frequency of 110 Hz. They differed in that harmonics 6, 12, 18, and 24 were greatly emphasized in one sound, harmonics 1, 8, and 9 in another and harmonics 1, 8, 16, and 24 in the third. Every effort was made to have all nine sounds at the same frequency and intensity on the recorded test.

Two variations were made of the original SI test, a "low"-SIL-and a "high"-SIH-frequency version. For both of them the voice announcements and the five second

* Exchange Scientist from US Navy Electronics Lab., San Diego, California, USA.

** Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

interval between items remained the same. For the "high frequency or SIH" variant the 4-tone test item was played back and re-recorded at double speed, so that all frequencies in all sounds were one octave higher but the duration of tones and intervals between was maintained.

Three tests were made up of nine different complex steady-state sounds. On one test, SIL, all nine sounds had fundamental frequencies of 55 Hz; on test SI all had fundamental frequencies of 110 Hz; and on test SIH all had fundamental frequencies of 220 Hz. Three of the nine sounds were completely meaningless and remained so on all three tests. Three sounds were segments of the intoned vowels a, i and u when the fundamental frequency was 110 Hz. At 55 or 220 Hz the fundamental and all other parts of the complex tones were shifted in frequency one octave and the resultant sounds were perhaps vowel-like but were not a, i and u. Three sounds were the tones of musical instruments when the fundamental frequency was 220 Hz. At 110 and 55 Hz all components were shifted and the resultant tones were music-like at 110 Hz but only musically-derived at 55 Hz.

## *TEST PROCEDURE*

Two groups—groups A and B—of six Royal Navy ratings with normal hearing served as subjects for two 45-minute periods a day for two weeks. During the first week group A listened to test SI where the fundamental frequency—and all formant frequencies—of the vowels were correctly located, while group B listened to test SIL—all nine sounds meaningless. During the second week the groups listened to the test they had not heard the first week.

Subjects were asked to identify the sounds by an arbitrarily assigned number. The only clue given to them was that certain sounds might sound voice-like or like musical instruments, or like nothing they had heard before.

## *CONCLUSIONS*

There were fewer errors in assigning the correct "identity number" to vowel sounds when the frequencies were 110 Hz or "correct". When all sounds were shifted downward one octave errors among the three classes of sounds were equally divided. A further test needs to be run to find out what happens when the musical instrument sounds are at the "correct frequencies".



## DISCUSSION

*Black:*

I wonder whether Dr. Webster is tempted to place a relationship between his results (position to human-like sounds) and a motor theory of speech perception.

*'Hart:*

I am not surprised to hear that the sustained vowel sounds were not identified significantly more easily than the musical sounds, as the vowel sounds, sustained for two seconds are deprived of so many attributes they would have in normal speech, that they can no longer be considered anything other than mere sounds.

*Webster:*

Ad Black: We showed that steady state vowel sounds with no pitch inflection and not identified as speech sounds were perceived better, than either meaningless or music-like sounds. A present experiment will see if musical sounds are easier than meaningless and speech-like sounds. If musical sounds and speech sounds are equally easy, than I'd say overlearning may explain it. If the speech sounds are easier, then the motor theory has some support.

Author's Note:

Since this paper was presented, the „Music" (SIH) group has been run. The listeners did not identify musical tones better than other tones. This implies there is something unique in the way listeners perceive speech-like sounds but not in the way they perceive music-like sounds.