Spectral Characteristics of Digit-Simulating Speech Sounds

By D. P. BORENSTEIN

(Manuscript received July 11, 1963)

A spectral analysis has been performed on a number of spoken vowel sounds, in particular those sounds causing digit registration in a TOUCH-TONE receiver. The analysis, implemented by computer methods, provides a definitive picture of the nature of digit simulation in TOUCH-TONE calling.

I. INTRODUCTION

A digit simulation in TOUCH-TONE calling (Ref. 1, pp. 9–12, 15–16) is, by practical definition, a speech segment capable of causing digit registration in a TOUCH-TONE signaling system. Spectral analyses have been performed on a number of speech segments, each of which was selected solely on the basis of having the above property. Briefly. a valid TOUCH-TONE signal requires the simultaneous presence of two code frequencies for a certain minimum length of time, and with some minimum signal-to-noise ratio. It was therefore theoretically anticipated (Ref. 1, pp. 10–12) that each of these speech segments would be linked by two other common characteristics: (1) a frequency spectrum having two sharply dominant peaks, and (2) a high degree of periodicity for some minimal length of time.

There is good reason to believe that speech segments of this general nature are likely to be troublesome in any signaling system based on the transmission of voiceband tones over speech channels.

Due to the inherent rarity and relatively brief duration of the voice-produced digit simulation, some special procedures were required both in obtaining and analyzing these speech segments. The remainder of this article comprises a description of these procedures, followed by a presentation and discussion of the resulting spectral analyses.

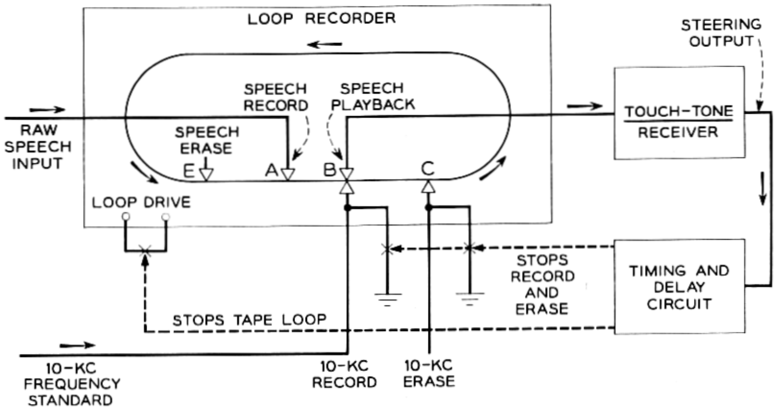

Fig. 1 — Apparatus for recording digit-simulating speech segments.

II. COLLECTING THE SPEECH SAMPLES

The digit-simulating speech segments were obtained by recording raw speech onto magnetic tape loops with the two-track recording arrangement shown in Fig. 1.

Using a 60-inch loop of tape at a speed of 15 in/sec, speech is continuously recorded at point A, played into a standard TOUCH-TONE receiver at point B, and, if there is no receiver output, erased at point E after traversal of the loop. Simultaneously, on a second track, a 10-kc pilot frequency is continuously recorded and erased at points B and C, respectively. If at any time there is a TOUCH-TONE receiver output, indicating the presence of a digit-simulating speech segment just past point B, the timing network is triggered. The timing network then performs two operations: (1) it disables the 10-kc record and erase after a delay of 35 ms, and (2) it stops the tape transport after a delay of 2 seconds (half the loop traversal time).

This process yields a 60-inch length of tape consisting of about 29 inches each of pre- and post-simulation speech plus a 1.5-inch (110 ms at 15 in/sec) segment which contains both the actual simulating speech sample and the 10-kc pilot frequency. In this manner, fourteen such samples were obtained, at the average rate of about one per ten hours of raw speech — an indication of the extreme rarity of simulation with the present TOUCH-TONE receiver.

III. ANALOG-TO-DIGITAL CONVERSION AND PRINTOUT

By means of encoding equipment developed by the Acoustics Research Department, the fourteen digit-simulating speech segments were converted from analog form to an eleven-bit digital signal. The sampling rate of 10 kc was gated directly from the pilot track of the original analog tape, thus eliminating sources of error due to tape flutter during the original recording process. Once the digital tape was obtained, the conversion process was reversed to obtain an accurate X-Y recording of each of the fourteen speech waveforms. Visual inspection of these waveforms, two of which are shown in Fig. 2, confirms their periodic nature (the periodicity of the samples shown in Fig. 2 would be still more evident were it not for the fact that most speech fundamentals are considerably attenuated by telephone apparatus).

IV. SPECTRAL ANALYSIS

The fourteen speech samples, in eleven-bit digital format, were then subjected to a "pitch synchronous"[2] Fourier analysis on the IBM-7090 computer. The pitch synchronous analysis consisted essentially of a conventional Fourier analysis performed on each successive fundamental pitch period in the speech sample. These pitch periods, in turn, were determined on the computer by counting the number of sampling intervals (each being 100 μsec) between successive maxima in the waveform and then interpolating between samples for greater accuracy. This method of Fourier analysis is ideally suited to waveforms that maintain an almost-periodic structure over an appreciable length of time.

For each speech segment analyzed, the computer output consisted of a sequential set of bar graphs, one for each fundamental pitch period of the speech waveform. Each graph, in turn, is a plot of harmonic amplitude (the Euler coefficient) in db versus harmonic number. In addition, each graph gives the "instantaneous pitch" (i.e. the reciprocal of the period) of each fundamental period analyzed. Figs. 3 and 4 show the

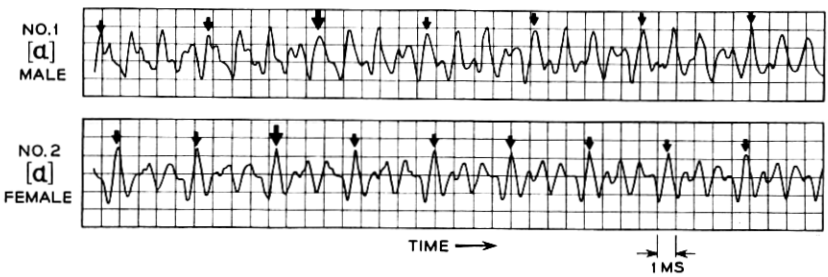

Fig. 2 — Analog waveforms of two digit-simulating speech segments. Also shown are the sex of each speaker and the particular phoneme causing the simulation. Fourier spectra of digit simulations 1 and 2 are shown in Figs. 3 and 4, respectively. Arrows indicate periodicity, with the large arrow showing the approximate start of the digital simulation.

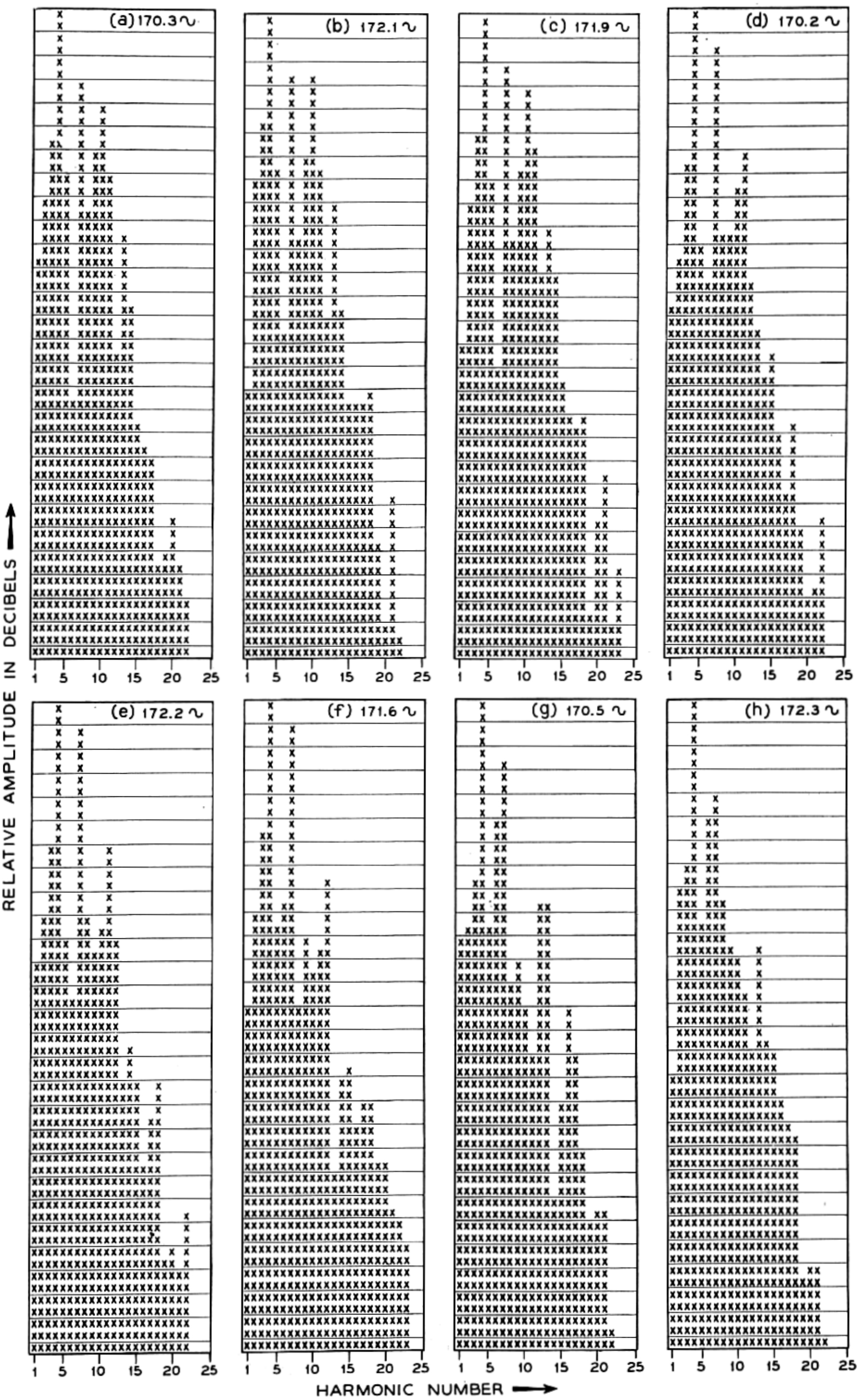

Fig. 3 — Set of Fourier spectra for digit simulation No. 1 (as shown in Fig. 2). Each X represents a 1-db relative amplitude increment. Spectra are in alphabetical order with respect to time.

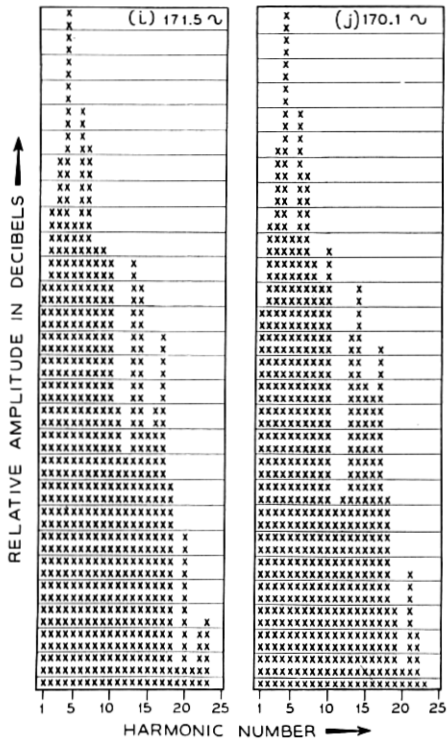

Fig. 3 — (continued)

two sets of spectra corresponding to the two speech segments whose time domain waveforms appear in Fig. 2.

V. DISCUSSION OF RESULTS

Several aspects of the spectra shown in Figs. 3 and 4 are worthy of note.

First of all, it is seen that these two speech segments (as well as the twelve others not shown here) do indeed satisfy the two properties anticipated in the introduction. The high degree of periodicity of these speech waveforms is spectrally confirmed by noting that in both sequences of spectra the harmonic structure remains extraordinarily uniform. (This result also confirms, by hindsight, the original validity of a period-by-period Fourier analysis.) By noting the fundamental pitch (thus the period) of each segment, it is seen that this highly stable harmonic structure is maintained for at least the 23 milliseconds which coincides with the duration requirements of the TOUCH-TONE receiver.

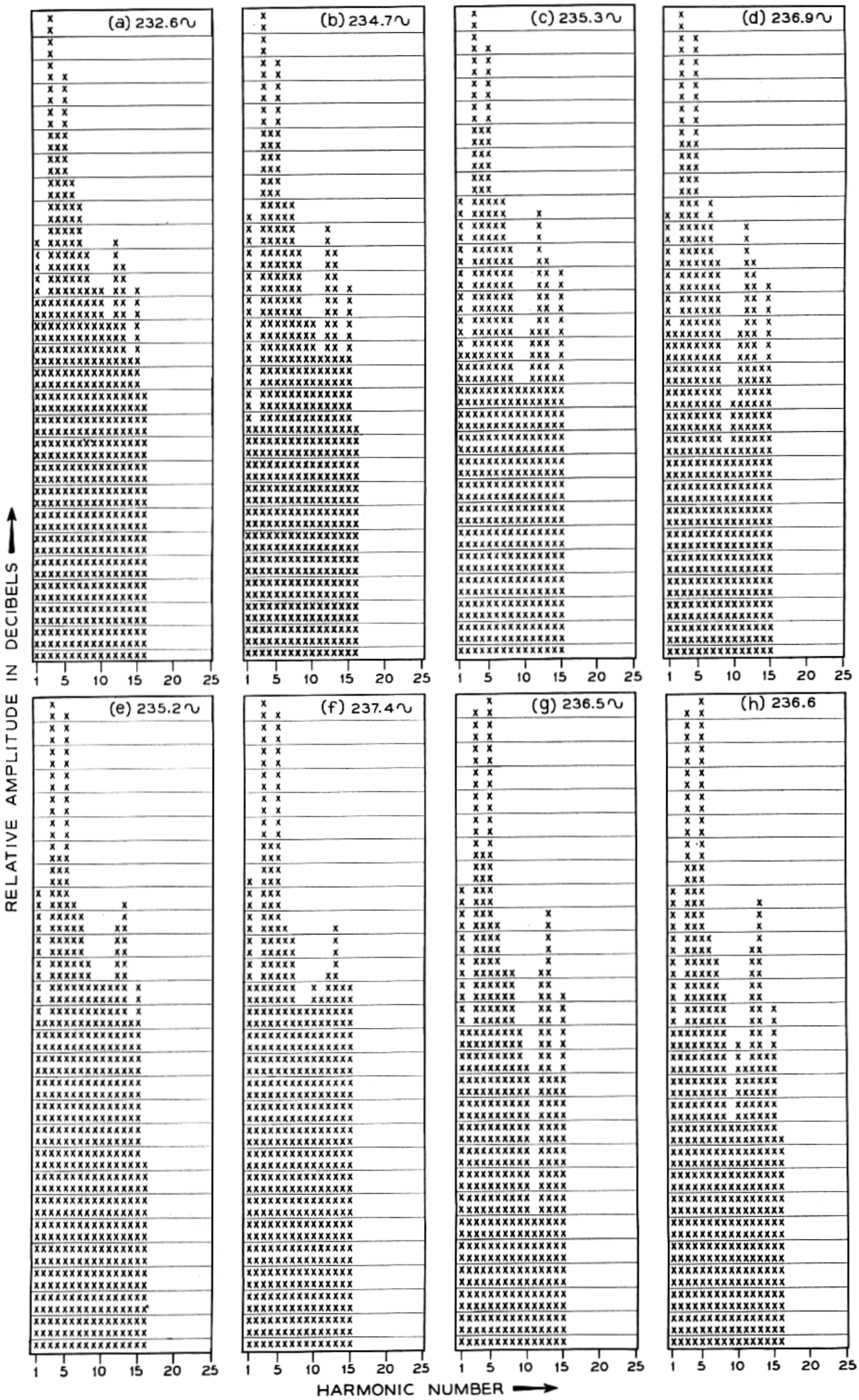

Fig. 4 — Set of Fourier spectra for digit simulation No. 2 (as shown in Fig. 2).

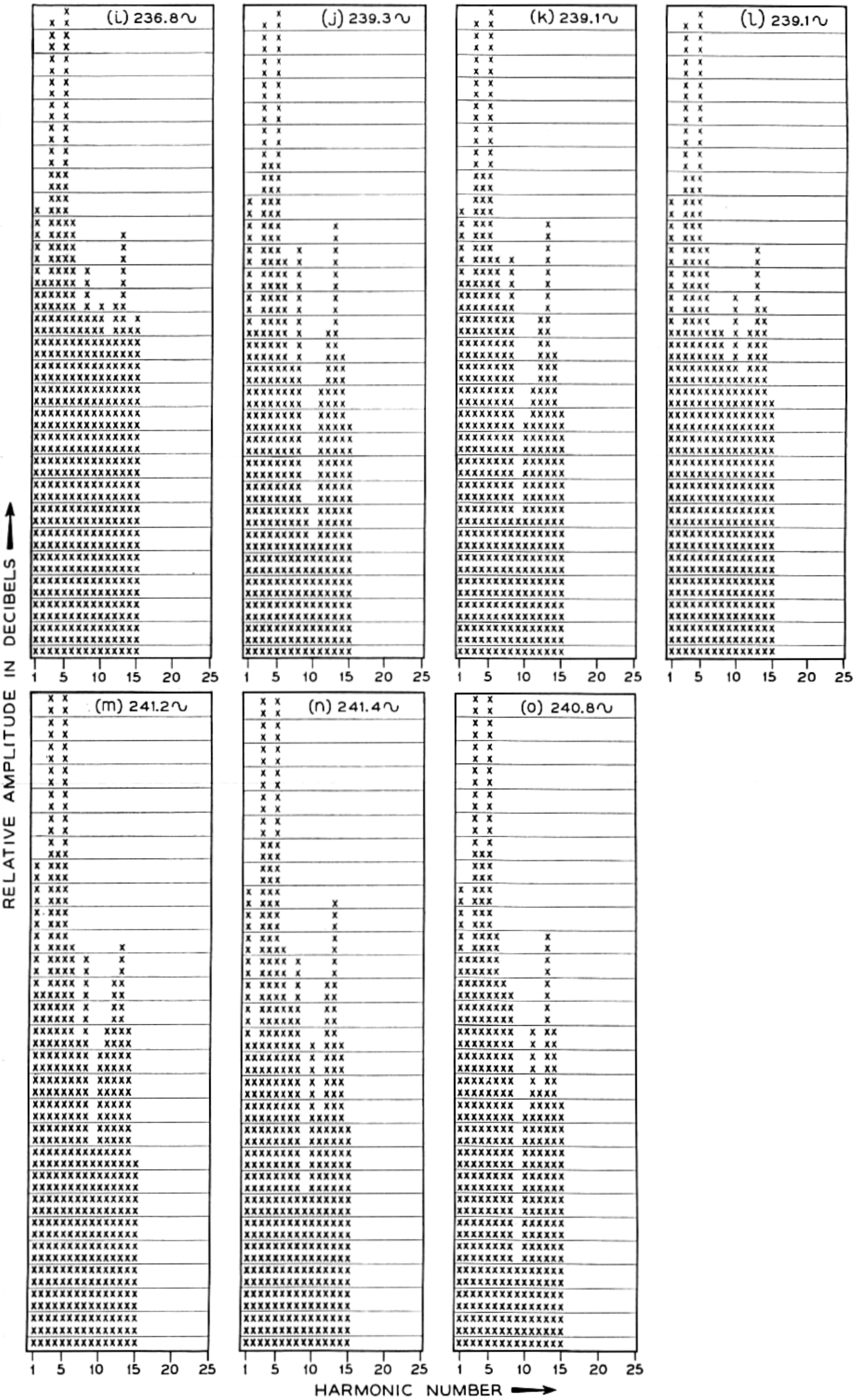

Fig. 4 — (continued)

Secondly, one finds immediate justification for the fact that these speech segments caused digit simulation. By multiplying the fundamental pitch of any segment by the orders of its two dominant harmonics, a valid TOUCH-TONE calling signal* is derived. Thus a voice-produced digit simulation is spectrally analogous to a valid TOUCH-TONE signal accompanied by noise, with the sole exception that in the former case both the "noise" and signal components are integral multiples of a discrete fundamental frequency. Indeed, this sole distinction between a digit simulation and a valid signal might possibly be used to provide further simulation protection in future voice-frequency signaling applications. Specifically, a receiver might be designed to be sensitive to the presence of selected harmonics and/or sub-harmonics of valid signal frequencies, and thereby to reject many speech phonemes which would ordinarily cause simulation.

In the portions of Figs. 3 and 4 where the harmonic structure is noticeably changing with time (namely at the beginning and end of each series of spectra) pitch-synchronous Fourier analysis can be regarded as only an approximation of spectral density. For some applications, however, the approximation is still useful. In the first place, one can obtain a practical "feel" for the rate of change of pitch and harmonic structure in vowel-type speech sounds. Also, from the standpoint of digit simulation, by examining the spectra one can ascertain just how and when a speech segment becomes a digit simulation. For example, in the early spectra of Fig. 3, although pitch requirements for digit simulation are satisfied, the 10th harmonic competes with the 7th harmonic for limiter capture, and receiver recognition is prevented by limiter guard action (i.e., insufficient signal-to-noise ratio). (See Ref. 1, pp. 10–11, 13.)

On the other hand, although early spectra of Fig. 4 show an acceptable harmonic structure for digit simulation, the pitch is slightly too low for receiver recognition. In a similar manner, one can determine how and when a digit simulating wave-form starts to degenerate.

Admittedly, the speech segments chosen here are both rare and few in number. Thus, one cannot draw conclusions of statistical significance from this study. However, there is no reason to believe that any other group of frequencies of the same capacity in the voice band would not be simulated by the voice about as often as were these TOUCH-TONE calling frequencies. Therefore, such vowel-type speech segments may be

---

* A valid TOUCH-TONE calling signal consists of one frequency from each of two groups: a low group — 697, 770, 852 and 941 cps ±2.5 per cent — and a high group — 1209, 1336, 1477 and 1633 cps ±2.5 per cent.

looked upon as potential digit simulations in almost any proposed voice-frequency signaling application.

REFERENCES

1. Battista, R. N., Morrison, C. G., and Nash, D. H., Signaling System and Receiver for TOUCH-TONE Calling, Trans. IEEE, **65**, March, 1963.
2. Mathews, M. V., Miller, Joan E., and David, E. E., Pitch Synchronous Analysis of Voiced Sounds, J. Acoust. Soc. Am., **33**, Feb., 1961, pp. 179–186.