# FACTORS OF THE GLOTTAL WAVE
# THAT CONTRIBUTE TO THE NATURALNESS OF SPEECH

S. Maeda* and O. Fujimura

Effects of different factors involved in the glottal wave upon naturalness of speech signals have been investigated by vowel synthesis experiments.

In previous studies,[1)][2)] effects of the shape of the glottal pulse on the voice quality were discussed considering in particular perceptual effects of the glottal zeros. In this report, a synthesis-listening experiment of five Japanese vowels is described. When the pulse shape and the pulse interval of the glottal wave fluctuate, the fundamental periods of the speech signal must be defined for the particular method of pitch detection employed. This point has been examined by use of vowel synthesis and perceptual evaluation, and it has been tentatively concluded which method best defined the pulse positions from a practical point of view.

## Experimental Procedures

Fig. 1 shows a block diagram of the vowel analysis and synthesis system. The speech and subglottal signals are recorded simultaneously. The speech signal is analyzed by the sound spectrograph and the formant frequencies ($F_1 \ldots F_5$) and bandwidths ($B_1 \ldots B_5$) are estimated by visual inspection. These values fed into the computer are used to control the resonance circuit of a terminal analog synthesizer.

The subglottal signal is low-passed with a cutoff frequency of 1-kHz, and is sampled at a rate of 10,000/sec in linearly distributed 11-bit levels. First, the fundamental periods of the subglottal wave are determined by a peak detection method. In this method, the position of each pulse of the glottal wave is defined as the time moment of the summit of a salient peak, and the fundamental period is defined as each time interval between the consecutive pulse positions. Next,

* Research Assistant in Research Laboratory of Communication Science, the University of Electro-Communications

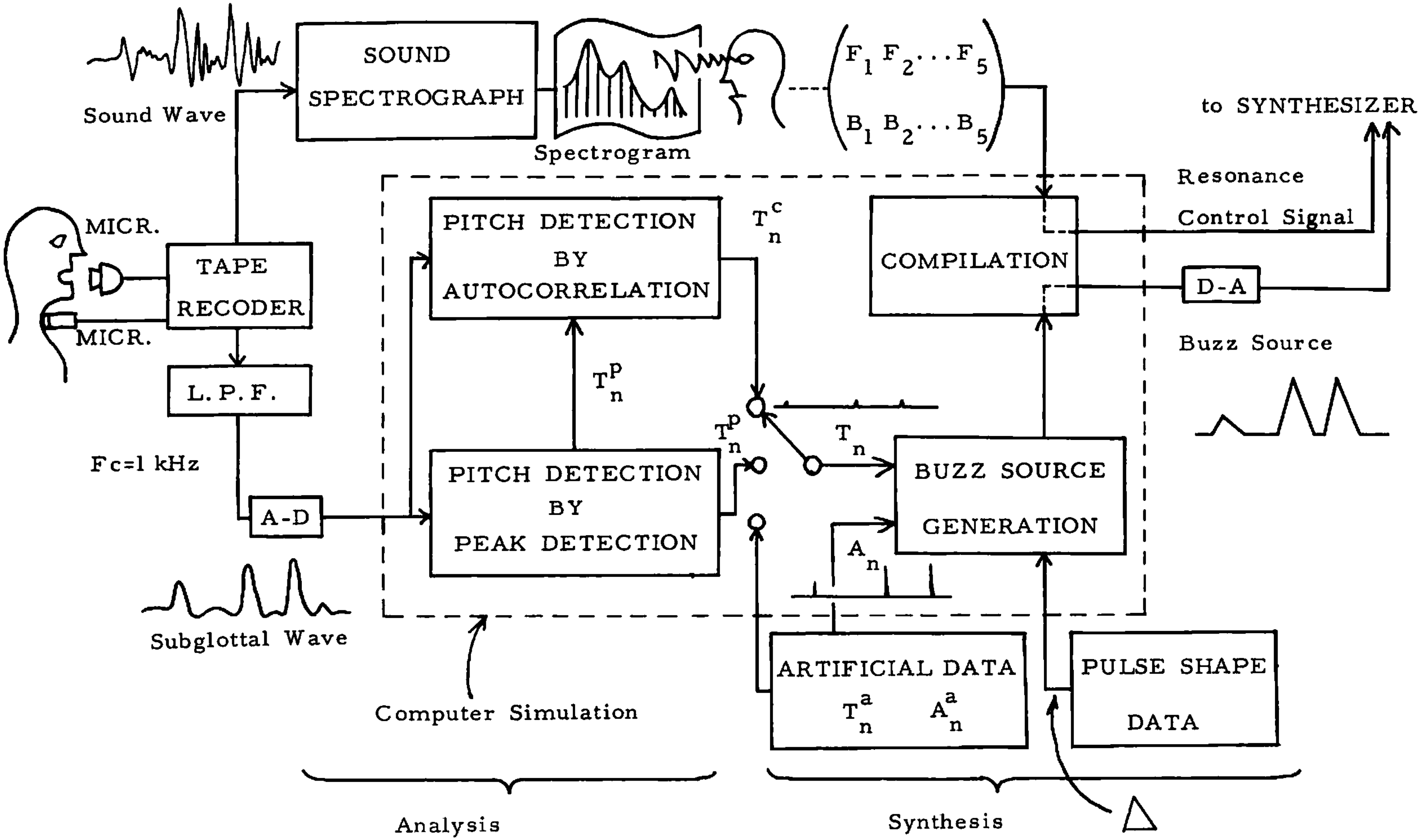

Fig.1 - Block diagram of the vowel analysis and synthesis system.

an improved pitch detection is performed by an autocorrelation method using the first results obtained by the peak detection as the center of the range of search for the maximum autocorrelation coefficient.

Fig. 2 illustrates examples of the successive values of the fundamental periods as determined by the methods described above, for a stretch of vowel utterance. The curve (a) shows the result of the peak detection method. The curves (b) and (c) were obtained by use of different time windows for the autocorrelation method.

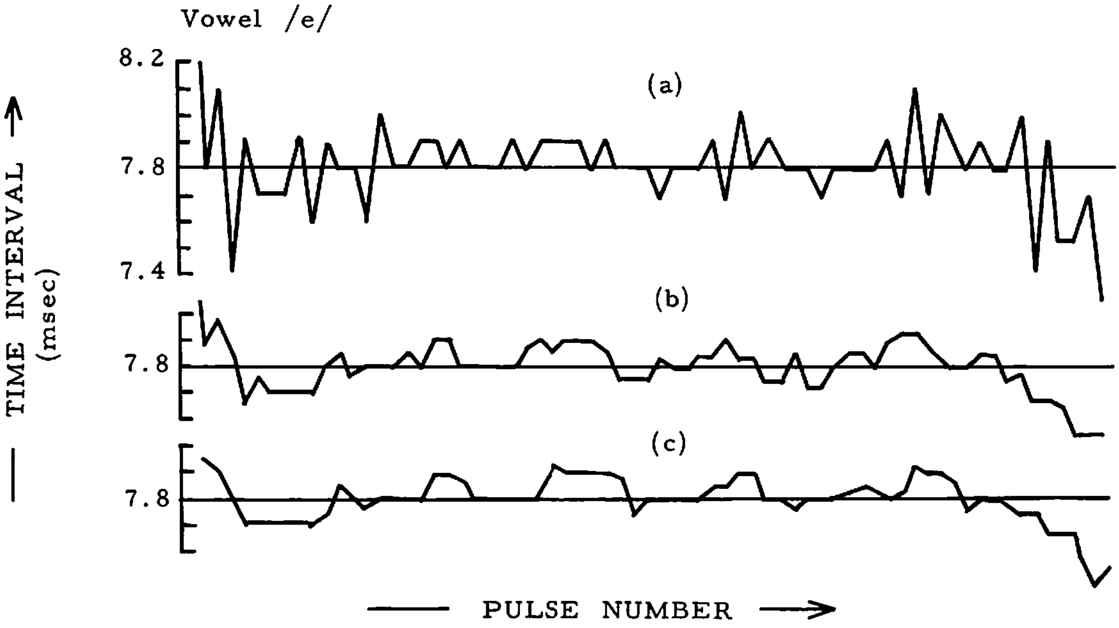

Fig. 2 - Time intervals between consecutive pulses determined by (a) peak detection method, (b) autocorrelation method using a ± 3-msec time window, and (c) autocorrelation with a ± 8-msec window.

In the "BUZZ SOURCE GENERATION" in Fig. 1, the buzz pulses of a fixed shape are generated at the time moments determined by the methods above. The buzz source signals are fed into a hardware terminal analog synthesizer. The glottal pulse shape is approximated by a triangular pulse with fixed values of rise and fall times, and also by a fixed half-sine pulse whose amplitude spectrum grossly falls at about -12-db/oct. The amplitude is also fixed except for a gradual rise and fall at both ends of an utterance.

Naturalness of the synthesized vowels was evaluated by listening tests, for different values of the effective cutoff frequency $f_e$, which represented the critical roll-off frequency for the spectrum envelope of the excitation pulse.

Prior to the test, the listening crew (five untrained male subjects) were informed that the test tape contained samples of five Japanese vowels that had been recorded under various conditions. They were instructed to evaluate naturalness of each stimulus as a vowel sample uttered in isolation, and select one of the three grades: natural, not quite natural but acceptable, and unnatural.

## Results and Discussions

Fig. 3 illustrates the estimated values of the naturalness plotted against $f_e$, (a) for vowel /a/ and (b) for the mean of five vowels. The curve generally forms a broad peak around 180-250-Hz, and thus there is found for naturalness an optimum value of $f_e$. For a glottal pulse that has conjugate zeros in the spectrum (for example, a symmetric triangular pulse), the observed naturalness is selectively low for particular vowels (see (c) and (d) in Fig. 3). This is explained as an influence of the glottal zeros upon the formant peaks. Perceptually, these glottal zeros, especially conjugate zeros, give rise to a contamination of the phonemic quality, and the synthesized vowels fail to simulate natural vowel samples.

Informal listening tests of the synthesized vowel samples have revealed that the determination of the pulse positions by the method (a) in Fig. 2 (peak detection) results in a coarse voice, far from the original voice quality. The sample synthesized by the method (b) was acceptable in naturalness and the sample by (c) was the most natural.

The irregularity of the glottal wave of the natural voice must include variation of the pulse waveform from period to period. This influences the peak position, but the positiin of each pulse that gives the maximum correlation coefficient for the consecutive pulses may remain comparatively less affected.

Consequently, the synthesized vowel in agreement with the peak detection data and still under the condition that the pulse shape be kept constant could have had an excessively fluctuating periodicity when determined by the autocorrelation

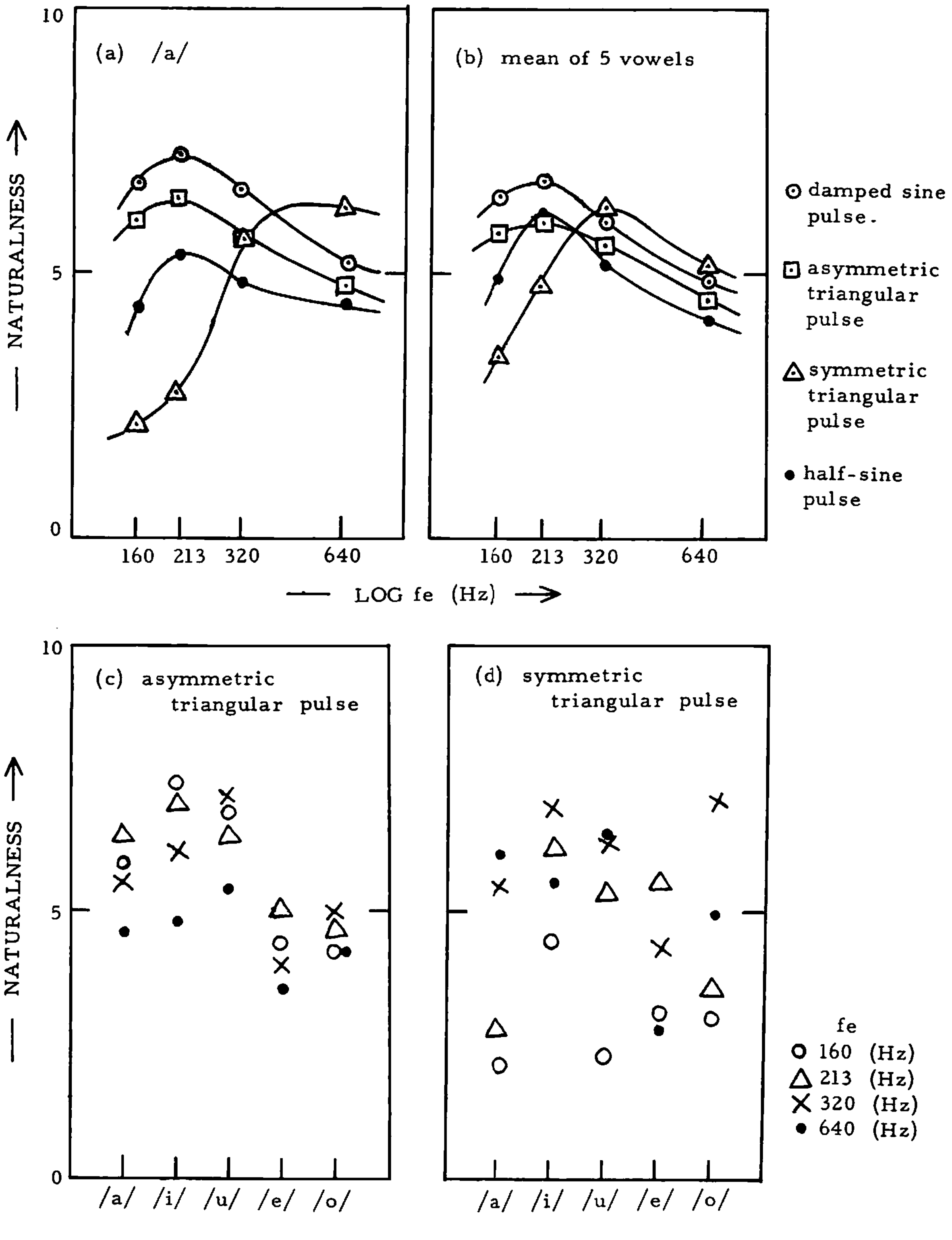

Fig. 3 - Ratings of naturalness.

method, and presumably also by the human auditory mechanism. Approximations are being explored in order to simulate these fluctuations by simple means.[3)4)5)]

## References

1) Flanagan, J. L.: "Some Influences of the Glottal Wave upon Vowel Quality," Proc. of the 4th International Congress of Phonetic Sciences, Helsinki (September, 1961).

2) Maeda, S. and Fujimura, O.: "Factors of Voice Fluctuation," Proc. of the Spring Convention of the Acoustical Society of Japan, p. 277 (1967).

# Kato, Y., Ochiai, K., Fujimura, O., and Maeda, S.: "A Vocoder Excitation with Dynamically Controlled Voicedness," 1967 Conference on Speech Communication and Processing (IEEE) Conference Preprints, 284-288, Cambridge, Mass. (November, 1967).

4) Fujimura, O.: "Speech Coding and the Excitation Signal," Digest of Technical Papers, 1966 IEEE International Communications Conference, p. 49 (1966).

5) Fujimura, O.: "An Approximation to Voice Aperiodicity," IEEE Transactions on Audio and Electroacoustics Au-16, 68-72 (1968).