Received 29 April 1970 9.4, 9.5, 9.6

Effect of Glottal Pulse Shape on the Quality of Natural Vowels

A. E. Rosenberg

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey 07974

A pitch-synchronous analysis was carried out over the vowel portions of the CVC utterances HAYED, HOD, HODE and the sentence FEW THIEVES ARE NEVER SENT TO THE JUG recorded by a male speaker. For every pitch period, the analysis provides formant frequencies and the waveform of the vocal-cord excitation. The excitation waveform was replaced by a simulated excitation waveform, with which the utterances were resynthesized. In Expt. I, six simulated waveforms with pulse shapes differing in the number and location of slope discontinuities were investigated. Listening tests indicated that simulated excitations with pulse shapes with a single slope discontinuity at closure are preferred. In Expt. II, simulated excitations with 16 combinations of opening and closing times of a preferred pulse shape were investigated. Listening tests indicated that very small opening or closing times, or opening times approximately equal to or less than closing times, are not preferred. In general, it was demonstrated that good-quality synthetic speech can be generated by using simple excitation waveforms specified uniformly over an utterance. The use of tournament testing strategies for perceptual evaluation of speech samples is also described.

INTRODUCTION

The nature of the vocal-cord excitation has long interested speech researchers. One of the most important problems has been the specification of source excitations for speech synthesizers. It is known that the shape and periodicity of the vocal-cord excitation are subject to large variations. The question we have been concerned with is the nature and extent of such variations which are significant for the preservation of naturalness in speech. In a previous paper,[1] the importance of pitch-period variations was discussed. It was found that listeners could not discriminate natural-speech samples from samples in which most of the fine details of pitch variations had been smoothed. The present experiment is concerned with the effect on naturalness of the variation of glottal pulse shape. The emphasis is therefore on the temporal properties of the excitation rather than the spectral properties. There are three stages to the approach used in this experiment: an analysis of an utterance, including inverse filtering to obtain an excitation waveform; resynthesis after substitution of this natural excitation by an abstracted waveform; and subjective evaluation of the new utterance.

In inverse filtering, the transfer characteristics of the vocal tract for a particular utterance are represented by a network. Application of the speech wave to the inverse of this network will then produce the excitation as the output. In the acoustical network analog of the vocal tract, this excitation is the volume velocity at the vocal cords or glottis. Specifying the proper transmission characteristics is by no means trivial. The first successful application of this technique was recorded by R. L. Miller.[2] Miller demonstrated the pulse-like nature of the waveform: that excitation is associated primarily with rapid closure of the glottis and that the slope at closure increases with increasing vocal effort. More extensive results were obtained by Mathews *et al.* by means of a computerized pitch-synchronous analysis technique.[3–5] The results they obtained showed that, with increasing pitch or decreasing intensity, the relative open time increases and the waveform takes on a more sinusoidal appearance. Conversely, with decreasing pitch or increasing intensity, the relative open time decreases and the slope at closure increases. They also found more variation in glottal pulse shapes among different talkers than among different utterances of the same speaker. The damping of formant poles was found to be higher during the open time of the glottis, but output amplitude was independent of the average flow or open time.

J. N. Holmes also reported an investigation of the excitation waveform by means of inverse filtering.[6,7] Holmes found large differences in waveshape between talkers, between utterances, and with variations in pitch

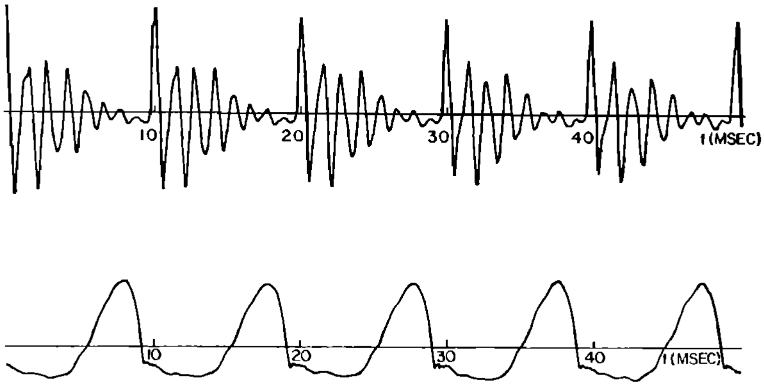

FIG. 1. Speech output (top) and analyzed excitation waveform (bottom) for vowel in HOD.

and intensity. He also showed that the excitation of higher formants does not necessarily coincide with closure. He felt that the failure to simulate such complexities might be a significant factor for explaining the unnatural quality of synthesizers. He carried his investigation further by exciting a synthesizer with a source waveform that closely simulated the source waveforms he obtained by inverse filtering. The speech output was compared to the output of the synthesizer with a simple source waveform obtained by passing repetitive pulses through a simple shaping network. He found a definite preference for the source closely simulating the waveform obtained by inverse filtering.

S. B. Michaels *et al.*[8,9] have also reported exciting a synthesizer with various source waveforms to observe the effect on quality. They used a source based on high-speed motion pictures of glottal movements, together with calculations transforming area functions to volume velocity functions. This source was compared with triangular-shaped sources and sources in which relative phases of harmonic components were altered. Results were not reported in the abstracts.

The present study uses basically the same pitch-synchronous analysis technique used by Mathews *et al.* with the addition of pitch-synchronous resynthesis to produce speech utterances with simulated excitations.

For the analysis, high-quality recordings of speech utterances are made directly into a digital tape recorder with a careful specification of transmission parameters of the recording process. Pitch periods are marked on the speech waveform (either by visual inspection or by an automatic process which detects the onsets of each period). The analysis proceeds by treating each period as if it were obtained from an absolutely periodic process, then applying Fourier analysis to obtain spectral coefficients at harmonics of the pitch frequency. The peaks of the amplitude spectrum are used to obtain an initial guess for the four lower formant-frequency locations. A trial spectrum is calculated by representing the formant resonances by complex conjugate pole pairs and specifying pole-zero representations for recording parameters, radiation, higher-order formant poles, and an initial source excitation. The formant-frequency locations are varied systematically until the best match in the least-squares sense is obtained between the observed and trial spectra. Inverse filtering follows by cancelling out the calculated formant poles from the observed spectrum. An inverse Fourier transformation of the residual spectrum yields the excitation waveform. It is well to keep in mind the principal assumptions in this analysis. First, each period is treated as if it were obtained from an absolutely periodic process. The effects of system variations within individual periods are therefore ignored. Second is the implicit assumption of a clear separation between the source and vocal-tract characteristics, ascribing all poles to the tract and all zeros to the source.

In the present experiment, two speech utterances recorded by a low-pitched male speaker, the consecutive CVC syllables HAYED, HOD, HODE and the complete sentence FEW THIEVES ARE NEVER SENT TO THE JUG are subjected to the analysis described above. The analysis, carried out over the vowel portions of the utterances, is executed on a General Electric 635 computer. It provides a period-by-period tabulation of pitch frequency, formant frequencies, bandwidths, and excitation waveforms resulting from inverse filtering. An example of a few periods of the speech and analyzed excitation waveform for the vowel in HOD is shown in Fig. 1. Note the clear pulselike character of the excitation waveform. The duty cycles (ratios of open time to period) are somewhat less than 50%. There is just one sharp slope discontinuity occurring at closure followed by the onset of the principal excitation

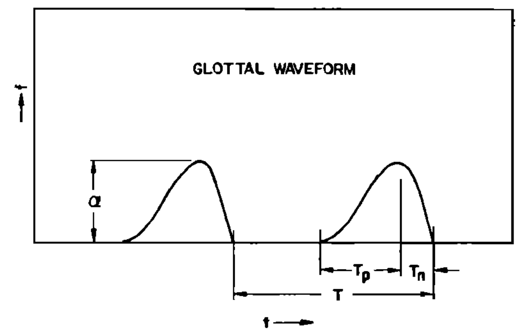

FIG. 2. Simulated pulse-shape parameters.

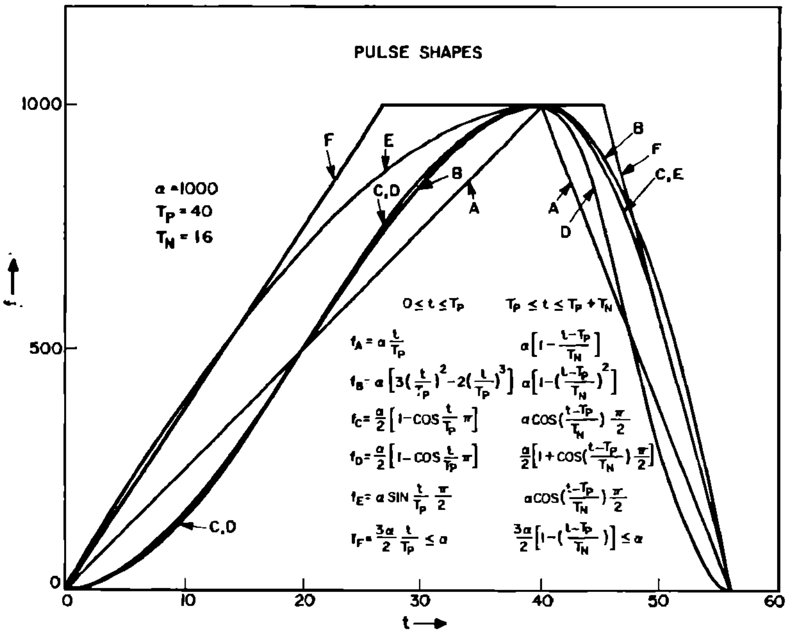

Fig. 3. Pulse shapes for Expt. I.

in the speech waveform. The kind of excitation seen here is particularly "clean." In comparison with glottal waveforms obtained by Miller and Mathews,[5] it seems to be fairly typical of low-pitched males speaking moderately loud, which is the case for all our experimental utterances.

There were two experimental aims. First, it was decided to simulate the basic shape of this natural excitation but to vary the relative opening and closing times to determine the effect of their variation on subjective quality. Second, it was decided to reduce the characterization of fine detail in the excitation to a specification of the number and location of slope discontinuities on the pulse. Excitations were simulated with pulse shapes differing mainly in these respects, with more or less fixed relative opening and closing times, to determine the effect of these differences on quality. In an attempt to eliminate loudness differences as a subjective parameter, the same rms level over each vowel portion of the natural utterances was maintained over these portions in the utterances with simulated excitations.

The waveform in Fig. 2 illustrates the experimental parameters. A particular pitch period is denoted by T. Opening time T_P is the portion of the pulse with positive slope; closing time T_N is the portion of the pulse associated with negative slope. Pulse amplitudes are held constant throughout all simulated excitations and do not enter as experimental parameters. For each experimental utterance, it is the relative opening and closing times T_P/T and T_N/T which are specified and held fixed throughout the utterance. These parameters will be given as percentages.

Two experiments were carried out: in one, the simulated excitations consisted of basically different pulse shapes but with the relative opening and closing times fixed at 40% and 16%, respectively; in the other, there was one basic pulse shape for the simulated excitations but with a range of four relative opening and closing times, from 9% to 60% for opening times and from 4% to 50% for closing times, for a total of 16 different combinations.

Subjective evaluation for each experiment was carried out by means of A–B comparison tests. The stimuli were presented to listeners seated in a sound booth via TDH-49 phones connected binaurally in phase. For each stimulus presentation, the listener was asked to compare the A–B samples (presented twice), and indicate which one sounded more natural. The tests were conducted under computer control. The samples, which were prerecorded on analog tape, could be fetched and presented under program supervision. A complete description of this system has been published.[10] The use of such a system provides the possibility of using sequential testing strategies; that is, the sequence of stimulus presentations can be a function of the subject's previous responses. The particular testing strategy used in these experiments is the elimination tournament. A description and discussion of this technique is given in Appendix A.

I. EXPERIMENT I

For each of the two utterances, there were seven different stimuli for the experimental investigation of the effect on quality of basically different pulse shapes: the natural utterance and six utterances with simulated excitations. The pulse shapes for the simulated excitations are described in Fig. 3. T_P/T and T_N/T are fixed at 0.40 and 0.16, respectively. With the exception of the trapezoid, the pulses consist of two functions connected

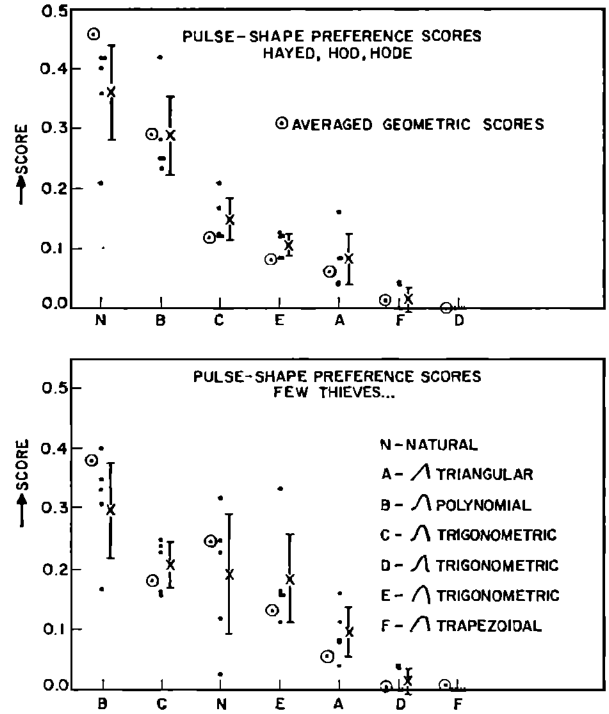

FIG. 4. Preference scores for Expt. I.

continuously at $t=T_P$. With the exception of the triangle, the connection is also made with continuous slope equal to zero. The trapezoid is a truncated triangle whose apex is at $t=T_P$. As mentioned previously, the basic difference between pulse shapes is the number and location of slope discontinuities. D has no slope discontinuities; B and C have one slope discontinuity at closing; E has slope discontinuities at opening and closing; the triangle A has an additional discontinuity at the peak; and the trapezoid F has two additional discontinuities. B and C have virtually the same shape, but B is composed of polynomials while C is composed of trigonometric functions.

Figure 4 shows the preference scores for each pulse shape obtained for each of the two utterances. Arithmetic scores are shown for each of the five listeners together with their means and standard deviations and, in addition, mean geometric scores. The distinction between arithmetic and geometric scores is discussed in Appendix A. The pulse shapes are placed along the abscissa in order of their mean arithmetic scores from highest to lowest. With the two utterances compared, except for the natural samples, the scores are approximately the same with just about the same ordering. The order of decreasing preferences is the same as the order of increasing number of slope discontinuities, with the exception of the pulse shape with no slope discontinuities, which ranks with the shape with the greatest number of slope discontinuities as poorest. B ranks consistently higher than C although they both have one slope discontinuity at closure. (B does have a sharper slope discontinuity than C and therefore a greater amount of high-frequency spectral energy.) It is interesting to note that the natural sample ranks first for the CVC syllables but only third for the sentence and that, among all the samples, the greatest amount of disagreement among subjects is associated with it. We note that the ranking based on geometric scores is the same as the ranking based on arithmetic scores, with the exception of an interchange of rank for the natural and "C" samples for the sentence utterance.

II. EXPERIMENT II

In the second series of experiments, pulse shape type C (a combination of two trigonometric functions with a single slope discontinuity at closure) was used as the basic pulse shape. This shape was modified by varying the relative opening and closing times over a range of four values each. The repertoire consisted of 16 stimuli, each one corresponding to a particular combination of relative opening and closing times.

The pooled results are shown in Fig. 5 as the rankings for the two types of material provided by average preference scores for all listeners. The rankings are displayed in the form of a square matrix with rows associated with relative opening times and columns associated with relative closing times. We note that there is no appreciable difference in the rankings between the two utterances, particularly for the six highest-ranking combinations, which are shown enclosed by a heavy border. It is perhaps easier to state which combinations are not preferred, using the enclosed "preferred" combinations as a basis. That is, combinations including the smallest value of T_P/T or the smallest value of T_N/T or combinations such that T_P/T is approximately equal to or less than T_N/T are not preferred. The combination which best matches the shape found in analyzed excitation waveforms ($T_P/T=0.40$, $T_N/T=0.16$) is approximately at the center of the "preferred" area.

A different viewpoint can be obtained by replacing the rankings in the matrices shown in Fig. 5 by the actual scores and then treating the scores as mass points, calculating first and second moments for the resulting distributions of mass. The first moment is the center of the mass of the distribution. The second moments can be represented by an "ellipse of concentra-

HAYED,HOD,HODE

T_P/T ↑ \ T_N/T →	0.04	0.09	0.19	0.40
0.60	9	2	3	5
0.33	8	1	6	12
0.18	10	4	7	11
0.10	16	15	14	13

FEW THIEVES..

T_P/T ↑ \ T_N/T →	0.04	0.09	0.19	0.40
0.60	9	4	2	6
0.33	12	1	3	14
0.18	7	5	8	10
0.10	15	13	11	16

FIG. 5. Pooled ranking tables for Expt. II.

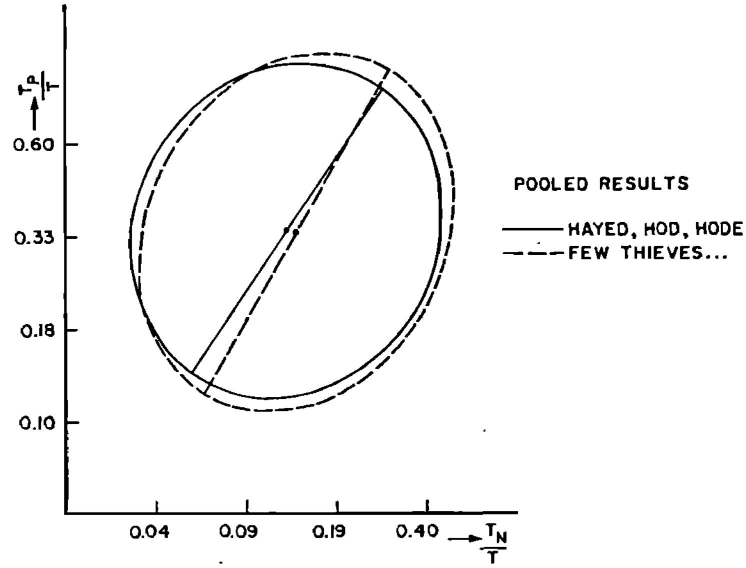

Fig. 6. Mean preferences and ellipses of concentration—pooled results.

tion" about the center of mass.[11] The ellipse encloses a uniform distribution of mass which has the same second moments as the original distribution. The relative size of ellipses indicates relative amounts of dispersion, while the eccentricity of an ellipse indicates the amount of linear correlation between the two parameters. The pooled results for the two types of material are represented in this way in Fig. 6. The small differences found here corroborate the observation made from the ranking tables that there is little to distinguish between the results for the two utterances. Figure 7 displays the results for five listeners for one utterance and one tournament strategy. We note that the centers of mass are closely clustered and that the ellipses are all about the same size. There are perhaps two listeners with ellipses oriented differently from the others. In general, however, differences are small and we would conclude that there is substantial agreement among listeners.

III. DISCUSSION

Much of the previous effort to determine the effect of excitation on speech quality has proceeded from a spectral point of view. Although the present experiments specify the excitation explicitly in the temporal domain, there are some pertinent comments which can be made by associating temporal and spectral aspects of excitation.

One question regarding the excitation spectrum is the specification of its rate of decay. Considering pulse functions, we note that those that have discontinuities have an average asymptotic decay of 6 dB/oct; continuous functions with discontinuous first (and higher) derivatives have an average asymptotic decay of 12 dB/oct, and so forth. Examination of source spectra of natural speech has shown that decays of the order of 12 dB/oct are typical.[2,12] All the source shapes used in the present experiments except D are continuous

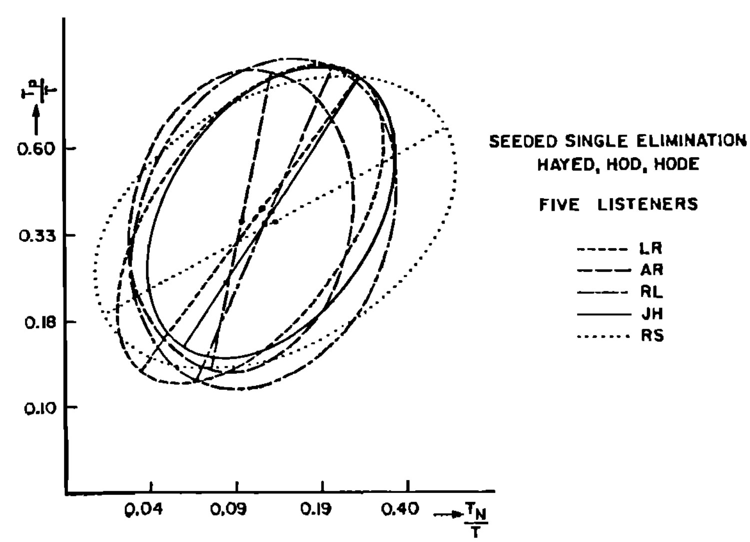

Fig. 7. Mean preferences and ellipses of concentration—five listeners.

with first-derivative discontinuities. Therefore they all have spectra with this characteristic asymptotic decay. Pulse shape D has only second- and higher-order discontinuities, so that it has an asymptotic decay of 18 dB/oct. Speech using this shape as an excitation is consistently judged as poor.[13]

Another question of importance from the spectral point of view is the role of irregularities or dips in the source spectra. Such dips are associated with complex zeros of the Laplace transforms of source pulses, that is, values of $s=\sigma+j\omega$ for which the Laplace transform $F(s)$ is equal to zero. For small values of σ, the dips in the spectrum will be quite sharp. Since the spectrum of the speech output can be considered as the product of the vocal-tract spectrum and the source spectrum, it is possible that a sharp minimum in the source spectrum can considerately flatten or cancel a formant peak in the vocal-tract spectrum, and thus considerably alter the speech sound. Flanagan has synthesized vowel sounds with source specified such that the spectral zeros coincide with formant poles, resulting in a significant perceptual effect.[14] It is not clear, however, that similar situations are likely to occur in natural speech. Dunn *et al.*[15] have shown that for a symmetric (isosceles) triangle and for other symmetric pulses the zeros lie on the $j\omega$ axis ($\sigma=0$), but that a slight perturbation from symmetry produces a large shift in σ away from zero, thereby considerably reducing the spectral dips. Given such instability with respect to σ, it seems likely that in natural speech, where there is always continuous variation in source characteristics, sharp dips are quite transitory or even nonexistent. For efficient encoding of synthetic speech, however, it is desirable to reproduce as few as possible of the details of variation without degrading the quality of the speech. Thus, if a constant pulse shape is used as an excitation, care should be taken to avoid symmetric shapes which produce pronounced spectral dips. In the present experiments, the preferred shapes B and C are intrinsically asymmetric so that the zeros can never fall on the $j\omega$ axis for any combination of relative opening and closing times. We have found that for shape C most of the spectral irregularities are obtained for shapes in which the relative opening time is approximately equal to or less than the relative closing time. In Expt. II, we have shown that such shapes rank low in preference.

Perhaps the single most important result of these experiments has been the demonstration that good-quality synthetic speech is obtained with excitation functions which can be specified by simple polynomial or trigonometric functions uniformly throughout the vowel portions of an utterance. The absence of temporal detail does not degrade the quality. The results of Expt. I have shown that the most preferred pulse shape has but a single slope discontinuity at closing. In Expt. II, we have found that there is a fairly large tolerance for different combinations of relative opening and closing times but that very small opening or closing times, or opening times less than or approximately equal to closing times, are not ranked favorably.

[1] A. E. Rosenberg, "A Computer-Controlled System for the Subjective Evaluation of Speech Samples," IEEE Trans. Audio Electroacoust. AU-17, 216–221 (1969).

[2] R. L. Miller, "Nature of the Vocal Cord Wave," J. Acoust. Soc. Amer. 31, 667–677 (1959).

[3] M. V. Mathews, J. E. Miller, and E. E. David, Jr., "Pitch Synchronous Analysis of Voiced Sounds," J. Acoust. Soc. Amer. 33, 179–186 (1961).

[4] M. V. Mathews, J. E. Miller, and E. E. David, Jr., "An Accurate Estimate of the Glottal Waveshapes," J. Acoust. Soc. Amer. 33, 843(A) (1961).

[5] J. E. Miller and M. V. Mathews, "Investigation of the Glottal Waveshape by Automatic Inverse Filtering," J. Acoust. Soc. Amer. 35, 1876(A) (1963).

[6] J. N. Holmes, "An Investigation of the Volume Velocity Waveform at the Larynx during Speech by Means of an Inverse Filter," Proc. Speech Commun. Seminar, Stockholm (1962) (Speech Transmission Lab., Roy. Inst. Technol., Stockholm).

[7] J. N. Holmes, "The Effect of Simulating Natural Larynx Behavior on the Quality of Synthetic Speech," Proc. Speech Commun. Seminar, Stockholm (1962) (Speech Transmission Lab., Roy. Inst. Technol., Stockholm).

[8] S. B. Michaels, H. I. Soron, W. J. Strong, and P. Lieberman, "An Accurate Estimate of the Glottal Waveshapes," J. Acoust. Soc. Amer. 23, 843(A) (1961).

[9] S. B. Michaels and W. J. Strong, "Analysis–Synthesis of Glottal Excitation," J. Acoust. Soc. Amer. 38, 935(A) (1965).

[10] A. E. Rosenberg, "A Computer-Controlled System for the Subjective Evaluation of Speech Samples," IEEE Trans. Audio Electroacoust. AU-17, 216–221 (1969).

[11] H. Cramer, *Mathematical Methods of Statistics* (Princeton U. P., Princeton, N. J., 1958).

[12] J. L. Flanagan, "Some Properties of the Glottal Sound Source," J. Speech Hearing Res. 1, 1–18 (1958).

[13] In informal listening, it has also been found that discontinuous pulses, such as a rectangular pulse with a 6-dB/oct asymptotic decay, are also judged as poor in quality.

[14] J. L. Flanagan, "Some Influences of the Glottal Wave upon Vowel Quality," Proc. Int. Congr. Phonetic Sci., 4th, Helsinki (1961).

[15] H. K. Dunn, J. L. Flanagan, and P. J. Gestrin, "Complex Zeros of a Triangular Approximation to the Glottal Wave, J. Acoust. Soc. Amer. 34, 1977(A) (1962).

Appendix A: Tournament Strategies[A1]

Suppose the primary experimental objective in an A–B comparison experiment is to find the most preferred sample in a repertoire of N samples. In the absence of sequential strategies, it would be necessary to compare each sample with every other sample for a total of $\frac{1}{2}(N^2-N)$ comparisons. In the language of tournaments (treating the samples as contestants), this is a round-robin tournament. In every tournament there are contestants who emerge as losers, and in view of the stated objective it is inefficient to have them participate in every round once this becomes clear. We can decide to eliminate such contestants from further play after they have lost a specified number of matches. Such elimination tournaments employ sequential stra-

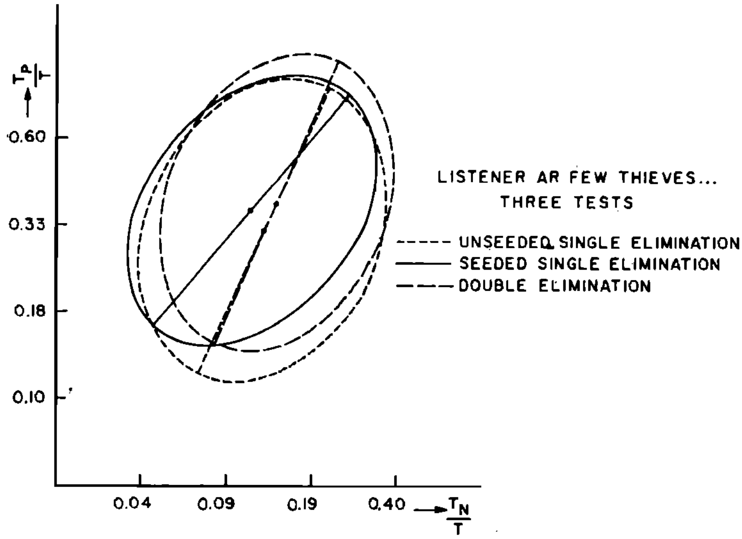

FIG. A-1. Mean preferences and ellipses of concentration—three testing strategies.

tegies, since the pairing of contestants in all rounds after the first depends on their won–loss records. In a single elimination tournament which eliminates a contestant after one loss, $N-1$ matches are required; in a double-elimination tournament, which eliminates a contestant after two losses, $2N-2$ or $2N-1$ matches are required. Thus, the number of matches in an elimination tournament is linearly proportional to the number of contestants, while in a round-robin tournament it is proportional to the square of this number.

In a round-robin tournament, preference rankings are established on the basis of the number of matches won by each contestant. In an elimination tournament, natural rankings are established by the greatest number of rounds each contestant survives. However, the rankings are ambiguous because of ties. One way to reduce ambiguity is to increase the number of losses which eliminate a contestant from the tournament. Another possibility is to repeat the tournament a specified number of times. Both of these methods have been tried in this set of experiments.

The question arises how to determine rankings over a series of tournaments. The method used here is to calculate scores for each contestant in each tournament and then rank the contestants by the scores they have accumulated over the series of tournaments. Two scoring techniques are used without any attempt to justify them rigorously. The first technique is simply to rank the samples according to the number of matches won. This is referred to as arithmetic scoring, since each score is obtained by simple addition. A second technique allots 1 point plus the opponent's score to the winning contestant in a match. This technique is referred to as geometric scoring, since the scores increase approximately geometrically with every winning match, and it reflects the approximately geometrically increasing difficulty of surviving each successive round in an elimination tournament. Implicit here is the assumption of (stochastic) transitivity. That is, if, on the average, A will defeat B and B will defeat C, then A will defeat C. The geometric scoring allots C's score to B and both C's and B's score to A, even if A has not been matched with C. It is assumed, provided each series is sufficiently long, that the ultimate ranking is independent of the initial pairings of samples in individual tournaments.

Experiment I was carried out by means of a series of two double-elimination tournaments for each of five listeners and each of the two utterances. The results that were displayed (Fig. 4) were cumulative arithmetic and geometric preference scores. There is perhaps a larger spread of the geometric scores than of the arithmetic. This is to be expected from the greater weight geometric scoring confers on the more preferred samples. Even so, it was noted that there is little difference between the rankings obtained from each scoring method.

Experiment II was carried out by means of three different tournament methods, so that some sort of comparison and evaluation of techniques is possible. First, there was a series of five single-elimination tournaments for five listeners in which the pairings of samples at the beginning of each tournament were randomized; second was a series of five single-elimination tournaments for five listeners in which the pairings at the beginning of each tournament after the first were a result of a seeding[A2] of the samples from the results of the previous tournament; third was a series of two double-elimination tournaments for four listeners in which the initial pairings were randomized. Two listeners participated in all three types of tournaments. There were two independent variables in this experiment. One method of displaying the results was by means of two-dimensional plots showing the first and second moments of the distributions of scores as a function of these variables. The first moment, the center of

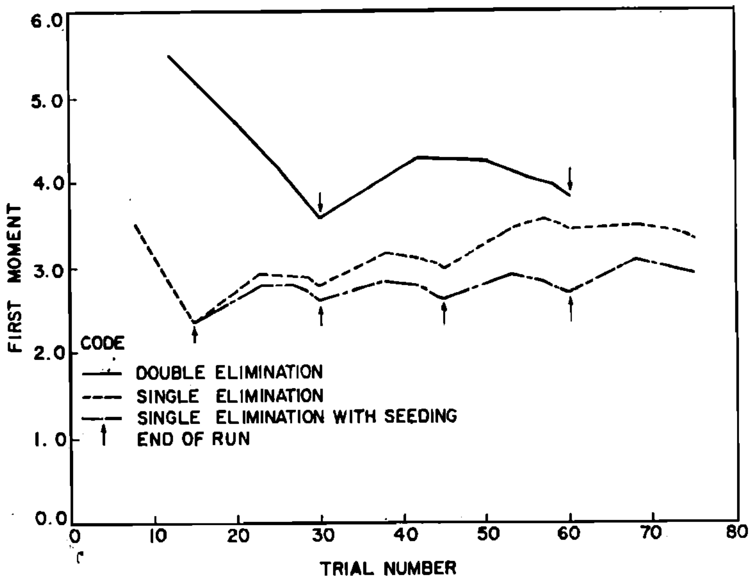

FIG. A-2. First moment of score distribution as a function of trial number for three testing strategies.

mass, is a point, and the second moments can be represented as an "ellipse of concentration" as described previously. Figure A-1 is similar to Figs. 6 and 7 and shows this type of plot for the results of one listener and utterance for each of the three different types of tournament methods. It can be seen that the locations of centers of mass and the sizes and orientations of the ellipses are quite similar. We can conclude that the results obtained are not dependent on the type of tournament method used to obtain them.

Given that the results obtained from different tournaments are not significantly different, it would be of interest to determine if one technique is more efficient than another in arriving at the result. Efficiency is a term which must be defined relative to the kind of information which we desire to extract from a test. With the present techniques, we have set our goal to obtain as clear a distinction as possible between the winner and losers after the least number of matches. For example, at the conclusion of a single-elimination tournament with 16 contestants, the winner will have scored four wins, the runner-up three wins, two players two wins each, and four players one win each; the remaining four will have scored no wins. In succeeding tournaments with the same players, we would like to see the better players of the first tournament accumulate wins as fast as possible. Suppose at any point in a series of tournaments we plot a histogram of the number of matches won (arithmetic scores) versus the players. Suppose we rearrange the histogram in descending order of scores and calculate the first moment of this distribution. The smaller this moment is—that is, the closer the center of mass is to the origin—the greater is the concentration of high scores and the distinction between winners and losers. Figure A-2 shows the relation between this moment and the progress through the different types of tournaments for one subject and type of utterance. The arrows indicate the ends of individual tournaments. There are 16 players, so that after about 60 matches four single-elimination tournaments or two double-elimination tournaments are complete. It is clear that the seeded single-elimination has the smallest moment and is consequently the most efficient in the sense defined. The differences, although consistent, do not seem great however.

Another consideration of importance is how consistent are the rankings from successive tournaments of the same type. Without entering into detail, we have examined and quantified the consistency of ranking for the three highest-ranking samples for the two listeners who participated in all three types of tournaments. It turns out that the double-elimination tournament has a slight advantage according to this measure. This is not surprising, since in the double-elimination tournament a sample can be excluded only after two losses and there is therefore less chance of eliminating a good sample. The rankings from successive tournaments can be expected to be more stable and consistent for this type than for single-elimination tournaments.

[A1] Discussions of tournament testing strategies can be found in monographs by H. A. David, *The Method of Paired Comparisons* (Hafner Publ. Co., New York, 1963) and J. W. Moon, *Topics on Tournaments* (Holt, Rinehart and Winston, New York, 1968). The emphasis in these monographs is on round-robin rather than elimination tournaments.

[A2] Seeding in a tournament is the process whereby contestants are paired initially so that matches between higher-ranking players are preserved for later rounds of the tournaments.