Canadian Acoustics /
Acoustique canadienne,
1988, 16 (1):17-23

PERIODICITY PERTURBATION IN NATURAL ENGLISH VOWELS *

Jan L. Adlington and Anton J. Rozsypal

Department of Linguistics, The University of Alberta
Edmonton, Alberta, Canada T6G 2E7

ABSTRACT

To establish normative values for natural production, jitter (glottal period perturbation) and shimmer (amplitude perturbation) were measured for nine Canadian English vowels, produced by eight male and eight female speakers in the sentence frame "Please say /hVd/ not /hVd/." The speech signals were digitized at a 20 kHz sampling rate. Following extraction of the vowel, the duration and peak amplitude of each period were measured using a semi-automatic peak-picking procedure with quadratic interpolation. Jitter and shimmer were determined as distance from a two-point linear trend line centered around the current period. Period measures were normalized by dividing this distance by the local mean period duration averaged across three periods; a similar measure was employed for shimmer. For both jitter and shimmer, unexpectedly large differences among speakers were found. The relation between jitter and shimmer within the vowel was also investigated by cross-correlating the signed jitter and shimmer perturbations of individual vowel periods. Significant correlations appeared for less than one quarter of the vowel tokens.

SOMMAIRE

Afin d'établir des valeurs normatives pour la production naturelle de la parole, des mesures de perturbation de la période glottale ou du "trémolo" ("jitter") et de perturbation d'amplitude ou de "miroitement" ("shimmer") ont été effectuées pour neuf voyelles de l'anglais canadien produites par huit locuteurs et huit locutrices dans la phrase suivante: "Please say /hVd/ not /hVd/." Les phrases ont été rendues numériques à une fréquence d'échantillonnage de 20 kHz et les voyelles en ont été extraites. La durée et l'amplitude maximale de chaque période vocalique ont été mesurées à l'aide d'une procédure semi-automatique avec interpolation quadratique pour déterminer chaque sommet. Le trémolo et le miroitement ont été définis pour chaque période comme étant la distance d'une ligne indiquant la tendance linéaire locale. Les mesures de période ont été normalisées en divisant cette distance par la moyenne locale de la durée périodique pour trois périodes. Une procédure analogue a été utilisée pour le miroitement. Pour le trémolo et le miroitement, contrairement à notre attente, de grandes différences inter-locuteurs ont été observées. Nous avons aussi étudié le rapport entre le trémolo et le miroitement au sein de la voyelle en comparant les valeurs algébriques de leurs perturbations pour des périodes vocaliques individuelles. Des corrélations significatives ont été observées pour moins d'un quart des voyelles étudiées.

1. INTRODUCTION

Small, cycle-to-cycle perturbations of the glottal period (jitter) and of peak amplitude (shimmer) have been extensively studied in attempts to non-invasively assess the functional status and health of the larynx (Heiberger & Horii, 1982). For this purpose, measurements are commonly made from the central portion of sustained vowels. However, more normal phonatory samples are required where the perceived quality of a healthy speaker's voice, or

* This paper was presented at the Acoustics Week '87, organized by the Canadian Acoustical Association, at Calgary, Alberta, in October 1987. It is based on parts of the M.Sc. thesis submitted by J.L. Adlington to the Faculty of Graduate Studies and Research of the University of Alberta.

the naturalness of a synthetic source, is at issue. The present study examines jitter and shimmer magnitudes in nine vowels produced in a sentential context. Since it was conceived as preliminary to a perceptual study of voice quality, perturbations were analyzed from the acoustic pressure wave rather than from a laryngograph signal. One aim was to confirm the results of a naturalness-rating experiment on synthetic speech (Rozsypal & Millar, 1979) which found the optimal amount of jitter to vary with the vowel quality. A secondary aim was to explore the relation between jitter and shimmer with correlations between the signed perturbation values for individual periods within each vowel token.

2. METHOD

The speakers were eight male and eight female adults, with no known speech or hearing defects. Their ages ranged from 19 to 38, with a mean of 27 years. Five male and five female subjects had never smoked.

Jitter and shimmer perturbations were measured for nine monophthongal Canadian English vowels, /i, ɪ, ɛ, æ, ɚ , ʌ, u, ʊ, ɑ/, spoken in the neutral vowel context /hVd/. To test for intonation effects, the vowels appeared in two stressed positions in the sentence frame "Please say /hVd/ not /hVd/." Thirty-six tokens (9 vowels x 2 positions x 2 replications) were produced in random order by each speaker; no vowel appeared twice in the same sentence. Subjects were instructed to speak at a comfortable loudness.

The speakers were digitally recorded in an acoustically isolated recording booth, using a high-quality microphone (Sennheiser MKH 405), a 12-bit analog-to-digital converter (Tecmar Lab Master), and an IBM AT microcomputer. A 20 kHz sampling rate was used. The amplitude sampling range was restricted to about 10 bits, providing a signal to quantization noise ratio of about 60 dB. The signals were bandpass filtered between 50 Hz and 7800 Hz; care was taken to ensure that all interfering signals were suppressed below detectable level.

A signal editor was used to isolate the vowel segment in each test word. Transitional portions were eliminated by digital gating at the vowel onset and offset, where amplitude changed rapidly over a small number of periods, or where waveshape changes suggested the start of a transition. Gating continued until playback produced a vowel percept.

The peak amplitudes and period durations within the gated vowels were found by a semi-automatic peak-picking program which allowed the operator to roughly position a forty-point wide bar cursor on a peak. The program then searched within the bar interval for the greatest sample point in the peak, and calculated the maximum of a parabola passing through this point and the ones immediately preceding and following. The y coordinate of the maximum was stored as peak amplitude; the x coordinate allowed the period duration, in ms, to be computed. This interpolation was implemented to improve the 50 μs temporal resolution given by the sampling rate. As shown by Titze, Horii, & Scherer (1987), interpolation with peak-picking can resolve jitter down to 0.1%, with fewer than 100 samples per period; about 500 samples per period would otherwise be needed to minimize the error caused by the finite duration of the sampling interval. As our calibration tests indicate, this measure could also reduce the amplitude measurement error by up to five quantization intervals.

The peak amplitude was always determined from the greatest sampling point in the period. Period durations were found by tracking a single, prominent peak throughout the vowel. Because of waveshape variations, the points from which these parameters were calculated did not always coincide. Jitter values for each of the vowels were calculated according to the following formula:

$$\text{Mean Jitter} = \frac{1}{N-2} \sum_{i=2}^{N-1} \left[\frac{\left| T_i - \frac{T_{i-1} + T_{i+1}}{2} \right|}{\frac{T_{i-1} + T_i + T_{i+1}}{3}} \right] \times 100\%$$

where T_i is the duration of the i^{th} period, and N is the number of periods included in the analysis. The numerator measures the absolute distance of T_i from an arithmetic average of the preceding and following period durations. This formula has been selected because it removes the effect of linear frequency trend on the jitter value. The denominator normalizes this value for frequency, since jitter without normalization is highly dependent on the frequency renge of phonation (Lieberman, 1963; Horii, 1979). Because frequency varies with intonation, a local average over three adjacent vowel periods was used. Shimmer was defined by an analogous formula, with peak amplitudes, A_i , replacing the period durations, T_i . Amplitude was measured on a linear scale.

Another analysis cross-correlated the signed (i.e. positive or negative) jitter and shimmer values, that is, deviations from the linear trend, for each period at lags of zero and one period. At a lag of zero, the period duration was correlated with the height of the peak within the same period; at a lag of 1, it was correlated with the peak of the following period.

3. RESULTS

3.1 Jitter and Shimmer Magnitudes

Mean jitter and shimmer magnitudes for each subject, averaged across the 36 tokens each produced, are presented in Table 1. Across all tokens, mean jitter ranged from 0.11% to 17.26%, and mean shimmer, from 0.73% to 84.43%. The exceptionally large values were taken to reflect the presence of a phenomenon additional to "normal" jitter and shimmer, often (but not always) identified as double periodicity. Based on distributional criteria, arbitrary upper limits on acceptable "normal" jitter and shimmer were established at 4.0% and 15.0%, respectively. Values above these levels were set to 1.0% for jitter and 3.5% for shimmer, the approximate overall means of the dependent variables. This step effectively eliminated them from the analysis. The outlying values were then examined separately.

From the total of 576 tokens, 29 jitter and 16 shimmer values were classed as outliers. These were not found to be dependent either on vowel quality, sentence frame position, or replication. As an example of the unpredictability of the outlier phenomenon, Figure 1 presents plots of the signed jitter and shimmer for each period in three repetitions of /æ/ produced by female speaker F5. "AEE1" (jitter = 17.20%, shimmer = 84.43%) shows the exaggerated, regular pattern typical of double periodicity; the inconsistent variations of "AEE2" (jitter = .55%, shimmer = 5.01%) and "AEW2" (jitter = .63%, shimmer = 2.54%) represent the more normal case. Outliers were characteristic of certain subjects, however, and multiple outliers occurred only with subjects F6, M2, and M3. With only one exception, the outlier tokens for shimmer were also outliers for jitter.

For both jitter and shimmer, three-way ANOVAs with repeated measures on vowels and position were performed, following adjustment of the outliers. In both cases, strong subject effects were found (jitter: $F(14,287) = 16.851$, $p<.001$; shimmer: $F(14,287) = 14.135$, $p<.001$). To examine this source of variance, hierarchical cluster analyses were produced using the Ward method. For jitter, the subjects clustered into two groups of six subjects each. Group membership was not determined by sex (nor, incidentally, by

Subjects	JITTER [%] All Values	JITTER [%] Outliers Adjusted	SHIMMER [%] All Values	SHIMMER [%] Outliers Adjusted
M1	.54	.54	3.34	3.34
M2	2.00	1.12	5.08	4.73
M3	4.81	1.27	14.27	5.46
M4	.46	.46	2.24	2.24
M5	1.21	.78	3.89	3.89
M6	.57	.57	2.15	2.15
M7	.54	.54	2.45	2.45
M8	.70	.70	4.11	4.11
F1	.43	.43	2.13	2.13
F2	.49	.49	2.33	2.33
F3	.79	.79	2.28	2.28
F4	1.11	1.11	3.14	3.14
F5	1.30	.85	5.05	2.80
F6	2.57	1.38	6.22	4.51
F7	.72	.72	2.64	2.64
F8	.80	.80	2.45	2.45

Table 1. Mean jitter and shimmer for all subjects. The effect of outlying tokens can be seen on subjects M2, M3, M5, F5, and F6.

Source	SS	df	MS	F	p
Groups	7.726	1	7.726	76.895**	.001
Subjects(G)	1.005	10	.100	1.127	
Vowels	3.889	8	.486	2.485*	.018
Groups x Vowels	.477	8	.060	.305	.962
Vowels x Subj(G)	15.653	80	.196	2.210**	
Position	.452	1	.452	4.513	.060
Groups x Position	.004	1	.004	.040	.845
Position x Subj(G)	1.002	10	.100	1.127	
Vowels x Position	.767	8	.096	1.012	.434
Groups x Vowels x Pos'n	.754	8	.094	.995	.447
Vowels x Pos'n x Subj(G)	7.576	80	.095	1.071	
Replication	.036	1	.036	.231	.641

Table 2. Analysis of variance summary table for jitter, with 12 subjects in two groups of six.

Source	SS	df	MS	F	p
Subjects	14.623	8	1.828	1.936	
Vowels	62.046	8	7.756	4.621**	.001
Vowels x Subj	107.416	64	1.678	1.778**	
Position	9.283	1	9.283	6.140*	.038
Position x Subj	12.094	8	1.512	1.602	
Vowels x Position	.710	8	.089	.120	.998
Vowels x Pos'n x Subj	47.457	64	.742	.786	
Replication	.066	1	.066	.030	.867

Table 3. Analysis of variance summary table for shimmer, for one group of nine subjects.

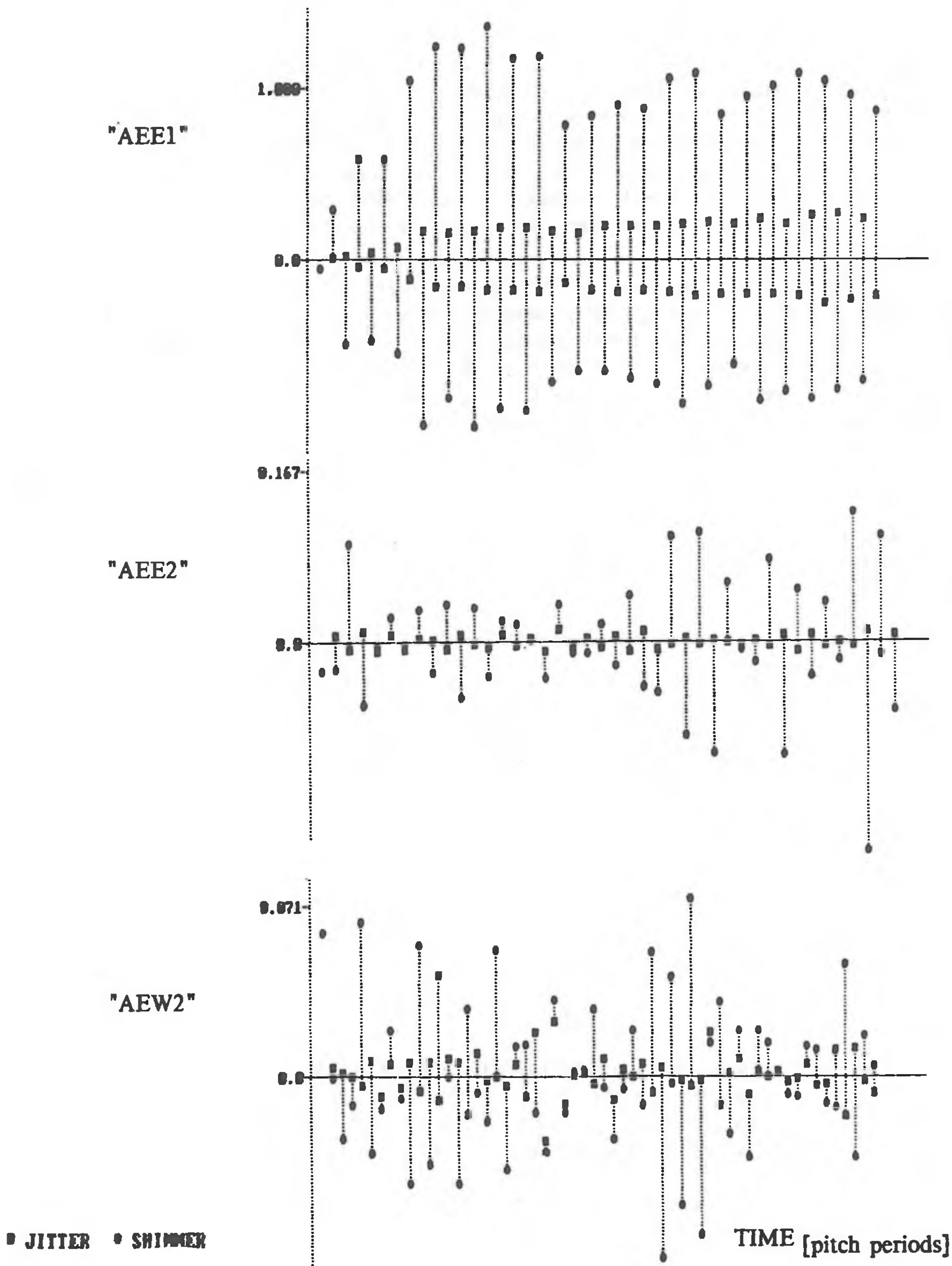

Figure 1. Signed jitter and shimmer for each period in three /æ/ segments produced by female speaker F5. "AEE1": sentence-final position, first replication; "AEE2": sentence-final, second replication; "AEW2": within sentence, second replication. The points represent relative perturbation values prior to the conversion to percentage. Note the differences in the ordinate scale.

smoking): group 1, the most homogeneous, contained two female and four male subjects, while group 2 contained four females and two males. The four remaining subjects were best treated as individuals. For shimmer, one group of nine subjects (six female, three male) was apparent. A second group of three subjects was loosely defined and not further analyzed, and four subjects were not grouped. Three of the ungrouped subjects, for both jitter and shimmer, were F6, M2, and M3, the speakers responsible for most of the outliers: even following removal of the extreme values, their measures appeared abnormal, in magnitude and consistency.

A new ANOVA was then performed on jitter for twelve subjects, with group membership as a factor. The results, presented in Table 2, show a significant vowel effect at the .05 level, and a vowel by subject interaction at the .01 level. From most to least jitter, the vowels were ordered /ι, u, ω, i, ϵ, æ, Λ, α, ɝ/, though the significant interaction indicates this would not necessarily hold for any given subject. A Tukey test showed only that the extreme cases, /ι/ and /ɝ/, differed significantly ($p<.05$).

Results of a shimmer ANOVA for the nine grouped subjects are given in Table 3. Vowels, and the vowel by subject interaction, were significant at $p<.01$, and position was significant at $p<.05$. From most to least shimmer, the vowels were ranked in the following order: /æ, Λ, ω, α, ϵ, ι, ɝ, i, u/. Here, a Tukey test showed /æ/ and /Λ/ to differ from /i/ and /u/ ($p<.05$). Vowels in sentence-final position had significantly more shimmer than those within.

3.2 Jitter and Shimmer Correlations

At the .01 level, signed period-to-period jitter and shimmer perturbations were significantly correlated for 140 of the 576 tokens (24.3%). Significant positive correlations at a lag of 0, or negative correlations at a lag of 1, accounted for 115 of these cases (20.0% overall, 82.3% of the significant tokens). This number includes 26 of the 30 tokens for which jitter or shimmer was an outlier. In contrast, negative correlations at lag 0 or positive correlations at lag 1 appeared in only 25 cases (4.3% overall, 17.7% of the significant tokens). Thus, where peak amplitudes and period durations were positively correlated, long periods tended to follow high peaks. Examples of correlated and uncorrelated measures can be seen in Figure 1: "AEE1" is an outlier with highly significant jitter/shimmer correlations (lag=0, r=.97); "AEE2" is a "normal" token with a significant r (lag=0, r=.54); and "AEW2" is a "normal" token without correlations (lag=0, r=.08; lag=1, r=-.16).

4. DISCUSSION

This study's focus on the productions of healthy speakers, to be related to perceived voice quality, emphasizes the differences which exist among members of the normal population. The between-subject differences found suggest that jitter and shimmer should perhaps be considered among the parameters which characterize individual speakers. The measures from subjects F6, M2, and, particularly, M3, also indicate that large perturbations can be habitually produced with non-pathologic origins.

The significant vowel by subject interactions, found after grouping subjects, show that vowel differences must be interpreted cautiously. Results ranging from no vowel effects, to vowel by sex interactions, have been reported in the literature. Kasprzyk and Gilbert (1975), measuring jitter in five sustained vowels, concluded that differences did not exist among the vowels studied; Ramig and Ringel (1983), with shimmer in three comfortable duration vowels, also failed to note any vowel effect. In contrast, a study conducted by Sorensen and Horii

(1983) found that females produced more jitter and less shimmer than males for /i/ and /u/, but not for /a/. For main effects, Horii (1980) and Sorensen and Horii (1983) reported /i/ to have significantly more jitter than /u/ and /a/, and /a/ to have more shimmer than /i/ or /u/, while Wilcox and Horii (1980) found more jitter for /i/ and /u/ than for /a/. Across studies, /i/ is often seen to have more jitter and less shimmer than /a/, although the differences are not always significant. The present study is consistent with these observations. However, the variability among subjects suggests that the population of normal speakers is not homogeneous, nor can it be easily subdivided on the basis of sex, and that varying results can be expected. Speculation on the origins of vowel differences can also only be formulated in terms of tendencies.

The cross-correlation data reveal the general independence of jitter and shimmer within specific periods. While the mechanical and neurophysiological origins of the perturbations are undoubtedly complex (Baer, 1980), the same causative factors had been expected to influence and relate jitter and shimmer, to an extent greater than was found.

5. ACKNOWLEDGMENTS

The authors would like to thank Wm. J. Baker for his valuable suggestions and help with statistical data processing.

Travel assistance provided by the Faculty of Graduate Studies and Research and the Alma Mater Fund at the University of Alberta allowed presentation of this paper at the Acoustics Week '87 at Calgary, Alberta.

References:

Baer, T. (1980). "Vocal jitter: a neuromuscular explanation," in *Transcripts of the Eighth Symposium on the Care of the Professional Voice*, edited by V. Lawrence and B. Weinberg, The Voice Foundation: New York, 19-22.

Heiberger, V. & Horii, Y. (1982). "Jitter and shimmer in sustained phonation," in *Speech and Language: Advances in Basic Research and Practice*, Vol. 7, edited by N.J. Lass, Academic Press: New York, 299-332.

Horii, Y. (1979). "Fundamental frequency perturbation observed in sustained phonation," J. Sp. Hear. Res. **22**, 5-19.

Horii, Y. (1980). "Vocal shimmer in sustained phonation," J. Sp. Hear. Res. **23**, 202-209.

Kasprzyk, P.L. and Gilbert, H.R. (1975). "Vowel perturbation as a function of tongue height," J. Acoust. Soc. Am. **57**, 1545-1546.

Lieberman, P. (1963). "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges," J. Acoust. Soc. Am. **35**, 344-353.

Ramig, L.A. and Ringel, R.L. (1983). "Effects of physiological aging on selected acoustic characteristics of voice," J. Sp. Hear. Res. **26**, 22-30.

Rozsypal, A.J. and Millar, B.F. (1979). "Perception of jitter and shimmer in synthetic vowels," J. Phonet. **7**, 343-355.

Sorensen, D. and Horii, Y. (1983). "Frequency and amplitude perturbation in the voices of female speakers," J. Comm. Dis. **16**, 57-61.

Titze, I.R., Horii, Y. and Scherer, R.C. (1987). "Some technical considerations in voice perturbation measurements," J. Sp. Hear. Res. **30**, 252-260.

Wilcox, K.A. and Horii, Y. (1980). "Age and changes in vocal jitter," J. Geront. **35**, 194-198.