ACOUSTIC CORRELATES OF VOCAL ROUGHNESS: A RESUME AND EXTENSION*

RONALD W. WENDAHL

We have previously reported on the importance of jitter (aperiodicity) and shimmer (rapid and abrupt changes in glottal amplitude) to the perception of auditory roughness. We demonstrated that both jitter and shimmer show linear increments in listeners' judgments of roughness. Our results, obtained from patient phonations and computer synthesis was in agreement with investigators who used either high speed photography or traditional glottography. We were not able to console our data with investigators who described the 'harsh voice' as characterized by rapid and extreme changes in 'fundamental frequency', averaging an octave in extent. We did not try to console our data with those who reported in terms of 'perturbations' because we believed that the definition of the term did not take into account those changes in period associated with an inflected vocal signal.

In this paper we shall try to console the data to some extent, and to show how the various parameters interact to produce various voice qualities, both normal and rough.

We take the position that the vocal folds are more closely coupled to the vocal tract than some investigators have thought. Dr. Stevens is in the audience and I would like him to comment on our position if he would be so kind.

To resolve some of the data, we did a study on ten 7 and 8 year old girls, ten adolescent males very carefully selected and recorded by Hollien, and ten 18-20 year old females. In this study, the Rainbow Passage was used to elicit the vocal responses. To quickly summarize a three-year study, we did find 'voice breaks' of an octave (and greater) in extent. The breaks occurred in all groups. However, they occurred either in transitional periods between consonants and vowels or when a downward inflection resulted in vocal fry. The adolescent males showed greater amounts of jitter in what would have been the steady-state portion of the vowels, but this jitter was more closely approximated by our groups who had minor laryngeal pathologies. Never, in these portions of speech, did we find the previously described 'octave voice breaks'.

* The writer acknowledges the contributions of the following: Belinda Cole, Robert F. Coleman, Martha Kvols, and Diane Linklater. This work was partially supported by the National Institutes of Health, Grant number 5 B 01 NB 04759-0.

Thus, we could resolve the data of other investigators who used the Rainbow Passage to elicit their material. It seemed to us that they had not taken the time to segment their analyses.

We returned to the data of Sherman and Linke. They had used passages loaded with low and high vowels. Their subjects had been pre-selected to have 'Harsh' voice qualities. They found that low vowels caused an impression of greater harshness than did high vowels. These results were quickly substantiated by synthesis, but the conclusions drawn by the original investigators were suspect because the same results were obtained with constant duration. Inasmuch as duration was apparently not the important determinant of listener judgments of roughness, we wanted to know what was. The clinical interpretations would be vastly different if duration were not a significant factor.

1. PROCEDURES

Of the three sets of procedures and sub-conditions to be reported, there were several common factors. The basic stimuli were generated by an electrical laryngeal analog (LADIC) which is capable of simulating glottal waveforms and glottal amplitudes on a cycle to cycle basis. A sawtooth waveform with rapid growth and slow decay was chosen. The model program called for a rising inflection from 100 to 135 Hz. with 100 msec. allowed for amplitude modulation at 100 Hz. and then each successive period increasing by one Hz. until 135 Hz. was reached. A 300 cent jitter program was superimposed on the model which called for variations in adjacent time periods between 6.27 and 12.22 msec. After the final wave, a five second silent interval was programmed and LADIC then recycled to provide a sufficient number of stimuli for later test construction.

A second common factor was that LADIC was always fed into vowel filters and the output of the filters was the product that was stored on an Ampex P.R. 10. Playback was achieved from the P.R. 10 thru a McIntosh 45 watt amplifier driving matched K.L.H. model 10 electrostatic speakers. Room conditions for playback were not ideal. The room was a converted broadcast studio with fair sound treating. The stimuli were all presented at approximately 75 dB SPL as measured in the center of the room with the listeners seated.

At this point the conditions and subconditions will be treated separately since they involved different listeners and different statistical treatments.

1. Study One — the relationship between first formant changes and listener evaluations of roughness. In this study the vowel simulator was initially set to formant frequencies of 500, 1,500, 2,500, 3,500, and 4,000 Hz. The first three formants were set to half-power bandwidths of 50, 75 and 100 Hz. respectively. No measure was made of the bandwidth of the fourth and fifth formants with the assumption that in no way would they affect the results. Following this, all formants were left at the

preset frequencies and the first formant was varied from 100 to 700 Hz. in 100 Hz. steps. The results of this study verified Sherman and Linke's data; vowels with high first formants were significantly judged to sound more rough than vowels with low first formants.

2. Study Two — formant 1 was held constant at 250 Hz.; formants 3, 4, and 5 were held constant as in study one; formant two was varied from 1,000 to 1,800 Hz. in 200 Hz. steps. The results of this study did not show any statistical differences between conditions. None were anticipated.

3. Study Three — this study deviated from the first two in that the first three formant frequencies were varied according to the norms for males reported by Peterson and Barney and formant bandwidths were also systematically varied. Only the vowels /i/, /ae/ and /ε/ were used. The bandwidths were also varied. For condition one, the bandwidths were set for the neutral vowel at 27, 44, and 107 Hz. (narrow condition), 94, 104, and 242 Hz. (median condition) and 153, 150 and 400 Hz. (wide condition). If these numbers seem arbitrary it is because the narrowest and widest represented the limits of the equipment and the median was as close as we could get between the extremes. For the narrow, median and wide bandwidth conditions the vowels were judged most to least rough in the following order, /æ/, /ε/, and /i/.

In all cases the formant bandwidth affected judgments of roughness with the median bandwidth sounding more rough than the narrow or wide.

Both parametric and non-parametric treatments were used to analyze the results. The two statistical treatments showed almost identical data. The analysis of variance showed the vowel effect almost four times that of the bandwidth as were the respective *F* ratios. However, both effects were significant at less than the .01 percent level of confidence. The narrow bandwidth was perceived as less rough than the median or wide condition; however, little differences were found for the vowel /æ/. The reasons for the bandwidth data are obscure and it is hoped that the audience might offer suggestions.

2. CONCLUSIONS

From these data we conclude that the first formant frequency is of great importance in determining listener judgments of roughness. We disagree with Sherman and Linke that vowel duration accounts for the differences between judgments because all of our stimuli were of the same duration.

These findings have many clinical implications. Number one is that voice therapists should use low vowels for the patient to be able to easily recognize his problems. 'Low' should be interpreted carefully however, because what we are really talking about is the amount of constriction and the associated damping effects. Thus, the practical therapist should begin with the lowest vowel with the least constriction. In

English, this vowel would be /æ/. The therapist should consider the patient's ability to recognize his own voice quality. From this point on, the therapist should go from low to high vowels and then diphthongs, gradually having the patient going to more extreme changes in articulatory positions. It is suggested that the last step in therapy would be to couple a low vowel with an unvoiced consonant before going to connected speech.

Department of Speech
University of Houston
Houston, Texas

DISCUSSION

BLACK (Columbus, Ohio)
I am frequently asked by students to define VOCAL FRY. This, I find difficult to do. Any help that you can give will be appreciated.

WENDAHL
We define vocal fry, or as Hollien calls the phenomenon, pulse register, in acoustic terms as well as physiological, The 'register' is of low frequency pulses and the number of pulses comprising the single event is not relevant. What is required for the perception is that the vocal tract be allowed to damp at least 35 dB between glottal excitations.