A PHOTOPHONELOGRAPHIC ANALYSIS OF HOARSE VOICE QUALITY

RONALD W. WENDAHL

The voice quality deviation associated with adolescent voice change and with harshness has often been said to be the perceptual correlate of voice breaks of... "extreme and abrupt changes in fundamental frequency, typically one octave in extent occurring both in upward and downward directions."[1]

The purpose of this paper is to challenge the hypothesis that one octave voice breaks will be perceived as voice quality deviations and to attempt to resolve pitch and vowel quality deviations from what the author would like to call voice quality deviations.

The work reported by Bowler on "A Fundamental Frequency Analysis of Harsh Voice Quality" appears to be in agreement with all previous articles and perhaps best summarizes the prevailing thought on the acoustic correlates of harshness. The following represents an abstract of Bowlers statements.

1. Harsh quality is characterized by extreme and abrupt changes in fundamental frequency, typically of one octave in extent and occurring in both upward and downward directions.... The rate of occurrence of breaks in the harsh portions for all subjects (5.0 breaks per second) suggests that frequency breaks are a common accompaniment, if not a distinguishing characteristic, of perceived harsh voice quality.
2. Harsh quality is characterized by a relatively high incidence of low frequencies as compared to non-harsh quality.
3. Harsh quality is characterized by lower than average median frequency values.
4. Harsh quality is most frequently perceived on falling inflections and termination of phonation.
5. Harsh quality is characterized by intermittancy of phonation.
6. Harshness appears to be associated with aperiodicity.

Moore and vonLeden[2] have described another kind of harshness which they equate with the term "Glottal Fry" and point out that the vocal folds have a... "vibration pattern (that) is unique with this harsh quality". They describe this vibratory pattern as follows: "Instead of the customary opening, closing, and closed phases of the vibratory cycle, the cords separate and approximate twice in rapid succession

[1] Bowler, N. D., "Fundamental Frequency Analysis of Harsh Voice Quality," Int. Voice Conference, Chicago, 1957.

[2] Moore, P., and vonLeden, H., "Dynamic Variations of the Vibratory Pattern in the Normal Larynx," *Folia Phoniatrica*, 10, pp. 205–238 (1958).

and then remain in contact for a relatively long period." Moore and vonLeden refer to the vibratory pattern as a "syncopated rhythm". It will be of interest later in this paper to note that a fundamental frequency analysis made by Moore and vonLeden showed that the longer of the openings had a fundamental of 150 cps. and the shorter to have a fundamental of 225 cps.

The conclusions drawn by Bowler and by other investigators concerning one octave voice breaks left each investigator in a logical dilema. How could it be true that one octave voice breaks could be present in portions of vocal productions judged to be non-harsh, and how could harshness be perceived in the absence of such breaks? Fairbanks, *et al.*[3] explained why the abrupt changes in fundamental frequency observed on osciliograms of seven and eight year old girls were not perceived as voice breaks by placing considerable emphasis on the location of the breaks within the total pitch range of the individual. They state: "As long as the voice breaks are in the lower portion of the pitch, range, down from and up to the mode, they tend to be overlooked as a common phenomenon of childhood, although they may be heard if an attempt is made. As the adolescent male pitch level lowers toward the adult male pitch level, probably quite rapidly in most cases, the voice breaks, if they persist, are found after a time above rather than below the mode pitch level. With such a relationship to the mode they are then heard as anachronistic returns to the childhood level against a background of predominently low-pitched, male, quasi-adult phonation and thus become more obvious to the listener." Bowler proposed that the octave voice breaks may not be perceived because they typically occur at the termination of phonation where the vocal intensity is relatively low. When he found no voice breaks during utterances judged by trained listeners to be harsh, he hypothesized that the judges were not, so-to-speak, keeping up with the phonations of the subjects and were, perhaps, making their judgments on a word behind. Of course if this were true it would appear to negate his judgmental procedure.

It seemed to the writer at the onset of this study that if octave voice breaks occurred at all in vocal productions they would either be perceived as pitch changes or vowel quality changes but not voice quality deviations. The author further believed that the phenomenon described by Moore and vonLeden would not result in an acoustic phenomenon which would be labelled as harsh unless the closed phases between successive cycles were of an irregular duration.

It was further hypothesized that since the vocal tract is a series resonant system whose damping characteristics are such that a period of time is required for a single pulse to die down that it would be entirely possible to stimulate the vocal tract with two pulses in rapid succession which would appear at the output as a single damped wave if an osciliographic type of analysis was used. In other words, failure to demonstrate vocal perturbations from an acoustic analysis would not necessarily mean that there were none.

[3] Fairbanks, G., Herbert, E., and Hammond, M., "An Acoustical Study of Vocal Pitch in Seven and Eight Yearld-Old Girls," *Child Development*, XX. pp. 71–78 (1949).

The present study was designed in two parts. The first part was based upon 120 phonations, 60 of which were produced by persons with known laryngeal pathologies and 60 by persons without known laryngeal involvments. The second part represented a preliminary attempt to synthesize with electronic apparatus the phenomenon reported by the investigators mentioned in the introduction of this paper.

The details of the analysis of part one will be eliminated for the sake of brevity. Suffice it to say that the persons with laryngeal pathologies were all judged to have severe voice quality deviations and none of the individuals in the normal control group were judged harsh.

In order to circumnavigate the problems associated with voice breaks occurring at the termination of downward inflections, isolated vowels were phonated by the subjects. These vowels were subjected to an osciliographic analysis of the same type reported by Fairbanks and all subsequent investigators.

The osciliographic analysis indicated not one single voice break of one octave in extent, either in an upward or downward direction. Voice breaks were of course found and significantly more voice breaks were found in the vocal productions of the pathological cases than in normal speakers, but these voice breaks were always less than one octave in extent.

For part two, the author designed a basic saw-tooth generator capable of shifting duty cycle at the zero-point of any given cycle. The saw-tooth generator served as an input to a vowel synthesizer (a five formant series resonant circuit) and the output of the vowel synthesizer was fed to a tape recorder. A Techtronik dual-trace osciloscope was plugged into both the output of the saw-tooth generator (the "larynx") and the output of the vowel synthesizer (the "mouth").

The first phenomenon synthesized was of a 100 cycle tone changing on alternate cycles to 200 cps. thus synthesizing the one octave voice breaks so often described in the literature. The perceptual correlate of the output was, of course, musical, not harsh, and as predicted the fundamental was perceived as 100 cps (matched to a pure tone by five trained judges) while the vowel quality changed from /a/ as in father to /ɔ/ as in caught.

As various vowels were synthesized it became readily apparent that certain osciliograms taken from the output of the whole system could be used as a valid measure of the vocal input while others could not. The damping characteristics of the resonant system was such that certain vowels clearly showed the characteristics of the input and others did not.

The repetition rate of the saw-tooth generator was then changed to vary from 150 to 225 cps on alternate cycles. Five subjects were again used to make pitch matches with a pure tone generator. All reported the task difficult but four out of five matched the pitch to 75 cps. One judge matched the pitch to 36 cps, approximately one octave lower. The tone was, indeed, of low pitch but completely melodious and not harsh. Dual-trace oscilographic analysis again indicated that the typical phonelographic approach to fundamental frequency analysis may not always yield valid results.

It will be noted that the frequencies synthesized of 150 and 225 cps are exactly those reported by Moore and vonLeden. In as much as the author was not familiar with this datum the correspondence of frequencies was a fortunate accident. At this moment, it may seem that the two sets of data are in disagreement. Moore described this phenomenon as vocal fry and associated the syncopated rhythm with harshness. The present study described the phenomenon as musical but low in pitch. Moore and vonLeden show a long pulse with a relatively wide opening followed by a short pulse with a more narrow opening. Wendahl shows a long pulse of greater amplitude followed by a shorter pulse of lesser amplitude.

Why then, you may ask, are the results apparently different. It may be noted that the pictures of vocal fold action described by Moore and vonLeden show a variable spacing between pulses, while in the case of the present study the time between the repetition of the double pulses is constant. Had it been possible within the confines of the equipment available during this part of the study to vary the off-phase as well as the duty cycle it is doubtful that the two sets of data would be in any disagreement.

The next set of data to be discussed will be tentative. The number of judges used is too small for generalizations to be made and so even though exact numbers will be quoted it is to be remembered that these numbers may change as a greater number of subjects is added to the experimental data. Nevertheless, results had not yet been obtained which were judged harsh. To achieve harsh perceptions, frequency modulated saw-tooth wave, best described as a sinusoidally varying frequency vibratto, was used.

At a basic repetition rate of 100 cps, where frequency was varied from 98 to 102 cps, the quality sounded harsh at 10 vps. When the frequency was varied from 95 to 105 cps, 6 vps resulted in harsh judgements. When 90 to 110 cps was used, four vps resulted in harsh or rough judgments.

At the present moment the instrumentation has not yet been sufficiently developed to allow frequency breaks in any desired programmed manner, but within the year it is hoped that a great deal of data will become available on the perception of harshness.

Houston Speech & Hearing Center
Texas Medical Center