ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

RELATIONSHIP BETWEEN ACOUSTIC MEASURES OF VOCAL PERTURBATION AND PERCEPTUAL JUDGMENTS OF BREATHINESS, HARSHNESS, AND HOARSENESS

Fred D. Minifie, Daniel Z. Huang, and Jordan Green

Speech and Hearing Sciences, University of Washington, JG-15
Seattle, WA 98195 USA

ABSTRACT

This paper compares the effects of acoustic perturbations on perceptual judgments of voice quality in disordered and synthesized vowels. Five levels of jitter, shimmer, and glottal noise, respectively, were systematically added to synthesized vowels. Perceptual judgments of breathiness, hoarseness and harshness showed that glottal noise was correlated more strongly with perceptual judgments of Breathiness and Hoarseness than with Harshness; and was a better predictor of all three perceptual dimensions that were jitter and shimmer.

I. INTRODUCTION

The use of acoustic indicators of laryngeal function during voice production by patients with laryngeal pathology has been appealing to many clinicians.[1] The primary appeal is that these are noninvasive objective measures that can be compared to data obtained from persons with normal voices.[2] The four most commonly used acoustic measures are: 1) fundamental frequency (F0) of vocal fold vibration, 2) pitch period perturbations (jitter), 3) pitch period amplitude perturbations (shimmer), and 4) normalized noise energy (NNE) which is an indicator of the amount of glottal noise in the vocal signal. Each of these acoustic parameters is presumed to have a direct relationship to normal and aberrant laryngeal behavior during voice production. For example, increased aperiodicity in the voice signal (jitter) is presumed to be related to increased harshness in the voice; increased NNE to increased breathiness, and so forth.[3,4] These assumed relationships have led voice clinicians to consider using acoustic measures to obtain diagnostically significant information, or as indices of vocal function that could enable them to monitor improvement in the voice during treatment. However, voice scientists have recently discovered that certain behavioral changes during normal voice production can significantly alter these acoustic parameters.[5] For example, changes in jitter, shimmer, and glottal noise that result from changes in the vocal effort level employed by the talker may be far greater than changes in those parameters that result from various types of laryngeal pathology.[6] Changes in vocal tract shape associated with the production of different vowels will also significantly alter the magnitudes of jitter, shimmer, and NNE obtained during sustained vowel productions.[5] In fact, many of the changes in the vowel signal that result from prosodic patterning (intonational changes, amplitude changes), and those resulting from microphone selection and placement, have caused voice scientists to seek a stabilized standardization of voice recording conditions if acoustic perturbation measures are to be used in evaluation of voice.[7,8] At the very least, voice clinicians and voice scientists ought to use caution in interpreting the results obtained with from acoustic perturbation measures. The professional community needs a better foundation for interpreting the information obtained from acoustic analysis of the voice signal.

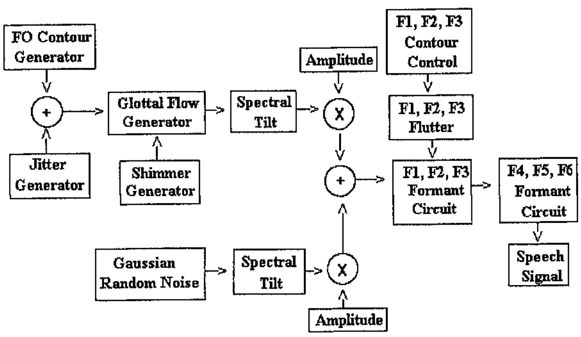
As a next step in developing an appropriate foundation for interpreting acoustic perturbation data, this study was designed to investigate the relationships among the acoustic perturbation measures of jitter, shimmer, and glottal noise, and perceptual judgments of the voice qualities of breathiness, hoarseness, and harshness. These relationships were investigated using both synthesized vowels and vowels produced by patients before and after laryngeal surgery.

II. METHODS AND PROCEDURES

Experiment 1: Synthetic Vowel Stimuli

The purpose of this experiment was to create a series of synthetic vowels in which only one acoustic perturbation dimension at a time would be varied, in order to determine the specific perceptual consequences of jitter, shimmer and NNE, respectively. Four hundred and sixty two three-formant synthesized vowels were generated using the *Dr. Speech Science* software developed by Tiger Electronics.[8] The synthesized vowels generated were /i,I,E,ae,a,U,u/. To account for gender differences, synthesized vowels were created using both male and female formants. Formant frequencies for the male and female tokens of these vowels were based on the normative data of Peterson and Barney (1952). The vowels averaged 400 ms in duration. Half of the vowels were produced using a flat intonational contour and half were produced using a rise-fall contour. See Figure 1 for a diagram of the speech synthesis system used in Experiment 1.

Each vowel was synthesized with a specified amount of jitter, shimmer, or NNE. The amount of an acoustic variable imposed on a vowel was predetermined to fall within one of five acceptable levels (ranges) for that acoustic dimension. (See Table 1 for values of jitter, shimmer, and NNE at each of the five levels created for Experiment 1.) Level one represented minimal amounts of that acoustic dimension and level 5 represented maximum amounts of that acoustic variable. Values within the ranges were chosen to represent the range from normal vowel production (level 1) to severely disordered vowel production (level 5).



The Model of Normal and Pathological Voice Production

Table 1. Acceptable Ranges for Jitter, Shimmer, and NNE

Level	Acceptable Range
	% Jitter
1	(0.00-0.00)
2	(0.68-0.83)
3	(1.35-1.65)
4	(2.03-2.50)
5	(2.70-3.30)
	% Shimmer
1	(0.00-0.00)
2	(3.60-4.40)
3	(7.20-8.80)
4	(10.80-13.20)
5	(14.40-17.60)
	NNE
1	(-18)-(-20)
2	(-14)-(-16)
3	(-10)-(-12)
4	(-6)-(-8)
5	(-2)-(-4)

In order to isolate the perceptual effects of each acoustic variable, only one acoustic change was imposed on an iteration of each vowel. Given that jitter, shimmer and NNE are not mutually exclusive phenomena, the manipulation of one variable inadvertently would have some effect on the other variables. For instance, when synthesizing a vowel with extensive jitter, trace levels of NNE and shimmer would also be detected. Although the contaminations were small, they will be addressed in subsequent papers.

Experiment 2: Disordered Voice Samples

Two hundred and eight digital recordings of the vowel /ae/ produced by 30 voice patients during various stages of their treatment (Hirano voice samples--Kurume '87) were also utilized. The durations of sustained vowels varied widely (from about 200 ms to 3000 ms in length). Approximately 10 percent of the vowels were unanalyzable due to acoustic aberrations. For example, some of the recordings were of vowels containing unacceptable noise from nonvocalic sources, some vowels had silence gaps interrupting the vowel, etc. Each measurable sustained vowel produced by these patients was acoustically analyzed to obtain measures of jitter, shimmer and NNE, using the acoustic analysis algorithms in the *Dr. Speech Science* software.

Listening Task--Perceptual Judgments

Perceptual judgments of breathiness, hoarseness and harshness were obtained for each vowel stimulus in Experiment 1 and Experiment 2 from 10 graduate students majoring in speech-language pathology who were experienced in making vowel quality judgments. The listening task took approximately 5 hours (3 listening sessions of about 100 minutes each). Subjects were required to perform the listening tasks in a sound proofed booth (IAC 1200 series). All stimuli were presented through 2 loudspeakers at a comfortable loudness level.

Judges were required to judge only one voice quality at a time. They listened to 462 randomly ordered synthesized vowel samples and 208 disordered voice samples in each 100-minute listening session. a 1-minute break occurred after each set of 52 stimuli, and a 5-minute break after 231 consecutive token presentations.

Listener Training. All listeners in the perceptual judging task received training at the beginning of the first listening session. In addition, judges were provided with a definition of the voice quality to be rated at the beginning of each session. The definitions were adapted from Bassich and Ludlow, (1986). Perceptual training included the presentation and rating of representative vowel tokens from the real voice samples used in Experiment 2. The tokens used for training were selected by an experienced voice clinician to be representative of the range of severity for each of the voice qualities (breathiness, hoarseness and harshness) to be rated. During training, students were familiarized with the perceptual rating form.

Perceptual Ratings. Vertically arranged continuous rating scales of 10 cm in length were used to obtain ratings of breathiness, hoarseness, or harshness. The polar positions on the scales were identified with dimensional adjectives (e.g., "no breathiness" to "extremely breathy"). After the brief training period during which judges practiced rating 8 sample vowels, the blocks of 52 experimental vowel tokens were presented. Each token was presented two times in succession, separated by 500 ms of silence, followed by a three-second judging interval. Immediately following the second presentation of a vowel judges were instructed to indicate the level of severity of the voice quality on the rating sheet provided. The rating sheets contained a separate vertical line for each vowel. The top of the vertical line was severe and the bottom end of the scale was normal. When judging tokens subjects were required to make a horizontal mark across the vertical line at the point most representative of the severity of a specified voice quality.

III. RESULTS AND DISCUSSION

Mean perceptual ratings from the 10 experienced listeners were used as measures of central tendency for each perceptual dimension. The major finds of this study were as follows. 1. Listeners perceived greater changes in the voice quality dimensions of breathiness, hoarseness and harshness for "real" voice samples wherein jitter, shimmer and NNE were all varying than for synthetic vowels in which only one acoustic perturbation dimension was varied. This trend was particularly evident for the breathiness and hoarseness conditions, and less so for the harshness condition. 2. Intercorrelations between the acoustic perturbation measures of jitter, shimmer and NNE and the perceived vowel qualities of breathiness, hoarseness and harshness showed stronger correlations between NNE and the perceptual judgments than for either jitter or shimmer changes. This difference was particularly evident in the synthesized vowel conditions, although the same trend was observed for the "real" disordered voice samples.

Two additional statistical analyses were completed. First, the values of jitter, shimmer and NNE for each vowel were used as independent measures to predict each of the dependent perceptual ratings of vowel quality using a multiple correlation procedure. Secondly, an analysis of variance was computed to evaluate the perceptual differences between the various levels of jitter, shimmer, and NNE. The results of this experiment show the effects of each of the acoustic perturbation functions on the perceptual identification of voice qualities. Since the tabled data are more extensive than can be accommodated in this proceedings, interested readers are encouraged to write to the first author to obtain copies of the tables for these two analyses. These results are presented on the Poster Display.

IV. CONCLUSIONS

Caution must still be exercised in interpreting the results of acoustic perturbation analyses of disordered voices. Small but significant correlations exist between perceived breathiness and hoarseness and the amount of noise in the vocal signal. Jitter is the best predictor of perceived harshness.

REFERENCES

[1] H. Kasuya, S. Ogawa, and Y. Kikuchi, "An Acoustic Analysis of Pathological Voice and its Application to the Evaluation of Laryngeal Pathology," *Speech Communication,* Vo. 5, No. 2, June 1986.

[2] R. J. Baken, *Clinical Measurement of Speech and Voice,* A College-Hill Publication, 1987.

[3] M. Hirano, H. Matsushita, and S. Hiki, "Acoustic Analysis for Voice Disorders: a Basic Conception for the use of Acoustic Measurements for the Diagnosis of Voice Disorders, *Pract. Otol. (Kyoto)*, Vo. 69, 267-271, 1976.

[4] E. Yumoto, Y . Sasaki, and H. Okamura, "Harmonics-to-Noise Ratio and Psychophysical Measurement of the Degree of Hoarseness," *J. Speech Hear. Res.,* Vo. 27, 2-6, 1984.

[5] R. F. Orlikoff and Z. Huang, "Influence of Vowel Production on Acoustic and Electromyographic Perturbation Measures," Paper presented at ASHA Convention, 1991.

[6] Z. Huang, F. D. Minifie, and X Lin, "Measures of Vocal Function during Changes in Vocal Effort Level," Paper presented at ASHA Convention, Anaheim, CA 1993.

[7] I. R. Titze and W. S. Winholtz, "Effect of Microphone Type and Placement on Voice Perturbation Measurements," *J. Speech Hear. Res.,* Vo. 36, 1177-1190, 1993.

[8] I. R. Titze and H. Liang, "Comparison of F0 Extraction Methods for High Precision Voice Perturbation Measurements," *J. Speech Hear. Res.,* Vo. 36, 1120-1133, 1993.

[9] G. E. Peterson and H. L. Barney, "Control Methods used in the Study of Vowels," *J. Speech Hear. Res.,* Vo. 9, 68-99, 1952

[10] C. J. Bassich and C. L. Ludlow, "The Use of Perceptual Methods by New Clinicians for Assessing Voice Quality," *J. Speech Hear. Dis.,* Vo. 51, 133, 1986.