The Impact of Breathiness on the Intelligibility of Speech

Louise Thompson, B.A., Emily Lin, PhD., and Michael Robb, PhD

Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Abstract

The aim of this study is to determine how deterioration of voice quality, such as breathiness, may impact on the intelligibility of speech. Acoustic analysis was conducted on sustained vowel phonation as well as discrete segments taken from recorded sentences, retrieved from a database of voice disordered speakers. Measures included: frequency of the first two formants (F1, F2), singing power ratio, the amplitude difference between the first two harmonics (H1-H2 amplitude difference), voice onset time, and energy ratio between consonant and vowel (CV ratio). A series of two-way (glottal closure x vowel) repeated measures Analysis of Variances conducted on these acoustic measures showed a significant glottal closure (complete vs. incomplete) or glottal closure by vowel interaction effect for the F2 frequency, H1-H2 amplitude difference, and singing power ratio. Based on findings in literature that reported a dominant first harmonic as a useful predictor of breathiness, the measure of H1-H2 amplitude difference was selected as a factor for investigation of the impact of voice quality on the perception of vowel intelligibility and clarity. Fixed-length vowel segments at five levels of H1-H2 amplitude difference were presented to 10 male and 10 female inexperienced listeners between the ages of 19 and 34 years. It was expected that the tokens with a dominant first harmonic, indicative of a more breathy voice, would be associated with a lower rate of correct vowel identification and a lower rate of being perceived as "clearer". The finding of a change of the perceptual ratings as a function of the H1-H2 amplitude difference will demonstrate the effect of voice quality on vowel intelligibility.

Introduction

Voice quality refers to the perceived auditory characteristics that mark an individual's speech (Gerratt & Kreiman, 2004). Voice quality such as breathiness, hoarseness, roughness, strain, weakness, and anomalies in pitch are common across the various types of voice problems (Sapienza & Ruddy, 2009). Pathological voice quality is associated with disorders resulting from mass or non-mass lesions, neurological problems, or functional problems (Sapienza & Ruddy, 2009). Voice quality is what most concerns people with voice disorders (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). Voice patients seek treatment because they do not sound normal, and judge the success of treatment on whether they sound better (Kreiman et al., 1993). Voice refers to the sound produced at the glottis by vocal fold vibration (Titze, 1994). With a pathologically breathy voice, the vocal folds vibrate but they are not completely closed during vocal fold vibration (Reetz & Jongman, 2009). It is the air leakage during the closed phase that gives the voice a breathy quality (Reetz & Jongman, 2009).

Methods

Stage One: Acoustic Analysis

Acoustic parameters from voice samples were measured for analysis. The voice samples consisted of segments taken from running speech and sustained vowels obtained from a database of voice recordings of individuals with varying degrees of voice pathology, classified according to whether they achieved complete or incomplete glottal closure during videostroboscopic examinations on the day of voice recording.

Voice Recordings Voice recordings were previously recorded from voice patients seen in the voice clinic in the Otolaryngology Department at the Christchurch Hospital. Digitized voice files of 26 voice patients, including 13 cases associated with complete and 13 cases with incomplete glottal closures were retrieved for acoustic analysis. The "complete glottal closure" group included 7 males (aged from 32 to 65 years; Mean = 46.7, SD = 12.2) and 6 females (aged from 29 to 54; Mean = 40.0, SD = 9.9) and the "incomplete glottal closure" group included 7 males (aged from 24 to 81; Mean = 48.4, SD = 20.3) and 6 females (aged from 43 to 68; Mean = 55.3, SD = 10.8). The recordings consisted of sustained vowels along with readings of the first six sentences in The Rainbow Passage (Fairbank, 1960). Words with similar syllabic and phonetic structure were selected from the sentences.

Instrumentation The acoustic recording system consisted of a headset microphone (AKG C420, Austria) and a mixer (Eurorack MX602A, Behringer) used as microphone preamplifier. The output of the mixer was connected to a 12-bit A/D converter (National Instrument DAQCard-AI-16E-4, USA) via a SCB-68 68-pin shielded connector box. The connector box contained a filter for the acoustic signals to be low-passed at 20 KHz. The A/D converter was housed by a laptop computer (Compaq 650 MHz Pentium 4, Taiwan) for direct digitization. Time-frequency analysis software TF32 (Milenkovic, 2001) was used to perform analysis of the acoustic signals.

Experimental Measures

I. Parameters measured from the mid portion of sentence-embedded vowels /a, i, o, u/ and sustained vowels /a, i/ included:

- Frequencies of the first two formants (F1 and F2) determine vowel space area which if large = more intelligible speech (Bradlow, Torretta, & Pisoni, 1996; Krause & Braida, 2004)
- H1-H2 amplitude difference (H1 H2): The amplitude difference between the first two harmonics higher (absolute) H1-H2 = more breathy voice (Bickley, 1982; Hillenbrand, Cleveland, & Erickson, 1994; Klatt & Klatt, 1990)
- Singing power ratio (SPR): The amplitude difference between the peak harmonics within the 2–4 kHz and 0–2 kHz frequency bands lower (absolute) SPR = greater vocal power (Kenny & Mitchell, 2006; Omori, Kacker, Carroll, Riley, & Blaugrund, 1996; Sundberg, 1987)

II. Parameters measured from consonants and vowels from sentence-embedded words "reach", "long", "arch", and "two" (CV ratio); and "pot", "people", "two", "take", and "colours" (VOT) included:

- Consonant-to-Vowel *Energy* Ratio (CV ratio): The ratio of the power of a consonant to that of the nearest vowel in the same syllable higher consonant energy = greater intelligibility (Gordon-Salant, 1986, 1987; Kennedy, Levitt, Neuman, & Weiss, 1998)
- Voice onset time (VOT): The time lapse between the release of a stop consonant and the onset of voicing for the following vowel higher VOT = better control of laryngeal structures and vocal tract articulators for more intelligible speech (Koenig, 2000; Monsen, 1978)

Stage Two: Perceptual Study

Participants and Participant's Task The participants were ten female and ten male adult native English speakers aged between 19 and 34 years with normal hearing and normal speech, language, and hearing history. Participants listened to sets of randomized stimuli and performed two forced choice tasks based on their perception of the stimuli. The 50 ms segments sectioned from sentence readings ("sentence-embedded vowels") were used in a "vowel identification" task for which listeners selected the vowel which best approximated their perception of what they heard. The 500 ms tokens taken from sustained vowels ("sustained vowels") were used for a discrimination task for which the listeners indicated which they perceived to be the "clearer" of two tokens presented in pairs. With both tasks, the listeners could repeat the sounds before selection. Instruction was provided on the interface screen of the programme and verbally by the experimenter. A Pearson Product Moment Correlation run on H1-H2 amplitude difference measures of the two types of speech stimuli (sustained & embedded) resulted in a correlation coefficient (r = 0.639) that suggested that there was a moderate positive correlation between the two types of stimuli.

Instrumentation Adobe Audition software was used to normalize the intensity of the sound samples. The listening tasks were carried out in University of Canterbury Communication Disorders Department sound treated booth. A Grason Stadler GSI 61 audiometer with TDH50 supra aural headphones was used to screen hearing. A locally developed computer algorithm written in C++ was installed in a desktop computer equipped with a high-quality sound card to present the stimuli and record responses.

Results

Stage One: Acoustic Analysis

- As shown in Figure 1, the second formant frequency (F2) measured from the male sustained vowels was significantly higher for the incomplete glottal closure group than the complete glottal closure group [glottal closure effect: $F(1,14) = 7.151$, $p = 0.018$; Glottal closure by vowel effect: $F(1, 14) = 7.292$, $p < 0.01$]. Post-hoc test results: F2 "incomplete closure" significantly > "complete closure" only for /a/ (suggesting smaller vowel space)
- The dominance of the first harmonic relative to the second harmonic (H1-H2 amplitude difference) measured from the male sentence-embedded vowels was significantly higher for the incomplete [illegible] glottal closure g

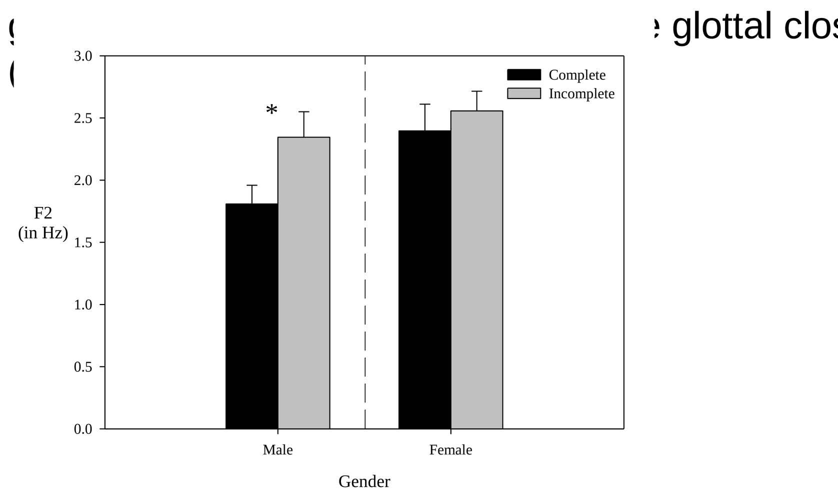

Figure 1. Means and standard deviation of means calculated from the *F2* acoustic measure obtained from "male and female sustained vowels" in relation to glottal closure. (Male complete n=18, incomplete n=14; Female complete and incomplete n= 14).

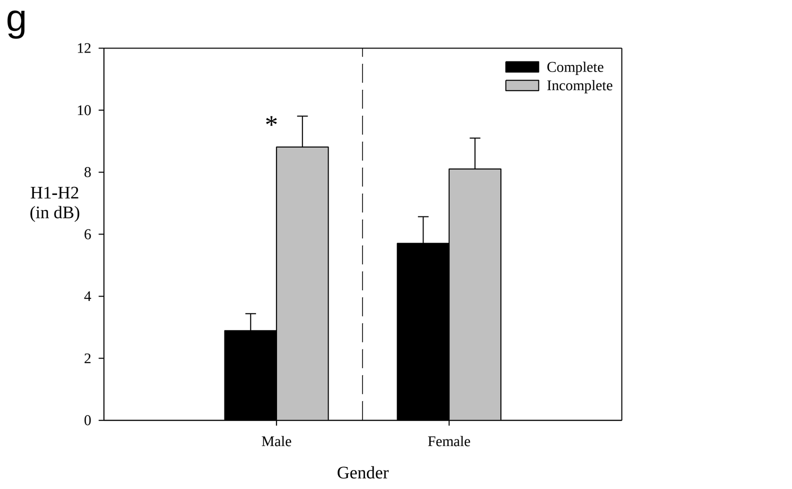

Figure 2. Means and standard error of means calculated from the *H1-H2* amplitude difference acoustic measure obtained from "male and female sentence-embedded vowels" in relation to glottal closure. (Male complete n=24, incomplete n=14; Female complete n=28, incomplete n= 24).

Significantly different pairs in each data set are marked with an asterisk ("*").

Stage Two: Perceptual Study It was expected that results would follow a pattern of the lowest level (level 1) of H1-H2 amplitude difference (supposedly less breathy) achieving the highest percentage correct and percentage clear scores and that these scores would decrease with increasing level of H1-H2 amplitude difference . As can be seen in Figures 3-6, results were somewhat ambiguous with some levels conforming to the trend and others not.

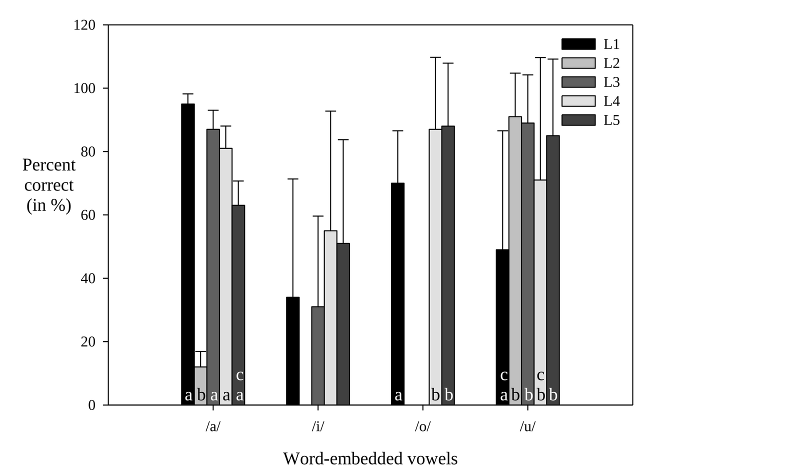

Figure 3. Means and standard deviations of percentage correct results calculated from the averaged listener responses (n=20) to a set of "male sentence-embedded vowels" used in the *"vowel identification" task.* Levels L1 to L5 represent increases in *H1-H2 amplitude difference* with L1 being the lowest level.

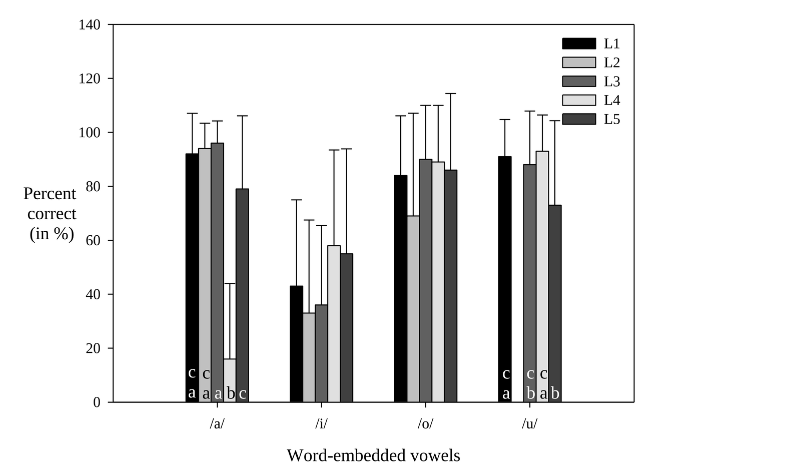

Figure 4. Means and standard deviations calculated from averaged listener responses (n=20) to a set of "female sentence-embedded vowels" used in the *"vowel identification" task.* Levels L1 to L5 represent increases in *H1-H2 amplitude difference* with L1 being the lowest level.

Significantly different levels in each data set are marked with different letters.

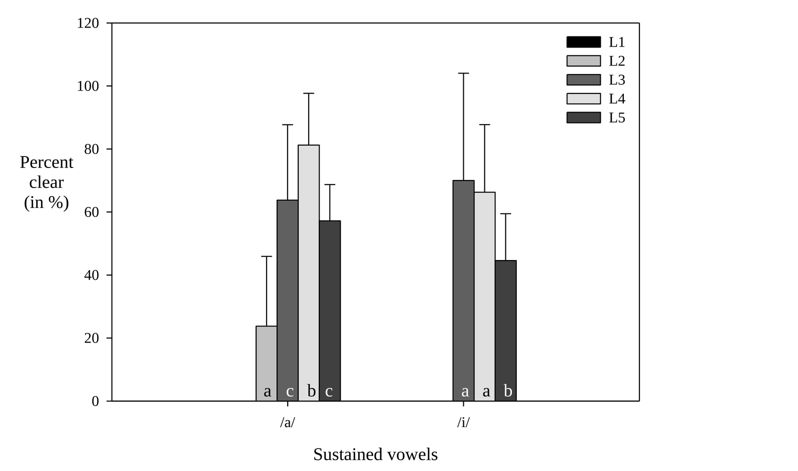

Figure 5. Means and standard deviations calculated from averaged listener responses (n=20) to a set of "male sustained vowels" used in the *"discriminate as clearer" task.* Levels L1 to L5 represent increases in *H1-H2 amplitude difference* with L1 being the lowest level.

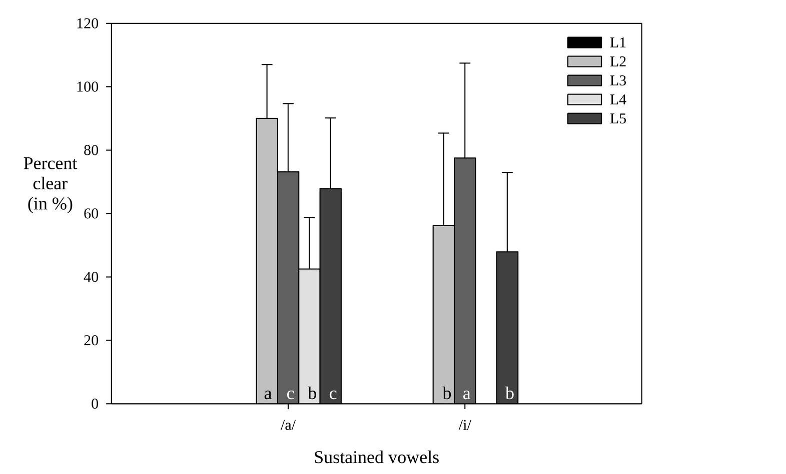

Figure 6. Means and standard deviations calculated from averaged listener responses (n=20) to a set of "female sustained vowels" used in the *"discriminate as clearer" task.* Levels L1 to L5 represent increases in *H1-H2 amplitude difference* with L1 being the lowest level.

Significantly different levels in each data set are marked with different letters.

Conclusion

The acoustic study results indicated that the F2 frequency, H1-H2 amplitude difference measures were higher for voice patients with incomplete glottal closure associated with breathy voices.

The perceptual study results require closer analysis at this point in time.

The findings of this study may contribute to determining how deterioration of voice quality, such as breathiness, impacts on the intelligibility of speech.

References

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers I, 1*, 71-82.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3-4), 255-272.

Fairbank, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York: Harper & Row.

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research, 36*(1), 14-20.

Gordon-Salant, S. (1986). Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *Journal of the Acoustical Society of America, 80*(6), 1599-1607.

Gordon-Salant, S. (1987). Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. *Journal of the Acoustical Society of America, 81*(4), 1199-1202.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769-778.

Kennedy, E., Levitt, H., Neuman, A. C., & Weiss, M. (1998). Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. *Journal of the Acoustical Society of America, 103*(2), 1098-1114.

Kenny, D. T. & Mitchell, H. F. (2006). Acoustic and perceptual appraisal of vocal gestures in the female classical voice. Journal of Voice, 20(1), 55-70.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*(2), 820-857.

Koenig, L. L. (2000). Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. *Journal of Speech, Language, and Hearing Research, 43*(1-5), 1211-1228.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America, 115*(1), 362-378.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality : review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*(1), 21-40.

Milenkovic, P. (2001). TF32. Madison, Wisconsin.

Monsen, R. B. (1978). Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research, 21*(2), 197-219.

Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., & Blaugrund, S. M. (1996). Singing power ratio: Quantitative evaluation of singing voice quality. *Journal of Voice, 10*(3 2), 228-235.

Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, production, acoustics, and perception*. Chichester, U.K. ; Malden, MA: Blackwell.

Sapienza, C. M., & Ruddy, B. H. (2009). *Voice disorders*. San Diego: Plural Publishing.

Sundberg, J. (1987). *The science of the singing voice*. DeKalb, Ill.: Northern Illinois University Press.

Titze, I. R. (2000). *Principles of voice production*. Iowa City: National Center for Voice and Speech.