Acoustic and Perceptual Evaluation of the Quality of Radio-Transmitted Speech

Shantanu Kirtikar, M. Arch., Emily Lin, PhD., and Michael Robb, PhD
Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand

Abstract

When speech signals are transmitted via radio, the process of transmission may add noise to the signal of interest (Biddulph, 1994; Coleman, 2004). This study aims to examine the effect of radio transmission on the quality of speech signals transmitted using a combined acoustic and perceptual approach.

Introduction

The purpose of the study is to identify the acoustic characteristics that need to be preserved to retain speech clarity while sending speech signals through common transmission devices such as radios. A selection of acoustic measures found in the literature to be related to voice quality or speech intelligibility were compared between the original and the radio-transmitted signals to identify the effect of radio transmission on speech quality. There are a range of acoustic measures designed to examine various aspects of spectral and waveform characteristics of speech signals. The acoustic feature most salient in distinguishing between the original and the radio-transmitted signals was selected for further perceptual investigation to delineate the relationship between speech perception and the acoustic changes that can be induced by radio transmission.

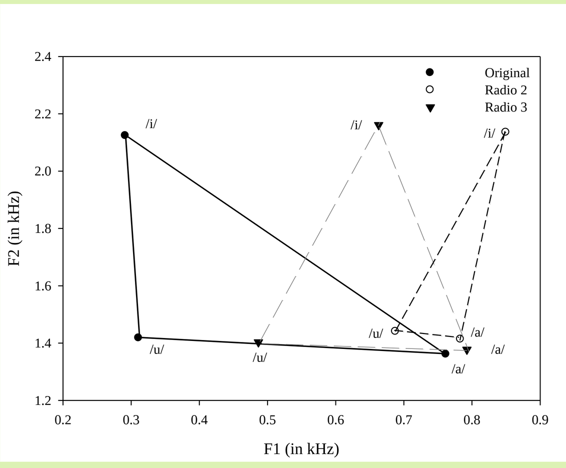
Methods

Acoustic Analysis. An acoustic recording of the Phonetically Balanced Kindergarten (PBK) word list read by a male speaker was played back in three conditions, one without radio transmission and two with two types of radio transmission (Radio 1 & Radio 2). The recording of the played back sound files and the radio transmitted signals was conducted in an anechoic room using a computer and USB audio interface, with a sampling rate of 44.1 KHZ and a 16-bit resolution. The vowel segments (/i, a, o, u/) embedded in the original and the re-recorded signals were analysed to yield measures of loci of the first two formant frequencies, and the vowel space measured (Robb & Chen, 2008), the energy ratio of the first two harmonics (H1-H2 amplitude difference; Klatt & Klatt, 1990), and the singing power ratio (Omori, Kacker, Carroll, Riley, & Blaugrund, 1996). Other measures included the mean and variance of the speech moment for the consonants /s/ and /sh/, voice onset time for the consonants /t/, /k/, /p/, /d/, /g/, /b/, and the energy ratio between consonant and vowel (CV energy ratio) for the vowels /i/, /a/, /o/ and /u/. The acoustic analysis of the speech samples was carried out by using automated speech analysis software, TF32.

Perceptual Study. Two perceptual studies were conducted. In the first study, vowel tokens of varying length, which was a third of the steady state of the vowel, were presented to 10 males (age: 22 to 40 years, Mean = 26.3 years, SD = 6.4) and 10 females (age range: 21 to 42 years, Mean = 30 years, SD = 7.5). All participants had normal hearing thresholds (Jerger & Jerger, 1980). Vowel tokens representing different H1-H2 amplitude difference levels were chosen. Each participant was asked to perform two tasks, vowel identification and clarity comparison tasks. In the second perceptual study, vowel tokens of constant length, from the steady state of the vowel, were presented to five participants (age range: 24 to 42 years, Mean = 33.4 years, SD = 6.7). Vowel tokens were organised on the basis of the H1-H2 amplitude difference levels and the singing power ratio levels.

Instrumentation and Procedures. The participants were seated in a sound booth, and the speech samples were played to them through a HP Intel ® Pentium M desktop, with a 1.73 GHz processor. These samples were presented to the participants via Sennheiser HD 215 headphones. In the vowel identification task, the participant s listened to one vowel segment at a time and selected from a list of five vowels, by clicking on an icon on the computer screen, which vowel they think they had just heard. In the clarity comparison task, the participants listened to one pair of two different recordings of the same vowel and indicate, also by clicking on an icon on the computer screen, which of the two presentations sounded clearer.

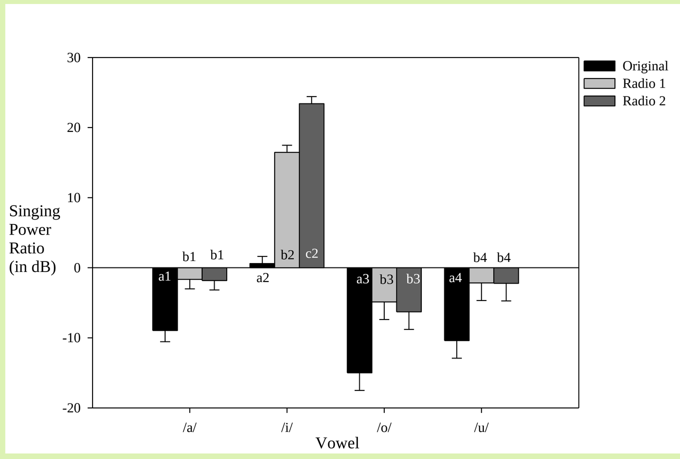
Results



Vowel space:

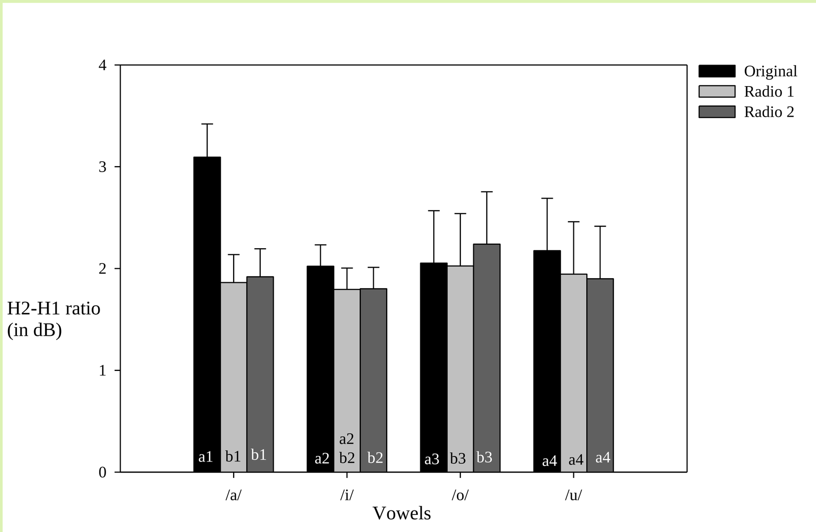
- Literature:
  - A reduced vowel space has been found to result in difficulty for a listener to identify vowels and words (Liu, Tsao, & Kuhl 2005)
  - Vowel space for stutterers, uncontrolled group: 158,379 Hz^2; treated group: 174,709 Hz^2 ; control group: 200,441 Hz^2 (Blomgren, Robb & Chen,1998)
- Findings in this study:

•Vowel space of original recordings: 634,200 Hz^2; Radio 1: 141,000 Hz^2 ; Radio 2: 473,100 Hz^2



Singing power ratio:

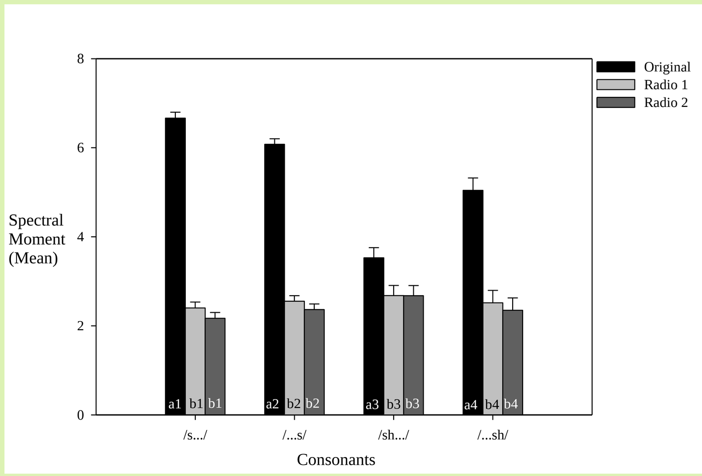
- Literature:
  - A SPR value reflecting more energy around the 2-4 kHz was found to be associated with the perception of a more resonant singing voice (Omori, Kacker, Carroll, Riley, & Blaugrund 1996),
  - Findings in this study:
  - Higher SPR (= lower energy around 2-4 kHz compared to energy around 0-2 kHz) for radio transmitted signals
- Two way (Vowel by Condition) ANOVA results
  - Vowel effect : F(3, 55) = 154.652, p < 0.001*
  - Condition effect : F(2, 55) = 41.156, p < 0.001*
  - Vowel by condition: F(6, 55) = 8.233, p < 0.001*



H1-H2 amplitude level difference:

- Literature:
  - The amplitude of H1 was found to be one of the factors in judging the breathiness of the voice (Hillenbrand, Cleveland, &Erickson 1994).
- Findings in this study:
  - H1-H2 amplitude difference affected by radio transmission

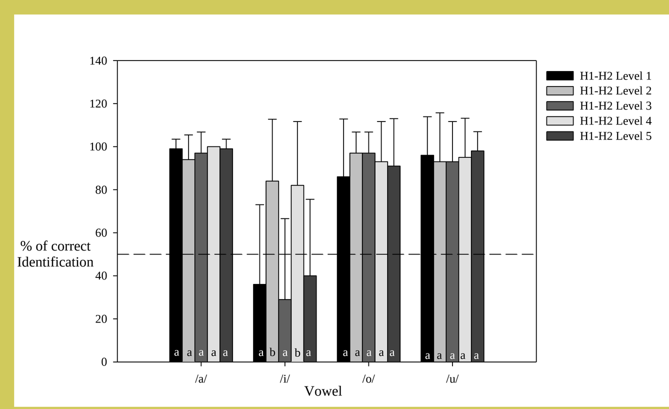
•Two way (Vowel by Condition) ANOVA results :
- Vowel effect : F(3, 56) = 23.585, p < 0.001*
- Condition effect : F(2, 56) = 3.174, p = 0.049*
- Vowel by condition effect : F(6, 56) = 8.432, p < 0.001



Spectral Moment, Mean:

- Literature:
  - Spectral moment mean plays a role in identification of consonants (Jongman, Wayland, & Wong, 2000).
- Findings in this study:
  - Radio transmission affected the concentration spectral energy.

•Two way (Vowel by Condition) ANOVA results :
- Consonant effect: F(3, 60) = 10.407, p <0.001 *
- Condition effect: F(2, 60) = 296.608, p <0.001,*
- Consonant by condition effect: F(6, 60) = 21.280, p <0.001*



Stimuli of Variable Length

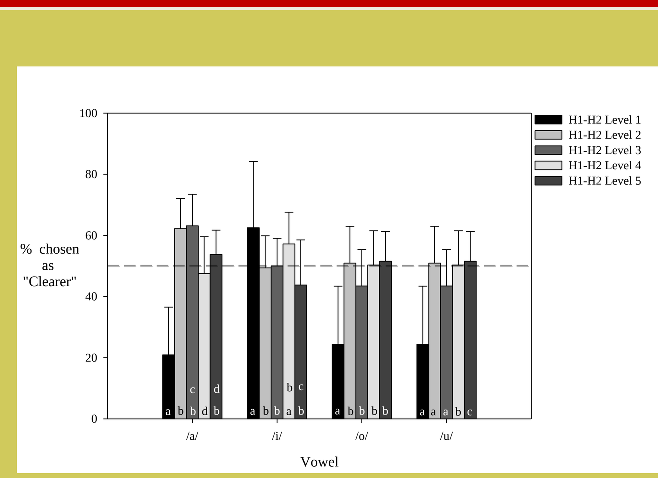
Vowel Identification

Tokens arranged in increasing **H1-H2 amplitude difference** level

Findings in this study:

- Vowel identification scores did not follow H1-H2 amplitude difference levels

•Two way (Vowel by Condition) ANOVA results :
- Level effect: F (4, 76) = 14.745, p < 0.001*
- Vowel effect: F (3, 57) = 42.053, p < 0.001*
- Level by vowel effect: F (12, 228) = 12.483, p < 0.001*



Stimuli of Variable Length

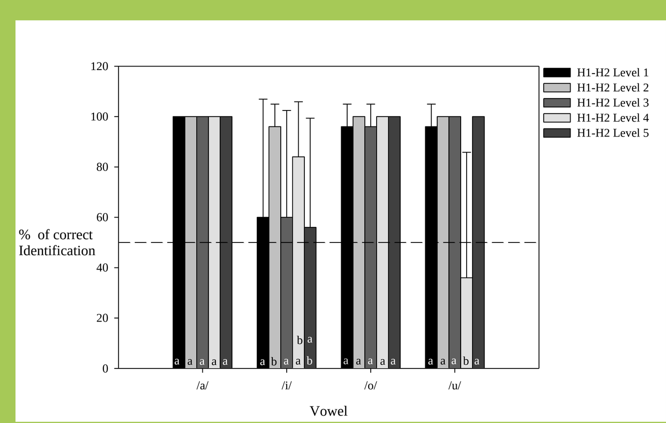
Vowel clarity

Tokens arranged in increasing **H1-H2 amplitude difference** level

Findings in this study:

- Vowel clarity scores did not follow the H1-H2 amplitude difference levels.

•Two way (Vowel by Condition) ANOVA results :
- Level effect: F (4, 76) = 16.539, p < 0.001*
- Vowel effect: F (3, 57) = 20.330, p < 0.001*
- Level by vowel effect F (12, 228) = 25.736, p < 0.001*



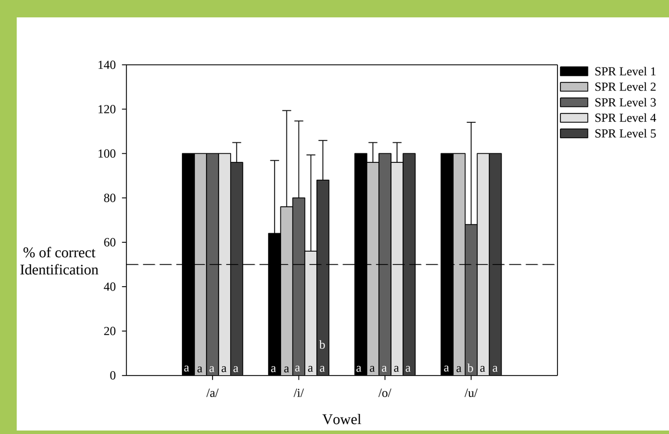
Stimuli of Constant Length

Vowel Identification

Tokens arranged in increasing **H1-H2 amplitude difference** level

Findings in this study:

- Vowel identification scores did not follow the H1-H2 amplitude difference level
- Two way (Vowel by Condition) ANOVA results
  - Level effect: F (4, 16) = 3.604, p = 0.028*
  - Vowel effect: F (3, 12) = 3.646, p = 0.045*
  - Vowel by level effect: F (12, 48) = 5.554, p < 0.001*



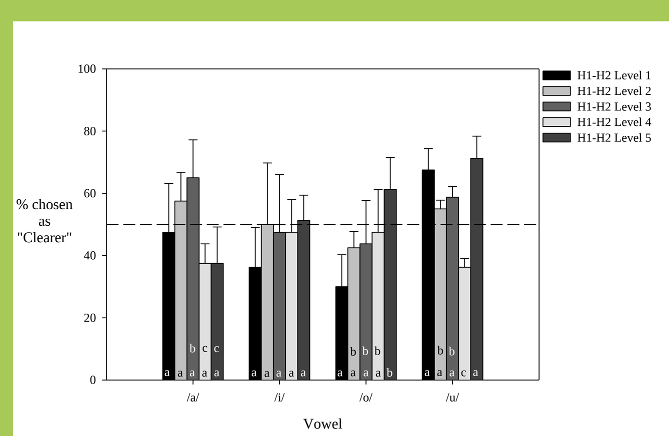
Stimuli of Constant Length

Vowel Identification

Tokens arranged in increasing **Singing power ratio**

Findings in this study:

- Vowel identification scores did not follow the singing power ratio level
- Two way (Vowel by Condition) ANOVA results
  - Level effect: F (4, 16) = 3.604, p = 0.028*
  - Vowel effect: F (3, 12) = 3.646, p = 0.045*
  - Vowel by level effect: F (12, 48) = 5.554, p < 0.001*



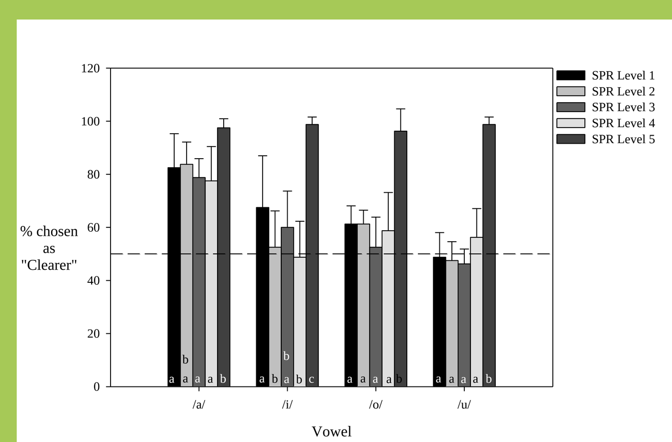
Stimuli of Constant Length

Vowel clarity

Tokens arranged in increasing **H1-H2 amplitude difference** level

Findings in this study:

- Vowel clarity scores tended to follow the H1-H2 amplitude difference level
- Two way (Vowel by Condition) ANOVA results
  - Level effect: F (4, 16) = 5.868, p = 0.004*
  - Vowel effect: F (3, 12) = 5.255, p = 0.015*
  - Vowel by level effect: F (12, 48) = 9.182, p < 0.001*



Stimuli of Constant Length

Vowel clarity

Tokens arranged in increasing **Singing power ratio**

Findings in this study:

- The highest SPR level (strong energy around 2-4kHz) is most often chosen as “clearer”
- Two way (Vowel by Condition) ANOVA results
  - Level effect: F (4, 16) = 76.375, p < 0.001*
  - Vowel effect: F (3, 12) = 9.357, p = 0.002*
  - Vowel by level effect: F (12, 48) = 4.340, p < 0.001*

Conclusion

It could be concluded from the findings that measures of energy ratio between different frequency regions, as well as the frequencies of the first two formant frequencies, were sensitive to the detection of the effect of radio transmission. The perceptual study showed that when duration was controlled, the clarity of vowels was affected by the H1-H2 amplitude difference level and the singing power ratio.

References

Biddulph, D. (Ed.). (1994). *Radio communication handbook* (6 ed.). Herts: Radio society of Great Britain.
Blomgren, M., Robb, M., & Chen, Y. (1998). A note on vowel centralization in stuttering and nonstuttering individuals. *Journal of Speech, Language, and Hearing Research, 41*(5), 1042-1051.
Coleman, C. (2004). *An introduction to radio frequency engineering*. Cambridge: Cambridge university press
Hillenbrand, J. M., Cleveland, R., & Erickson, R. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769-778.
Jerger, J., & Jerger, S. (1980). Measurment of hearing in adults. In M. M. Paparella & Shumrick (Eds.), *Otolaryngology*. Philadelphia: W. B,. Saunders.
Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108*(3 I), 1252-1263.
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.
Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America, 117*(6), 3879-3889.
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.
Robb, M. P., & Chen, Y. (2008). A note on vowel space in Mandarin accented English. *Asia Pacific journal of speech, language, and hearing, 11*(3), 175 - 188.