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

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Audiology

in the Department of Communication Disorders

at the University of Canterbury

By

Shantanu Kirtikar

University of Canterbury

2010

Contents

Acknowledgements

Firstly I would like to thank my supervisor, Dr. Emily Lin, for her guidance, support and patience in the preparation of this thesis, without which it would not have been possible. I also take this opportunity to thank Prof. Michael Robb for his role as co-supervisor in this thesis.

I would like to thank James Collett, a former Master student from the department of Mechanical Engineering, for the recording of the radio-transmitted signals. I would like to thank all the participants who have taken the time to be involved in the study.

And finally I take this opportunity to thank my wife Minoti, and children, Advait and Aniruddh, for their patience and support in this endeavour.

Abstract

Aim

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

Method

A standard 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. The vowel segments (/i, a, o, u/) embedded in the original and the re-recorded signals were analysed to yield measures of frequency loci of the first two formant frequencies (F1 and F2), amplitude difference between the first two harmonics (H1-H2), and singing power ratio (SPR). Other measures included Spectral Moment One (mean), Spectral Moment Two (variance), and the energy ratio between consonant and vowel (CV energy ratio). To examine how H1-H2 and SPR were related to the perception of vowel intelligibility and clarity, vowels at five levels of each of these two measures were selected as stimuli in the perceptual study. The auditory stimuli were presented to 20 normal hearing listeners, including 10 males and 10 females aged between 21 to 42 years, the listeners were asked to identify the vowel for each vowel stimulus in the vowel identification task and judge from a contrast pair which vowel sounded “clearer” in the clarity discrimination task. A follow-up study using vowel stimuli with a constant length and five H1-H2 or five SPR levels was conducted on five listeners to determine the relationship between the perception of speech clarity and H1-H2 or SPR.

Results

Results from a series of one-way or two-way analyses of variance (ANOVAs) or ANOVAs on Ranks and post-hoc test revealed that radio transmission had a significant effect on all of the selected acoustic measures except for the CV energy ratio. Signal degeneration due to radio transmission is characterized by changes of F1 or F2 frequencies toward a more compressed vowel space, a H1-H2 value indicating an increase of H1 dominance, a SPR value suggestive of an increase in the energy around the 2-4 kHz region, and a loss of differentiation between /s/ and /sh/ on the measures of Spectral Moments One and Two. Vowel duration was also found to play a major role in affecting the perception of vowel intelligibility and clarity. The follow-up study, with a control on vowel duration, found that SPR played a role in affecting the perception of vowel intelligibility and clarity.

Conclusion

It was 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 in detecting the effect of radio transmission.

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Thesis Overview

The purpose of the study was to identify the acoustic characteristics that need to be preserved to retain sound clarity while sending speech signals through common transmission devices such as radios. With both acoustic and perceptual approaches employed for this investigation, this study includes two phases. In the first phase, an acoustic analysis is conducted on digitized signals which have been processed with and without radio transmission. Specifically, a selection of acoustic measures found in the literature to be related to voice quality or speech intelligibility are compared between the original and the radio-transmitted signals to determine the effect of radio transmission on these measures. In the second phase, the acoustic measures considered useful in differentiating between the original and radio-transmitted signals are selected for perceptual investigation to determine the relationship between the acoustic changes that can be induced by radio transmission and the perception of speech quality.

1.2 Rationale

As radio transmission may result in audible changes in the quality of the transmitted sound, it is conceivable that the acoustic signals of interest may be manipulated to change the perception of their quality. It has been shown in the literature that the performance of a speech sound recording, transmission, or playback system can be evaluated through a general test of the frequency response of the system and its components, speech recognition or discrimination tests, subjective jury judgement, and instrumentally testable metrics associated with auditory perception. However, it remains unclear what acoustic changes are key to the perceived quality of speech signals and why. A better understanding of the relationship between the acoustic measures of voice quality and speech intelligibility will help develop a parsimonious parameterization scheme for evaluating and improving the output of a speech transmission system.

Chapter 2. Literature Review

This chapter includes the background information about radio waves and radio transmission, the concepts of speech perception and intelligibility, and a critical review of the literature related to the acoustic and perceptual approaches in assessing speech and voice qualities.

2.1 Radio Transmission

Radio waves are a part of the electromagnetic radiation spectrum, which also includes light, microwaves, gamma rays, and X-rays, amongst others. Radio frequency ranges from 3 kHz to 300 GHz (Biddulph, 1994). Radio waves, which are often classified by their wavelengths, all travel at the speed of light, which is approximately 3 x10^{10} centimetres per second (Biddulph, 1994). In the mid 1890's, Gugliemo Marconi developed a radio telegraph system demonstrating that radio waves could be used as a means of communication (Biddulph, 1994). Since then, the technology of radio waves has been developed and integrated into our lives, including radio transmissions, radar, TV, and more recently, cell phone technology (Bertoni, 2000; Coleman, 2004). In the field of audiology, radio transmission systems are commonly used to send messages from a speaker to a hearing impaired listener via a dedicated system which allows the listener to receive the signal with reduced interference, and is commonly referred to as an "FM system" (Dillon, 2001). All of these applications use radio waves of different frequencies.

Although there are a wide range of frequencies in radio waves, the use of these frequencies is restricted, with different frequency bands allocated for various applications. In New Zealand, the Ministry of Economic Development manages the radio spectrum, allocating different bands of the spectrum for use by various parties and regulating the transmission of radio waves based on the *Radiocommunications Regulations 2001* under the *Radiocommunications Act 1989* (Ministry of Economic Development, 2009). The FM systems in audiology use several different bands in the spectrum, including bands at 37 MHz, 43 MHz, 173 MHz, and 183 MHz as well as bands between the ranges of 72-76 MHz and 216-217 MHz (Crandell & Smaldino, 2002; Dillon, 2001).

Radio waves by themselves do not carry any information. The information to be transmitted has to be added onto the radio waves, which act as a carrier of this information (Bertoni, 2000). When sound is presented to a microphone, it converts the sound into an electrical current of changing intensities corresponding to the time-varying magnitudes of sound pressure within the sound wave. Different types of microphones use the sound energy to drive different mediums, including an aluminium ribbon, carbon granules, or quartz crystal, to generate the current (Biddulph, 1994). This electrical representation of the sound wave is used to encode information onto the radio signal by modulating the radio wave (Biddulph, 1994). Adding a modulating frequency (F_m) to a carrier radio frequency (F_c) results in changes to the amplitudes of the time waveforms. The sound transmission method involving the use of three frequencies, namely, $F_{c,}$ ($F_m + F_c$), and ($F_m - F_c$), is referred to as amplitude modulation (Yost, 2007). If the frequency of the carrier radio frequency is modulated instead of the amplitude, the transmitted wave is referred to as a frequency modulated wave. In a frequency modulated wave, the carrier frequency, instead of amplitude, is modulated resulting in a wave with its frequency changing over time (Yost, 2007). Radio signals can also be phase modulated (Coleman, 2004). Frequency, amplitude, or phase modulation of a known carrier wave in accordance to the input sound waves are the basic techniques used to add information to radio waves for sound transmission.

Once the information is added to the radio wave, the signal is propagated into space via a transmitting antenna. The process of adding information at the source is essentially reversed at the receiving end, where the modulated radio signal is picked up by a receiving antenna and amplified. The information used to modulate the radio wave is extracted as the signal by a receiving microphone (Biddulph, 1994). As the receiving system extracts relevant information from the modulation of the signal, the transmission process is normally independent of the strength of the carrier signal and thus the information is relatively resilient to signal distortion (Dillon, 2001). The use of different carrier frequencies also allows several signals from different propagation sources to be transmitted at the same time through the same space (Coleman, 2004). However, it is still inevitable that the original signal may be altered during the process of transmission resulting in a difference in the frequency content between the original signals and the output signals of the transmission system (Wilber, 2002).

Transmission noise, which is added due to signal distortion introduced in the transmission process, is different from the background noise that may have already been present in the original signal (Crandell & Smaldino, 2002). Transmission noise, which is essentially unwanted signals, may get added to the signal of interest due to several factors. For example, the noise may come from natural sources like electric storms, the Sun, thermal activity (“Johnsons Noise”), or circuit noise. The noise may also be from man-made sources such as switches, thermostats, and other electrical devices in the vicinity (Biddulph, 1994; Coleman, 2004). Above all, noise is most likely to be added to the signal if the strength of the carrier frequency is reduced to a certain extent (Dillon, 2001). The addition of noise or the alteration of the frequency content may interfere with the signal of interest. When transmitting human speech, the integrity of speech intelligibility is of main concern. To understand the process of speech intelligibility, it is essential to first outline the process of speech perception.

2.2 Speech Perception

Speech perception concerns how speech sounds are decoded from the nerve conducted signals which have been passed through the auditory system, normally those transduced from the acoustic sound waves. According to the "Motor Theory of Speech Perception" proposed by Liberman and Mattingly (1985), the human auditory system is designed to detect and understand gestures made by the speaker. The ability of a person to perceive speech is dependent not only on the target speech signal itself but also on the context of its presentation. In particular, when the listener is able to use or access information preceding or following the stimulus of interest, the ability to identify the stimulus correctly increases. For example, in a study where nonsense words, words in sentences, and digits (numerals) were presented to listeners, Miller, Heise, and Lichten (1951) found that the context of the presentation was an important factor in the accuracy of speech perception. Along with acoustic cues, the listener uses contextual cues and knowledge of the language to understand speech and this ability is thought to be an integral and vital part of the human experience.

There is also evidence showing that the human auditory system has evolved to become capable of extracting useful acoustic and linguistic cues even when presented in a degraded condition (Northern & Downs, 2002). For example, in a study of speech signals, Shannon, Zeng, Kamath, Wygonski, and Ekelid (1995) investigated how normal hearing listeners perceived signals which had their frequency contents degraded but temporal and amplitude information retained in the signal envelope. These signals were processed using one to four processing bands. All of the eight participants in the study were found to be able to correctly identify 90% of the presented words in the condition where only two bands of information are preserved. This finding suggested that a person's ability to understand speech was robust and able to withstand substantial degradation of the speech signal. The auditory system seems to be designed to resolve speech sounds in a manner to make speech intelligible (Moore, 2008).

2.3 Speech Intelligibility

Speech intelligibility, which refers to how easily speech can be understood by the listener, is generally measured as the percentage of speech signals identified correctly. There are a variety of measures available to assess the capability of a sound transmission system in maintaining speech intelligibility. These measures range from those obtained from listeners undergoing comprehensive but laborious speech recognition tests to those derived from model or theory-based machine calculations. For example, in the early 20th century, a speech audibility measure, Articulation Index (AI), was developed to understand the impact of signal distortion on speech intelligibility and predict the performance of auditory communication devices (French & Steinberg, 1947; Fletcher & Galt, 1950; Allen, 2005). The Speech Transmission Index (STI), developed by Houtgast and Steeneken (1973), is obtained through a procedure involving a complex amplitude modulation scheme to generate test signals and a Modulation Function to allow for a comparison between the original and the transmitted signals. With a similar model-based comparison scheme, the Perceptual Evaluation of Speech Quality (PESQ) was established by the *International Telecommunications Union, ITU P. 862 standard* as a means for evaluating the quality of speech transmission (Di Persia, Milone, Rufiner, & Yanagida, 2008). The PESQ method has been validated through an additional calibration procedure relating the machine-derived scores to subjective test results. For all perceptual assessment tools, the principle measure used in evaluating the intelligibility of transmitted speech is to calculate the proportion of signal presentations that the listener can correctly identify. The frequency of correct identification for transmitted speech appears to be useful for assessing the performance of a sound transmission system in retaining or improving the speech intelligibility of the transmitted signals.

While external noise may result from components of the transmission system, the inherent properties of the signal itself may also impact on speech intelligibility. For example, when asked to speak in a manner making their speech clearer, speakers were found to speak at a slower rate than conversational speech (Picheny, Durlach, & Braida, 1985). However, when recordings of conversational speech were slowed down to match the production rate of “clear speech”, it did not result in greater intelligibility than the conversational speech (Uchanski, Choi, Braida, Reed, & Durlach, 1996). These finding suggest that speech intelligibility is related to the rate-induced articulatory or voice change rather than the change in speech rate alone. Furthermore, in a study evaluating the speech intelligibility of six normal speakers in different environments, Cox, Alexander and Gilmore (1987) found that speech produced at a faster rate was not necessarily considered less intelligible. As speech intelligibility involves a process of transducing and decoding the acoustic signals through the auditory system, one way of understanding the signal characteristics related to speech intelligibility is to study speech and voice characteristics acoustically.

2.4 The Acoustic Theory of Speech Production

The acoustic characteristics of sound depend on the properties of the source of the sound (Yost, 2007). According to the “source-filter model” proposed by Johannes

Muller in 1848, the vibration of the vocal folds in the larynx acts as the source of the sound in phonation. The rate of vibration of the vocal folds determines the fundamental frequency (F0) and overtones of the voiced sounds. Speech sounds produced with vocal fold vibration are referred to as "voiced" and those produced without vocal fold vibration "unvoiced". When speakers change the fundamental frequency of their voice, it changes the intonation, which is of considerable significance to the understanding of the content and the interpretation of the speech signal (Fry, 1979). A voiced sound can also be "nasalized" and classified as "nasals", which are generated with the opening of velopharyngeal port in the naso-pharynx during speech production (Allen, 2005; Fry, 1979). Specifically, nasal sounds are created when the soft palate or velum, which is located in the rear portion of the roof of the mouth, is lowered. Velar lowering opens up the velopharyngeal port to allow the expiratory air, which normally flows only into the oral cavity during production of oral sounds, to flow partially into the nasal cavity. The vocal tract, including throat, mouth, lips, teeth, tongue, and nose, forms the filter that shapes the sound waves passing from the source (Miller, 1981).

The parts involved in the change of vocal tract configuration are referred to as articulators. Some articulators can be moved and shaped, like the tongue, the lips, and portions of the throat. Fixed articulators include hard palate, alveolar ridge, and teeth (Davenport & Hannahs, 1998). In the production of a sonorant (i.e., vowels, semi-vowels, liquids), vocal tract shaping, which is related to variations in the shaping and placement of various articulators, can create regions of resonant peaks at various frequencies. These resonant peaks are referred to as formants (Miller, 1981). Formants can be measured from a spectrogram or a spectrum.

In acoustic analysis, a sound wave can be presented in three forms: time waveforms (amplitude over time), spectrogram (frequency over time), and spectrum (amplitude over frequency). An instrument called a spectrograph invented in Bell Laboratories in 1946 (Cooper, 1980) can be used to allow for measurement and visualisation of the acoustic properties of speech sounds (Davenport & Hannahs, 1998). The output of the spectrograph is referred to as a spectrogram. In a spectrogram, the speech is presented in the time and frequency domains, with the strength of any given frequency region reflected by the darkness or the colour of the corresponding traces. As a 2-D plot of frequency and time, the spectrogram can be used to infer the positioning and the movement of the articulators or the change of the manner of articulation over time (Davenport & Hannahs, 1998; Fry, 1979).

The acoustic characteristics of speech vary with the anthropomorphic features of speaker, as each individual may have different F0 and vocal tract shapes. It has been estimated that the average F0 is 120 Hz for males, 225 Hz for females, and 265 Hz for children (Fry, 1979). The speaking F0, which ranges from 60 Hz to 500 Hz (Fry, 1979), has been found to be an important cue in distinguishing between male and female speakers. For example, Hillenbrand and Clark (2009) examined 1,116 h/V/d words with 12 vowels spoken by 48 female and 45 males and found that correct identification of the gender of the speaker was achieved in 96% of cases. Since differences in F0 result in different harmonic structures for different speakers and vocal tract configuration may vary even when two individuals are producing the same phoneme (Miller, 1981; Northern & Downs, 2002), it is most likely that phonemic discrimination may be affected by voice quality when the sound energy distribution across frequencies is distorted.

2.5 Acoustic Measures of Speech

In speech production, the overall spectral characteristics, as well as the amplitude, of speech vary over time, either rapidly or slowly (Stevens, 1980). There are a variety of acoustic measures designed to examine various aspects of these spectral and amplitude characteristics during speech production. These acoustic measures can be grouped into three categories: i) those that measure vowels, ii) those that measure consonants, and iii) those that compare between vowels and consonants. A selection of the acoustic measures commonly used to study voice quality or intelligibility-related sound differentiation is described as follows.

2.5.1 Vowels

The acoustic measures concerning vowel discrimination include mainly the frequency loci of the first two formants. Voice thinness or breathiness can be reflected in the prominence of the first harmonic, which can be gauged from the amplitude difference between the first two harmonics (H1-H2). Voice projection power may involve measurement of the energy ratio between spectral peaks in the lower (0-2 kHz) and higher frequency (2-4 kHz) regions, such as the measurement of singing power ratio.

2.5.1.1 Formants One and Two (F1 and F2) and Vowel Space

Formants are shown as clusters of concentrated energy in the spectrum of a sonorant. Although the formant patterns depend on the characteristics of the entire vocal tract, changes to the length between the throat and the lips or to the configuration of the horizontal portion have a significant influence on the frequency of the second formant (F2). Changes to the vertical part of the vocal tract, including tongue height, influence the frequency of the first formant (F1). It has been shown that formant patterns provide important information for vowel identification and differentiation.

2.5.1.1.1 Formant Frequencies and Vowel Identification

The importance of the first two formants in vowel identification has been investigated. For example, Parikh and Loizou (2005) studied the effects of multi-talker babble and speech shaped noise on F1 and F2 in relation to acoustics and perception. Results from a total of 20 male and 23 female speakers, who were asked to produce 11 vowels in an h/V/d configuration, revealed that with a low signal-to-noise ratio of -5db, listeners identified vowels primarily based on F1 but used partial information of F2. Based on a comparison of the same vowels produced by different speakers, Hillenbrand, Getty, Clark, & Wheeler (1995) found acoustic evidence that the production of vowels changes over a period of a few years. In this study, an acoustic analysis was conducted for /hVd/ words, produced by a normal hearing group, including adults (45 males, 48 females) and children (10-12 years old, 27 male and 19 female). The authors noted that as the production of vowels was found to vary in time, automated vowel identification systems need to take this into account. Further, when the words were presented to a group of 20 young adult listeners, it was found that although F1 and F2 values alone were not a good indicator of the vowel being produced, the presence of F1 and F2 in the signal, without higher formants, allowed the listener to identify the vowel. It was concluded that vowel identification required more cues than just F1 and F2 values.

2.5.1.1.2 Vowel Space and Vowel Differentiation

Vowel space is a theoretical construct of vocal tract area made up by plotting F1 and F2 values for three vowels (/i/, /a/ and /u/). Vowel space can be derived from vowels produced by a speaker or averaged from vowels produced by a group of speakers, with the area that falls within the selected vowels in the F1-F2 plot calculated as the vowel space. The size of the vowel space has been related to vowel differentiation, with a reduced vowel space being associated with difficulty for a listener to identify vowels and words.

Liu, Tsao, and Kuhl (2005) studied 20 dysarthric speakers, with cerebral palsy with a varying degree of deficit in speech, producing 18 words containing the vowels /i/, /a/, and /u/ in Mandarin. They found that these participants produced a smaller vowel space when compared to a control group consisting of speakers with normal speech. The frequency of F1 was significantly lower for /a/ and higher for /i/ for dysarthric speakers as compared to the control group. It was also found that the vowel space could be used as an indicator of the intelligibility of the speech token. When judged by a group of listeners with normal hearing, who were unfamiliar with dysarthric speech, reduced vowel space was reported as being significantly correlated to the intelligibility of the vowel ($r = 0.632$, $p < 0.005$) and also to the intelligibility of the word ($r = 0.684$, $p < 0.005$). In the second part of the study, Liu et al. (2005), synthesised 6 tokens of each of the three vowels /i/, /a/, and /u/, and systematically varied their vowel working space, based on the value of the space produced by the dysarthric speaker judged to have the clearest speech in the first part of the study. Twelve male and 10 females with normal hearing and speech were asked to identify the vowel that they heard and also rate them on a seven point scale, to judge whether the token that they heard was a good representation of the vowel or a poor one. While the vowels with the smaller vowel spaces were rated as poorer exemplars, the rate of correct identification remained high for /i/ (97-100%), followed in order by /u/ (92-97%) and /a/ (92 - 93%).

In another study, Blomgren, Robb and Chen (1998) studied the vowel space produced by three groups of adult males producing 12 C/V/*t* tokens. The consonants used were /p/, /b/, /s/, and /z/. Each token was produced three times by the speakers at a speaking rate and volume that they were comfortable with. The first group of speakers consisted of 5 individuals who had untreated stuttering, with three participants having severe stuttering, one with a moderate-severe impairment, and the last with a moderate impairment. The second group included another five individuals who had treated stuttering, including four participants with speech now considered as normal, and one still with mild impairment. The control group consisted of 5 normally speaking individuals. The average vowel space was found largest for the control group (vowel space area = 0.200441 kHz^2), followed in order by the treated group (0.174709 kHz^2) and the untreated group (0.158379 kHz^2). While the differences were not statistically significant, the trend towards a smaller vowel space in untreated stutterers as compared to the control group is evident. One of the possible explanations provided for the lack of statistical significance in the finding was that the sample size was small (Blomgren et al., 1998).

In a study of speech of individuals with Parkinson disease (PD), McRae, Tjaden, and Schoonings (2002) investigated the relationship between vowel space and speech rate. A total of 13 adults, including 4 females and 9 males of ages ranging from 59 to 82 years (mean age = 70), served as subjects. The patients had mild to moderate idiopathic PD, all but one of whom were judged to have mild to moderate hypokinetic dysarthria. The control group included 13 age and sex-matched normal controls. The participants were asked to read out the ‘Farm passage’ at three different rates, first at their normal reading rate, then at twice of their normal rate, and then at half of their normal rate. It was found, as expected, that in the control group, the area of the vowel space was smaller for the speech with the fast rate of production and larger when speech was produced slowly. The vowel space for the participants with PD was found to be smaller, although without statistical significance, than the controls when compared at the same rate of production. It was also noted that the perceived severity of the speech disorder was the lowest in the normal speech rate in persons with PD.

The relationship between vowel space area and speech intelligibility has been found not only in individuals with speech impairment but also in normal speakers. For example, Bradlow, Torretta and Pisoni (1996) examined the acoustic and perceptual relation of “100 Harvard sentences” (Bradlow et al., 1996, pg 258) produced by 10 male and 10 female speakers of American English. Each talker’s recordings were transcribed by 10 listeners with normal hearing. The listeners were also speakers of American English. From the 100 sentences, a subset of 18 sentences, containing six tokens of the vowels /i/, /a/, and /o/, was selected for the measurement of vowel space area. This subset of sentences had intelligibility scores that correlated with the overall intelligibility of the speaker across the 100 sentences. Although the vowel space was found to be generally larger for the more intelligible speech, the authors did not always find a correlation between the two across all the 20 speakers. A correlation between vowel space area and intelligibility was found only for the 10 most intelligible speakers. This finding suggests that there is limitation to the use of vowel space, when viewed in isolation, as a measure of speech intelligibility.

2.5.1.2 Amplitude Difference Between Harmonics One and Two (H1-H2)

In a study of the voice quality of speech by Klatt and Klatt (1990) the amplitude of the first harmonic (H1) was considered to be one of three possible cues relating to the voice being judged as breathy. They found that the amplitude of the second harmonic (H2), as compared with F1 amplitude or the overall RMS amplitude, provided a better reference point against which H1 can be compared because the amplitude of H2 was less likely to be affected by changes in F1. They compared the amplitude of H1 with that of H2 in utterances produced by ten female speakers and six males speaking two sentences and their reiterated imitations. They found that the average H1-H2 for males was 6.2dB, which was 5.7 dB less than that for females. When a panel of eight listeners were asked to rate the breathiness of the reiterated imitation of the sentence, on a seven point scale (with 0 being least breathy), using the vowel extracted from the reiterated imitation, it was found that the average rating was 3.1 for males and 3.9 for females, suggesting that the female voices were judged as being more breathy. However, not all females were judged as having more breathy speech than males, as two female speakers had lesser breathy voices than the average for males and one male had a voice that was found to be more breathy than the average for the female voice. In addition, correlational analyses indicated that H1 amplitude alone could not be considered a measure of breathiness, suggesting that there were other factors involved.

The findings regarding the value of H1 amplitude in judging the quality of voices were generally supported. For example, Hillenbrand, Cleveland, and Erickson (1994) studied four vowels, /a/, /i/, /ae/, and /o/, produced by eight male and seven female speakers in three conditions of breathiness, including normal, moderate, and extremely breathy. These were judged by 19 female and one male. The amplitude of H1 was found to be one of the significant factors in judging the breathiness of the voice, with a correlation of 0.66. As the association between an increase of H1 amplitude and a higher open quotient has also been demonstrated in synthesized voice (Klatt and Klatt, 1990), it appears that the prominence of H1, which may be most reliably estimated by the energy ratio between the first two harmonics, provides a useful cue for the detection of breathy voice or the differentiation between female and male voices.

While voice quality affects the amplitude of the first harmonic, speech intelligibility has also been linked to the structure of harmonics, as shown in a study by Hu and Loizou (2010). It has been observed that the noise reduction algorithms, which are often used to make speech in noise more intelligible for cochlear implant users, may distort the structure of the harmonics in the speech sample. Hu and Loizou (2010) tested the hypothesis that the preservation of the harmonics in the speech signal prior to vocoding would improve intelligibility. The hypothesis was supported as they found that reintroduction of the harmonic structure and its partials to a signal that has been pre-processed improved the intelligibility of the signal by 41%. Therefore, acoustic measures of aspects of the harmonics may provide information useful for understanding the relationship between voice quality and speech intelligibility.

2.5.1.3 Singing Power Ratio

The singing power ratio (SPR) is the ratio of the largest spectral peak between 0 and 2 kHz to the largest spectral peak between 2 and 4 kHz, as defined by Omori, Kacker, Carroll, Riley, and Blaugrund (1996), who developed it as an objective measure of the quality voice while singing. Their study examined the phonation and singing of the vowel /a/ by 37 trained singers, including 21 females and 16 males, and a control group, including 10 males and 10 females with no singing experience. Perceptual, as well as spectral, analysis was carried out on the recordings. The mean SPR derived from the sung /a/ tokens was found to be lower for the non-singers (-22.7) than for the non-professional singers (-14.2) and the professional singers

(-13.1). The mean SPR derived from phonated /a/ was also lowest for the non-singers (-22.7), followed in order by the non-professional singers (-22.6) and the professional signers (-19.7). A higher SPR, which indicates more energy in the higher harmonics, was found to be associated with the perception of a more resonant singing voice. A significant correlation between SPR and the ringing quality was also found ($r = 0.4285$, $p < 0.01$). Further, the samples were acoustically manipulated to reduce the peak between 2 and 4 kHz in 6 dB steps, from the original to a 24dB reduction and subjected to perceptual analysis. It was found that the ringing quality, judged by 5 listeners and rated on a scale from 7 (strongest ringing) to 1 (weakest) was consistently rated lower as the amplitude of the spectral peak dB in the 2-4 kHz region was reduced. It was also found that the SPR for persons trained in singing was higher as the duration of training increased, showing a significant increase of SPR for those with training of more than or equal to 4 years. Based on these findings, it was concluded that the SPR could be considered as a metric of singing quality.

Another study, by Watts, Barnes-Burroughs, Estis, and Blanton (2006), examined whether individuals who had no training in singing but were judged to have singing talent had a SPR that was different from those judged as being untalented at singing. Thirty-nine females, aged between 20 and 35, were asked to sing a passage from a song familiar to all of them. The perceptual quality of their singing was assessed as either talented or untalented by two experienced singing teachers. Only those judged unanimously by both were included in the study. Twelve participants were rated as talented and 21 as untalented singers. The talented singers showed a mean SPR of 22.6 dB, which was significantly different from that for the untalented group (30.7 dB), indicating that the talented singers exhibited a relatively greater energy in the 2-4 kHz frequency region. It was also noted that the voice associated with a stronger energy in the 2-4 kHz frequency region is perceived by the listener to have a more ‘ringing’ quality to it and is perceived as being more pleasant.

It has been found, however, SPR did not relate to good quality singing when applied to some singing techniques. Kenny and Mitchell (2006) studied the acoustics and perception of singing voice produced in an optimal and a suboptimal “open throat” technique. The “open throat” technique is reported as producing a reliably identifiable improvement in the perception of the quality of singing, when judged by experienced music teachers. Six experienced female singers who aged between 23 and 30 years with a background of having studied singing for seven years or more were asked to sing parts of two classical songs in three formants. Subjects were asked to sing once in an optimal “open throat” technique, then with a suboptimal “open throat” technique, and then with a loud suboptimal “open throat” technique. A singing teacher attended the recordings and instructed singers as needed to ensure that they were singing using the correct technique. The singing quality was then ranked by 15 teachers who were considered experts in teaching singing. Acoustic analysis was also performed, and the SPR was calculated using the long term average spectrum of the whole singing passage. The perceptual ratings given by the listeners were generally in agreement that the optimal “open throat” technique produced better quality singing. However, the perceptual ratings did not correlate with the ranking determined based on the SPR (Kenny & Mitchell, 2006).

The study by Omori et al. (1996), concluded that the SPR was a method of judging “singing voice quality” (pg: 235) of the vowel /a/ being sung, while the study by Kenny and Mitchell (2006) has attempted to use the SPR obtained from a singing passage in judging the quality of sound produced by a specific singing technique of “optimal” singing and “sub-optimal” singing, which use different throat configurations. The methodological difference between the two studies in that one used vowels and the other a singing passage may have contributed to the difference in the finding regarding the strength of the relationship between SPR and singing quality. It has been speculated that the perception of the quality of a sung voice has more dimensions to it than just the difference in the spectral peaks, and so the SPR (and the energy ratio) in itself was probably an inadequate metric to differentiate singing technique (Kenny & Mitchell, 2006).

2.5.2 Consonants

There are several types of consonants, including stops, nasals, liquids, and glides, and fricatives. Two common acoustic measures of speech quality related to consonant production are employed in this study, including spectral moments and consonant-to-vowel (CV) energy ratio.

2.5.2.1 Spectral Moments

Spectral moment analysis is a way of measuring the frequency distribution of the spectral energy for a consonant. The first moment indicates the “mean” frequency where spectral energy is concentrated and the second moment the spread of the energy around the center frequency (Jongman, Wayland, & Wong, 2000). Forrest, Weismer, Milenkovic, and Dougall (1988) studied the classification of voiceless obstruent (/k/, /t/, /p/, & /f/) in word-initial positions. They examined 14 words produced by 5 male and 5 female speakers. The Bark transformed and linear spectra of these words were analysed using 10, 20, and 40 msec duration segments. It was found that the mean (Spectral Moment 1), skewness (Moment 3), and kurtosis (Moment 4) were useful for classifying the place of articulation in voiceless stop consonants but there was less success with spectral moment analysis in the classification of fricatives.

Jongman et al. (2000) found that the first and second spectral moments were useful in identifying the place of articulation for fricatives. They examined 144 CVC words, with one of the 8 selected fricatives as the initial consonant (onset), followed by one of the 6 selected vowels as the nucleus and the consonant /p/ as the ending consonant (coda). The speakers were 10 male and 10 female speakers of American English. A window of 40 msec duration was used for the spectral moment analysis. The reason why mean (Spectral moment 1) and variance (Moment 2) were found useful in classifying fricatives in Jongman et al.'s (2000) study may be due to the increase of the sample size, including an increase in the number of the words and the speakers used in the study, as compared with Forrest et al. (1988).

2.5.2.2 Consonant-to-Vowel (CV) Energy Ratio

The relation between consonants and vowels can be studied in several ways including studying the transition from one to the other, the duration of the two, and the relative amplitudes of the consonants and the vowels. One well known acoustic characteristic of speech is that the amplitude of vowels is greater than the amplitude of the adjacent consonants (Stevens, 1980). The consonant-vowel energy ratio is defined as "ratio of the power of a consonant to that of the nearest vowel in the same syllable" (Picheny, Durlach, & Braida, 1986). While studying the correlation of acoustic features to intelligibility of both adult and children's speech, Hazran and Markham (2004) examined the role of CV energy ratio. The study used a representative sample of 124 CVC tokens, encompassing all vowels and consonants, spoken twice by each speaker. The CVC words were produced by 45 speakers, including 18 adult females, 15 adult males, six boys and six girls, all speaking British

English. There were 45 listeners, divided into three groups, adults, 11 to 12 year olds and 7 to 8 year olds, all having normal hearing. These listeners judged the intelligibility of the talkers. Acoustic measurements were carried out on the speech tokens produced by the 45 speakers and the CV energy ratio was calculated for 12 words commencing with nasals, 12 with fricatives, 12 with plosives, and 14 words with plosives at their ends. No statistically significant correlation was found between the intelligibility of the words and the CV ratios for nasals, fricatives, or plosives.

While the intelligibility of a word does not seem to be related to its CV energy ratio (Hazan & Markham, 2004), an alteration in the CV energy ratio has been found to affect the perception of fricatives (Hedrick & Odhe, 1993) and the perceived clarity of the speech (Picheny et al., 1986). For example, to study the perception of relative amplitude of consonants and adjacent vowels, Hedrick and Ohde (1993) presented to ten normal hearing adults synthesized vowel and consonant combinations using two consonants, /s/ and /sh/, in combination with three vowels, /a/, /i/, and /u/. They found that when the relative amplitude between the vowel and consonant changed, the perception of the consonant changed in the /s/ to /sh/ continuum, with the perception of the consonant shifting towards /s/ if the relative amplitude of the consonant was high and towards /sh/ if it was low. The identification of the consonant was also found to be affected by the adjacent vowel, suggesting that both the primary property of the consonant and the context dependent property seem to play a role in speech identification.

To examine the difference between deliberately “clear” and normal “conversational” speech, Picheny, Durlach, and Braida (1985) studied 50 nonsense sentences produced by three speakers. These sentences were presented to 5 listeners, all of whom had a sensorineural hearing loss. The listener heard the sentences at three levels, one at maximum comfortable volume, the second at a comfortable volume, and the third at a level 10 dB below the comfortable volume. In addition to the difference in presentation levels, two different frequency-gain variables were also included. In one variation, the sound was divided into four frequency bands from 160 Hz to 8,000Hz, each band with its own volume control. The second variation of the frequency-gain characteristic was a flat response, with no alterations to the frequency-gain characteristics. The study found that “clear” speech was, as the name suggests, more intelligible than conversational speech, and this difference in intelligibility was independent of the talker, listener, the volume, and frequency-gain response characteristics. Deliberately changing the speaking style resulted in more intelligible speech. In a subsequent study, Picheny et al. (1986) studied the differences in acoustic properties of these two forms of speech: “clear” and “conversational” speech. As the tokens they were studying were running speech, where the amplitude of the vowel was variable, they only examined the amplitude of the consonant. They found that when the participants produced “clear” speech, the intensity of stop consonants was increased by 10 dB. They also found that while producing “clear” speech, the overall amplitude of speech was between 5 and 8 dB more than that in conversational speech. These findings suggest that changes in CV energy ratio may have an impact on speech intelligibility.

Chapter 3. Research Outline

Previous research indicates that listeners perceive distinct words based on the configuration of acoustic energy present spectrally and temporally. The various acoustic parameters elucidated above are ways of examining the configuration and distribution of this energy, and its impact on perception. In the same way that speech produced by an individual contains spectral and temporal patterns which have a relation to the characteristics of the speech and its clarity, radio-transmitted speech will also have its own characteristics and so the measures that indicate the characteristics of a person's speech should be sensitive to the effect of radio transmission. However, there is a paucity of studies applying these acoustic measures in the analysis of radio-transmitted speech.

3.1 Statement of the Problem

The lack of understanding of the relationship between a specific acoustic parameter and speech quality makes it hard to improve the quality of radio transmission in a systematic and efficient way. The main research questions proposed in this study are: (1) Which acoustic measures related to voice quality and speech intelligibility are susceptible to radio transmission? (2) How are the acoustic markers of sound quality change related to the perception of speech clarity?

3.2 Aims and Importance of the study

This study is aimed at determining what and how acoustic changes due to radio transmission may have an impact on the intelligibility or perceived clarity of the transmitted speech sounds. There is a paucity of research that evaluates the quality of radio-transmitted speech by using acoustic measures used in the evaluation of speech as reviewed above. A better understanding of the acoustic characteristics of radio-transmitted speech in comparison with the original signal will help develop a reliable and efficient way of assessing and improving the quality of radio transmission devices. A perceptual study clarifying the impact of the acoustic change that can be induced by radio transmission on the perception of speech quality will validate the monitoring of these acoustic parameters in the application for design improvement.

3.3 Hypotheses

It is hypothesized that a selection of acoustic measures, including the first two formant frequencies and vowel space, H1-H2, singing power ratio, the first two spectral moments (mean and variance), and CV energy ratio, will be useful for determining the effect of radio transmission on speech quality.

It is further hypothesized that the accuracy of identification and the clarity of the speech signal would be related to the level of H1-H2 and the level of the SPR.

4. Method

4.1 Participants

A convenience sampling method was used to recruit 20 normal hearing adults for the perceptual study. The inclusion criteria was that the participant be a native New Zealand speaker with (1) no history of speech or hearing abnormality and (2) audiometric thresholds within the normal range, -10 dB to 20 dB HL, as per the University of Canterbury protocols, based on the data by Jerger and Jerger (1980). Each participant's hearing was screened to 20 dB HL in the frequency range of 250 Hz to 8,000 Hz, in increments of octave intervals. Ten males (age range: 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) participated in the study. The participants received a fixed amount of compensation for the travel expense incurred by the participation in the study. Before the experiment, the participants were asked to read the subject information and sign consent forms related to the project, approved by the University of Canterbury's Human Ethics Committee (Appendix I)

4.2 Participant's Task

Each participant was asked to perform two tasks, vowel identification and clarity comparison tasks. In the vowel identification task, the participant was asked to listen to one vowel segment at a time and select from a list of five vowels, by clicking on an icon on the computer screen, which vowel they thought they have just heard (Appendix II). In the clarity comparison task, the participant was asked, for each trial, to listen 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 (Appendix III).

4.3 Materials

The vowel stimuli were extracted from the digitized acoustic recordings of the playback of a CD containing the recordings of a male adult's recitation of a Phonetically Balanced Kindergarten (PBK) word list (Appendix IV), also known as the Phonetically Balanced Test of Speech discrimination for Children, PBK50 (Brandy, 2002; Meyer & Pisoni, 1999), and the radio-transmitted sounds of the playback of this CD. These recordings had been recorded while playing back the original CD in three conditions, one without radio transmission (original) and two with radio transmission (Radios 1 and 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. These words were subjected to acoustic analysis.

Once acoustic analysis of the words was completed, as described later, two sets of five tokens of each of the four vowels were selected for the perceptual study, forming two groups, of 20 tokens each. In the study of vowels of variable duration, the selection of vowel stimuli was based on their associated H1-H2 values, with each token representing a point in one of the five equally spaced H1-H2 levels, from level 1 (highest H1 dominance) to level 5 (lowest H1 dominance). For the identification task in the perceptual study, the 20 tokens were randomly repeated five times each, for a total of 100 presentations for each set. The samples consisted of exemplars from the Original recording, Radio 1, and Radio 2. As the study was examining the relation of H1-H2 level to the accuracy of identification and perceived clarity, the word from which the exemplar had been extracted was not given any further consideration. In the comparison of vowels task, each of the five selected tokens was compared to the other four tokens in the set, leading to 20 groups. Each group was randomly presented twice, leading to 160 presentations. The random order of presentation was obtained from the Random Number Generator website (Haahr, 2009). The loudness of all the tokens was normalised to 98% using the Adobe Audition 3 software.

In the study of vowels of constant duration, five levels of H1-H2 ratio (level 1: strongest H1 dominance) and five levels of SPR (level 1: lowest energy around the 2-4 kHz region) were used as the basis for selecting tokens from the same word for each vowel as stimuli. The vowel /a/ from “cart”, /i/ from “beef” /o/ from “falls” and /u/ from “grew” were extracted. A segment of consistent duration, i.e. of 78 msec. was extracted from the steady state of the vowel and used for both the acoustic analysis, and also for the perceptual study. The exemplars selected were from the Original presentation, Radio 1 and Radio 2.

4.4 Instrumentation

The acoustic analysis of the speech samples was carried out by using automated speech analysis software, TF32, installed on a desktop computer. The equipment used to screen the audiometric thresholds of the participants included a Granson-Stadler GSI-61 clinical audiometer, with current calibration, and Telephonix TDH-50P supraaural headphones. Audiometric testing was carried out in a sound booth, with ambient sound level at 32dBA, measured using a Centre 322 Sound Level Meter. The perceptual study was also conducted in the same sound booth. The participants were seated at a desk, 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.

4.5 Procedure

As mentioned earlier, the participants were asked to perform two tasks. The instructions were presented to the participants from a written script, as seen on the screenshot in Appendix III, so as to keep the instructions consistent across all participants. Each participant was seated in the acoustically treated room, with the headphones in place.

In the first session, which involved a clarity differentiation task, participants were asked to listen to two sets of a series of pairs of vowels and indicate which of the two was clearer. In the second session, which involved a vowel identification task, the participants were asked to listen to two series of vowel stimuli and to identify, one at a time for each vowel token presented, from a list of 5 cardinal vowels, the one that they heard. In this task, there were four vowels, with each vowel shown at five different H1-H2 or SPR levels and each token repeated five times. The participants controlled the speed at which the presentations were made and had the option of repeating a presentation if they chose to do so. The loudness of presentation was set at a constant level for all participants. Prior to the actual testing, five samples of both the comparison task and the identification task were presented to the participants to familiarise them with the task. The presentation of the vowel tokens was ordered in a pre-determined random order. The tasks were presented sequentially, with one set of comparison followed by one set of identification. The process was then repeated with the second set.

4.6 Data Analysis

Two sets of data obtained were obtained. The first set consisted of acoustic data derived from the analysis of selected vowels and consonants from the three recordings, i.e. Original, Radio 1 and Radio 2. The second set consisted of scores for correct identification of vowels and scores for clarity of vowels, obtained during the perceptual study.

4.6.1 Acoustic Data

The author conducted all measurements for the acoustic study. The words were analysed by software for speech analysis run on a desktop computer. The word were be viewed graphically on the computer screen, in both spectral and waveform. Measurements were done by positioning the cursor manually, to extract the values of interest. The vowels /i/, /a/, /o/, and /u/ were used to extract measures of F1 and F2 frequencies, H1-H2, and SPR. The mean and variance of spectral moments for /s/ and /sh/ were examined, when they occurred at both the beginning of the word, and at its end. The CV ratio was examined for words embedded with vowels /i/, /a/, /o/, and /u/.

4.6.1.1 F1 and F2 Frequencies and Vowel Space

The ratio was calculated based on the F1 and F2 data that was obtained while calculating the vowel space. Data for the vowel /o/ was also included. The values for F1 and F2 were plotted for i/, /a/, and /u/. To obtain the values for F1 and F2 the following procedure was followed: Time slice of words containing the vowels /i/, /a/, and /u/ was be measured. The TF32 software view was set to display the spec's with the "preemphasis" and the LPC turned on. With the word being displayed in the time-frequency domain, a time slice at the commencement of the steady-state in the vowel was selected. The cursor was then located on the peak of F1 and the data copied to excel worksheet. Similar procedure was followed to obtain data for F2. Once F1 and F2 for all the tokens were obtained and the mean F1 and F2 values for each of the three vowels in each of the three conditions, i.e. Original, Radio 1 and Radio 2 was calculated. The vowel space for each of the three triangles was calculated as described by Robb & Chen (2008). First, the Euclidian distance (ED) between the points of the triangle formed by the F1 to F2 values of the vowels was calculated as follows:

$$ED_{u/a} = \sqrt{} (F1_u + F1_a)^2 + (F2_u + F2_a)^2$$

The area of each triangle was calculated as follows:

$$Area_{iua} =$$

$$\sqrt{} (ED_{ia} + ED_{iu} + ED_{ua}) (ED_{ia} + ED_{iu} - ED_{ua}) (ED_{ia} - ED_{iu} + ED_{ua}) (ED_{iu} + ED_{ua} - ED_{ia})$$

4.6.1.2 H1-H2

To obtain the values for H1 and H2 the following procedure was followed: Segments of words containing the vowels /i/, /a/, /o/ and /u/ were measured. The TF32 software view was set to display the spec's with the the LPC and "preemphasis" turned off and the Harmonics,"H" and the LTA turned on. With the word being displayed in the time-frequency domain, a time slice at steady-state in the vowel was selected. The cursor was then located on the peak of H1 and the data copied to excel worksheet. Similar procedure was followed to obtain data for H2.

4.6.1.3 Singing Power Ratio

The highest spectral peak between 0 Hz and 2 kHz and between 2 kHz and 4 kHz was identified and a ratio between the dB reading for both of them was calculated. To obtain the values for the peak between 0-2 kHz and 2-4 kHz, the following procedure was followed. Segments of words containing the vowels /i/, /a/, /o/, and /u/ were measured. The TF32 software view was set to display the spectrum, with the “preemphasis” and the Harmonics, “H” turned on. With the word being displayed in the time-frequency domain, a time slice between the commencement of the vowel and its end was selected. The cursor was then moved to locate the harmonic peak between 0-2 kHz and the data copied to excel worksheet. The cursor was then moved to locate the peak between 2-4 kHz and the procedure repeated.

4.6.1.4 Spectral Moments

Segments of words containing /s/ and /sh/ were measured when they occurred in either the beginning or the end of the word. The TF32 software view was set to display the spectrogram with the LTA and the Harmonics, “H” turned on. The checkbox for “Moment” was selected. With the word being displayed in the time-frequency domain, the cursor was moved to select the middle one third of the /s/ and /sh/ sounds. The readout of values for the mean and variance was then copied to a spreadsheet.

4.6.1.5 Consonant-to-Vowel Energy Ratio

To obtain the values for the amplitude of the consonant and vowel following procedure was followed: Segments of words containing the vowels /i/, /a/, /o/, and /u/ were measured. The TF32 software view was set to display the endpoints on the RMS trace. Based on visual inspection of the waveform and the display in the time-frequency domain, the cursor was placed on the peak amplitude of the consonant and the reading will be copied to an excel spreadsheet. The procedure was then repeated to obtain data of the peak amplitude of the vowel.

4.6.2 Perceptual Data

The perceptual data for the primary study consisted of four sets of scores, obtained by each of the 20 participants. Two sets were based on the identification of the vowel, and two sets based on identifying which presentation of vowel that the participants thought was clearer. The scores obtained in the primary study were all related to vowels organised on the basis of the H1-H2 ratio.

In the follow-up study, the perceptual data also consisted of four sets of scores, obtained by each of the five participants. Two sets of scores were for the identification of the vowel and two for identifying which presentation of vowel that was perceived as being clearer. One set of scores for identification and judgment of clarity was related to vowels organised on the H1-H2 measure, and the other set was based on the vowels organised based on the SPR measure. These eight scores, i.e. four scores for the first perceptual study, using tokens of variable duration, and four scores from the second perceptual study, using tokens of fixed duration, were subjected to statistical analysis.

4.7 Statistical Analysis

The statistical analysis software Sigmastat® 3.5 for Windows was used to perform all statistical analysis. The significance level was set at 0.05. Three independent variables (factors) were considered. The condition factor includes three levels: Original, Radio 1, and Radio 2. The vowel factor includes the four corner vowels or general vowel types, /i/, /a/, /o/, and /u/. The consonant factor includes /s/ and /sh/ in the initial and final word position. A series of Kruskal-Wallis one-way ANOVA on Ranks was conducted to determine whether there was a condition effect on F1 and F2 frequencies. Where there was insufficient number of data for measures from the original signals, a t test is conducted to compare Radios 1 and 2 on these measures and the single measure from the original signals was plotted against the mean or median scores of measures obtained from the 10 playback volume levels used for the transmission of Radios 1 and 2 signals to allow for a visual comparison of the three conditions. A series of two-way Analyses of Variance was conducted to determine the vowel (or consonant) effect, condition effect, and vowel (or consonant) by condition interaction effect on H1-H2, SPR, Spectral Moment One (mean), Spectral Moment Two (variance), and CV energy ratio.

Similar statistical analysis was performed on the perceptual data. Two-way (vowel by H1-H2 or SPR level) Analyses of Variance were conducted on the vowel identification or clarity scores obtained for vowel segments with variable or constant lengths and different H1-H2 or SPR level, each with 5 levels. A series of Pearson's Product Moment correlation procedures was also conducted between vowel duration and the vowel identification and vowel clarity scores for the dataset that includes stimuli with variable lengths.

4.8 Reliability

The inter-judge reliability of acoustic measures was carried out by re-measuring approximately 20% of the words for all the acoustic analysis. For the F1 and F2 frequencies, H1-H2, SPR, and CV energy ratio measures, five out of 23 words were re-measured. For Spectral Moment One (mean) and Two (variance), five out of 26 words were re-measured. A series of Pearson Product Moment correlation procedures was carried out on the original data and the re-measured data. The measure-remeasure reliability was found to be high for all measures, including F1 and F2 frequencies (r = 0.946 for the F1/F2 frequency ratio), H1-H2 (r = 0.855), SPR (r = 0.982), Spectral Moment One (r = 0.989), Spectral Moment Two (r = 0.907), and CV energy ratio (r = 1.0). The correlation values for all the acoustic measures are high, indicating that there was consistency in the measurements of both, the first set of measurements that were used in the acoustic analysis, and the re-measured data (Portney & Watkins, 2009).

5. Results

Results from a comparison of the acoustic measures of sound segments embedded in words recorded for three conditions (Original, Radio 1, and Radio 2) and from the perceptual study of the effect of H1-H2 and SPR on the perception of vowel identification and clarity are presented separately in this chapter.

5.1 Acoustic Measures

Results of the statistical tests conducted on the acoustic measures of vowels and consonants are presented for vowels and consonants in separate sections.

5.1.1 Vowels

Results regarding the acoustic measures for vowels are reported as follows, including F1 and F2 frequencies, vowel space, H1-H2, and SPR. The words selected for various statistical tests of the acoustic measures of vowels are listed in Table 1.

Table 1. Vowel sets used for grouping data in the statistical tests on vowel-related measures.

/i/	/a/	/o/	/u/
/i/ (12 words):	**/a/ (3 words):**	Fall (/ɔ/)	Grew
Please	Cart	Plow (/aᴜ/)	Few
Need	Park		
Teach	Darn		
Tree	**Others:**		
Me	Bath		
Beef	Class		
Bead	Laugh		
Weed	Path		
Bee			
Feed			
Freeze			
Knee			

5.1.1.1 F1-F2 Frequencies and Vowel Space

For vowel /i/, results from a series of Kruskal-Wallis one-way (3 conditions: Original, Radio 1, Radio 3) ANOVA on Ranks performed on the F1 and F2 values of the vowels segmented from 12 words (see Table 1), along with the median F1 and F2 for the same tokens recorded at 10 volume levels with Radios 1 and 2, showed a significant condition effect (H = 27.411, df = 2, $p < 0.001$) on F1 but no significant condition effect on F2 (H = 1.577, df = 2, p = 0.454). Post-hoc pair-wise comparisons with Tukey test revealed that, for the vowel /i/, the Original had a significantly lower F1 than both Radios 1 and 2 but Radios 1 and 2 did not differ significantly on F1.

As for vowels /a/ and /u/, since they were sampled from only three words and two words respectively (see Table 1), the Original dataset was excluded from further statistical tests due to insufficient sample size. With all the /a/ tokens (3 words X 2 radios X 10 volume levels) combined, t tests showed that Radios 1 and 2 were not significantly different on F1 (t = -1.205, df = 56, p = 0.233) but Radio 1 had a significantly higher F2 than Radio 2 (t = 5.255, df = 56, $p < 0.001$). For measures obtained from the /u/ segmented from the word “grew” (1 word X 2 radios X 10 volume levels), t tests showed that Radio 1 had a significantly higher F1 than Radio 2 (t = 7.926, df = 17, $p < 0.001$) but Radios 1 and 2 did not significantly differ on F2 (t = 1.25, df = 17, p = 0.228).

The median scores of the F1 and F2 values derived from the vowels, /i/ (embedded in the randomly selected word “freeze”), /a/ (“park”), and /u/ (“grew”), recorded at 10 volume levels with Radios 1 and 2 respectively are plotted, in Figure 1, against the values derived from the original signals. As shown in Figure 1, the vowel space area is largest for the Original (0.1791 kHz^2), followed in order by Radio 2 (0.1642 kHz^2) and Radio 1 (0.0401 kHz^2). A t test conducted on all the vowel space data (2 radios X 10 volume levels) as calculated from the F1 and F2 of these vowels revealed that Radio 1 had a significantly smaller vowel space than Radio 2 ($t = -17.427$, $df = 17$, $p < 0.001$). Using the same corner vowels derived from different words ("beef", "cart", "fall", and "grew"), the same observation could be made about the between-condition difference in the vowel space area (see Appendix VI). These findings suggest that changes of F1 and F2 frequencies resulting in a smaller vowel space area is a sign of signal degeneration that can be induced by radio transmission.

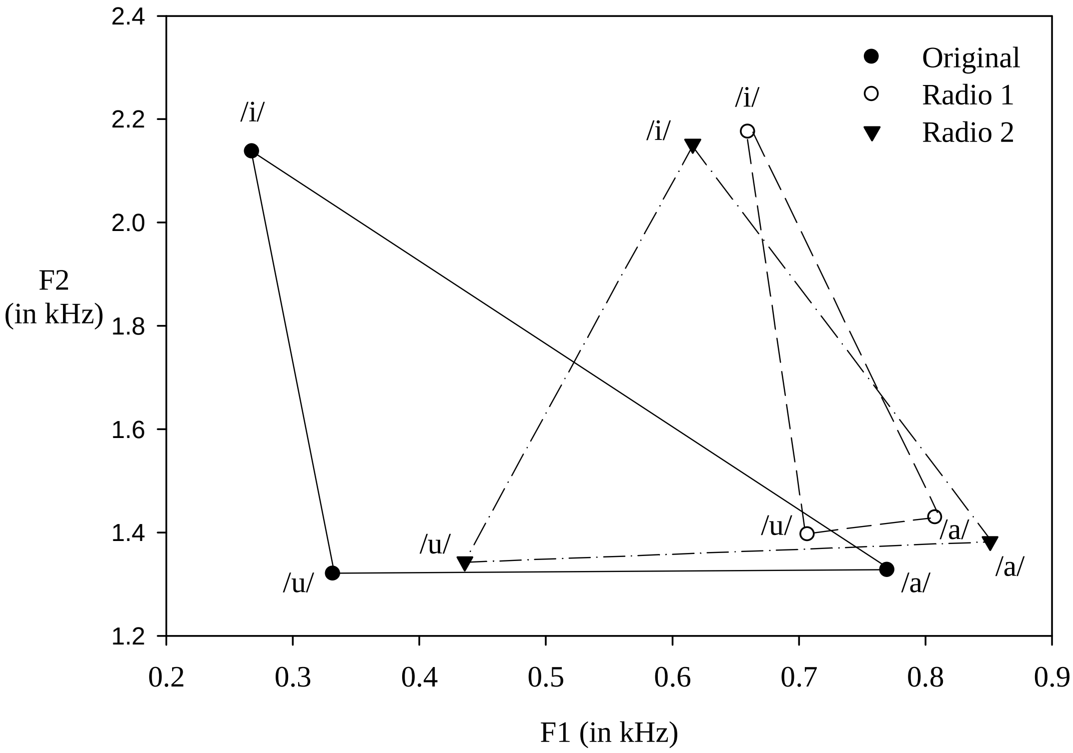


Figure 1. Schematic representations of vowel working space.

Schematic representations of the vowel working space obtained from recordings for three signal conditions (Original, Radio 1, and Radio 2). The data points shown in the graph are the median F1 and F2 values derived from the selected vowels /i/ ("freeze"), /a/ ("park"), and /u/ ("grew") transmitted through Radios 1 and 2 at 10 playback volume levels and those from the original signals.

5.1.1.2 H1-H2

To allow for a statistical comparison on H1-H2 between the three signal conditions (Original, Radio 1, and Radio 2), the H1-H2 values of vowels segmented from 23 words were used. These included recordings of Radios 1 and 2 signals at the 10 playback volume levels and those of the Original signals. The data were grouped into four vowel sets, /a/, /o/, /i/, and /u/ (see Table 1), and submitted to a two-way (vowel by condition) ANOVA. Results from the ANOVA conducted on the H1-H2 measures showed a significant condition effect [F(2, 459) = 4.314, p = 0.014], vowel effect [F(3, 459) = 41.151, p < 0.001], and vowel by condition interaction effect [F(6, 459) = 9.594, p < 0.001].

As shown in Figure 2 (on the next page), post hoc pair-wise comparisons using the Tukey test indicated that, for the vowel /i/, the only significant between-condition difference on H1-H2 was found between the Original and Radio 2, with the Original showing a significantly lower mean H1-H2 (suggesting a thicker or less breathy quality) than Radio 2. For both the low-vowel (/a/ and /o/) sets, the Original had a significantly higher mean H1-H2 (suggesting a thinner or more breathy quality) than both Radios 1 and 2. No statistically significant difference was found between conditions for the vowel /u/ set most likely due to the small sample size used for that vowel.

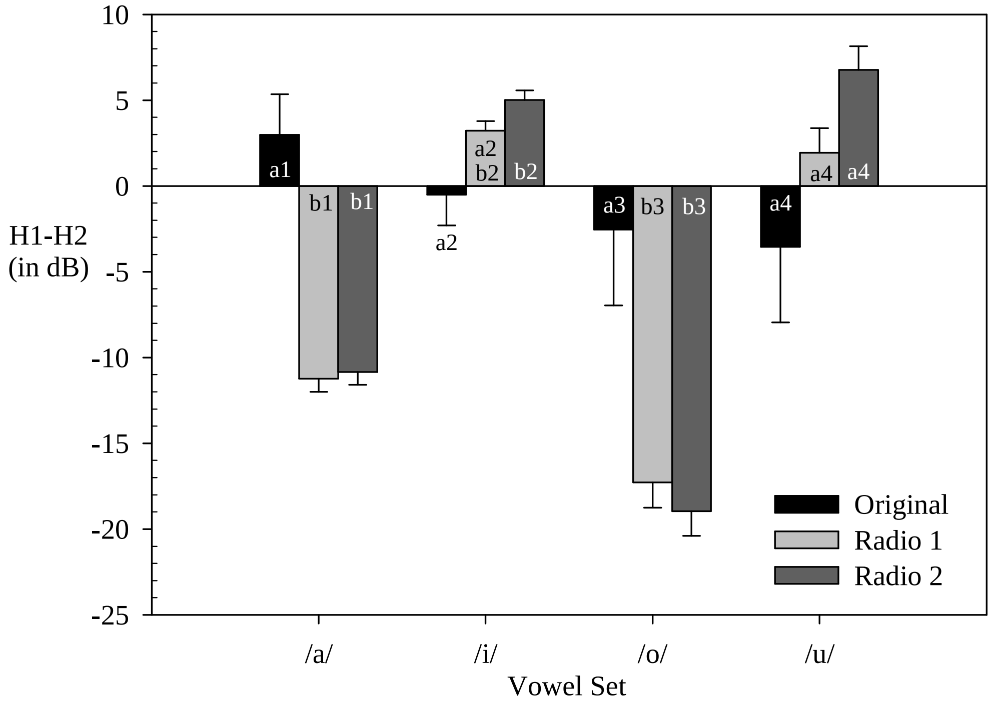


Figure 2. Means and standard errors of the measure of the amplitude difference between the first two harmonics (H1-H2). The mean scores for each of the three conditions (Original, Radio 1, and Radio 2) found to be statistically different from one another within each vowel set (/a/, /o/, /i/, and /u/) were marked with different letters (Note: The numbers are attached to the letter to indicate that no comparison is being made between different vowels.)

5.1.1.3 Singing Power Ratio

The measurement of SPR involves measurements around the frequency regions where the first two formant frequencies are typically located and thus may be sensitive to the vowel difference in formant structure. Therefore, signal comparisons on SPR was more restrictively controlled, with only measures of the four corner vowels segmented from the randomly selected words (“beef”, “cart”, “fall”, and “grew”), as previously mentioned in Section 5.1.1.1, being used for comparing the three signal conditions (Original, Radio 1, and Radio 2).

As compared with Radios 1 and 2, the Original tends to have a lower SPR value except for the vowel /u/ (see Figure 3), suggesting that the original signals have relatively greater energy around the 2-4 kHz region and thus higher voice projection power. Results from a series of t tests revealed that Radio 1, as compared with Radio 2, had a significantly lower mean SPR for the vowels /i/ (t = -2.612, df = 17, p = 0.018) but significantly higher mean SPR for the vowel /ɔ/ (t = 3.697, df = 17, p = 0.002). Radios 1 and 2 did not differ significantly on SPR for the vowels /a/ (t = -1.087, df = 17, p = 0.292) or /u/ (t = -1.171, df = 17, p = 0.258).

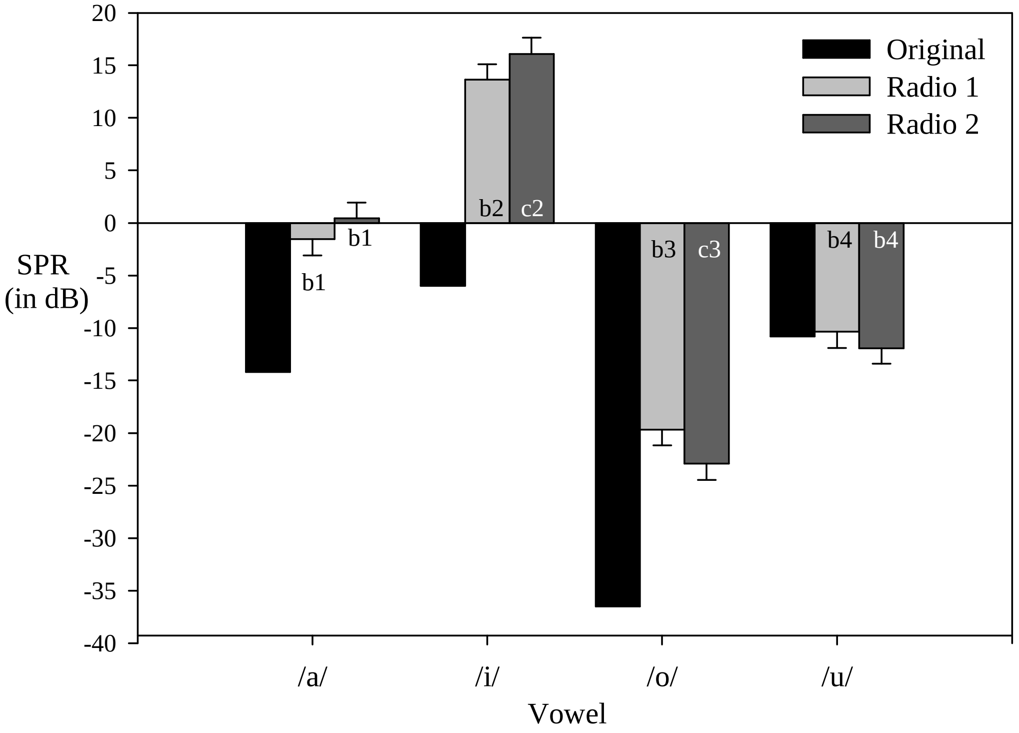


Figure 3. Means and standard errors of the measure of singing power ratio (SPR). The mean scores for the two conditions (Radio 1, and Radio 2) found to be statistically different from each other within each vowel, /a/ (“cart”), /ɔ/ (“fall”, shown as “/o/” in the graph), /i/ (“beef”), and /u/ (“grew”) were marked with different letters (Note: The numbers are attached to the letter to indicate that no comparison is being made between different vowels.)

5.1.2 Consonants

Results from a series of two-way (consonant by condition) ANOVAs performed on the acoustic measures involving consonants were summarized in Table 2. These acoustic measures included Spectral Moment One (mean), Spectral Moment Two (variance), and CV energy ratio. As shown in Table 2, a significant condition (3 levels: Original, Radio 1, and Radio 2) effect was found for all of these measures except for CV energy ratio. Details of the results for these measures are presented in separate sections as follows.

Table 2. Two way (vowel by condition) ANOVA results for the acoustic measures related to consonants, including Spectral Moment One (mean), Moment Two (variance), and consonant-to-vowel (CV) energy ratio.

	Consonant		**Condition**		**Consonant by Condition**	
Spectral Moment 1 (Mean)	F(3, 60) = 10.407,	p < 0.001*	F(2, 60) = 296.608,	p < 0.001*	F(6, 60) = 21.280,	p < 0.001*
Spectral Moment 2 (Variance)	F(3, 60) = 17.634,	p < 0.001*	F(2, 60) = 66.150,	p < 0.001*	F(6, 56) = 1.267,	p = 0.286
CV Energy Ratio	F(5, 55) = 1.651,	p = 0.188	F(2, 55) = 4.395,	p = 0.612	F(6, 55) = 0.180	p = 0.981

*****Significant at 0.05 level

5.1.2.1 Spectral Moments

Results of the two-way (consonant by condition) ANOVA conducted on the measures of Spectral Moment One (mean) obtained from the consonants (/s/ or /sh/) segmented from words initiated or ended with these sounds (i.e., /s.../, /...s/, /sh.../, and /...sh/) and recorded in three signal conditions (Original, Radio 1, and Radio 2) revealed a significant consonant effect, condition effect, and consonant by condition interaction effect (see Table 2). Post hoc pair-wise comparisons using the Holm-Sidak method revealed that for both Radios 1 and 2, no between-consonant comparison among the four consonant contexts (i.e., /s.../, /...s/, /sh.../, and /...sh/) were significantly different. In contrast, for the Original, the mean Spectral Moment One measures for each of the consonant contexts were found to be all significantly different from one another ($p < 0.05$). As can be seen in Figure 4, the original signals show a clear differentiation on the Spectral Moment One measure between consonant contexts, with /s/ exhibiting higher Spectral Moment One values than /sh/. Figure 2 also shows that, across all consonant contexts, the original signals have a significantly higher Spectral Moment One than both Radios 1 and 2 while Radios 1 and 2 do not differ significantly from each other.

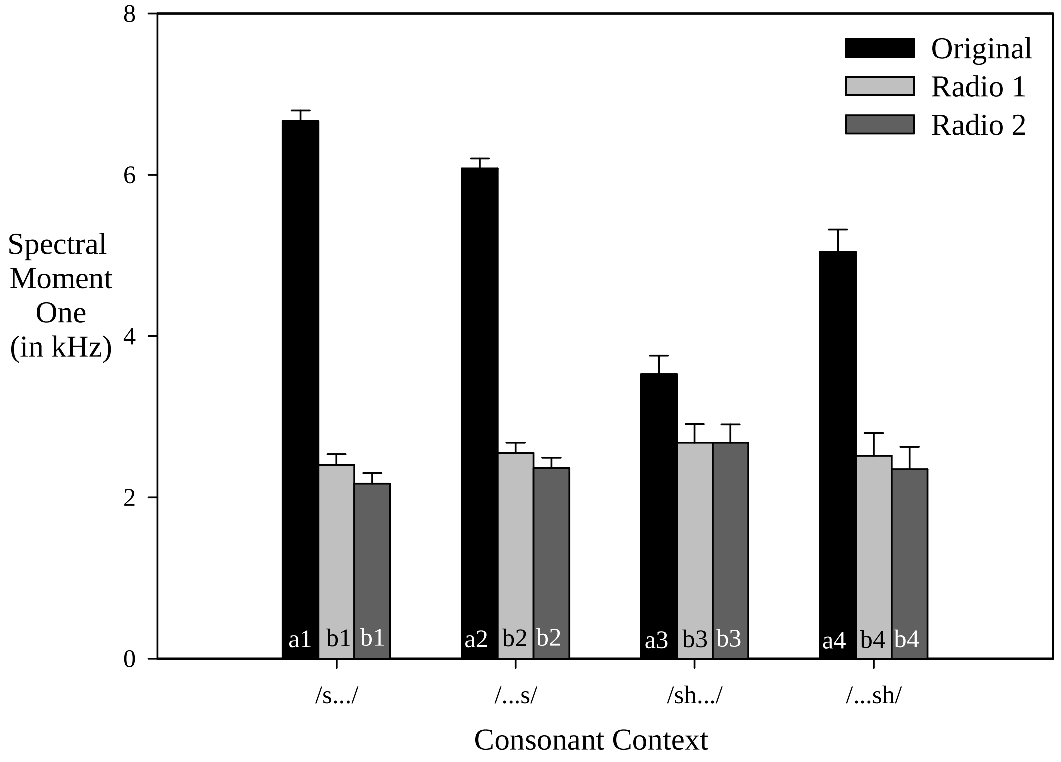


Figure 2. Means and standard errors of Spectral Moment One (mean) for the fricative in each of the four contexts, including /s/ preceding a vowel (/s…/), /s/ following a vowel (/…s/), /sh/ preceding a vowel (/sh…/), and /sh/ following a vowel (/…sh/) and recorded in three signal conditions (Original, Radio 1, and Radio 2). Significantly different between-consonant condition comparisons within each vowel set were marked with different letters (Note: The numbers are attached to the letter to indicate that no comparison is being made between different vowels.)

As for Spectral Moment Two (variance), results of the two-way (consonant by condition) ANOVA conducted on measures obtained from the same set of consonants as previously described that were used in extracting Spectral Moment One revealed a significant consonant effect and condition effect but no significant consonant by condition interaction effect (see Table 2). Post hoc pair-wise comparisons using the Holm-Sidak method revealed that all pair-wise comparisons were statistically significant, with Spectral Moment Two being significantly higher for the Original (1.354 kHz), followed in order by Radio 2 (0.781 kHz) and Radio 1 (0.6 kHz). The pair-wise comparisons between consonant contexts, as shown in Figure 5, were all significant except for that between /s…/ and /…s/. It can be seen from Figure 5 that regardless of the position, /s/ has a higher Spectral Moment Two than /sh/. For /sh/, the mean Spectral Moment Two value is higher in the initial position than in the final position.

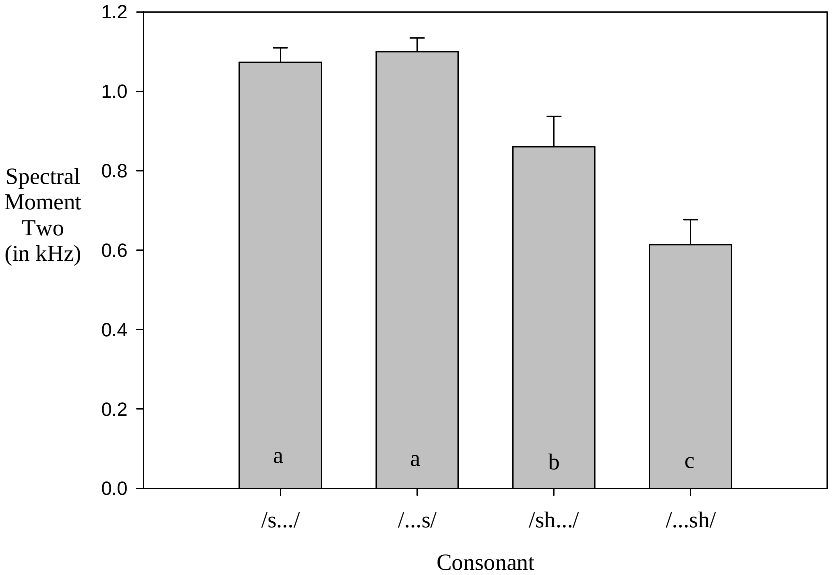


Figure 3. Means and standard errors of the Spectral Moment Two (variance) for the fricative in each of the four contexts, including /s/ preceding a vowel (/s…/), /s/ following a vowel (/…s/), /sh/ preceding a vowel (/sh…/), and /sh/ with all signal conditions combined. Significantly different means were marked with different letters.

Although there was no significant consonant by condition interaction effect, Figure 6 presents the mean Spectral Moment Two for each subgroup to show the magnitude of the between-condition difference in each consonant context. It can be observed from Figure 6 that, unlike Spectral Moment One, the trend showing /s/, in either initial or ending position, to exhibit a higher Spectral Moment Two than /sh/ is well maintained across the three signal conditions (Original, Radio 1, and Radio 2).

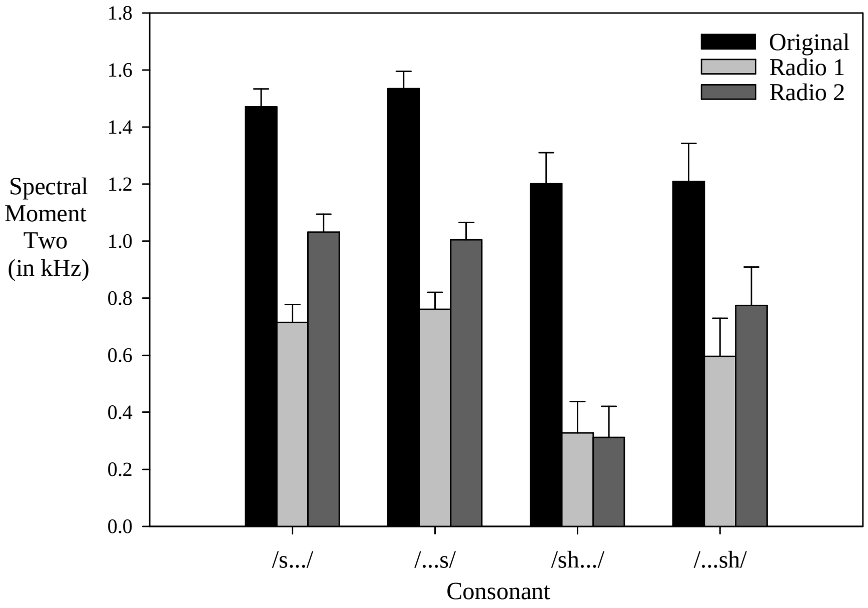


Figure 6. Means and standard errors of the Spectral Moment Two (variance) for the fricative in each of the four contexts, including /s/ preceding a vowel (/s…/), /s/ following a vowel (/…s/), /sh/ preceding a vowel (/sh…/), and /sh/ following a vowel (/…sh/) and recorded in three signal conditions (Original, Radio 1, and Radio 2). Significantly different between-consonant condition comparisons within each vowel set were marked with different letters (Note: The numbers are attached to the letter to indicate that no comparison is being made between different vowels.)

5.1.2.2 CV Energy Ratio

Results of the two-way (vowel effect by condition effect) ANOVA, conducted on the measures of CV energy ratio obtained from the Original recording, Radio 1, and Radio 2, for the vowels /a/, /i/, /o/ and /u/, indicate that the transmission of the signal through Radio 1 and Radio 2 had no significant vowel effect, condition effect or vowel by condition interaction effect on the vowels /a/, /i/, /o/, and /u/ as shown in Table 2. The consonant-vowel energy ratio was not found to be a sensitive measure to transmission through these radio systems.

5.2 Perceptual Measures

The results of the perceptual study is organised in two parts. The first part of this section reports the results of the study based on vowels with durations of variable lengths, extracted from the middle third, steady portion of the vowel. The second part of this section reports the results of the study based on vowel segments of constant duration, extracted from the steady portion of the vowel.

5.2.1 Stimuli of Variable Lengths

The results from the perceptual study using vowel segments of variable lengths are organised in two parts. In the first part, the results of the study of the identification of the vowels organised on the basis of H1-H2 levels is reported for two datasets (dataset H1H2-VL1 and dataset H1H2-VL2) separately. In the second part, the results from the study of the differentiation of vowel clarity are reported for these two datasets (dataset H1H2-VL1 and dataset H1H2-VL2) separately.

5.2.1.1 Vowel Identification for Stimuli of Variable Lengths

Results of the two-way (vowel by H1-H2 level) ANOVA conducted on the vowel identification scores obtained for dataset H1H2-VL1, which consisted of vowel tokens of variable lengths with different H1-H2 levels (with lower H1-H2 level indicating greater H1 prominence and thus being associated with a more breathy or thinner voice quality), revealed a significant H1-H2 level effect [F (4, 76) = 14.745, p < 0.001], vowel effect [F (3, 57) = 42.053, p < 0.001], and vowel by H1-H2 level interaction effect [F (12, 228) = 12.483, p < 0.001]. As shown in Figure 7, the percentage of correct vowel identification for /a/, /o/, and /u/ were consistently high, and close to a 100 % across all the five H1-H2 levels. The identification for /i/ for H1-H2 level 1, 3, and 5 were all lower than 50% while the scores for 2 and 4 were similar, at around 80%. No trend relating to an increase in the correct vowel identification scores as a function of H1-H2 level can be observed in any of the four vowels. Results from a series of Pearson's correlation procedures performed on the vowel identification scores in dataset H1H2-VL1 revealed that the vowel identification scores for /i/ were positively correlated with vowel duration (r = 0.533) and that no significant correlation was found for /a/ (r = 0.206), /o/ (r = -0.169), or /u/ (r = 0.062), suggesting that the vowel identification scores for /i/ might be influenced by the duration of the vowel present.

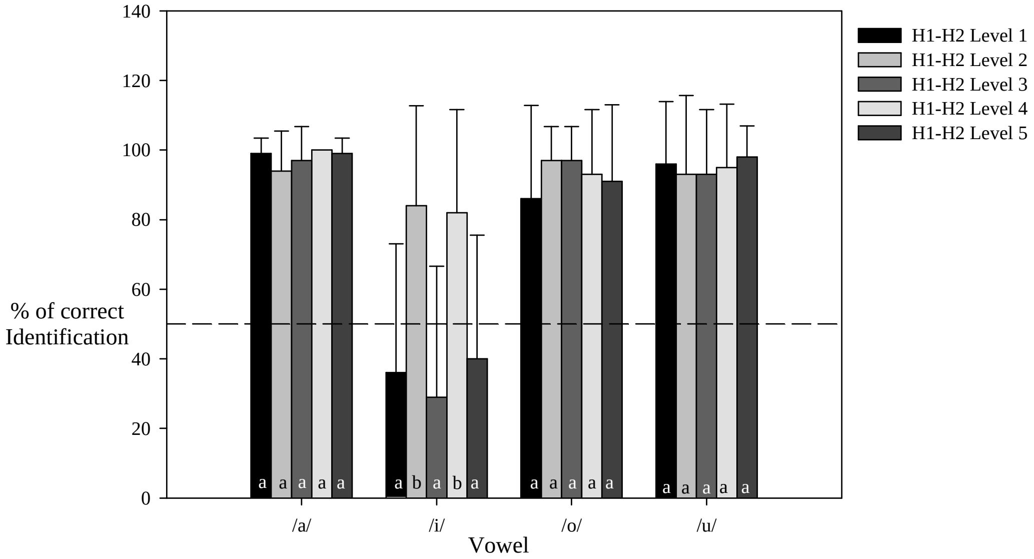


Figure 7. Vowel identification scores for stimuli with variable lengths and different H1-H2 levels used in the first stimulus set (dataset H1H2-VL1). (Note: The lower the H1-H2 level, the more prominent the H1 energy and thus the more breathy or thinner the associated voice quality). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

Results of the two-way (vowel by H1-H2 Level) ANOVA conducted on the vowel identification scores obtained for dataset H1H2-VL2, which also consisted of vowel tokens of variable lengths, were in agreement with the finding in dataset H1H2-VL1, showing a significant H1-H2 level effect [(F (4, 76) = 13.511, p < 0 .001], vowel effect [F (3, 57) = 40.752, p < 0.001], and vowel by H1-H2 level interaction effect [F (12, 228) = 16.006, p < 0.001]. As shown in Figure 8, the percentage of correct vowel identification for /a/, /o/, and /u/ were consistently high across all of the five H1-H2 levels, which was similar to the results for the first set. Like the finding in the first dataset, the identification for H1-H2 levels 1 and 5 for /i/ were lower that 50% but the scores for H1-H2 levels 2 and 4 were similar, around 80%. However, unlike the finding in the dataset H1H2-VL1, the H1-H2 level 3 in dataset H1H2-VL2 showed a relatively high percentage of correct vowel identification. This conflicting finding for H1-H2 level 3 reinforces the observation that the reduction of correct vowel identification rate for the vowel /i/ may not be related to the change in H1-H2 level alone. Results from a series of Pearson' correlation procedures performed on the vowel identification scores in dataset H1H2-VL2 also agree with the finding in the dataset H1H2-VL1, showing that the vowel identification scores for /i/ are positively correlated with vowel duration (r = 0.628) and there is no significant correlation between vowel duration and vowel identification scores for /a/ (r = -0.18), /o/ (r = -0.059), or /u/ (r = 0.0129). As the vowel /i/ is the only vowel that generally show a poor vowel identification score in both dataset H1H2-VL1 and dataset H1H2-VL2, it is not surprising that the impact of vowel duration on vowel identification, if there is any, would be more evident in this vowel. The positive correlation between vowel duration and vowel identification scores is also not surprising as the longer the vowel duration, the more time and sound content to make the identification task easier and thus the more likely that the vowel is correctly identified.

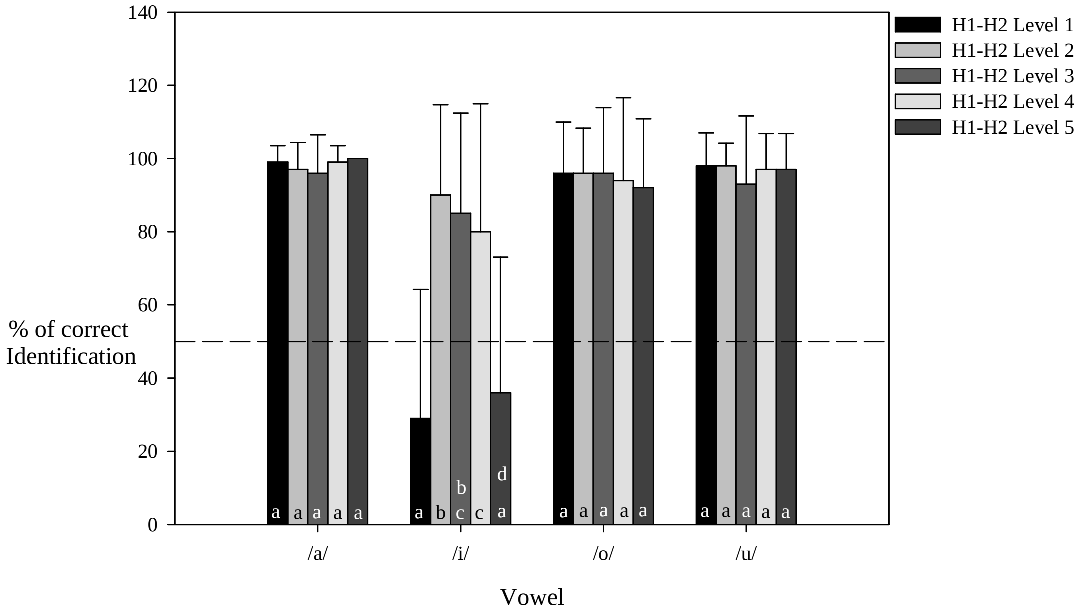


Figure 8. Vowel identification scores for stimuli with variable lengths and different H1-H2 levels used in the second stimulus set (dataset H1H2-VL2). (Note: The lower the H1-H2 level, the more prominent the H1 energy and thus the more breathy or thinner the associated voice quality). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

5.2.1.2 Vowel Clarity for Stimuli of Variable Lengths

Results of the two-way (vowel by H1-H2 level) ANOVA on the vowel clarity scores obtained for dataset H1H2-VL1, which consisted of vowel tokens of variable lengths, revealed a significant H1-H2 level effect [F (4, 76) = 16.539, p < 0.001], vowel effect [F (3, 57) = 20.330, p < 0.001], and vowel by H1-H2 level interaction effect [F (12, 228) = 25.736, p < 0.001]. As shown in Figure 9, for all vowels except for the vowel /i/, H1-H2 level 1, (which indicates a more prominent the H1 and thus a more breathy or thinner voice quality), is associated with the lowest vowel clarity score. This finding suggests that there exists, at least for the back vowels, /a/, /o/, and /u/, a H1-H2 threshold beyond which the voice may be perceived as less clear. Results from a series of Pearson's Product Moment correlation procedures performed on the vowel clarity scores in dataset H1H2-VL2 revealed that the vowel clarity scores for /a/ and /o/ was negatively correlated with vowel duration (r = -0.532, and r = -0.585 respectively). The vowel clarity scores for /u/ were found to be positively correlated with duration (r = 0.529) and the vowel clarity scores for /i/ showed no significant correlation with vowel duration (r = -0.0282). The inconsistent direction of the relationship between vowel duration and vowel clarity scores suggests that it may be more of the variations in the acoustic information contained in the signals, rather than the vowel duration itself, that is affecting the perception of vowel clarity.

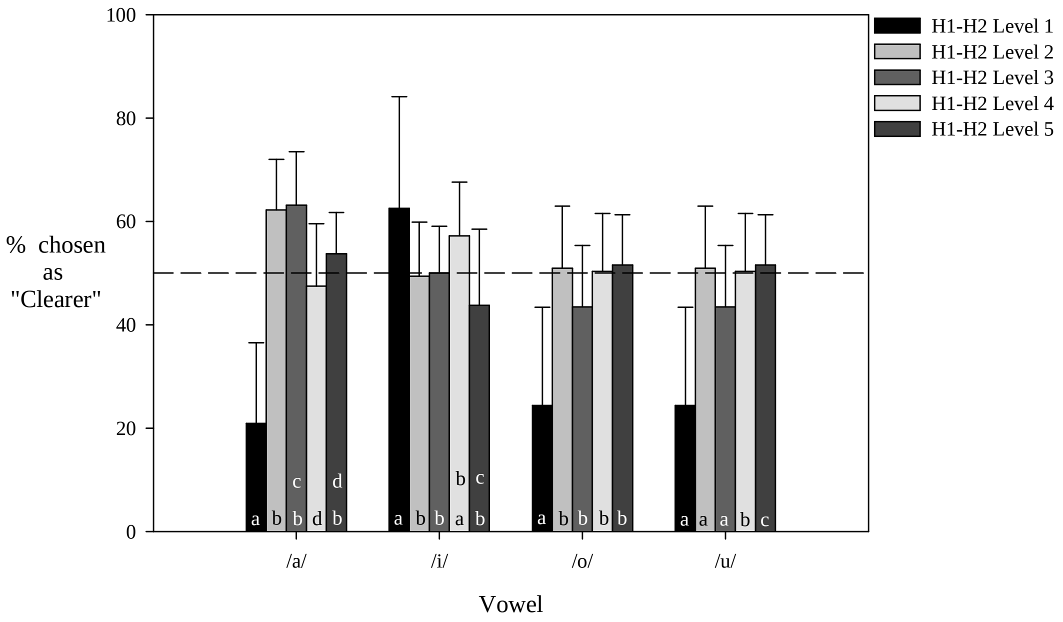


Figure 9. Vowel clarity scores for stimuli with variable lengths and different H1-H2 levels used in the first stimulus set (dataset H1H2-VL1). (Note: The lower the H1-H2 level, the more prominent the H1 energy and thus the more breathy or thinner the associated voice quality). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

Results for the two-way (vowel by H1-H2 level) ANOVA performed on the vowel clarity scores in dataset H1H2-VL2 also showed a significant H1-H2 level effect [F (4, 76) = 9.569, $p<0.001$], vowel effect [F (3, 57) = 20.700, $p < 0.001$], and vowel by level interaction effect [F(12, 228) = 4.967, $p < 0.001$]. However, the pattern of change in the vowel clarity scores as a function of H1-H2 level was not consistent with the finding in dataset H1H2-VL1. As shown in Figure 10, the vowel clarity scores for vowels /i/, /o/, and /u/ was the highest with the H1-H2 level 1, which is associated with a more prominent H1 and thus a more breathy voice quality. This finding does not support the hypothesis that the breathier or thinner the voice, the less clear the vowel is perceived. Similar to the finding in dataset H1H2-VL1 where no consistent relationship between vowel duration and vowel clarity scores was found, results from a series of Pearson's correlation procedures in dataset H1H2-VL2 also failed to show any significant relationship between vowel duration and vowel clarity scores for the vowel /a/ ($r = 0.106$), /i/ ($r = -0.243$), /o/ ($r = -0.381$), or /u/ ($r = 0.154$).

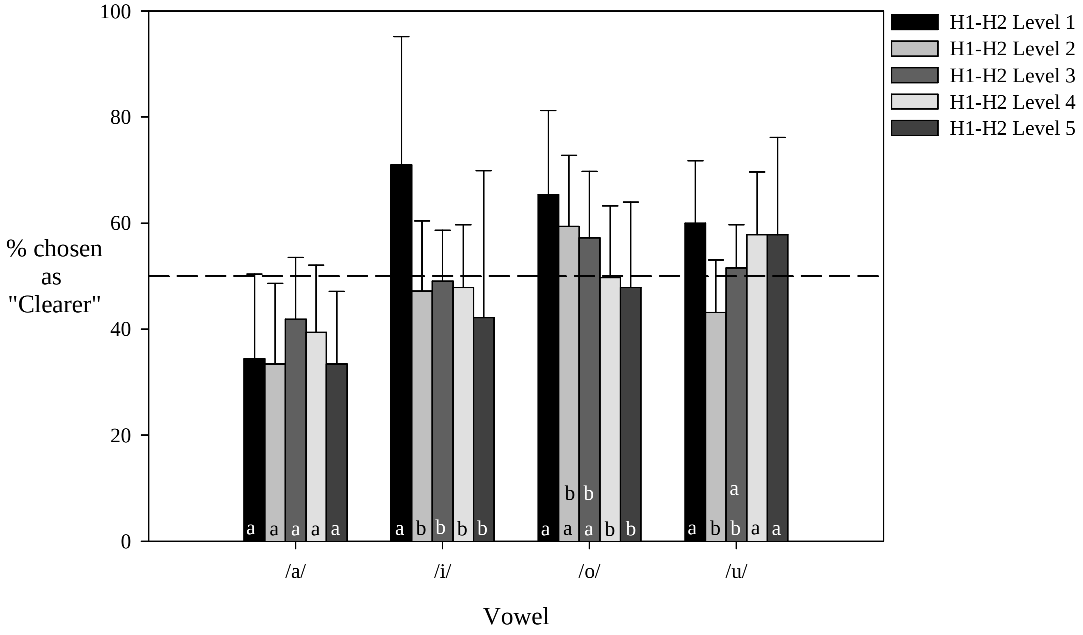


Figure 10. **Vowel clarity scores for stimuli with variable lengths and different H1-H2 levels used in the second stimulus set (dataset H1H2-VL2).** (Note: The lower the H1-H2 level, the more prominent the H1 energy and thus the more breathy or thinner the associated voice quality). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

5.2.2 Stimuli of Constant Lengths

The results from the perceptual study using vowel segments of constant lengths are organised in two parts. In the first part, the results from the study of the identification of vowels organised on the basis of H1-H2 level (dataset H1H2-CL) is reported, followed by the results from the study of the identification of the vowels organised on the basis of SPR level (dataset SPR-CL). In the second part, the results from the study of the differentiation of vowel clarity using the two datasets (dataset H1H2-CL and dataset SPR-CL) are reported.

5.2.2.1 Vowel Identification for Stimuli of Constant Lengths

Results for the two-way (vowel by H1-H2 level) ANOVA on the vowel identification scores obtained for dataset H1H2-CL revealed a significant H1-H2 level effect [F (4, 16) = 3.604, p = 0.028], vowel effect [F (3, 12) = 3.646, p = 0.045], and a vowel by level interaction effect [(F (12, 48) = 5.554, p < 0.001]. As shown in Figure 11, the percentages of correct vowel identification for /a/ and /o/ were relatively high while those for /u/ were also close to 100% except for H1-H2 level 4, which fell below 50%. For the vowel /i/, the vowel identification scores for H1-H2 levels 1, 3, and 5 were low. No clear trend can be discerned in these scores. The vowel identification scores for /i/ in this stimulus set (dataset H1H2-CL) follows a similar pattern to those for datasets H1H2-VL1 and H1H2-VL2, as shown in Figures 7 and 8 respectively, in that some H1-H2 levels with the vowel /i/ are associated with poor vowel identification scores. As vowel duration was controlled in this stimulus set (dataset H1H2-CL), the consistent finding that low vowel identification scores were found for the vowel /i/ at H1-H2 levels 1 and 5 in both stimulus sets with variable lengths (dataset H1-H2-VL1, H1H2-VL2) and constant lengths (dataset H1H2-CL) failed to support the speculation that the poor vowel identification scores with the vowel /i/ might be related to the length of the vowel segment presented.

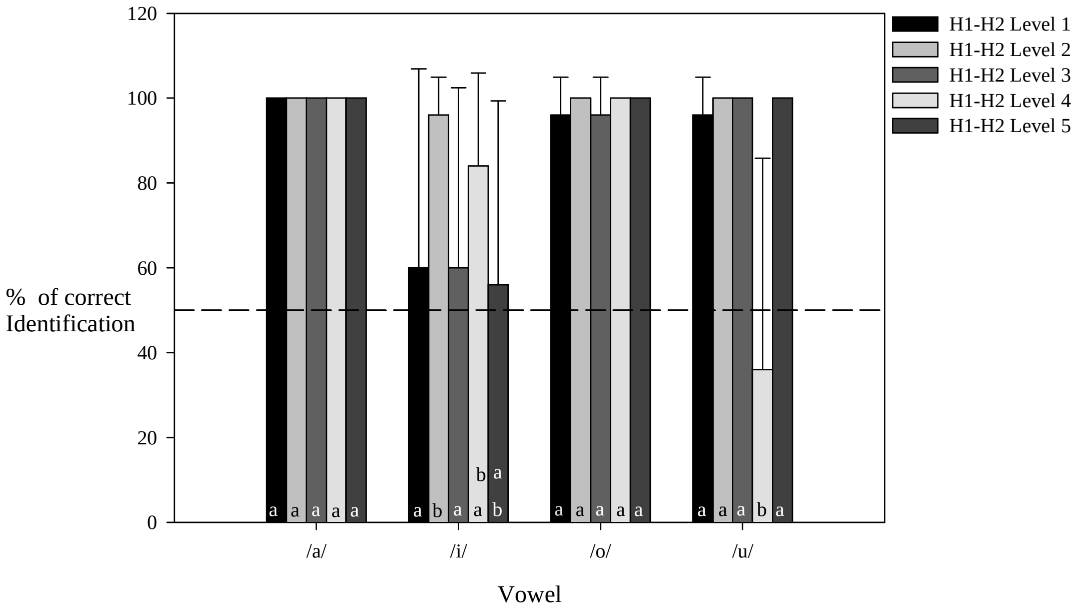


Figure 11. Vowel identification scores for stimuli with constant lengths and different H1-H2 levels (dataset H1H2-CL). (Note: The lower the H1-H2 level, the more prominent the H1 energy and thus the more breathy or thinner the associated voice quality). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

Results for the two-way (vowel effect by level effect) ANOVA on the vowel identification scores for the vowel segments with constant lengths and different SPR levels (dataset SPR-CL) revealed no significant SPR level effect [F (4, 16) = 0.884, p = 0.518] but a significant vowel effect [F (3, 12) = 4.093, p = 0.032] and vowel by SPR level interaction effect [F (12, 48) = 2.091, p = 0.036]. As shown in Figure 12, the vowel identification scores for /a/ and /o/ were either always correct or close to 100% correct, while those for /u/ were low only for SPR level 3, with vowel identification for segment at the rest of the levels being all correct. The vowel identification scores for /i/ generally show a trend of increasing with the SPR level (i.e., higher energy at 2-4 kHz and thus more voice projection power), apart from the score for SPL level 4, which is the lowest. These vowel identification scores are generally supportive of the hypothesis that an increased SPR level will result in clearer exemplars, leading to more accurate identification.

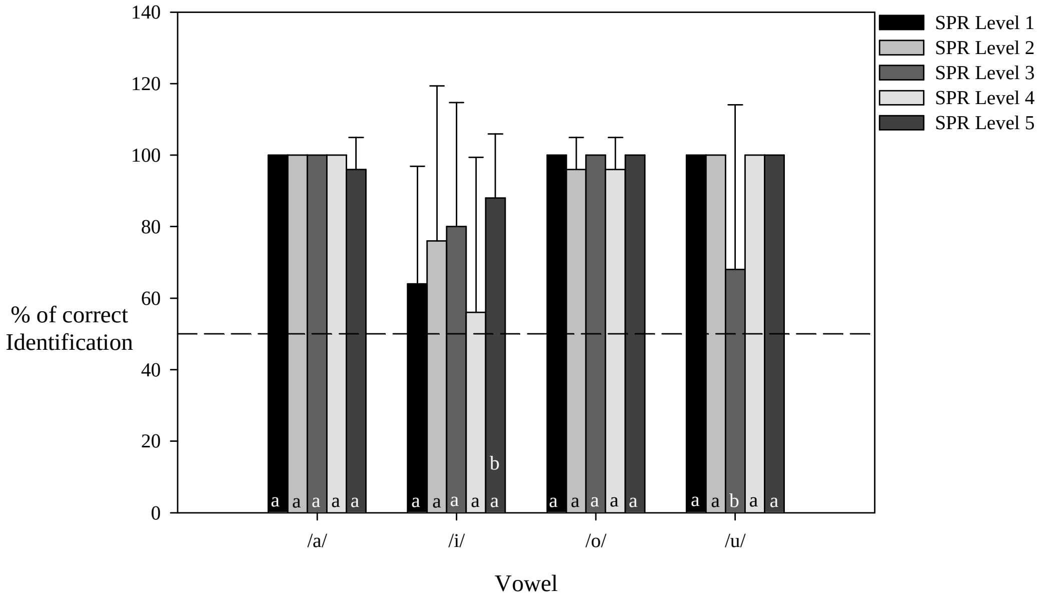


Figure 12. Correct identification scores for stimuli with constant lengths and different SPR levels (dataset SPR-CL). (Note: The lower the SPR level, the lower the energy around 2-4 kHz and thus presumably the lower the voice projection power).

5.2.2.2 Vowel Clarity for Stimuli of Constant Lengths

Results from the two-way (vowel by H1-H2 level) ANOVA on the scores for the differentiation of the clarity of the vowel segments with constant lengths, organised by H1-H2 level (dataset H1H2-CL), revealed a significant H1-H2 level effect [F (4, 16) = 5.868, p = 0.004], vowel effect [F (3, 12) = 5.255, p = 0.015], and vowel by level interaction effect [F (12, 48) = 9.182, p < 0.001]. As shown in Figure 13, there was a clear trend of increased vowel clarity scores with increased H1H2 levels for the vowel /o/. The only significant between-level difference for the vowel /o/ was between H1-H2 levels 1 and 5. While no clear trend is present across vowels, it can be observed that H1-H2 level 5 is associated with the highest vowel clarity score for the vowels /i/, /o/, and /u/, which is generally as per the hypothesis.

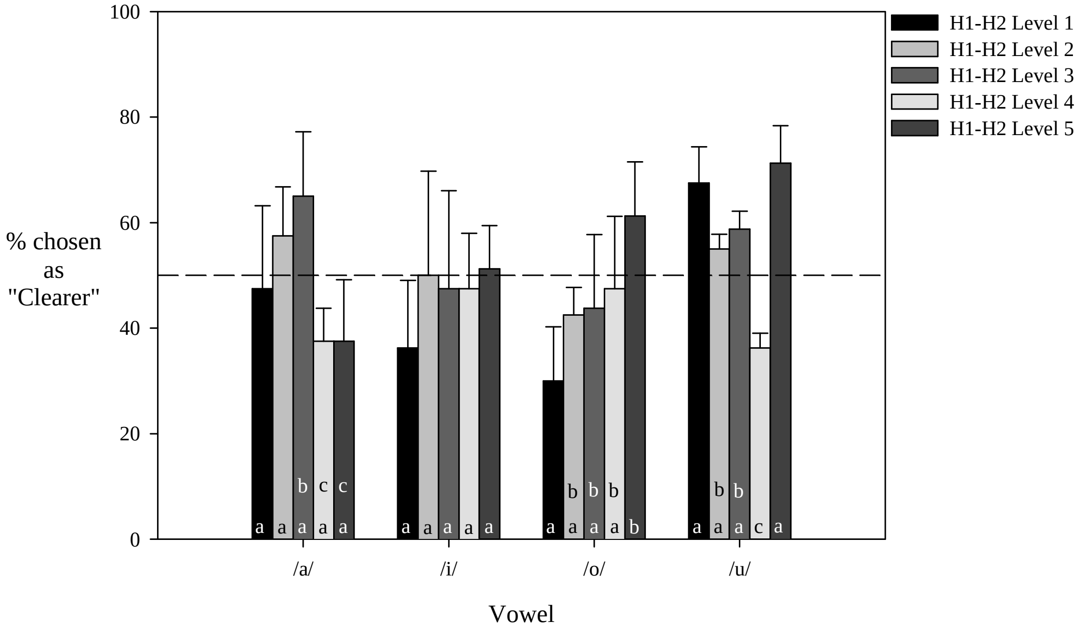


Figure 13. Clarity scores for stimuli with constant lengths and different H1-H2 levels (dataset H1H2-CL). (Note: The lower the H1-H2 level, the more prominent the H1 energy and thus the more breathy or thinner the associated voice quality). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

Results for the two-way (vowel by SPR level) ANOVA performed on the vowel clarity scores for stimuli with constant lengths and different SPR levels (dataset SPR-CL) revealed a significant level effect [F (4, 16) = 76.375, p < 0.001], vowel effect [F (3, 12) = 9.357, p = 0.002], and vowel by level interaction effect [F (12, 48) = 4.340, p < 0.001]. As shown in Figure 14, a very clear trend can be seen in which the vowel clarity score for level 5 is always clearly the highest of all, generally following the overall pattern suggested by the hypothesis.

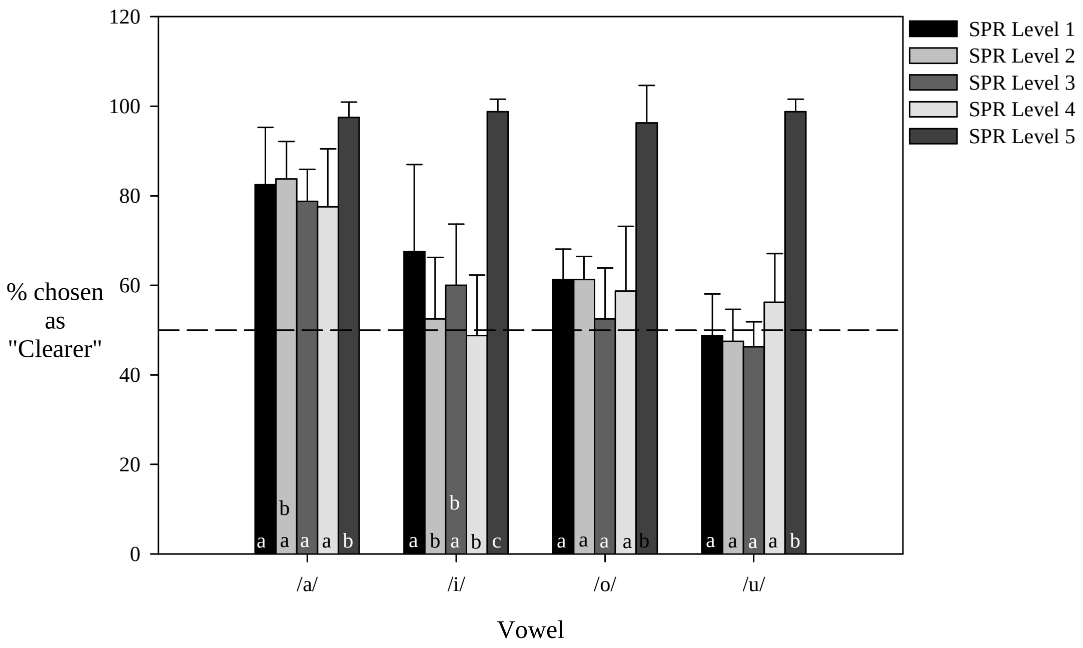


Figure 14. Vowel clarity scores for stimuli with constant lengths and different SPR levels (data set SPR-CL). (Note: The lower the SPR level, the lower the energy around 2-4 kHz and thus presumably the lower the voice projection power). Significantly different between-level comparisons within each vowel set are marked with different letters. No comparison is being made between vowels.

5.3 Summary of Main Findings

The main findings from the acoustic studies were:

1. The acoustic measures that are sensitive to the effect of radio transmission included: F1 and F2 frequencies and vowel space, H1-H2, SPR, and Spectral Moments One (mean) and Two (variance).
2. The measure of CV energy ratio is not sensitive to the radio transmission effect.

Specific findings for each acoustic measures are summarized as follows.

1. **F1 and F2 frequencies and vowel space area:**
    a. The radio-transmitted signals, as compared with the Original, were characterized by an upward shift of F1 frequency for the vowel /i/.
    b. An upward shift in F2 frequency for the vowel /a/, as well as an upward shift in F1 frequency for the vowel /u/, may be a sign of poorer radio quality.
    c. The compression of vowel space area may be a sign of signal degeneration due to radio transmission.
2. **H1-H2:** As compared with the original signals, radio-transmitted vowels are found to show an increase of H1-H2, which is associated with a more breathy or thinner voice quality, for the vowel /i/ and a decrease of H1-H2 for the vowels /a/ and /o/.
3. **SPR:** An increase in SPR, which indicates a lower energy level around the 2-4 kHz, is shown to be associated with radio-transmitted signals compared to the original signals. However, the direction of H1-H2 change due to radio transmission varied by vowel.

4. **Spectral Moment One (mean):** The differentiation between /s/ and /sh/ based on the Spectral Moment One values (i.e., higher values for /s/) may be lost after radio transmission.
5. **Spectral Moment Two (variance):** Although radio-transmitted signals, as compared with the original signals, show a general decrease in the measure of Spectral Moment Two, the differentiation between /s/ and /sh/ in this measure (i.e. higher values for /s/) appears to be less susceptible to signal degeneration related to radio transmission and thus more likely to be maintained.

The main findings from the perceptual studies were:

1. Correct identification of the vowel /i/ seems to be most difficult at least in the framework of this study.
2. No clear trend linking H1-H2 level to the percentage of correct vowel identification even if the vowel length was controlled for.
3. There is a tendency for the vowels to be perceived as “clearer”, most likely when a threshold is reached, with the H1-H2 level moving toward the end associated with a less prominent H1 or the SPR level moving toward the end associated with a stronger energy around the 2-4 kHz region.

6. Discussion

The aim of the first part of this study was to examine the effect of radio transmission on a selection of speech and voice-related acoustic measures, including vowel space, F1 and F2 frequencies and vowel space, H1-H2, SPR, Spectral Moments One and Two, and the CV energy ratio. Results from the acoustic study have shown that all of the selected measures except for CV energy were sensitive to the effect of radio transmission. The second part of the study aimed to examine if the perception of radio-transmitted speech, in terms of vowel intelligibility and voice clarity , was sensitive to changes in the H1-H2 or SPR level. This chapter includes a discussion on the acoustic and perceptual results in relation to the research questions and hypothesis and in comparison to previous research. It will also discuss the limitations of this study and directions for future research.

6.1 Acoustic Study

The vowel space area has been shown to be related to speech intelligibility based on the previous finding that with a larger vowel space, the speech token is perceived as being more intelligible (Bradlow et al., 1996). Liu et al. (2005) has shown that the vowel space for persons with cerebral palsy tended to be smaller and more centralised, with alterations to the F1 and F2 frequencies for all the three vowels examined: /i/, /a/, and /u/. A similar pattern of centralisation and reduction of vowel space was reported by McRae et al. (2002) when they studied the pattern of vowel space in speech produced by speakers with Parkinson’s disease. They also reported that a reduction in vowel space was found to be related to an increased perception of severity of speech impairment.

The present study showed that radio-transmitted signals had significantly shifted and reduced vowel space triangles, rather than being centralised. In particular, one corner vowel, /a/, unlike the other vowels, /i/, /u/, and /o/, did not show much change in position, albeit with a slight upward shift of F2 frequency. Despite the same change in the size of the vowel space, the pattern of alteration of shape and size of the vowel space was quite different from that produced by speakers with cerebral palsy or Parkinson's disease. More specifically, the effect of radio transmission was more evident for /i/ and /u/ and not so pronounced for /a/. The relatively higher degree of frequency shifts of the two high vowels, /i/ and /u/, resulted in transposed vowel triangles rather than centralised ones, as shown in Figure 1. Transmission through the radio systems was affecting the shape of the vowel space more extensively than the two pathological speech patterns.

To gauge the impact of the transmission-induced changes to the vowel space on speech intelligibility, the effect of radio transmission on the percentage change of the space was compared with other studies. While studying the effect of stuttering on the vowel space, Blomgren et al, (1998) reported that the mean vowel space area for treated stuttering was 87% of the mean of the area for control group and the mean of the area for the untreated stutterers was 79%. A comparison of the area of vowel space produced by speakers of American English and Mandarin Accented English revealed that the average area produced by the first group was larger by approximately 30% (Robb & Chen, 2008). In comparison, the present study found that compared to the area of the original recording, the vowel space area was reduced to 22% of the original for Radio 1 and 75% for Radio 2. It can be concluded that vowel space area is a measure sensitive to the effect of radio transmission through the two radio systems examined.

This study did not examine the perceptual correlate of the reduced and transposed vowel triangle. However, while studying the relation between vowel space and intelligibility, Bradlow et al. (1996) found a significant correlation between intelligibility and the range of F1 frequency. It was also found that alterations to F2 frequency had a lower impact on intelligibility. However, appropriate F2 information was considered important in the perception of clarity of speech. This was somewhat different than what was found in a study by Monsen and Shaughnessy (1978). In an attempt to improve speech intelligibility, vowels produced by hearing impaired children were included as stimuli. The participants in their study were taught articulation techniques in producing /a/, /i/, and /u/ correctly. After training had improved clarity, the F1 value of the speech that they were producing had not altered, but the F2 value had increased by between 328 and 520 Hz which brought it closer to the range for normal speech. Given that the F1 values for /i/ and /u/ were dramatically altered, it is conceivable that the vowel clarity scores, if based on vowel space, would have shown a pattern of increasing with smaller alterations to it.

On the other hand, previous studies have reported that the vowel space on its own is not a strong indicator of intelligibility. Neel (2008) reported that on its own the vowel space is able to predict a variance in the identification scores between 9% and 12%. Similar findings about the limitation of using the vowel space as a measure of intelligibility were also reported by Hillenbrand et al., (1995).

While relation between the largest harmonic peaks in the 0-2 and 2-4 kHz regions is sensitive to transmission through the systems, the results also indicate that H1-H2 is being affected by transmission. Previous studies have indicated that as the ratio of the first harmonic to the second harmonic increases, the perception of breathiness increases (Hillenbrand et al., 1994; Klatt & Klatt, 1990). As the quality of speech resulting from transmission had altered substantially from the original, it was hypothesised that the H1-H2 ratio would be sensitive to the effects of transmission, with the transmitted speech having increased H1-H2. We found that transmission through the radio systems had resulted in the increase of H1-H2 (i.e. decreased for H2-H1, as shown in Figure 2). The differences were more marked for /a/, /i/ and /u/ than for /o/. The perceptual correlates of H1-H2 ratio are discussed later.

Higher energy in the largest harmonic peak in the 2-4 kHz range as compared largest harmonic peak in the 0-2 kHz range, resulting in a higher SPR, has been shown to correlate to a voice having a more resonant singing quality (Omori et al., 1996). As the quality of speech resulting from transmission had altered substantially from the original, it was hypothesised that the SPR would be sensitive to the effects of transmission. The study by Omori et al. (1996) had examined the quality of /a/, and found that a higher SPR corresponded to voice having a more "ringing" quality to it. Our study found that the SPR of the original radio was lower for all the four vowels studied. This seems to indicate that the radio systems were either suppressing output in the 0-2k kHz region or were boosting the output in the 2-4 kHz region, thus causing a change to the SPR. Further, both the radio systems studied had similar effects on the SPR for three vowels except for /i/ where Radio2 produced a lower SPR that Radio 2. Previous research has also shown that the SPR can be used to judge the talent of a singer. In the study by Watts et al. (2006) the talented speakers were those who produced a voice with a better voice quality. The talented singers were found to produce lower SPR's as compared to singers who were judged as untalented (Watts et al., 2006). This seems to indicate that the speech transmitted by through the radios had a poorer voice quality than the original. However, it should be noted that a study by Kenny and Mitchell (2006) reported that they did not find a correlation between perception of singing quality and the SPR, when two different styles of singing were examined.

The spectral moments can be used as a quantitative means of studying how the energy distribution in the spectrum has been altered, as the mean and variance give an indication of the concentration of energy and its tilt (Forrest et al., 1988). This study also found that the first and second spectral moments were sensitive to transmission through the radios, and were affected by both the radio systems. The first and second spectral moment have been shown to be important cues in identifying place of articulation of fricatives (Jongman et al., 2000), and also a cue to the classification of stop-consonants (Forrest et al., 1988). As these measures are affected it is probable that perceptual studies would have indicated that both the radios had affected perception of fricatives and stop consonants, as both radio systems had similar effects on the mean and variance.

This study had hypothesised that the CV energy ratio would be affected due to radio transmission. However the results indicated that the CV energy ratio was not sensitive to transmission through the radio systems. The study by Hedrick and Ohde (1993) had found that the CV ratio was important in judging the place of articulation. It is likely that the radio systems affected the amplitude of both the consonants and the vowels in similar measures, leading to unaltered ratios. As this study did not carry out further perceptual study of consonant identification after transmission but based on the findings by Hendrick and Ohde (1993) it is likely that the consonant identification would have not been affected by transmission. This is further supported by findings by Hazran and Markham (2004) who reported that the CV ratio was not correlated to intelligibility for stop consonants.

6.2 Perceptual Study

The perceptual study was organised into two parts using different vowel tokens and selection criteria. In the first study, the entire vowel was used in both the identification task and also the clarity judgement task. The vowel durations varied from 76 msec to 283 msec, (mean = 189 msec, SD = 54 msec). Two sets of vowels were selected on this basis. In the second part of the perceptual study, duration of the vowel was controlled, with a 78 msec long segment being extracted from the middle of the vowel, which was close to its steady state (Blomgren & Robb, 1998). The first study was based on tokens selected on the basis of H1-H2 ratio measured from the vowel, and the second study, one set was based on the H1-H2 ratio, and the second set based on the SPR, both measured for the extracted vowel segment.

It was hypothesised that if the tokens were arranged in order of H1-H2 in the first study and H1-H2 and in the second study, they would be scored by the participants, in relation to this order, with the tokens ranked highest getting the highest scores for both identification and clarity while the lowest ranked would get the lowest scores. The first part of the perceptual study conducted the identification and perception of clarity task twice each. The first and second set was chosen on exactly the same basis and had similar H1-H2s forming level 1 to 5. It was expected that the results for both would yield similar patterns. The second study used one set based on H1-H2 levels and the second set based on levels, and it was expected that the results of the perceptual study would correspond to the levels of the H1-H2 and levels.

Since very high or a 100 % correct identification scores for /a/, /o/, and /u/ were obtained, no observable trends were noticed. This also indicates that although the frequency composition was altered due to radio transmission, as seen in the effect on the acoustic measures, the signal was not altered sufficiently to affect the identification of /a/, /o/ and /u/, which were almost always identified correctly.

The scores for identification of /i/ in the first study were not as consistently high as the other three vowels for both the sets, and here, using variable vowel lengths, the results obtained were not in keeping with the hypothesis that the scores would increase with the H1-H2 ranking. In an attempt to understand the reason for this variation from the expected outcome, statistical studies were carried out on the scores. The study indicated that the scores obtained for identification of /i/ in the test where the vowel length varied, was correlated to the duration of the vowel. The tokens with longer durations were being perceived as clearer. There have been several studies that have found that there is a relationship between duration of speech and perceived clarity. When speech is deliberately produced to be “clear” as against “conversational”, the duration of clear speech was found to be twice as much as the conversational token (Picheny et al., 1985). A second study by Picheny, Durlach, and Braida (1989) found that this is achieved by lengthening the duration segments in the speech as well as the pauses introduced and also by production of modified acoustic characteristics. However, simply extending the length of the recording of conversational speech did not render it clear but, rather, reduces the intelligibility (Uchanski et al., 1996). In the first part of our study, this relation between longer duration and a perception of clarity seems to have been more influential in the intelligibility and clarity score than the ranking based on H1-H2 levels.

In this study, it was also noted that the F1 frequency for /i/ was affected quite severely, as shown in Figure 1. As discussed earlier, Bradlow et al. (1996) have reported the relation between F1 frequency and identification accuracy, with identification being affected by changes in the F1 frequency. A similar result can be seen in the case of the identification of /i/. However, the identification scores for/u/ did not follow this pattern in spite of its F1 being altered in an almost similar manner. It is likely that the identification of /i/ was confused with /e/, and no such comparable sound was available for possible confusion for /a/, /o/, and /u/. So it is likely that in spite of not being clear, vowels with a limited range of possibilities to be confused with would tend to show a higher vowel identification score.

In the task in judging the clarity of the vowels, the results did not follow the hypothesised pattern of highest ranked H1-H2 scores being judged as clearest and lowest ranked, as least clear. In this part of the study, little correlation was found between the vowel clarity scores and the duration of vowel. The score for /u/ was also found to be correlated with duration in one of the tasks involving the judgment of clarity. Further, the two sets of scores obtained of the judgment of clarity of each of the vowels, which were chosen on the basis of H1-H2, did not yield similar trends either. Previous research (Hillenbrand et al., 1994; Klatt & Klatt, 1990) has shown that when the H1 is increased, the voice is perceived as being more breathy. Their studies also found that the level of aspiration noise in the token was the factor that was most influential in the judgment of a voice being breathy voices. The findings of our study seem to suggest that in vowels where duration is not controlled, the H1-H2 does not affect the perception of clarity. It was further hypothesised that if the duration of the vowel was controlled, and the H1-H2 level used to rank the tokens, the scores obtained for identification and clarity would correlate to the ranking. It was found that the scores did not follow the rankings and the results of our study seem to indicate that H1-H2 did not correlate to clarity. While H1-H2 has a role to play in the perception of breathy voices, as discussed earlier, it does not necessarily mean that all breathy voices perceived as being unclear or unintelligible. While breathy voices in

English are associated with a pathological condition, in other languages breathiness is part of the phonation, as in the case of Gujerati (Ladefoged, 1983). Also, a breathy voice may be part of normal speech, as female voices are generally more breathy than male voices, and there are occasions in social interaction, when females deliberately make their voices more breathy (Henton & Bladon, 1985). So, while H1-H2 plays a role in the perception of a breathy voice, it does not seem to have a role in the perception of clarity of speech.

In the task of identification, based on vowels ranking based on the SPR scores for /a/, /o/ and /u/ were very high or almost always correct, and the influence of ranking by SPR on the scores was not apparent. The score for /i/ was more varied and although a clear trend linking the scores obtained for identification with their ranking based on SPR level was not obtained, four out five scores followed the hypothesised trend. However the score for clarity of the vowels followed a clearer pattern, with the score for the highest level being substantially better than the adjacent scores in all the four vowels tested and supported the hypothesis.

The SPR was devised by Omori et al. (1996) as a means to evaluate the quality of a singing voice. They found that the SPR could be representative of the 'ringing voice quality' (pg. 235), that let a singers voice carry to the listener in spite of the presence of other competing sounds such as the sounds of musical instruments. Our study seems to suggest that the speech signal with an SPR indicating greater energy in the 2-4 kHz region was able to carry over the background noise caused by transmission and thus was judged as being the clearest. The manner of improvement of the vowel identification and clarity scores as a function of SPR was not gradual suggesting that the existence of a threshold at which dramatic improvements in clarity occurred.

6.3 Practical Applications

In personal FM systems used in audiological applications, the acoustic signal is converted to an electrical representation of the sound system and transmitted via radio signals to the listener, with the intention of increasing the signal to noise ratio (Smaldino, Crandell, Kreisman, John, & Kreisman, 2009). The audiological assessment of these FM systems includes measures such as the gain, output, harmonic distortions, produced by these systems (Smaldino et al., 2009) but these measures do not assess the perceptual sound quality. The quality of the reproduced sound during digital process of coding and decoding sound for transmission is assessed based on the 'bit' rate used, the delay that is built up, the complexity of the signal and its quality (Painter & Spanias, 2000). Although the technology to transmit signals with 'lossless' quality exists, i.e. where almost exact replication of the signal is achieved, achieving this lossless transmission requires a far higher coding and processing capacity than the commonly used algorithms in coding and transmission can achieve which results in a loss of quality (Painter & Spanias, 2000). To assess the quality of processed speech, subjective tests which are both expensive and time consuming are used (Painter & Spanias, 2000). The measure for speech intelligibility uses the Perceptual Evaluation of Speech Quality (PESQ), which is established by the International Telecommunications Union, ITU P. 862 standard, as a means of evaluation of speech being transmitted via communication channels (Di Persia et al., 2008). The results in this study suggest that F1 and F2 frequencies and vowel space, H1-H2, SPR, and Spectral Moments One and Two were sensitive to the radio transmission effect and thus could be developed into supplementary quality measures of transmitted speech.

6.4 Limitations and Future Directions

Although this study has shown that a selection of acoustic measures, including F1 and F2 frequencies and vowel space area, H1-H2, SPR, and Spectral Moments One and Two, to be sensitive to the effect of radio transmission and may be useful for assessing the quality of radio-transmitted speech, there are a few limitations in the study that need to be highlighted. Firstly, the word list used was not specifically designed or selected for acoustic analysis. The list did not contain an equal number of samples of the four vowels examined, nor were there a planned or number of consonants being examined in the study. The vowels were not consistently placed in relation to the same consonant, and this could have had an impact on the acoustic study, as it has been reported that the consonant adjacent to the vowel has an effect the formant transition and the identification of the vowel (Lindblom & Studdert-Kennedy, 1967). Further the list of words containing the target vowels and consonants was small, with only 23 tokens for all four vowels. Before the results of the acoustic analysis can be generalised, a similar study with a larger and more inclusive list of words specifically designed for analysis should be used. For example, to study fricatives, Jongman et al. (2000) used a list comprising eight fricatives, with 6 vowels, to form 48 CVC words which were then examined, allowing for a more controlled study, which will allow for a more acceptable generalisation.

This study examined several acoustic parameters and used the H1-H2 and SPR as the basis for organising the perceptual study. Given that the vowel space and the mean and variance of the spectral moments (i.e., Spectral Moments One and Two) were also sensitive to radio transmission, acoustic measures obtained for these could have also been used to rank the data for perceptual analysis of identification and clarity. Future studies could restrict themselves to one of these acoustic parameters, which when examined using a specifically designed word list to allow for a more controlled study and thus a more acceptable generalisation.

The perceptual study also had a few limitations. In the first study, where the vowel duration was not controlled, the H1-H2 measure was obtained from vowels embedded in different words. As a result, the influence of the consonant on the vowel was not controlled. As identification of vowels is influenced by the adjacent consonant and formant transitions (Lindblom & Studdert-Kennedy, 1967), this effect was not controlled. The cohort size for the perceptual study, with vowel duration not controlled, was small, with twenty participants. The cohort size for the perceptual study with vowel duration controlled was smaller still, with five participants. This means that before the results of the perceptual study can be generalised, a follow-up study with a larger sample size should be conducted.

In this study, the identification and clarity judgment tasks involved only vowels. If future studies are conducted using synthesised words or designed word lists for the task, it will allow for a more specific targeting of vowels and consonants being studied, and so allow for greater generalisation of the results. Not only will use of better designed word lists be required for generalisation but also examining the output of a range of different radio systems, transmitting devices and sound-reproducing devices should be undertaken, so that the effect of transmission can be studied in several systems, prior to generalisation.

The studies of SPR in the literature have generally concentrated on the sung voice and the perception of singing quality. In this study, SPR was employed as a means to measure the clarity of speech. Before the relation of SPR to speech clarity that is suggested in this study is generalised, a larger and more comprehensive study needs to be undertaken. This study used only a small cohort for this part of the study, and, as discussed earlier, the speech samples were not designed specifically to ascertain the relation of SPR to clarity. A follow-up study to this study could involve the systematic manipulation of the SPR to verify its relation to speech clarity, especially for establishing if there was a particular threshold beyond which the scores for clarity would show marked improvements.

6.5 Conclusion

In this study, the effects of radio transmission on a selection of speech and voice-related acoustic measures were tested. The impact of H1-H2 and SPR on the perception of vowel identification and differentiation of vowel clarity has also been explored. The study found that measures of F1 and F2 frequencies, vowel space area, H1-H2, SPR, Spectral Moment One, and Spectral Moment Two were sensitive to radio transmission. For vowels, changes of F1 and F2 frequencies resulting in the compression of vowel space, increased H1 dominance (as reflected in H1-H2), and decreased spectral energy around the 2-4 kHz region (as indicated by SPR) are signs of signal degeneration that can be induced by radio transmission. For consonants, the loss of differentiation in Spectral Moment One (mean) between /s/ and /sh/ may result from radio transmission while consonant differentiation cues provided by Spectral Moment Two (variance) may be more resistant to the transmission effect. As for the relationship between these acoustic parameters and the perception of speech clarity, this study did not find a consistent effect of H1-H2 on the perception of speech clarity.

However, a clearer trend was identified suggesting that SPR was related to the perception of speech clarity. Further studies with a better control and more complete inclusion of the phonetic contexts of the segment and a large sample size are needed to allow for a better generalization of the study findings in identifying the key acoustic parameters in testing the quality of a speech transmission output.

References

Allen, J. (2005). Consonant recognition and the articulation index. *Journal of the Acoustical Society of America, 117*(4 I), 2212-2223.

Bertoni, H. L. (2000). *Radio propagation for modern wireless systems*. Upper Saddle River: Prentice Hall PTR.

Biddulph, D. (Ed.). (1994). *Radio communication handbook* (6 ed.). Herts: Radio society of Great Britan.

Bladon, R. A. W., & Lindblom, B. (1981). Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America, 69(5),* 1414-1422.

Blomgren, M., & Robb, M. (1998). How steady are vowel steady-states? *Clinical Linguistics and Phonetics, 12(5),* 405-415.

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.

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.

Brandy, W. T. (2002). Speech Audiometry. In J. Katz (Ed.), *Handbook of clinical audiology*. Philadelphia: Lippincott Williams & Wilkins.

Coleman, C. (2004). *An introduction to radio frequency engineering*. Cambridge: Cambridge university press.

Cooper, F. S. (1980). Acoustics in human communication: Evolving ideas about the nature of speech. *Journal of the Acoustical Society of America, 68(1),* 18-21.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Intelligibility of average talkers in typical listening environments. *Journal of the Acoustical Society of America, 81(5),* 1598-1608.

Crandell, C. C., & Smaldino, J. (2002). Room acoustics and auditory rehabilitation technology. In J. Katz (Ed.), *Handbook of clinical audiology*. Philadelphia: Lippencott Williams and Wilkins.

Davenport, M., & Hannahs, S. J. (1998). *Introducing phonetics & phonology*. London: Arnold.

Di Persia, L., Milone, D., Rufiner, H. L., & Yanagida, M. (2008). Perceptual evaluation of blind source separation for robust speech recognition. *Signal Processing, 88(10),* 2578-2583.

Dillon, H. (2001). *Hearing Aids*. Sydney: Boomerang Press.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84(1),* 115-123.

French, N. R., & Steinberg, J. C. (1946). Factors governing the intelligibility of speech sounds. *The journal of the acoustical society of America, 19(1),* 90-119.

Fry, D. B. (1979). *The physics of speech*. Cambridge Cambridge university press.

Haahr, M. (2009). Random.org, from **http://www.random.org/sequences/**

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America, 116(5),* 3108-3118.

Hedrick, M. S., & Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America, 94(4),* 2005-2026.

Henton, C. G., & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication, 5(3),* 221-227.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, and Psychophysics, 71(5),* 1150-1166.

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

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American english vowels. *Journal of the Acoustical Society of America, 97(5 I),* 3099-3111.

Houtgast, T., & Steeneken, H. J. M. (1973). Modulation transfer function in room acoustics as a predictor of speech intelligiblity. *Acustica, 28(1),* 66-73.

Hu, Y., & Loizou, P. C. (2010). On the importance of preserving the harmonics and neighboring partials prior to vocoder processing: Implications for cochlear implants. *Journal of the Acoustical Society of America, 127(1),* 427-434.

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.

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.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. Bless & A. J (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 351 -360). San Diego: College Hill press.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21(1),* 1-36.

Lindblom, B. E., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America, 42(4),* 830-843.

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.

McRae, P. A., Tjaden, K., & Schoonings, B. (2002). Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 45(1),* 35-50.

Meyer, T. A., & Pisoni, D. B. (1999). Some computational analyses of the PBK test: Effects of frequency and lexical density on spoken word recognition. *Ear and hearing, 20(4),* 363-371.

Miller, G. A. (1981). *Language and speech*. San Francisco: W. H. Freeman and company.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41(5),* 329-335.

Ministry of Economic Development. (2009). Radio spectrum management, from **http://www.rsm.govt.nz/cms**

Monsen, R. B., & Shaughnessy, D. H. (1978). Improvement in vowel articulation of deaf children. *Journal of Communication Disorders, 11(5),* 417-424.

Moore, B. C. J. (2008). Basic auditory processes involved in the analysis of speech sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1493),* 947-963.

Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research, 51(3),* 574-585.

Northern, J. L., & Downs, M. P. (2002). *Hearing in children*. Philadelphia: Lippencott Williams & Wilkins.

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.

Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE, 88(4),* 451-512.

Parikh, G., & Loizou, P. C. (2005). The influence of noise on vowel and consonant cues. *Journal of the Acoustical Society of America, 118*(6), 3874-3888.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28(1),* 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29(4),* 434-446.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research, 32(3),* 600-603.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research* (3rd ed.). New Jersey: Pearson Prentice Hall.

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.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270(5234),* 303-304.

Smaldino, J., Crandell, C., Kreisman, B., John, A., & Kreisman, N. (2009). Room acoustics and auditory rehabilitation technology. In J. Katz (Ed.), *Handbook of clinical audiology*. Philadelphia: Wolters Kluwer/Lippencott Williams and Wilkins.

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America, 68(3),* 836-842.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research, 39(3),* 494-509.

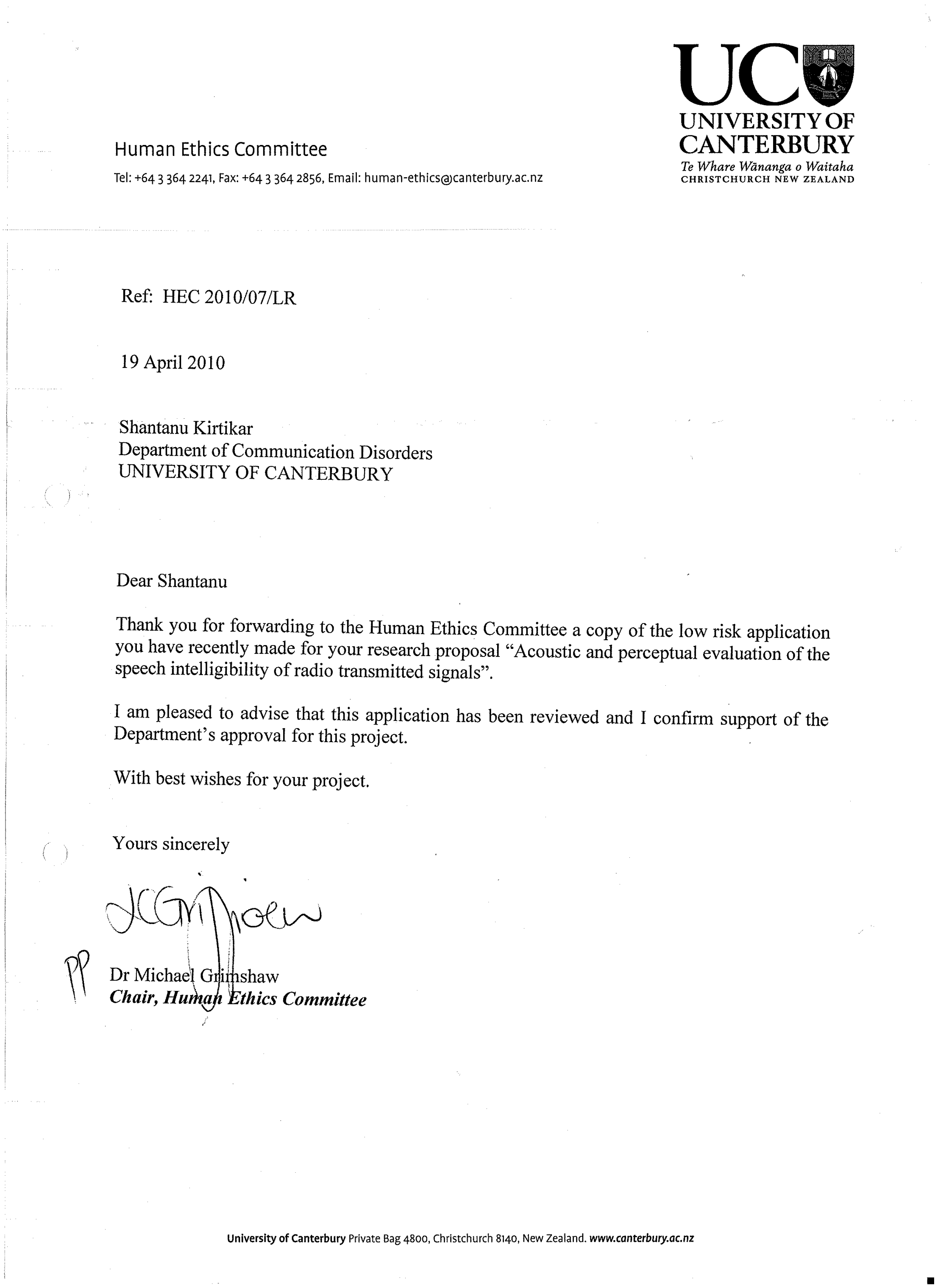
Watts, C., Barnes-Burroughs, K., Estis, J., & Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. *Journal of Voice, 20(1),* 82-88.

Wilber, L. A. (2002). Transducers for audiologic testing. In J. Katz (Ed.), *Handbook of clinical audiology*. Philadelphia: Lippincott Williams & Wilkins.

Yost, W. A. (2007). *Fundamentals of Hearing. An introduction* (5th edition ed.). San Diego: Elsevier.

Appendices

Appendix I. Human ethics approval



UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Human Ethics Committee
Tel: +64 3 364 2241, Fax: +64 3 364 2856, Email: human-ethics@canterbury.ac.nz

Ref: HEC 2010/07/LR

19 April 2010

Shantanu Kirtikar
Department of Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Shantanu

Thank you for forwarding to the Human Ethics Committee a copy of the low risk application you have recently made for your research proposal "Acoustic and perceptual evaluation of the speech intelligibility of radio transmitted signals".

I am pleased to advise that this application has been reviewed and I confirm support of the Department's approval for this project.

With best wishes for your project.

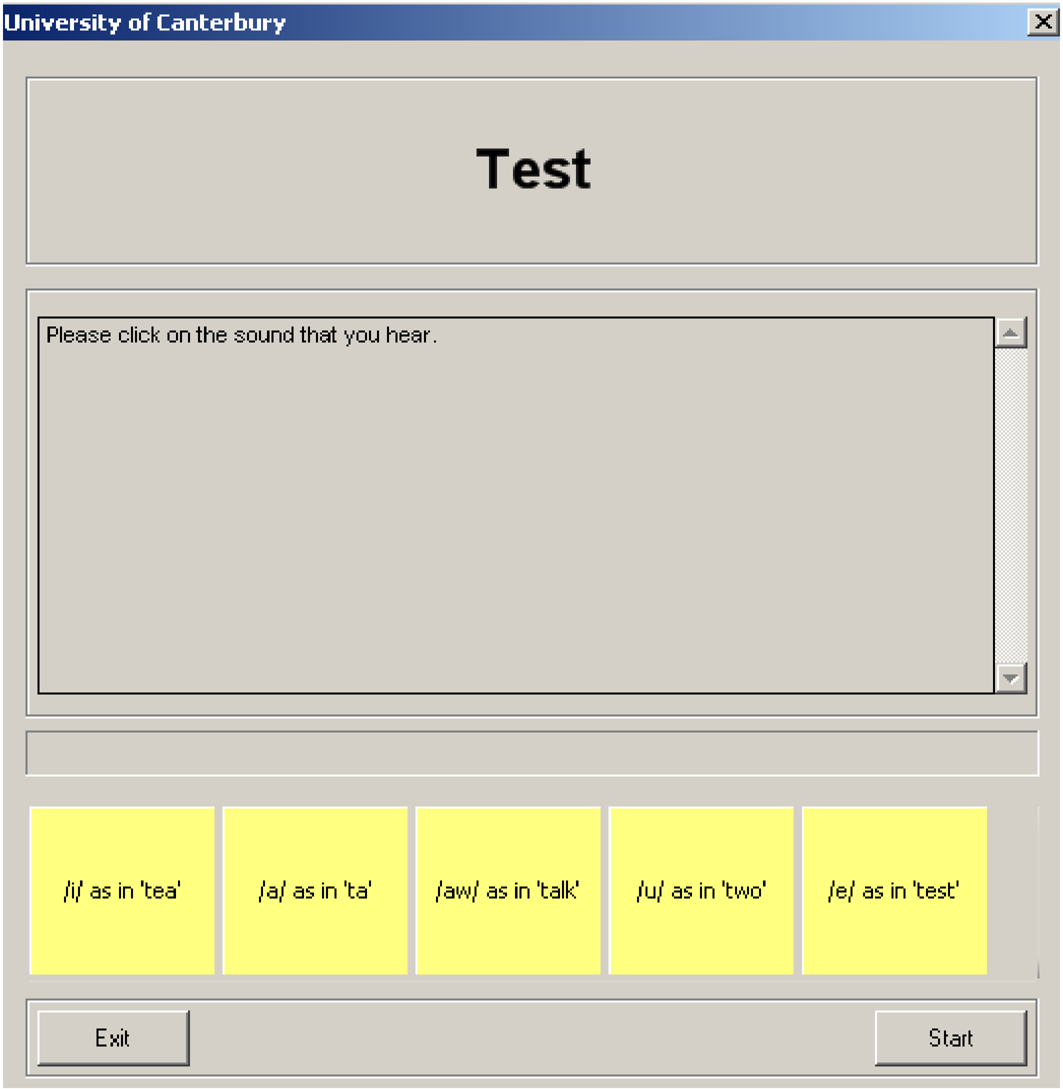
Yours sincerely

pp

Dr Michael Grimshaw
Chair, Human Ethics Committee

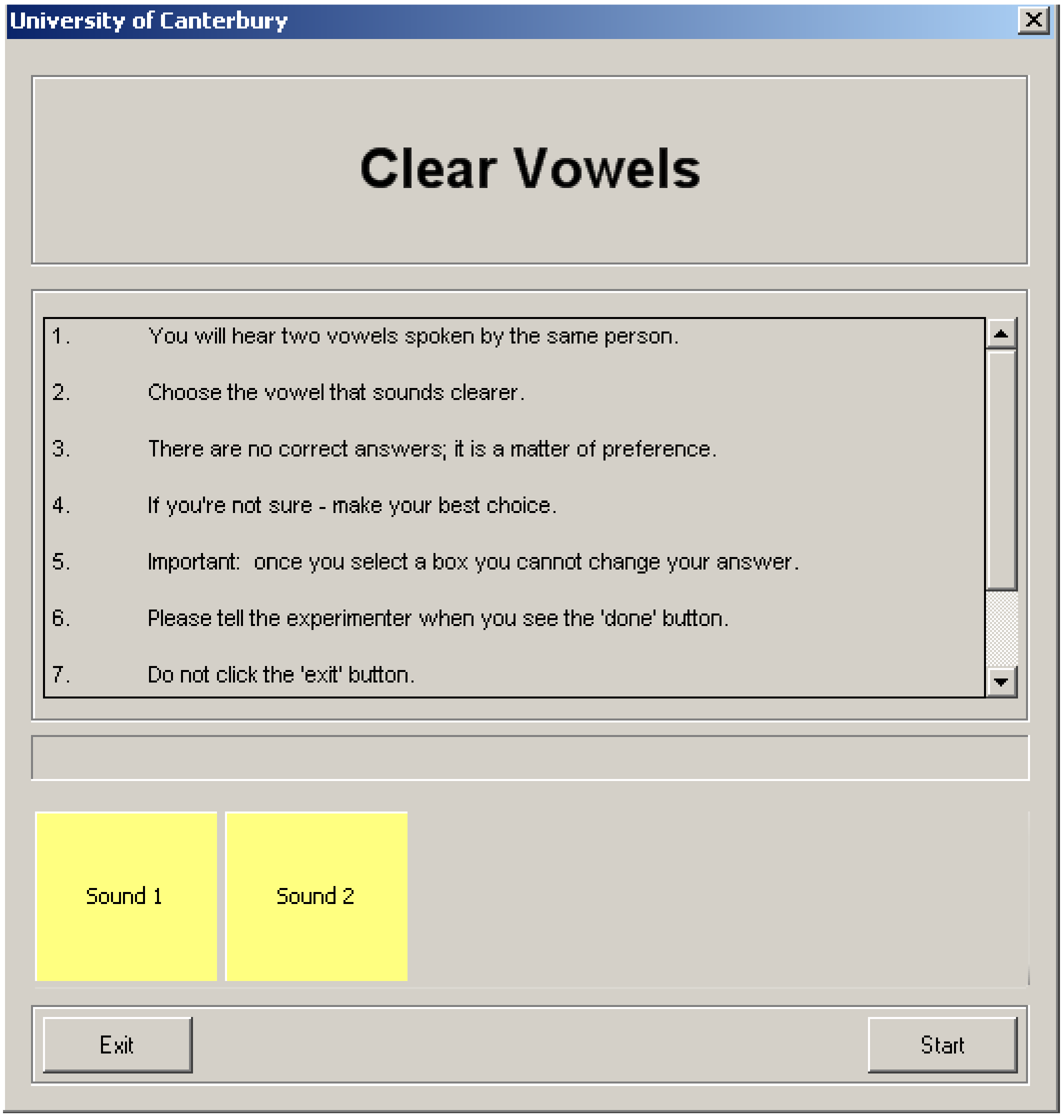
University of Canterbury Private Bag 4800, Christchurch 8140, New Zealand. ***www.canterbury.ac.nz***

Appendix II. Vowel Identification Task



Please click on the sound that you hear.

Appendix III. Vowel Clarity task



Clear Vowels

1. You will hear two vowels spoken by the same person.
2. Choose the vowel that sounds clearer.
3. There are no correct answers; it is a matter of preference.
4. If you're not sure - make your best choice.
5. Important: once you select a box you cannot change your answer.
6. Please tell the experimenter when you see the 'done' button.
7. Do not click the 'exit' button.

Appendix IV. Phonetically Balanced Kindergarten (PBK) lists

List 1

1. Please
2. Great
3. Sled
4. Pants
5. Rat
6. Bad
7. Pinch
8. Such
9. Bus
10. Need
11. Ways
12. Five
13. Mouth
14. Rag
15. Put.
16. Fed
17. Fold
18. Hunt
19. No
20. Box
21. Are
22. Teach
23. Slice
24. Is
25. Tree
26. Smile
27. Bath
28. Slip
29. Ride
30. End
31. Pink
32. Thank
33. Take
34. Cart
35. Scab
36. Lay
37. Class
38. Me
39. Dish
40. Neck
41. Beef
42. Few
43. Use
44. Did
45. Hit
46. Pond
47. Hot
48. Own
49. Bead
50. Shop

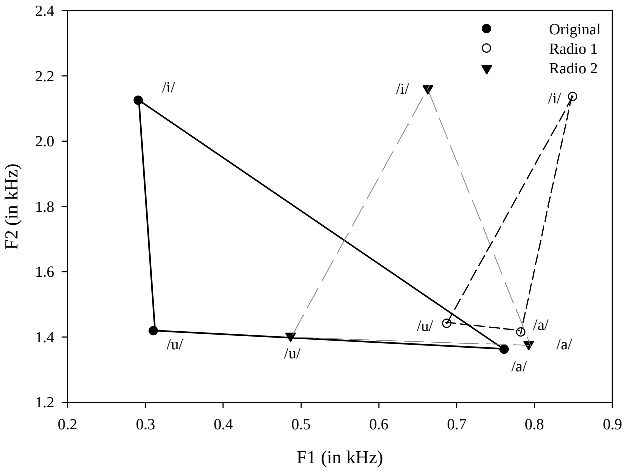
List 2

1. Laugh
2. Falls
3. Paste
4. Plough
5. Page
6. Weed
7. Grey
8. Park
9. Wait
10. Fat
11. Axe
12. Cage
13. Knife
14. Turn
15. Grab
16. Rose
17. Lip
18. Bee
19. Bet
20. His
21. Sing
22. All
23. Bless
24. Suit
25. Splash
26. Path
27. Feed
28. Next
29. Wreck
30. Waste
31. Crab
32. Peg
33. Freeze
34. Race
35. Bud
36. Darn
37. Fair
38. Sack
39. Got
40. As
41. Grew
42. Knee
43. Fresh
44. Tray
45. Cat
46. On
47. Camp
48. Find
49. Yes
50. Loud

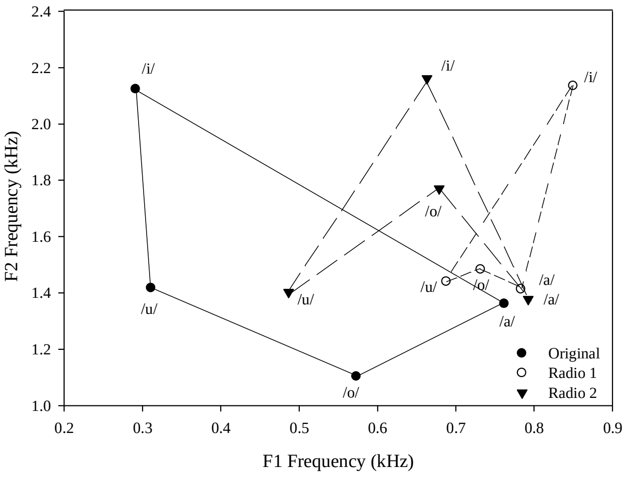
Appendix V. Mean F1 and F2 values for the vowels /i/, /a/, and /u/ in Hz

	Original		**Radio 1**		**Radio3**	
	F1	**F2**	**F1**	**F2**	**F1**	**F2**
/i/	291.6	2123.3	849.7	2134.8	663.2	2161.1
/a/	761.9	1360.9	783.2	1413.2	792.6	1377.2
/u/	311.0	1417.5	687.9	1440.4	486.6	1402.9

Appendix VI. The F1-F2 plot showing the vowel space (enclosed in a triangle) for the vowels /i/ (“beef”), /a/ (“cart”), /ɔ/ (“fall”), and /u/ (“grew”) with the mean F1 and F2 values derived from tokens transmitted through Radios 1 and 2 at 10 playback volume levels and the single measures of F1 and F2 values for original signals. The upper graph shows the vowel space with three vowels (/i/, /a/, and /u/), with the vowel space area shown to be highest for the Original (0.1586 kHz^2), followed in order by Radio 2 (0.1183 kHz^2) and Radio 1 (0.0353 kHz^2). The lower graph shows the vowel space with all of the four corner vowels (/i/, /a/, /u/, and /ɔ/).



a. Upper graph



b. Lower graph