EFFECTS OF THE PROMINENCE OF FIRST HARMONIC ON THE PERCEPTION OF BREATHINESS AND VOWEL IDENTITY

By

Samuel Sloane

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

THESIS SUPERVISORS

PRIMARY: DR. EMILY LIN

ASSOCIATE: PROFESSOR DONAL SINEX

University of Canterbury

2013

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my primary supervisor Emily Lin for all of the time, effort and expert advice she has given me. I would also like to acknowledge my family for supporting me every step of the way.

Abstract

Background: Human communication relies on adequate speech intelligibility to enable the comprehension of verbal messages. Dysphonia (i.e., aberrant voice) may not only result in distraction during communication but also interfere with speech intelligibility leading to a communication barrier. One voice quality commonly found in dysphonia is breathiness, which is related to the presence of excessive airflow during phonation due to incomplete glottal closure. Breathiness has been associated with the prominence of the first harmonic (H1) in the acoustic analysis of voice.

Objectives: This study aimed to determine whether excessiveness in the first harmonic (H1) dominance, which has been associated with breathy voice, may result in the perception of breathiness and compromise vowel intelligibility.

Methods: Participants included 10 female and 10 male normal-hearing adults, aged between 19 to 40 years. Participant's tasks included a "breathiness rating" and a "vowel identification" task. For the "breathiness rating" task, a direct magnitude method was employed for the participant to rate a 500-ms long vowel (/i/ and /a/) segmented from sustained vowel phonation. For the "vowel identification" task, the vowel stimuli were segmented out from running speech ("Rainbow passage") and the participants were asked to listen to one vowel stimulus (/i/, /a/, or /o/; duration: 60 ms) at a time and indicate which vowel (i.e., /i/, /e/, /a/, /o/, or /u/) they perceived the stimulus to be. The vowel stimuli included processed and unprocessed voice recordings of individuals with and without voice disorders. Voices showing the lowest, median, and highest amplitude differences between the first two harmonics (H1-H2) were chosen from a voice database for female and male voices respectively. The 18 selected vowel signals (3 vowels X 3 H1-H2 levels X 2 speaker genders) were processed through 12 signal manipulation conditions. The 12 signal

conditions involved increasing or decreasing the H1 amplitude of the original signals in six 2-dB interval steps in both directions.

Results: For the "breathiness rating" task, the five-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions X 2 listener genders) Mixed Model Analysis of variance (ANOVA) conducted on the breathiness scores for normal speakers and voice patients separately showed significant findings for various main and interaction effects, such as a significant speaker gender by signal condition by vowel interaction effect on the perception of breathiness [$F(12, 96) = 1.95$, $p = 0.038$] for normal voice. An increase of H1-H2 through signal manipulation led to an increase of perceived breathiness only when performed on the vowel /i/ produced by female normal speakers. As for the "vowel identification" task, a relationship between H1-H2 increment and vowel intelligibility was found but the relationship was affected by vowel type, speaker gender, and H1-H2 level. With all vowel types, speaker genders, and H1-H2 levels combined, a significant signal condition effect on the number of incorrect vowel identification was found ($\chi^2 = 188.585$, $df = 10$, $p < 0.001$). Generally, it appeared that an increase of H1-H2 would worsen the identification of /i/ but enhance that of /o/.

Conclusion: The relationship between H1 dominance and perceived breathiness was non-linear. Factors found to disrupt the linear relationship included speaker gender, vowel type, and the extent of H1 dominance. In addition, there was evidence that acoustic manipulation of the H1 amplitude would affect vowel intelligibility and the relationship between vowel intelligibility and H1-H2 values also vary by speaker genders and vowel types.

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Speech intelligibility is vitally important for effective spoken communication and thus receives considerable attention in the realm of communication disorders and sciences. It is commonly observed that voice disorders may result in deterioration of voice quality and compromise speech intelligibility. To monitor changes affecting oral communication, the human speech and voice can both be evaluated using subjective and objective measures. Although there have been many acoustic and perceptual studies of voice quality and speech intelligibility in the literature, the number of studies which directly attend to the impact of voice quality on speech intelligibility is relatively small. Therefore, there remains no clear understanding of the interaction between voice characteristics and speech intelligibility.

To identify the voice-related acoustic changes affecting speech intelligibility, this study manipulates the acoustic features related to breathiness, which is a type of voice quality commonly found in voice patients, and investigates the impact of this acoustic manipulation on the perception of breathiness and vowel intelligibility. Specifically, the effect of different levels of breathiness on vowel intelligibility and perception of voice quality will be examined in this study using both processed and un-processed speech samples. The aims of this study are to measure vowel intelligibility and voice quality perception using acoustically manipulated voice samples taken from both pathological and non-pathological English speakers.

This chapter presents the rationale and an overview of the thesis, the research question and its importance, and the aims and hypotheses.

1.1 Rationale and overview

Breathiness is a common descriptor of pathological voice quality (Kreiman, Gerratt, & Berke, 1994). Breathy voice is associated with glottal incompetence, which is a phonatory condition where vocal folds fail to close along their full length during the closed phase of a vibratory cycle. The excessive airflow escaping through the glottis may yield air turbulence when it coincides with voicing (Titze, 1994). While breathiness can be classified as a voice disorder, it is a feature of both pathological and non-pathological speech. In languages such as Gujarati, breathiness exists as a phonemic feature, which can change the meaning of a word. With respect to English, breathiness occurs not as a phonemic but an allophonic variant. It is thought that allophonic breathiness and pathological breathiness share the same acoustic feature and only differ in covering different ranges of the same continuum.

Since breathiness is not only found in pathological voice but also normal voice, the breathy voice quality can be considered to play a multiple role in the perception related to diagnostic, phonemic, and sociolinguistic discrimination. Since changes in both voice quality and speech intelligibility can be heard, the acoustic parameters reflecting an audible change in voice quality and/or speech intelligibility would provide a link between the production and perception of speech and voice. Therefore, an acoustic-perceptual study on the relationship between breathiness and vowel intelligibility may provide useful information for developing an acoustic tool to monitor voice problems related to glottal competence as well as improving the design of processes involved in speech synthesis, recognition, transmission, or training for speech enhancement.

An abnormally high prominence of the first harmonic (H1), namely, the fundamental frequency (F0) of vocal fold vibration, relative to other harmonics has been identified as the main acoustic marker of breathiness (de Krom, 1995). To determine how voice-related acoustic features may affect speech intelligibility, this study included both 1) a perceptual study examining how a linear change to the selected acoustic correlate of breathiness may affect the perception of breathiness and vowel quality and 2) an acoustic analysis of the stimuli employed in the perceptual study to investigate how changes in a selection of acoustic parameters as a result of the signal manipulation scheme used in this study are related to the different degrees of perceived breathiness and intelligibility.

1.2 Research Question and Importance

The current study aims to delineate the effects of diminished voice quality, especially those arising from heightened levels of breathiness, on speech intelligibility. Clarification of the relationship between voice quality and speech intelligibility would help in monitoring or manipulating the voice-related changes that impact speech intelligibility. The relationship between the production and perception of speech and voice can be determined through manipulating the acoustic parameters which govern audible changes in voice quality and speech intelligibility. A fair amount of relevant information about phonation can be elicited from acoustic analysis as both voice acoustics and vocal fold physiology do correspond (Radish Kumar, Bhat, & Prasad, 2009). Since the acoustic signs of breathy voice may be found in a number of speech or voice disorders, including functional dysphonia, vocal fold paralysis, dysarthria, laryngeal cancer, and benign vocal fold mass lesions (Castillo-Guerra & Ruiz, 2009), acoustic changes identified as being important to the perception of voice quality and speech intelligibility may be used to monitor post-treatment progress or develop screening, training, or signal processing tools. In addition, a better understanding of the relationship between voice quality and speech intelligibility would help in improving the signal processing strategies used in technologies related to speech communication.

1.3 Aims and Hypotheses

This study aimed to determine how changes to a chosen acoustic correlate, specifically, the prominence of H1 as indicated by the amplitude difference between H1 and the second harmonic (H2), could affect the perception of voice quality and vowel identity. There were two main hypotheses to be tested in this study:

Hypothesis One: It was hypothesized that changes to the magnitude of the amplitude difference between H1 and H2 (H1-H2) would lead to changes of the perception of breathiness and vowel identity. Specifically, an increase of H1-H2 would be expected to lead to an increase in the perceived breathiness and in the number of vowel misidentifications.

Hypothesis Two: It was also hypothesized that the relationship between voice quality and vowel intelligibility might be affected by vowel type and speaker gender as well as the speaker's vocal health status (i.e., voices from normal speakers versus those from voice patients).

2. LITERATURE REVIEW

2.1 Breathiness

Voice quality can be described as a set of distinctive auditory components which represent an individual's speech (Gerratt & Kreiman, 2004). It excludes pitch, loudness, and phonetic category and includes features such as roughness, breathiness, creakiness, and nasality (Titze, 2000). Breathiness in speech refers to a situation where excessive loss of air occurs due to incomplete glottal closure during the closed phase in the cycles of vocal fold vibration. Breathy voice is commonly found in voice patients but can also be found in normal speakers.

2.1.1 Breathiness in Voice Patients

Voice may be narrowly defined as the sounds generated through vocal fold vibration (Titze, 1994). When voiced speech sounds are produced, vocal fold vibration is initiated to modulate airflow from the lungs to enable phonation (Jing, 2009). Disruptions to the periodicity of vocal fold vibration can adversely affect the quality of voice. The primary concern for individuals with voice disorders is their voice quality (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). The perceived speech or voice abnormality leads voice patients to seek treatment and the success of treatment is often gauged subjectively based on the auditory perception of the patient's voice quality (Kreiman et al., 1993).

Breathy voice may or may not involve disruptions to the periodicity of vocal fold vibration. However, breathy voice is produced with vocal folds failing to achieve a full closure during the closed phase of the vibratory cycles of vocal folds (Reetz & Jongman, 2009). The resulting airflow escape through the glottis may turn into excessive aspiration noise when occurring simultaneously with voice. As stated in Simpson (2012), an incomplete glottal closure in phonation may result in a variety of acoustic changes, including decreased energy of the overtones at high frequencies, relatively stronger noise in the spectrum, increased Formant one (F1) bandwidth, and cycle-to-cycle variations of the amplitude and F0 of vocal fold vibration.

During phonation, the primary vocal tract excitation period occurs with the closure of the vocal folds. The maximum airflow declination rate, which reflects the rate of vocal folds recoiling back to their resting position, can be calculated based on the waveform derived from low-pass and inverse filtered airflow. A higher maximum airflow declination rate is associated with an increased amount of high frequency energy (Simpson, 2012). In cases of incomplete glottal closure, the maximum airflow declination rate is reduced resulting in lower energy in high frequencies and thus low frequencies becoming more prominent in the spectrum. The reduction of the harmonic energy at high frequencies, which were replaced by noise, suggests that less aerodynamic force is transferred to acoustic energy in phonation associated with glottal incompetence. Since this acoustic effect of an incomplete glottal closure has been demonstrated, it appears that the pathophysiology of voice disorders showing signs of breathy voice may be detectable and its severity traceable through acoustic measurement. However, as not all breathy voice is considered abnormal and not all speech with breathy quality is unintelligible, the relationship between the acoustic correlate of breathiness and the perception of breathiness and speech intelligibility may well be non-linear or categorical and even language-specific or gender-dependent rather than being linear or universal.

2.1.2 Breathiness in Normal Voice

Breathiness is inherent in non-pathological voice, serving as a phoneme in some languages. For example, in English, allophonic breathiness presents itself non-contrastively in the intervocalic /h/ such as that in the words "behind" and "ahead". In Indo-Aryan languages such as Nepali, Hindi and Gujarati, breathiness is used as a contrastive feature (Ladefoged, 1983). The !Xóõ language uses subtle changes in laryngeal position as a glottal feature to create a phonetic distinction between breathy and clear vowels (Bickley, 1982). It has also been noted that voice qualities used phonemically in some languages can be perceived by English speakers as being associated with pathological voice (Gordon & Ladefoged, 2001). These findings suggest that a study of the relationship between the acoustic correlate of breathiness and the perception of breathiness and speech intelligibility needs to take the speaker's vocal health and language and the listener's native language into consideration.

2.2 Gender Effect on Breathiness

The role of breathiness varies depending not only on language, but also on gender. Physiological changes for males at puberty initiates a divergence from the female voice. This is due to rapidly rising levels of testosterone causing changes in laryngeal cartilage and vocal folds (Sapienza & Ruddy, 2009). Gender-related difference in voice can be quantified in terms of differences in voice projection power, F0, and measures of vocal tract resonance (Sapienza & Ruddy, 2009). For example, the F0 of the male speaking voice is around 125 Hz compared to a higher average of 200 Hz for adult females (Titze, 2000). Klatt and Klatt (1990) conducted a study involving acoustic analysis, synthesis, and perceptual measures to investigate variations of voice quality among male and female voice. A panel of eight listeners judged, on a seven-point scale of breathiness, natural voice samples produced by six male and 10 female normal speaking talkers (Klatt & Klatt, 1990). The perceptual scores indicated that male voice was generally perceived to be less breathy than female voice (Klatt & Klatt, 1990).

The anatomical differences between men and women may be a contributing factor to this difference in voice quality. Vocal tract length is longer in men than in women. Women have larger posterior cartilaginous spaces than men and enlarged glottal space, resulting in a greater chance of forming a posterior glottal gap during phonation and thus creating a more breathy voice quality (Sapienza & Ruddy, 2009). Females are also predisposed to breathiness through a longer open phase of the glottal pulse which allows greater airflow into the vocal tract (Sapienza & Ruddy 2009). However, despite a high occurrence of incomplete glottal closure found in females through fiberstroboscopic examination **(**Södersten, Hertegard & Hammarberg, 1995**),** female voices have sometimes been found to be perceived as non-breathy (Södersten et al., 1995). It was speculated that this finding might suggest that a higher threshold was set for female voice for the perception of breathiness because breathiness was generally not perceived to be an unusual component in female voice and thus easily overlooked by the listener (Södersten et al., 1995).

In a study of acoustic correlates of breathiness, Hillenbrand, Cleveland and Erikson (1994) found higher breathiness ratings for men in comparison to women in the very breathy condition but not for normal or moderate conditions.

It was noted, however, that there was a high likelihood that listeners allowed for more breathiness in female voice while the same level of breathiness in males sounded abnormal (Hillenbrand et al, 1994). In an investigation into the theory of women having a relatively higher amount of breathiness in normal speech compared to men, Henton and Bladon (1985) questioned why women would exhibit breathy voices when breathiness is not used contrastively in English and has the potential to reduce speech intelligibility. They concluded that there may be a behavioural origin for this phenomenon. Although Henton and Bladon's (1985) study did not include perceptual judgements of breathiness nor intelligibility, the H1-H2 measure was used in the study to provide evidence of breathiness based on a high correlation found in previous studies (Bickley, 1982; Fischer-Jørgensen,1967) between this acoustic measure and the perception of breathiness.

2.3 Acoustic Measures Related to Breathiness

A variety of acoustic measures have been employed in the study of breathiness. These include acoustic perturbation measures related to the periodicity of vocal fold vibration, aspiration noise, intensity, and patterns related to the energy distribution over the harmonics or the overall spectral envelope (Castillo-Guerra & Ruiz, 2009). In studies employing only acoustic data (Simpson, 2012), a variety of measures have been used to quantify breathiness in both linguistic and non-linguistic contexts. As pointed out by Simpson (2012), some examples of the acoustic characteristics of breathiness include:

1. the relative lowering in or amplitude of F0 (Pandit, 1957),
2. presence of noise in the upper spectrum (Klatt &Klatt,1990; Ladefoged & Anton˜anzas Barroso,1985; Pandit, 1957),
3. presence of tracheal poles/zeroes (Klatt&Klatt,1990),
4. a stronger first harmonic H1 relative to the first three formants A1, A2, A3 (Abramson,Luangthongkum, &Nye, 2004; DiCanio, 2009; Fischer-Jørgensen, 1967; Garellek & Keating,2011; Hanson,1995; Hanson & Chuang,1999; Keating &Esposito,2007; Ladefoged,1983; Shrivastav&Sapienza, 2003),
5. the amplitude difference between H1-H2 (Abramson et al.,2004; Bickley, 1982; Fischer-Jørgensen,1967; Garellek &Keating,2011; Hanson, 1995; Hanson &Chuang,1999; Henton &Bladon,1985; Huffman, 1987; Klatt &Klatt,1990; Ladefoged &Antonanzas Barroso, 1985), and
6. a greater amplitude difference between the second and the fourth harmonic, H2–H4 (Keating &Esposito,2007; Keating, Esposito, Garellek, Khan,& Kuang, 2010).

A combination of other acoustic measures has also been associated with breathiness. With the aim of achieving automatic detection of perturbations associated with breathy voice, Castillo-Guerra and Ruiz (2009) compared both simulated and pathological utterances of the vowel /a/ compiled from a database of 108 pathological speakers and 19 normal speakers. A comparison of 9 separate indexes, each as a variant of the acoustic representation of breathiness, was performed. The indexes contained different acoustic voice features including noise, H1-H2 amplitude, formant structure, jitter, shimmer, and a selection of energy ratios. Three speech language pathologists with eight or more years of experience were asked to judge the levels of breathiness on a scale from 0 to 6, with 6 being severe breathiness. They found that 88.5% of variability of perceptual judgements could be accounted for by an index which contained a grouping of perturbations associated with harmonic/formant structure, harmonic ratio, and noise levels.

A common feature shared by these observed acoustic characteristics of breathy voice is that harmonics at the low frequencies, particularly the first harmonic, are more prominent relative to harmonics at higher frequencies. However, the variety of the acoustic markers of breathiness proposed in the literature suggests that other factors may affect the measurement of this H1 dominance and complicate its relationship with the perception of breathiness.

2.3.1 Prominence of the First Harmonic

Fischer-Jorgensen (1967), in a study of murmured vowels in Gujarati, found that breathiness was characterized by a high H1 amplitude. However, she qualified this finding by stating that a single acoustic feature alone would not consistently elicit the perception of breathiness from listeners.

The H1-H2 measure, that is, the amplitude difference between H1 and H2, was introduced as an acoustic correlate of breathy voice in a study by Bickley (1982). The study employed native speakers of !Xóõ and Gujarati to read a word list to generate clear and breathy versions of the vowel /a/ in their native language (Bickley, 1982). In nine out of the 10 samples of the !Xóõ breathy vowels, the amplitude of H1 was found to be greater than that of H2 (Bickley, 1982). The same H1-H2 pattern was found for all of the 80 Gujarati samples of breathy vowels (Bickley, 1982). It was found, based on the perception of two phonetically trained listeners, including a native English speaker and a native Gujarati speaker, that the vowels judged most breathy exhibited the greatest amplitude difference between H1 and H2 (Bickley, 1982).

A further perceptual experiment using acoustically altered Gujarati vowels samples was performed. Two acoustic parameters, H1 and the degree of aspiration noise, were adjusted to simulate clear to breathy versions of the vowels /a/, /i/, and /o/ (Bickley, 1982). They were combined with natural Gujarati consonants to make one syllable Gujarati words and presented to four native Gujarati talkers for identification. Again, vowels with the highest first harmonic amplitudes were consistently identified as breathy. To alleviate the glottal waveforms of sound radiation and vocal tract filtering, inverse filtering was carried out using both normal and breathy naturally produced Gujarati vowels. It was observed that glottal waveforms of the clear vowels had slow opening phases relative to closing, abrupt closures, and closed phases lasting a third of a cycle .In contrast, glottal waveforms of breathy vowels had less abrupt closures and shorter closed periods than the clear waveforms. This indicates that complete glottal closure was not achieved in breathy voice (Bickley, 1982). Bickley (1982) concluded that enhanced H1 is the strongest acoustic measure of breathy speech.

In reference to this study, Klatt and Klatt (1990) noted that the H1 amplitude required to produce a clear distinction between clear and breathy vowels was 15 dB, which is far greater than H1 amplitude in natural speech (Klatt & Klatt, 1990). They added that due to this exaggerated H1-H2 amplitude in relation to normally produced speech, this measure alone was not sufficient to perceptually separate clear and breathy voice (Klatt & Klatt, 1990). However, there are a large number of studies which have taken H1-H2 as one of the primary correlates of perceptual breathiness measurement (Abramson et al.,2004; Bickley, 1982; Fischer-Jørgensen,1967; Garellek &Keating,2011; Hanson, 1995; Hanson &Chuang,1999; Henton &Bladon,1985; Huffman, 1987; Klatt &Klatt,1990; Ladefoged &Anton˜anzas Barroso, 1985). For example, Ladefoged and Antonanzas-Barroso (1985) compared H1-H2 and aspiration noise as cues for the judgement of breathy voice amongst American listeners. The finding was that H1-H2 was the more superior measure of the two for the identification of breathy voice.

In a study of glottal flow measures comparing the phonemic use of breathy and non-breathy vowels in Hmong, Huffman (1987) had three native speakers repeat a word list consisting of four phonation types. No perceptual testing was performed in this study. Inverse filtering was used to "remove the resonance effects of the vocal tract and radiation at the lips in order to determine the characteristics of the acoustic signal that were attributed to the glottal source" (Huffman, 1987, pg 3). This allowed the extraction of glottal airflow from oral airflow recordings, enabling frequency domain measures of spectral tilt and harmonic amplitude to be obtained. It was found that while spectral tilt could not be reliably measured for most tokens, the H1-H2 measure resulted in a significant difference between breathy and normal phonation. Specifically, the mean H1-H2 value was 9.48 dB for breathy vowels and 2 dB for non-breathy vowels. The H1-H2 measure was not found to be affected by pitch, duration, or quality of the vowel tokens analyzed.

In a study of the variations of voice quality among male and female voice, Klatt and Klatt (1990) referred to H1-H2 as a primary acoustic correlate of breathiness. In their study, six male and 10 female normal speakers produced speech samples in which normal syllables were replaced by /ʔɑ/ or /hɑ/. The H1-H2 measure was obtained from the midpoint of a vowel segment to avoid the influence of adjacent consonants. There was a difference of about 5.7 dB difference between males and females on the averaged H1-H2, with the latter having a higher first harmonic amplitude (Klatt & Klatt, 1990). On average, female voice was found to be perceived to be more breathy than male voice (Klatt & Klatt, 1990). Out of 10 acoustic measures associated with breathiness, the H1-H2 measure was one of the only two measures that had shown a statistically significant correlation with the perceptual judgements of breathiness when natural voice material was used (Klatt & Klatt 1990). The other measure, was the degree of aspiration noise around the F3 region, with a rise in aspiration noise resulting in an increase in the perception of breathiness (Klatt & Klatt, 1990). In contrast, Ladefoged and Antonanzas-Barroso (1985) found that aspiration noise, when used alone, resulted in the highest reliably high breathy voice ratings although it was noted that a combination of acoustic alterations yielded signals associated with the highest breathy ratings and deemed more natural sounding.

The H1-H2 measure was one of the measures used by Hillenbrand et al. (1994) to evaluate the ability of acoustic measures to predict breathiness ratings. In their study, normal talkers produced normal, moderately breathy, and very breathy versions of sustained vowels which were rated by Speech-Language Pathology graduate students using a direct magnitude rating scale (Hillenbrand et al., 1994). A number of acoustic measures were investigated in the study, including cepstral peak prominence, peak to average ratio, spectral tilt, ratio of high to mid/low frequency energy and signal periodicity (Hillenbrand et al. 1994). The H1-H2 measure was found to be moderately correlated with breathiness ratings. Amongst several different variations of the H1 measure (e.g., H1 amplitude relative to Formant 1 amplitude and H1 amplitude relative to overall amplitude) experimented with in their study, the H1 amplitude relative to H2 amplitude was found to be the most sensitive measure of breathiness.

This finding was in agreement, to an extent, with that of Klatt & Klatt (1990). Furthermore, using synthetically altered samples, Kreiman & Gerratt (2010) found the average H1-H2 amplitude difference that could elicit the perception of breathiness in English speakers was 3.61 dB. This finding suggests that the H1-H2 measure might be a viable acoustic cue for discriminating between pathological and non-pathological breathy voice.

Other studies have found the H1-H2 measure to be a significant predictor of voice quality. For example, to determine which acoustic measures best reflect changes in glottal pulse and spectral shapes, Kreiman, Gerratt and Antonazas-Barroso (2007) employed 70 subjects (42 female and 28 male), 10 normal speakers, and 60 with speech pathologies, to produce sustained phonation of the vowel /a/. Out of 19 different spectral measures, H1-H2 accounted for the most perceptual and acoustic variance in glottal pulse shapes in English voices. Similarly, in a cross-linguistic sample of breathy versus modal phonation, Esposito (2010) found that H1-H2 distinguished these phonation types in eight out of the 10 languages or dialects included in their study.

In summary, based on findings from studies on the relationship between H1-H2 and breathiness, it appears that an increase of H1-H2 would lead to an increase in the perception of breathiness and a reduction in the perception of clarity. However, findings of a weaker link between H1-H2 and breathiness in other studies suggest that other factors may affect the H1-H2 measure.

2.3.2 Measurement Considerations

Some methodological differences between studies that have shown a strong link between H1-H2 and breathiness and those that have not suggest that there are factors to be considered that may have an impact on the measurement of H1-H2 as well as its relationship with the perception of breathiness. These include the susceptibility of the H1-H2 measure to acoustic changes related to vocal tract configuration and the speech samples selected for the derivation of the measure.

2.3.2.1 Vocal Tract Configuration

Since different vowels were used in studies of the relationship of H1-H2 and the perception of breathiness, the conflicting findings amongst these studies suggests that the selection of vowel type for deriving H1-H2 may affect the sensitivity of this measure in detecting breathiness. Other factors that may affect vocal tract configuration and thus the distribution of acoustic energy of the vocal output across frequencies after being filtered through the vocal tract include nasality and gender.

<u>Vowel Effect.</u> Vowel effect can be considered mainly a type of vocal tract effect. It has been well known that vowels identity is determined by the locations of the first two formant frequencies, which are determined by vocal tract resonance (Peterson & Barney, 1956). Formants are a concentration of acoustic energy resulted from the filtering characteristics of the vocal tract (Titze, 2000). The specific formant frequencies are determined by vocal tract shape and related resonance (Reetz & Jongman, 2009). The many harmonics created at the glottis by vocal fold vibration are filtered by the vocal tract (Titze, 2000). Incomplete glottal closure has been found to result in an increase in F1 bandwidth (Simpson, 2012). The frequency and amplitude of F1 can also vary by vowel. There is an inverse relationship between vowel height and the strength of F1. That is, where the tongue

position during vowel production is high, the frequency of F1 will be low (Reetz & Jongman, 2009). Since a high vowel (e.g., /i/ and /u/) has a lower F1 than a low vowel (e.g. /a/ and /o/), it is likely that the H1-H2 measure may vary by vowel. Specifically, the F1 of a low vowel is located further away in frequency from H1 and thus far less likely to raise the amplitude of the lower harmonics than that of a high vowel (Henton & Bladon, 1985). This being the case, it has been common for researchers to select low vowels for the measurement of H1-H2 to ensure that F1 does not interfere with H1 (Hanson, 1997; Henton & Bladon, 1985). However, since there may be a vowel difference in the effect of F1 on H1-H2, the sensitivity of H1-H2 in detecting the change in vowel breathiness may also vary by vowels.

In a perceptual study using synthesized versions of the vowel /a/ and 12 naive listeners (11 females and one male), Samlan, Story and Bunton (2013) compared the breathiness of two tokens presented concurrently on a continuum. They found that H1-H2, along with spectral tilt, had a non-linear relationship with the perception of breathiness. Specifically, they found that an increase in the prominence of H1 did not necessarily yield a higher breathiness perception rating. However, both H1-H2 and spectral tilt corresponded well to changes in the spectral slope of the acoustic signal at the low to moderate and the higher breathiness levels respectively. It was noted that previous studies had shown H1-H2 to be effective in phonemic breathiness distinction (Esposito, 2010). Samlan et al. (2013) concluded that H1-H2 was useful for detecting changes in vowel breathiness only for vowels with breathiness at low to moderate levels. However, neither Samlan et al.'s (2013) nor Esposito (2010) studies used alternative vowels. The use of the vowel /a/ alone may have had an impact on the results. It is not unlikely that inclusion of high vowels (/i/ or /u/) would have shown a different pattern in the relationship between H1-H2 and the perception of breathiness.

Nasality and Gender Effects. Nasality has been shown to impact on breathy voice perception through studies of the acoustic correlates of various aspects of voice quality and gender-related voice differences. For example, Klatt and Klatt (1990) speak of an H1-H2 amplitude increase with reference to nasality. They state that the perception of nasality and that of breathiness are not only both affected by H1-H2 but also differentiated based on the values of the other acoustic cues present in the signal. In particular, they found that the strongest perception of breathiness was inherent in the stimulus that incorporated a variety of natural acoustic cues such as aspiration noise, spectral tilt, H1-H2, and bandwidths of F1 and Formant two (F2). Nevertheless, adjustments applied to individual acoustic parameter such as H1-H2 and formant bandwidths were also found to result in the perception of nasality for many listeners in their study (Klatt & Klatt, 1990). These observations suggest that vocal tract configuration may interact with the source characteristics in affecting the perception of voice quality such as breathiness.

In a study analyzing voice tokens to measure the effectiveness of using H1-H2 for gender comparison of breathiness, Simpson (2012) used two perceptual groups, one consisting of 25 females and 25 males and the other 18 females and 14 males. Voice data for both groups was obtained from pre-existing data sets. In addition, recordings of 15 female and 7 male German speaking subjects from the second group, vocalized the open vowel /a/ in three contexts, two of which were segmented out from German words and one version an isolated, sustained /a/. They found that the gender difference on the harmonic expression of nasality was a confounding factor for the use of H1-H2 in comparing breathy voice across genders. They speculated that female voice was more susceptible to the effect of the nasal zero because H2 in female voice is close in frequency to the nasal zero and can thereby be attenuated resulting in a larger difference between H1 and H2. The nasal pole, may strengthen the prominence of H1 as well, further exaggerating the H1-H2 difference

(Simpson, 2012). The impact of the nasal pole on H1-H2 varies by gender and warrants further explanation.

Furthermore, since there is generally a gender difference on the spoken F0, F0 may also play a role in affecting the perception of breathiness or nasality. For example, the first nasal pole occurs in the region of 200Hz to 300Hz for both males and females. Therefore, nasality may affect H1 in females and bring about an increase in H2 for males. Consequently, the net effect of the first nasal pole would be an overall rise in H1-H2 for females and a decrease of H1-H2 for males (Simpson, 2012). This complication raises the question as to whether the relationship between H1-H2 and the perception of breathiness would vary by gender due to a myriad of gender-related differences in the structure of the vocal folds and the vocal tract.

2.3.2.2 Speech Sample Effect

There is also evidence from various acoustic studies suggesting that 1) the sensitivity of an acoustic correlate of breathiness in detecting breathiness may be affected by the speech samples used for the derivation of the measures, and that 2) redundant acoustic information may also play an important role in the perception of breathiness. One such example is spectral tilt, which can be derived from connected speech as well as sustained phonation.

Spectral tilt is a measure of the ratio of low frequency to high frequency energy in a speech signal. A steep, negative spectral slope is expected if H1 is dominant relative to the harmonics at higher frequencies. Attributed to the domination of mid to high frequency energy in aspiration noise inherent in voices associated with an incomplete glottal closure, breathy voice commonly exhibits more high frequency noise components compared to low frequencies (Hillenbrand & Houde, 1996). The degree of spectral tilt may be measured using a Breathiness Index (BRI) (Fukuzawa, El-Alsoouty & Honjo, 1988). The BRI measure is defined as the energy ratio of high to mid/low frequency noise (H/L) (Hillenbrand & Houde, 1996). In a study (previously mentioned in 2.3.1) looking at the relationship between the perception of breathiness and three acoustic features, including spectral tilt, aspiration noise, and H1-H2, Hillenbrand et al., (1994) analyzed sustained vowels obtained from non-pathological speakers trained to simulate 3 levels of breathiness: normal, moderately breathy and very breathy. The two measures of spectral tilt (i.e., BRI and H/L) correlated weakly with perceived breathiness while H1-H2 was moderately correlated. In a study of the perception of voice quality in males and females, Klatt and Klatt (1990) also found that while spectral tilt may contribute to the perception of breathiness when other acoustic cues are present, spectral tilt alone plays little or no role in the perception of breathy voice.

In a follow-up study to Hillenbrand et al.'s (1994) report, Hillenbrand and Houde (1996) investigated both sustained vowel (/a/) and vowels embedded in the sentences of the Rainbow passage produced by 20 pathological and 5 non-pathological speakers. Hillenbrand and Houde (1996) found that the two aforementioned measures of spectral tilt, including BRI and H/L, correlated moderately with breathiness ratings when the sustained /a/ was used to derive acoustic measures. However, with the vowel samples segmented out from connected speech, the correlation was considerably higher with spectral tilt accounting for 75% to 85% of variance. In other words, it suggests that the effectiveness of spectral tilt as a predictor of breathiness is reliant on the type of voice sample (Hillenbrand & Houde, 1996).

These findings also suggest that speakers may use a different control mechanism in producing sustained phonation and connected speech and thus the usefulness of an acoustic correlate of breathiness in predicting the extent of breathy voice quality depends on the methodology used in deriving the measure, including the speaker's task.

2.4 Sustained Phonation Versus Connected Speech

It has been noted that perception of voice quality differences are better recognised using data obtained from connected speech as opposed to sustained vowel phonation (de Krom, 1995). Two possible reasons that have been proposed for this advantage in the perceptual evaluation of speech are that 1) connected speech resembles natural speech compared with sustained phonation and that 2) it allows for more detailed description of deviant voice quality aspects associated with pathological voice (de Krom, 1995). Klingholtz (1990) questioned the validity of the use of sustained phonation likening it to the difference between a singing voice and speech.

Proponents for the use of sustained vowels, however, suggest that a simpler acoustic structure, which alleviates issues with phonemic context and stress, can be achieved with sustained phonation. Individual characteristics of voice source such as speaking rate, dialect and intonation, which may unduly influence the perception of voice quality can also be eliminated. Also, as vocal tract and articulators do not vary greatly over time, the use of sustained vowels ensures a more “controlled environment” (Parsa & Jamieson, 2001, p328) than with connected speech and allows for easy adjustment of noise and perturbation parameters (Parsa & Jamieson, 2001). This increases the likelihood of obtaining robust and reliable perceptual judgements of voice quality (Kreiman & Gerratt, 1994).

Parsa and Jamieson (2001) compared the efficacy of using acoustic cues derived from sustained phonation of vowels and connected speech in differentiating normal and pathological voice. Voice samples from a total of 53 normal and 175 pathological speakers were obtained from which nine separate acoustic cues were analysed. To enable the comparison of the two phonation types, the connected phonation was analysed by isolating voiced sections from 9 seconds of the rainbow passage. Using a three step criteria, voiced portions of the continuous speech were segmented into 30 ms frames and then transformed into the spectral domain using a Hamming window. The analysis of sustained phonation was based on a 1 second duration and the same transformation method. It was found that perturbation measures taken from sustained vowels were superior than those taken from connected speech in the differentiation between normal and pathological voice, pointing out that perturbations measures extracted from connected speech are likely to be compromised by intonation and other voice modulations and are therefore not as sensitive in the detection of vocal pathologies (Parsa & Jamieson, 2001).

In terms of the practical implications of the acoustic voice measures in clinical practice, the physiological differences between sustained vowels and connected speech should be considered (Law, Kim, Lee, Tang, Lam & Hasselt, 2012). More complex, dynamic and well controlled laryngeal and supra laryngeal muscular activity is necessary for phonation with connected speech than with sustained vowels. In addition, vocal fold control is heightened when complex articulation occurs during periods of voicing and devoicing of consonants. Thus, as de Krom (1995) stated, any deviant aspects of a voice will be more prominent and readily captured during connected speech (Law et al., 2012).

In contrast, sustained vowels are characterized by a relatively static configuration of the laryngeal and supralaryngeal muscles with minimal variation during their production. Hence, the production of sustained vowels is considered to be a simpler motor task compared with that of connected speech (Law et al., 2012). However, the use of connected speech has proven perceptual benefits. For example, as was stated in a previous section (see Section 2.3.2 Speech Sample Effect), more accurate predictions of breathiness ratings using spectral tilt were made when connected speech embedded vowels samples were used, compared with the sustained vowel /a/ (Hillenbrand & Houde, 1996). It appears that acoustic voice measures derived from sustained phonation may be sensitive in detecting the difference between normal and pathological voice while those from running speech may correspond better with the perception of the overall voice quality in voice.

2.5 Perceptual Studies of Voice Quality and Vowel Intelligibility

The analysis of breathy voice is conducted mainly via two main approaches: instrumental and perceptual (Castillo-Guerra & Ruiz, 2009). Perceptual studies often use trained specialists in areas such as speech and language, audiology and other related backgrounds. They may also use untrained listeners. However, the judgements made by listeners, regardless of background, are often inconsistent due to differences in experience, training, memory, judging protocol and rating scales (Castillo-Guerra & Ruiz, 2009). Rating scales have a significant impact on the outcomes of perceptual studies involving breathy voice. Previous perceptual studies of breathiness, some of which claim significant findings in modelling breathiness perception, involve factors such as the use of simulated data only and breathy phonation produced by non-pathological speakers. The use of different methods makes inter-study comparison difficult (Castillo-Guerra & Ruiz, 2009).

Patel, Shrivastav, and Eddins (2012) state that voice quality models cannot be adequately created using data obtained from rating scales or direct magnitude estimation. One reason for this is that the perceptual difference between concurrent scores along a rating scale is arbitrary and possibly not equivalent, leading to variability across raters. Unbiased perceptual data with clear measurement properties such as interval, ratio or ordinal levels are required to ensure successful models of voice quality are achieved. Rather than ask participants to assign numbers to quantify the quality of voice, they suggest the use of a comparison stimulus to enable listeners to perform a matching task. Matching tasks have been shown to reduce variability in judgments of voice quality (Kreiman & Gerratt, 2005;

Kreiman, Gerratt, & Ito, 2007; Patel, Camacho, Shrivastav, & Eddins, 2010 as cited in Patel et al., 2012).

Despite the aforementioned limitation, direct magnitude estimation has been employed in a number of studies to investigate breathiness perception (Shrivastav, Camacho, Patel & Eddins, 2011). According to Shrivastav et al. (2011), having the listener perform a magnitude estimation task for judging breathiness is preferable over a standard rating scale (e.g. multiple choice) as the latter would lead to ordinal data (Shrivastav et al., 2011). In this study, direct magnitude estimation method was used in the "breathiness rating" task in order to reduce testing time and avoid the fatiguing or testing effect. The experimental design employs the main factors of interest as within-subjects factors to circumvent the issues raised in previous studies regarding the inter-rater variability in performing the perceptual task using the direct magnitude estimation method.

3. METHOD

3.1 Participants

A quota sampling strategy was used to recruit 10 male and 10 female adults with normal hearing from university students and their acquaintances. Subject inclusion criteria were adult native speakers of English with no history of hearing loss. The pure-tone audiogram of each listener was ascertained using the Hughson-Westlake method of threshold detection. Air conduction thresholds in both ears were tested at 0.5, 1, 2 and 4 kHz. The hearing test was conducted in the sound treated booth in the hearing clinic at the University of Canterbury (Christchurch, New Zealand). A total of 21 volunteers went through the hearing screening test. One volunteer failed the pure-tone audiometry test (i.e., At least one pure-tone threshold was above the 20 dB hearing level) and was excluded from the study and referred to a hearing clinic. The 20 participants included in the study aged between 19 to 50 years (Mean = 28 years, SD = 7.9). Results from an independent t test revealed no significant gender difference on age ($t = 0.277$, $df = 18$, $p = 0.785$).

3.2 Participants' Tasks

Participants were asked to perform a "breathiness rating" task and a "vowel identification" task. In the "breathiness rating" task, participants were presented with one vowel stimulus (either /i/ or /a/; duration: 500 ms) at a time and asked to mark on a 10 cm long line, which represented a continuum ranging from 0 ('Not Breathy') to 100 ('Very Breathy'), to indicate how breathy they perceived the voice to be (see Figure 3.1). In the "vowel identification" task, participants were asked to listen to one vowel stimulus (/i/, /a/, or /o/; duration: 60 ms) at a time and identify which vowel (i.e., /i/, /e/, /a/, /o/, or /u/) they perceived the stimulus to be by clicking on the selection shown on a computer screen (see

Figure 3.2) Having five options for the three vowel stimuli presented allowed further insight into the overall reliability of participant responses.

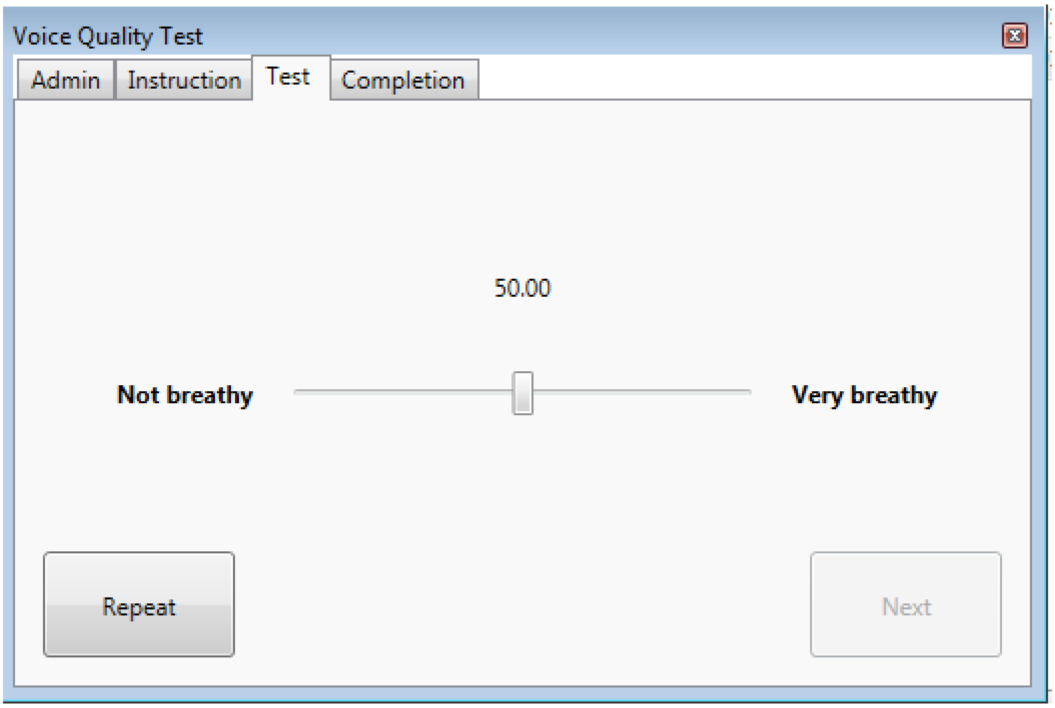


Figure 3.1 Computer interface for the “breathiness rating” task.

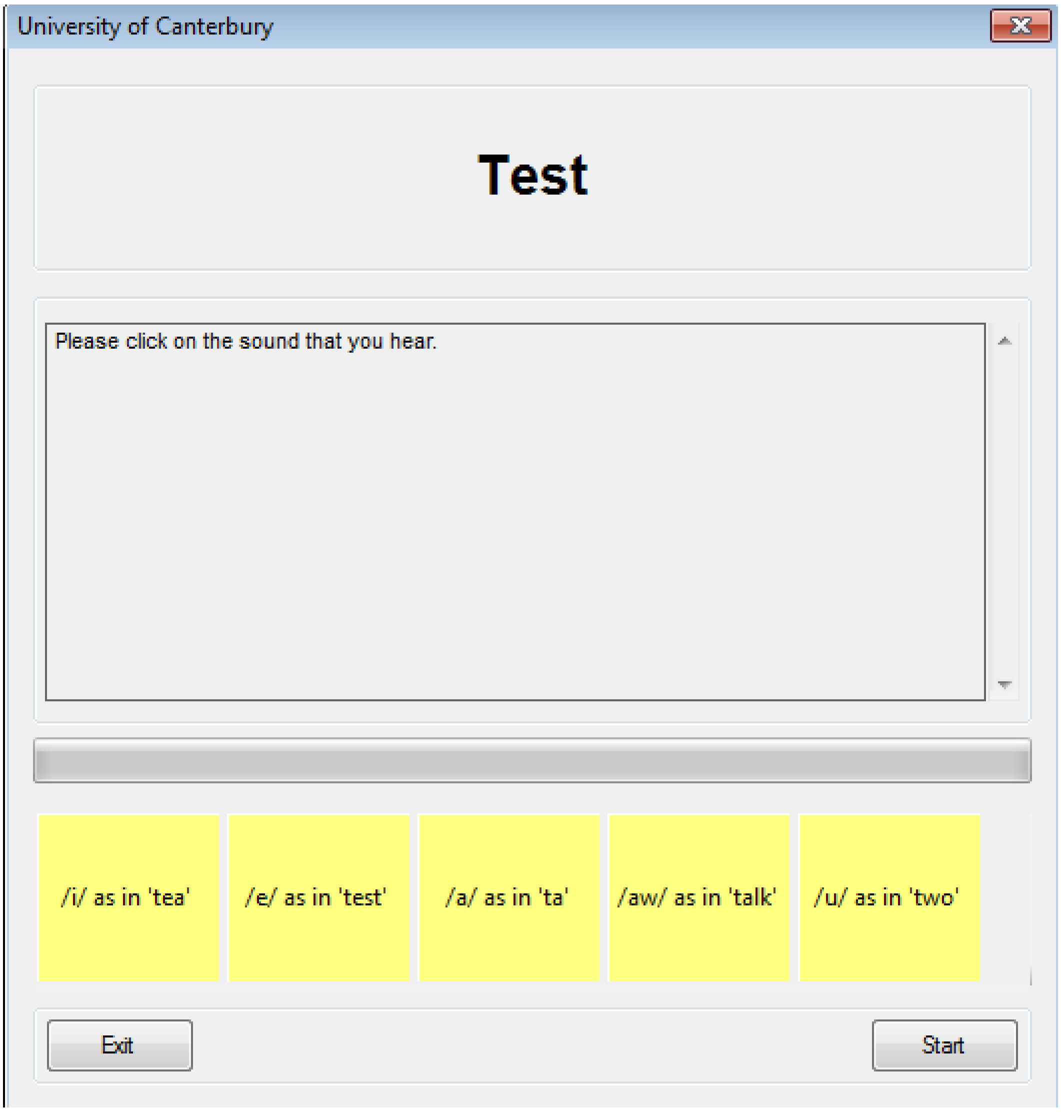


Figure 3.2 Computer interface for the "vowel identification" task.

3.3 Stimuli

The original vowels used to generate the stimuli used in this study were selected from the voice recorded from 30 speakers. The voice recordings have been made using a direct digitization method (Sampling rate: 44.1 kHz, 16-bit). The speakers included 20 voice patients (10 with vocal fold paralysis and 10 with functional voice disorders) and 10 normal speakers, with equal number of males and females in each group. These speakers aged between 22 and 86 years (Mean = 51.5 years, SD = 19.7). Two sustained vowels (i.e., /i/ and /a/ sustained for around 2 to 3 seconds at comfortable pitch and loudness level) and three vowels (/i/, /a/, and /o/) embedded in the connected speech (i.e., Rainbow passage) were obtained from each speaker. The stimuli used for the "breathiness rating" task were signals segmented from the sustained vowel phonation and those for the "vowel identification" task were signals segmented from the connected speech.

3.3.1 Breathiness Rating Task

To select the original signals for use in the "breathiness rating" task, the 500-ms mid portion of the sustained vowel (/i/ and /a/) phonation were used. Results from an acoustic analysis of these vowels revealed that the H1-H2 measures ranged from -9.5 to 21.8 dB, with a mean of 4.09 dB (SD = 6.69). (Note: A positive sign means H1 has a higher amplitude than H2). For female and male speakers in the normal and patient groups separately, the signals showing the lowest, median, and highest H1-H2 values were selected to represent three H1-H2 levels, namely, "Low", "Mid", and "High". Table 3.1 shows the H1-H2 amplitude difference for the original signals selected for use in the "breathiness rating" task. These 24 vowel signals (2 vowels X 3 H1-H2 levels X 2 speaker genders X 2 voice status) were submitted to 12 signal manipulation conditions. The 12 signal conditions involved increasing or decreasing the H1 amplitude, with a 2 dB increment per step, including "m12" (i.e., H1-H2 decreased from that of the original signal by 12 dB), "m10", "m8", "m6", "m4", "m2", "p2" (i.e., H1-H2 amplitude difference increased by 2 dB), "p4", "p6", "p8", "p10", and "p12". Consequently, there were 312 tokens (2 vowels X 3 H1-H2 levels X 2 speaker genders X 2 voice status X 13 signal conditions) included in the "breathiness rating" task. To avoid the potential fatiguing effect of a long testing procedure, each listener was asked to listen to samples of only one voice status (i.e., either normal speakers or voice patients), which contained a total of 156 tokens.

3.3.2 Vowel Identification Task

To select vowel tokens for use in the “vowel identification” task, the 60-ms mid portion of 90 vowels (3 vowels X 30 speakers) was segmented out from words in the third sentence of the Rainbow passage (i.e., “These take the shape of a long round arch, with its path high above beyond the horizon.”). The three vowels were long vowel /i/ (embedded in “these”), /a/ (“beyond”), and open /o/ (“long”). The words where the selected vowels were embedded were chosen because the vowels in these words did not immediately precede or follow a voiceless consonant and thus were produced in the same phonetic context where continuous voicing was normally expected. Results from an acoustic analysis of these vowels revealed that the H1-H2 measures ranged from -11.7 to 17.4 dB, with a mean of 0.96 dB (SD = 5.58). With normal and pathological voice samples combined, the signals showing the lowest, median, and highest H1-H2 values in females and males separately were used to represent three H1-H2 levels, namely, “Low”, “Mid”, and “High”.

Table 3.2 shows the H1-H2 amplitude difference for the original signals selected for use in the “vowel identification” task. These 18 vowel signals (3 vowels X 3 H1-H2 levels X 2 speaker genders) were submitted to 12 signal manipulation conditions as previously described. Consequently, there were 234 tokens (3 vowels X 3 H1-H2 levels X 2 speaker genders X 13 signal conditions) used in the “vowel identification” task. Each listener was asked to listen to samples of only one speaker gender, which contained a total of 117 tokens.

Table 3.1 The H1-H1 amplitude difference for the original signals used in the "breathiness rating" task.

H1-H1 level	Female		Male	
	/i/	/a/	/i/	/a/
Normal Speakers				
Low	-2.8	0.6	-3.7	0.2
Mid	2.7	3.2	-0.7	4.3
High	5.7	4.2	1.2	5.9
Voice Patients				
Low	0.9	-3.9	-9.5	-2.4
Mid	11.4	5.1	3.3	4.0
High	21.8	14.7	20.5	12.8

Table 3.2 The H1-H1 amplitude difference for the original signals used in the "vowel identification" task.

H1-H1 level	Female			Male		
	/i/	/a/	/o/	/i/	/a/	/o/
Low	-10.8	-3.9	-3.9	-10.0	-11.7	-4.0
Mid	1.7	2.5	1.3	-2.9	-0.2	0.1
High	8.5	17.4	14.8	12.6	13.1	10.5

3.4 Instrumentation

For signal manipulation, a locally developed algorithm written in MATLAB 7 (The Mathworks, Inc.) was used to perform adjustments to the H1 amplitude to match the 12 signal manipulation conditions as previously described. For each vowel, H1 was selected for amplitude scaling. The fast Fourier transform (FFT) was calculated. The modified amplitude for the H1 component and the original amplitudes for all other components were used to synthesize a new waveform. Original phases were used for all components. The process was repeated for the other scaled amplitudes. For signal playback during the perceptual experiment, a pair of headphones (Sony SHN9500) was connected to an Asus M51Vseries laptop. A locally developed computer algorithm was used to present the stimuli in a predetermined random sequence, provide a user interface, and record the participant's response onto a text file, which was later imported into an Excel spreadsheet.

3.5 Procedures

Participants were seated in the sound booth. During the experiment, after the headphones have been placed over the participant's ears and the volume adjusted to the participant's comfort level, the participant was asked to commence the assigned task. Participants were randomly assigned to each of four listener groups, with five females and five males in each group. As shown in Figure 3.3, listeners in each listener group were asked to listen to samples of only one speaker gender in the "breathiness rating" task and only one voice status (i.e., either normal speakers or voice patients) in the "vowel identification" task. This balanced design allowed for reducing the number of tokens each listener was required to listen to minimize the aforementioned potential fatigue effect.

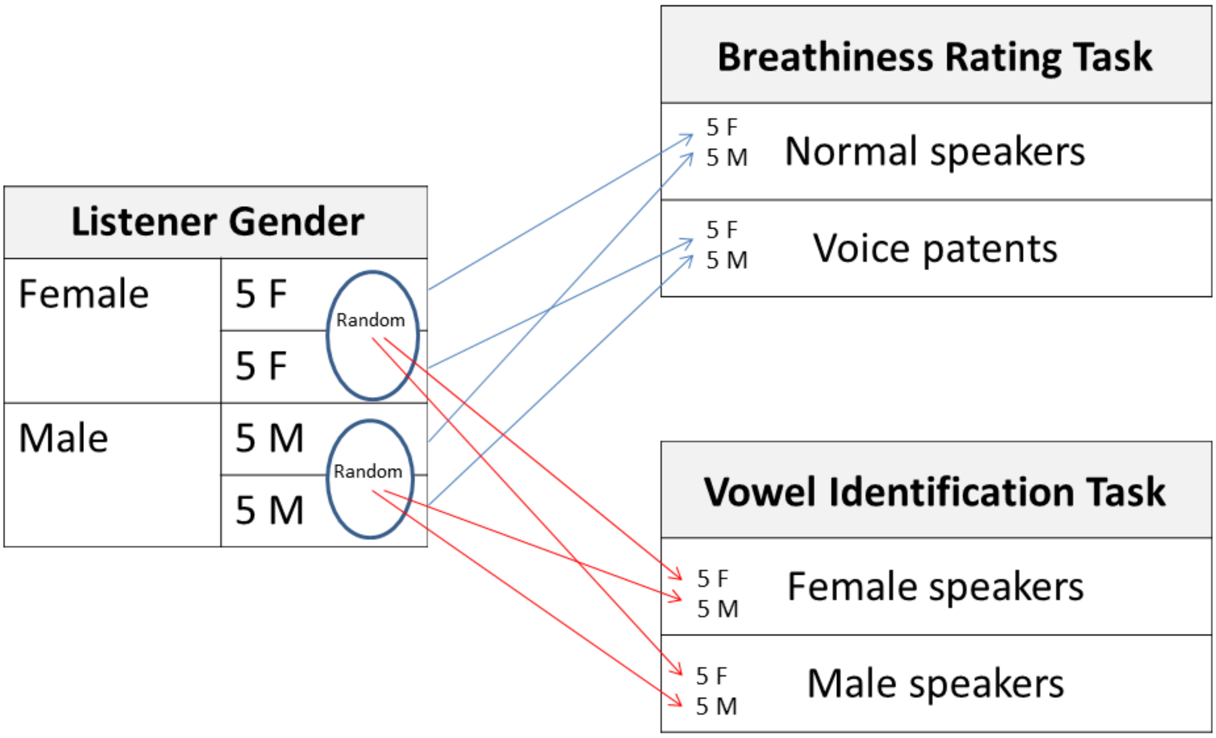


Figure 3.3 Assignment of listeners to the participants' tasks

3.6 Statistical Analysis

For the “breathiness rating” task, the breathiness scores for the voice of normal speakers and voice patients were analyzed separately. A five-way (2 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions X 2 listener genders) Mixed Model Analysis of Variance (ANOVA) was conducted for each data set, with vowel, speaker gender, H1-H2 level, and signal condition treated as within-subjects factors and listener gender as a between-subjects factor.

For the “vowel identification” task, the counts of incorrect vowel identification for female and male voice were analyzed separately. A four-way (3 vowels X 3 H1-H2 levels X 13 signal conditions X 2 listener genders) Mixed Model ANOVA was conducted on the count of incorrect vowel identification made by individual participants, with vowel, H1-H2 level, and signal condition treated as within-subjects factors and listener gender as a between-subjects factor. The significance level was set at 0.05. The SPSS statistical software (Version 19) was used for statistical analysis.

3.7 Reliability

The intra-subject reliability in rating breathiness was assessed by having each listener repeating the “breathiness rating” task on a random selection of 25% (39 out of 156) of the test tokens. A series of Pearson’s correlation procedures conducted on the test and re-test breathiness scores revealed that five (3 females and 2 males) out of 20 listeners failed to show a significant test-retest correlation (see Table 3.3). A series of paired t tests conducted on the test and re-test breathiness scores revealed that four (3 females and 1 male) exhibited a significant test-retest difference (see Table 3.4). Participants who showed both a significant correlation and no significant difference between test-retest breathiness scores were considered to meet the test-retest reliability criteria (i.e., rating breathiness in a reliable manner). Out of the 20 participants, 12 (60%) were found to pass the criteria. The correlations between test and retest breathiness scores in these 12 participants (i.e., NF1, NF5, NM1, NM2, NM3, NM4, NM5, PF1, PF2, PF5, PM1, and PM2) ranged from 0.4 to 0.88, with a mean value of 0.63 (SD = 0.14).

Table 3.3 Test-retest reliability for breathiness ratings: Pearson's correlations. Data for participants who failed to show a significant test-retest correlation were marked in bold face.

Female listeners				Male listeners			
Subject Code	Subject Code	r	p	Subject Code	Subject Initials	r	p
Voice samples from normal speakers							
NF1	EH	0.61	< 0.001	NM1	GR	0.69	< 0.001
NF2	**AS**	**0.16**	**0.342**	NM2	JD	0.80	< 0.001
NF3	**EG**	**0.19**	**0.243**	NM3	JH	0.40	0.011
NF4	AG	0.64	< 0.001	NM4	JW	0.58	< 0.001
NF5	RP	0.58	< 0.001	NM5	HP	0.55	< 0.001
Voice samples from voice patients							
PF1	AM	0.88	< 0.001	PM1	IS	0.68	< 0.001
PF2	KT	0.43	0.007	PM2	DB	0.75	< 0.001
PF3	SG	0.40	0.012	**PM3**	**JS**	**0.28**	**0.081**
PF4	**AD**	**0.09**	**0.588**	PM4	GV	0.65	< 0.001
PF5	EK	0.65	< 0.001	**PM5**	**BW**	**-0.08**	**0.634**

Table 3.4 Test-retest reliability for breathiness ratings: Paired t tests. Data for participants who showed a significant test-retest difference were marked in bold face.

Female listeners				Male listeners			
Subject Code	Subject Code	t (38)	p	Subject Code	Subject Initials	t (38)	p
Voice samples from normal speakers							
NF1	EH	1.213	0.233	NM1	GR	-0.782	0.439
NF2	AS	0.199	0.843	NM2	JD	-0.109	0.914
NF3	**EG**	**-4.061**	**< 0.001**	NM3	JH	1.593	0.120
NF4	**AG**	**-2.180**	**0.036**	NM4	JW	0.369	0.714
NF5	RP	-1.88	0.068	NM5	HP	1.321	0.195
Voice samples from voice patients							
PF1	AM	-0.398	0.693	PM1	IS	1.577	0.128
PF2	KT	0.086	0.932	PM2	DB	-0.488	0.628
PF3	**SG**	**-4.025**	**< 0.001**	PM3	JS	0.348	0.729
PF4	AD	0.177	0.861	**PM4**	**GV**	**3.067**	**0.004**
PF5	EK	0.626	0.535	PM5	BW	-0.168	0.868

4. RESULTS

This chapter presents statistical results for data collected from the “breathiness rating” and the “vowel identification” tasks.

4.1 Perceptual Ratings of Breathiness

For the perceptual ratings of breathiness, each listener was instructed to listen to signals originally obtained from either normal speakers or voice patients alone but not both. Therefore, the breathiness ratings on the voice samples obtained from normal speakers and voice patients were analyzed separately.

4.1.1 Normal Speakers

This section includes results for the breathiness ratings on normal voice based on the perception of 1) all listeners and 2) only the 12 listeners who have met the test-retest reliability criterion as defined in the previous chapter.

4.1.1.1 All Listeners

Tables 4.1 and 4.2 show the means of all female (Table 4.1) and male listeners’ (Table 4.2) breathiness ratings on the normal voice samples as grouped by speaker gender, H1-H2 level, and vowel type, and signal condition. The corresponding standard deviations are shown in Appendices 1 and 2. Table 4.3 shows the results of the five-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions X 2 listener genders) Mixed Model ANOVA conducted on the breathiness scores for the voice obtained from normal speakers. As shown in Table 4.3, there were significant speaker gender, H1-H2 level, speaker gender by H1-H2 level, vowel by H1-H2 level interaction, speaker gender by H1-H2 level by listener gender interaction, speaker gender by H1-H2 level by vowel interaction, speaker gender by signal condition by vowel interaction, and vowel by H1-H2 level by signal condition interaction effects.

Figures 4.1 to 4.4 illustrate the simple effects of signal condition, speaker gender, vowel, H1-H2 level, and listener gender across various levels of selected factors. Specifically, Figure 4.1 shows the means and standard errors of breathiness scores for each of the 13 each signal conditions across speaker genders and vowels, with all H1-H2 levels and listener genders combined. Figure 4.2 shows the means and standard errors of breathiness scores for each of the 13 signal conditions across vowels and H1-H2 levels, with all speaker genders and listeners combined. Figure 4.3 shows the means and standard errors of breathiness scores for each of the three H1-H2 levels across speaker genders and vowels, with all signal conditions and listener genders combined. Figure 4.4 shows the means and standard errors of breathiness scores for each of the three H1-H2 levels across speaker genders and listener genders, with all vowels and signal conditions combined.

Table 4.1 Means of **female listeners'** (n = 5) breathiness ratings (in %) on the **normal speakers'** voice samples as grouped by speaker genders, H1-H2 levels (L1: lowest H1-H2 amplitude difference, L3: highest), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	34.81	75.25	24.10	2.85	17.15	16.40	45.49	70.13	42.58	43.74	31.71	23.59
m10	36.49	81.76	29.62	5.44	12.55	16.82	54.64	50.41	62.63	61.67	57.25	30.13
m8	38.40	81.00	41.59	5.94	18.24	50.38	55.48	53.73	35.82	61.78	42.68	35.65
m6	23.93	80.50	29.96	12.89	37.24	18.95	49.32	74.96	52.05	56.18	41.08	41.09
m4	39.06	75.48	9.12	5.10	10.96	13.15	68.96	67.23	51.55	47.53	49.96	33.74
m2	35.40	81.17	23.01	29.04	15.17	27.86	72.64	56.99	50.71	51.07	40.59	16.90
Original	23.60	79.61	31.07	9.37	30.71	7.62	43.61	75.15	37.32	59.30	37.08	34.98
p2	33.89	80.77	22.76	15.57	23.85	21.00	50.49	68.00	43.01	58.09	52.89	27.11
p4	35.40	78.83	43.76	10.04	37.07	25.33	60.09	70.96	47.95	64.69	48.66	51.21
p6	29.88	79.84	46.53	18.79	29.88	41.17	62.68	75.31	43.60	60.58	42.77	44.30
p8	35.56	84.50	47.36	22.59	15.23	58.31	69.71	71.80	47.47	62.84	43.35	39.16
p10	47.38	75.81	50.22	9.62	23.60	48.70	70.38	82.26	35.53	45.19	47.49	46.95
p12	55.48	66.11	64.10	21.59	28.81	44.35	65.27	74.14	50.53	63.68	47.28	32.15

Table 4.2 Means of **male listeners'** (n = 5) breathiness ratings (in %) on the **normal speakers'** voice samples as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	41.93	61.92	59.75	32.97	33.47	45.65	72.30	66.11	57.16	63.43	62.10	53.39
m10	46.53	64.43	47.36	24.94	40.97	51.63	61.67	64.89	67.28	59.08	57.57	56.32
m8	44.68	66.39	50.21	18.33	38.42	36.32	70.47	68.24	67.78	61.92	55.15	59.50
m6	50.04	58.49	54.94	32.07	34.47	30.54	69.74	75.40	42.51	57.91	52.97	52.89
m4	42.59	56.32	46.87	20.05	35.59	33.49	65.61	72.38	62.46	50.54	51.37	51.30
m2	40.31	50.63	64.27	28.01	19.00	38.33	72.13	66.86	54.89	60.66	57.74	56.99
Original	39.21	58.91	59.02	29.81	27.78	44.27	65.10	78.41	63.01	61.34	49.46	53.22
p2	38.41	53.06	51.22	44.18	23.35	49.87	75.23	69.37	56.99	56.11	57.40	57.57
p4	37.41	56.42	40.42	26.19	56.48	47.95	57.57	69.30	46.36	50.54	53.84	37.91
p6	29.79	52.69	48.53	32.30	47.19	43.76	60.63	70.27	64.85	45.91	54.31	53.89
p8	48.70	63.97	47.78	32.47	19.83	42.27	63.60	63.68	44.02	55.23	48.03	56.40
p10	52.55	65.02	51.81	22.85	38.07	44.37	64.85	64.94	55.68	45.69	64.18	54.98
p12	59.75	60.19	46.11	34.82	44.94	45.86	61.84	65.25	40.58	68.47	52.05	54.81

Table 4.3 Results of the five-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions X 2 listener gender) Mixed Model ANOVA conducted on all listeners' breathiness ratings of **normal speakers'** voice samples.

Effect	F	Hypothesis df	Error df	p	η_p^2
Speaker Gender (SG)	17.83	1	8	0.003*	0.69
H1-H2 Level (L)	15.49	2	16	< 0.001*	0.66
Signal Condition (C)	0.65	12	96	0.795	0.08
Vowel Type (V)	3.08	1	8	0.117	0.28
Listener Gender (LG)	3.87	1	8	0.085	0.33
SG x L	4.02	2	16	0.039*	0.33
SG x C	1.19	12	96	0.303	0.13
SG x V	0.52	1	8	0.491	0.06
SG x LG	0.00	1	8	0.992	0.00
L x C	1.26	24	192	0.193	0.14
L x V	18.50	2	16	< 0.001*	0.70
L x LG	1.92	12	16	0.178	0.19
C x V	1.14	12	96	0.338	0.13
C x LG	1.13	12	96	0.349	0.12
V x LG	2.35	1	8	0.164	0.23
SG x L x C	1.11	24	192	0.334	0.12
SG x L x V	20.58	2	16	< 0.001*	0.72
SG x L x LG	4.00	2	16	0.039*	0.33
SG x C x V	1.95	12	96	0.038*	0.20
SG x C x LG	0.89	12	96	0.564	0.10
SG x V x LG	0.89	1	8	0.372	0.10
L x C x V	1.59	24	192	0.047*	0.17
L x C x LG	0.70	24	192	0.852	0.08
L x V x LG	3.14	2	16	0.071	0.28
C x V x LG	0.69	12	96	0.758	0.08
SG x L x C x V	0.82	24	192	0.704	0.09
SG x L x C x LG	1.12	24	192	0.324	0.12
SG x L x V x LG	1.65	2	16	0.224	0.17
SG x C x V x LG	1.26	12	96	0.255	0.14
L x C x V x LG	1.53	24	192	0.063	0.16
SG x L x C x V x LG	0.99	24	192	0.488	0.11

* Significant at the 0.05 level.

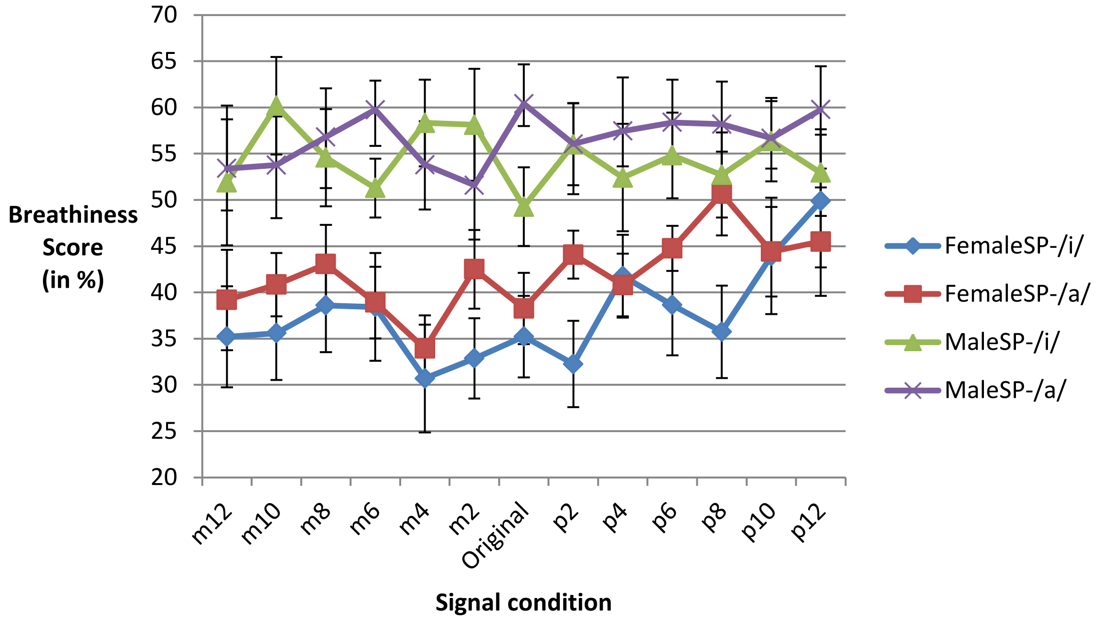


Figure 4.1 Means and standard errors of breathiness scores across 13 signal conditions for the vowels /i/ and /a/ obtained from female (FemaleSp) and male (MaleSp) **normal speakers**, with all H1-H2 levels and listener genders combined.

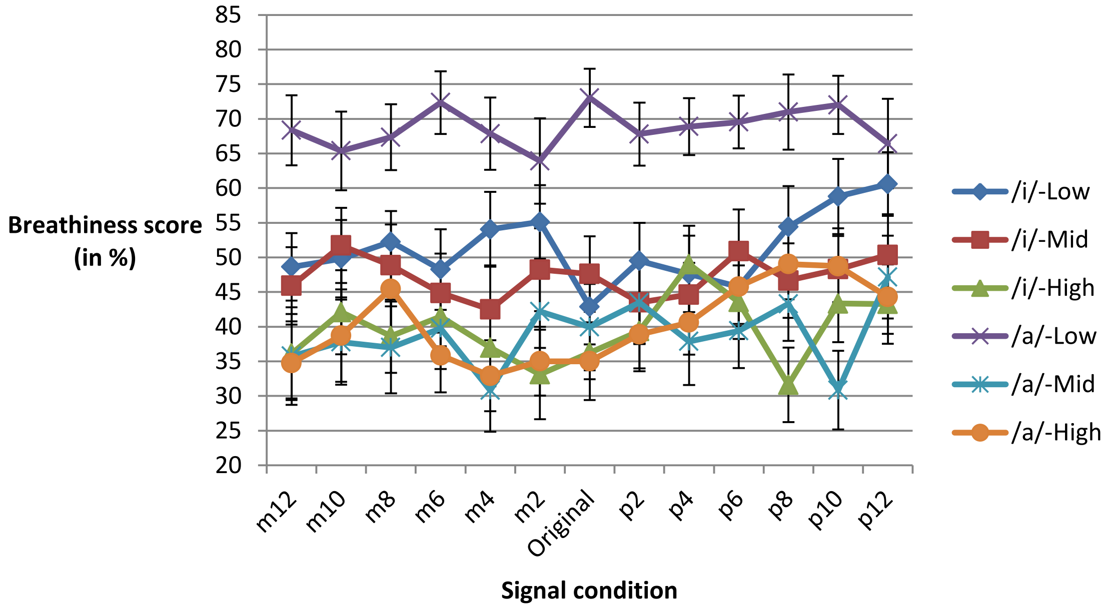


Figure 4.2 Means and standard errors of breathiness scores across 13 signal conditions for the vowels /i/ and /a/ obtained from **normal speakers** at three H1-H2 levels (Low, Mid, and High), with all speaker genders and listener genders combined.

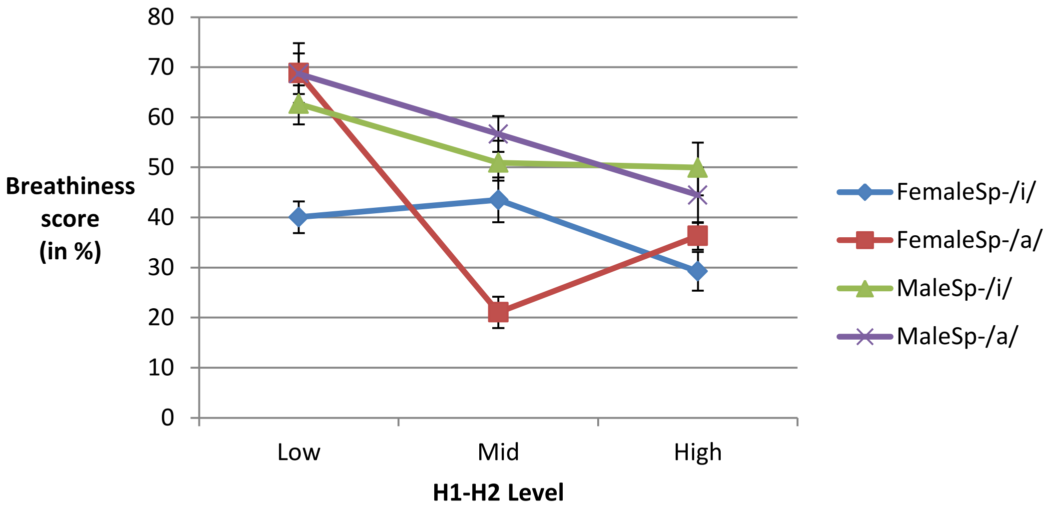


Figure 4.3 Means and standard errors of breathiness scores across three H1-H2 levels for the vowels /i/ and /a/ obtained from female (FemaleSp) and male (MaleSp) **normal speakers**, with all signal conditions and listener genders combined.

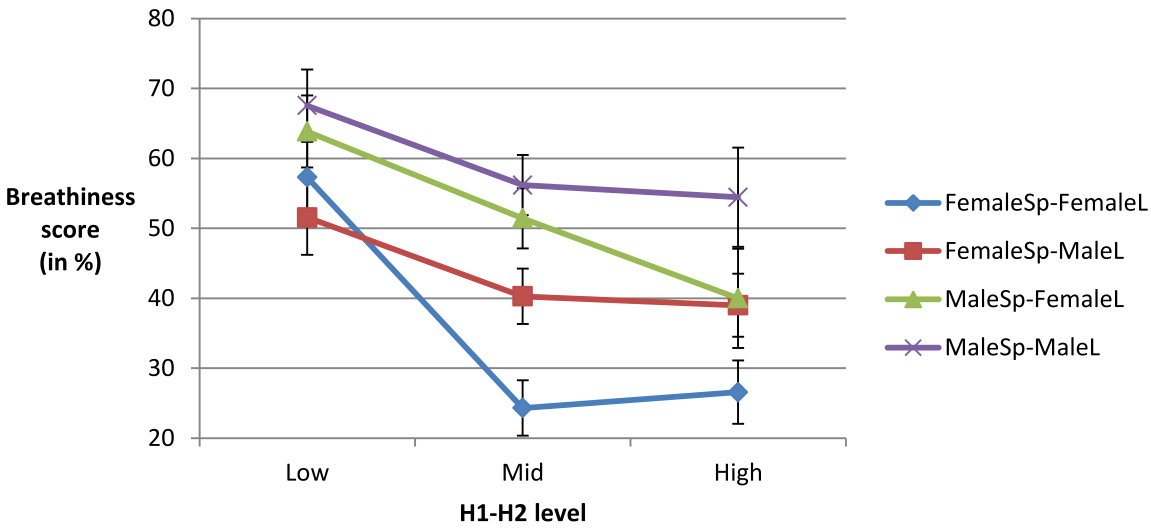


Figure 4.4 Means and standard errors of female and male listeners' (FemaleL and MaleL) breathiness ratings across three H1-H2 levels (Low, Mid, and High) for signals obtained from female (FemaleSp) and male (MaleSp) **normal speakers**, with all signal conditions and vowels combined.

Signal Condition

With all H1-H2 levels and listener genders combined, female voice samples showed a tendency of breathiness scores increasing with increasing H1-H2 amplitude difference (i.e., signal condition going from “m12” toward “p12”) while male voice samples showed no clear relationship between breathiness score and signal condition (see Figure 4.1). Specifically, for female /i/, there was a significant quadratic trend of breathiness scores changing as a function of signal condition [$F(1, 29) = 9.079$, $p = 0.008$]. With only the original signal condition and all the signal conditions above it (i.e., “p2”, “p4”, “p6”, “p8”, “p10”, and “p12”) included in the analysis of female /i/, there was a significant linear increase of perceived breathiness with increasing H1-H2 [$F(1, 29) = 7.968$, $p = 0.009$]. For the female /a/, although no significant linear trend was found between breathiness score and signal condition, the “p8” signal condition showed a significantly higher mean breathiness score than the “m4” signal condition, which was associated with a lower H1-H2 value than the “p8” signal condition (see Figure 4.1).

H1-H2 Level

With all speaker genders and listener genders combined, there was a significant trend of breathiness scores decreasing, for the /a/ only, with an increase of H1-H2 level (i.e., H1-H2 values decreasing from “Low” to “High”) in all signal conditions ($p < 0.05$). However, as shown in Figure 4.2, a significant difference between H1-H2 levels was found only between the “Low” H1-H2 level and the other two H1-H2 levels (“Mid” and “High”).

With all signal conditions and listener genders combined, there was a significant trend of breathiness scores decreasing with an increase of H1-H2 levels across all speaker gender by vowel groups, including female /i/ [$F(1, 129) = 19.04$, $p < 0.001$], female /a/ [$F(1, 129) = 123.7$, $p < 0.001$], male /i/ [$F(1, 129) = 27.84$, $p < 0.001$], and male /a/ [F(1,

129) = 75.73, $p < 0.001$] (see Figure 4.3). The pairwise comparisons between H1-H2 levels on mean breathiness scores were all significant except that the difference between "Low" and "Mid" H1-H2 levels in female /i/ and that between "Mid" and "High" H1-H2 levels in male /i/ were not significant (see Figure 4.3).

With all signal conditions and vowels combined, there was also a significant trend of female listener to perceive less breathiness with an increase of H1-H2 levels, including female voice [$F(1, 19) = 9.303$, $p = 0.014$] and male voice [$F(1, 19) = 8.125$, $p = 0.019$]. In contrast, no significant H1-H2 level effect on breathiness scores was found in male listeners' perception of female voice [$F(1, 19) = 2.098$, $p = 0.181$] and male voice [$F(1, 129) = 1.972$, $p = 0.194$] (see Figure 4.4).

Vowel

With all H1-H2 levels and listener genders combined, /a/ showed a significantly higher mean breathiness score than /i/ in the female voice at the "p8" signal condition and in the male voice at the original signal condition (see Figure 4.1).

With all speaker genders and listener genders combined, /a/ showed a higher mean breathiness score than /i/ in all signal conditions at the "Low" H1-H2 level, with the between-vowel difference reaching a significant level in all signal conditions except for the "m10", "m4", "m2", and "p12" signal conditions (see Figure 4.2). At the "Mid" and "High" H1-H2 levels, the between-vowel difference did not reach a significant level except that the mean breathiness score for /a/, compared to /i/, was significantly higher in the "p8" signal condition at the "High" H1-H2 level and significantly lower in the "m4" and "p10" signal conditions at the "Mid" H1-H2 level (see Figure 4.2).

With all signal conditions and listener genders combined, both female and male voices yielded a significantly higher mean breathiness score for /a/ than for /i/ at the "Low" H1-H2 level. However, the mean breathiness score for /a/, compared to /i/, was significantly lower at the “Mid” H1-H2 level and significantly higher at the “High” H1-H2 level in female voice but the reverse was true in male voice (see Figure 4.3).

Speaker Gender

With all H1-H2 levels and listener genders combined, male voice was generally associated with a higher mean breathiness score than female voice. Specifically male voice showed a significantly higher mean breathiness score than female voice in all signal conditions except that no significant gender difference was found in the “p4” and “p12” signal conditions for the vowel /i/ and in the “m10”, “m8”, “m2”, “p8”, and “p10” signal conditions for the vowel /a/ (see Figure 4.1).

With all signal conditions and listener genders combined, male voice showed a significantly higher mean breathiness score than female voice in all H1-H2 levels for both vowels except for the vowel /a/ at the “Low” H1-H2 level (see Figure 4.3).

With all signal conditions and vowels combined, male voice showed a significantly higher mean breathiness score than female voice regardless of or H1-H2 level or listener gender (see Figure 4.4).

Listener Gender

With all signal conditions and vowels combined, the mean breathiness scores given by male listeners were significantly higher than those by female listeners for female voice at the “Mid” and “High” H1-H2 levels and for male voice at the “High” H1-H2 level (see Figure 4.4)

4.1.1.2 Listeners Showing Reliable Breathiness Scoring Only

A four-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions) Mixed Model ANOVA was conducted on the seven reliable raters’ (i.e., NF1, NF5, NM1, NM2, NM3, NM4, and NM5) breathiness ratings for the voice obtained from normal speakers. As shown in Table 4.4, there were significant speaker gender, H1-H2 level, vowel by H1-H2 level interaction, and speaker gender by vowel by H1-H2 level interaction effects. Figures 4.5 shows the means and standard errors of breathiness scores across H1-H2 levels for the vowels /i/ and /a/ obtained from female and male normal speakers, with all signal conditions combined.

H1-H2 level

With all signal conditions combined, there was a significant trend of the breathiness scores decreasing with an increase of H1-H2 level for all speaker gender by vowel groups, including female /i/ [$F(1, 90) = 19.453$, $p < 0.001$] and /a/ [$F(1, 90) = 69.218$, $p < 0.001$] and male /i/ [$F(1, 90) = 29.998$, $p < 0.001$] and /a/ [$F(1, 90) = 57.93$, $p < 0.001$].

Vowel

With all signal conditions combined, the vowel /a/ showed a significantly higher mean breathiness score than /i/ for female voice at the “Low” and “High” H1-H2 levels and for male voice at the “Low” and “Mid” H1-H2 levels. For female voice at the “Mid” H1-H2 level, however, the vowel /a/ showed a significantly lower mean breathiness score than /i/ (see Figure 4.5).

Speaker Gender

With all signal conditions combined, male voice showed a significantly higher mean breathiness score than female voice in all vowel by H1-H2 level groups except for the vowel /i/ at the “Mid” H1-H2 level, where no significant speaker gender difference was found (see Figure 4.5).

Table 4.4 Results of the four-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions) Mixed Model ANOVA conducted on the seven reliable raters' breathiness ratings of **normal speakers'** voice samples.

Effect	F	Hypothesis df	Error df	p	η_p^2
Speaker Gender (SG)	11.230	1	6	0.015*	0.652
H1-H2 Level (L)	8.196	2	12	0.006*	0.577
Signal Condition (C)	0.491	12	72	0.914	0.076
Vowel Type (V)	2.158	1	6	0.192	0.265
SG x L	0.352	2	12	0.710	0.055
SG x C	1.192	12	72	0.305	0.166
SG x V	0.020	1	6	0.892	0.003
L x C	1.252	24	144	0.209	0.173
L x V	7.790	2	12	0.007*	0.565
C x V	0.894	12	72	0.557	0.130
SG x L x C	1.473	24	144	0.085	0.197
SG x L x V	9.390	2	12	0.004*	0.610
SG x C x V	1.885	12	72	0.051	0.239
L x C x V	1.557	24	144	0.059	0.206
SG x L x C x V	0.621	24	144	0.913	0.094

* Significant at the 0.05 level.

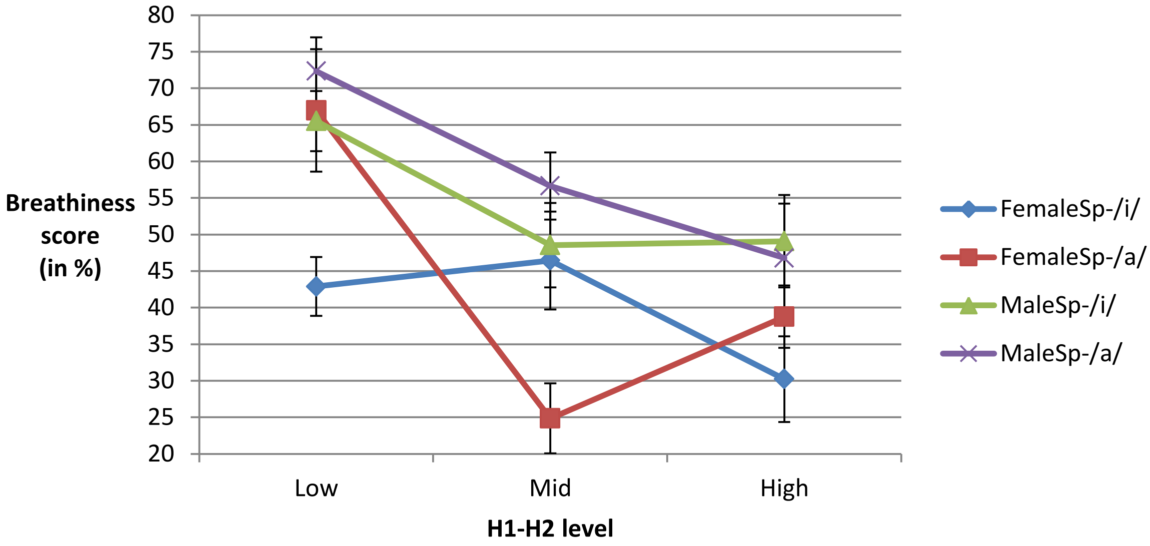


Figure 4.5 Means and standard errors of the seven reliable raters' breathiness ratings across three H1-H2 levels (Low, Mid, and High) for the vowels /i/ and /a/ obtained from female (FemaleSp) and male (MaleSp) **normal speakers**, with all signal conditions combined.

4.1.1.3 Summary

For normal voice, it was evident only in female /i/ that an increase of H1-H2 through signal manipulation, and only when the H1-H2 value was increased from that of the original signal, would result in an increase in perceived breathiness. The "Low" H1-H2 level appeared to result in a higher breathiness score than the other two higher H1-H2 levels, suggesting that within normal voice, an increase of H1-H2 amplitude difference did not lead to an increase in perceived breathiness. Moreover, normal voice with an H1-H2 value at the lower end of the normal range of H1-H2 value may actually be perceived as more breathy than those at the higher end of the range. As for the vowel effect on the perception of breathiness, the vowel /a/ was consistently perceived to be more breathy than /i/ only at the "Low" H1-H2 level. Male voice was generally perceived as more breathy than female voice. Male listeners tended to perceived higher breathiness than female listeners.

4.1.2 Voice Patients

This section includes results for the breathiness ratings on pathological voice based on the perception of 1) all listeners and 2) only the 12 listeners who have met the test-retest reliability criterion.

4.1.2.1 All Listeners

Tables 4.5 and 4.6 show the means of all female (Table 4.5) and male (Table 4.6) listeners' breathiness ratings on the pathological voice samples as grouped by speaker gender, H1-H2 level, and vowel type, and signal condition. The corresponding standard deviations are shown in Appendices 3 and 4. Table 4.7 shows the results of the five-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions X 2 listener genders) Mixed Model ANOVA conducted on the breathiness scores for the voice obtained from voice patients. As shown in Table 4.7, there were significant speaker gender, H1-H2 level, speaker gender by signal condition interaction, speaker gender by vowel interaction, vowel by signal condition interaction, H1-H2 level by signal condition interaction, speaker gender by H1-H2 level by listener gender interaction, and speaker gender by vowel by H1-H2 level by signal condition interaction effects.

With all signal conditions and vowels combined, no significant difference on breathiness scores between female and male listeners was found. With both listener genders combined, Figures 4.6 and 4.7 show the means and standard errors of breathiness scores for vowels /i/ and /a/ at three H1-H2 levels (Low, Mid, and High) across 13 signal conditions in female (Figure 4.6) and male voice respectively (Figure 4.7).

Table 4.5 Means of **female listeners'** (n = 5) breathiness ratings (in %) on the **voice patients'** voice samples as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	49.10	17.57	50.11	32.80	53.45	35.06	45.10	26.70	79.08	74.90	91.21	87.11
m10	34.68	23.93	46.28	34.56	62.43	22.84	30.63	37.15	72.80	81.26	81.99	82.60
m8	34.81	24.91	41.42	31.30	55.26	35.31	36.48	27.28	74.81	80.92	87.03	82.94
m6	19.41	27.49	35.90	44.27	55.48	20.06	31.97	39.58	82.05	86.03	87.28	82.51
m4	37.74	10.77	44.18	28.95	63.52	35.98	30.04	53.64	72.30	85.96	76.07	71.91
m2	36.15	12.65	48.37	59.33	53.89	33.05	37.82	28.77	82.30	78.49	76.32	79.83
Original	15.82	15.06	42.72	39.24	61.51	20.21	39.58	35.31	82.96	82.51	67.36	70.71
p2	31.63	17.24	42.26	39.33	50.96	40.17	45.23	38.83	74.31	71.13	73.05	69.04
p4	20.25	6.86	48.87	54.31	54.81	39.67	45.36	42.68	82.36	68.62	86.22	77.74
p6	23.22	17.32	35.42	38.24	44.84	38.84	38.74	57.28	78.13	74.70	46.95	76.74
p8	30.67	19.75	49.04	57.82	38.84	49.46	30.90	33.98	82.53	73.97	76.24	74.56
p10	34.81	31.55	39.16	51.47	41.42	41.92	39.83	56.57	68.45	81.80	76.99	69.54
p12	30.21	20.26	44.18	66.19	43.26	45.36	32.80	43.67	75.23	81.09	65.69	69.29

Table 4.6 Means of **male listeners'** (n = 5) breathiness ratings (in %) on the **voice patients'** voice samples as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	30.53	15.21	56.48	46.22	54.98	36.77	42.38	44.19	51.78	80.50	89.47	83.86
m10	20.28	11.65	76.90	71.34	69.21	27.11	36.90	45.80	51.01	81.06	79.09	77.53
m8	18.49	16.08	49.70	44.75	48.13	24.43	40.31	57.02	72.47	83.18	78.51	78.85
m6	16.99	19.66	51.88	58.77	45.19	26.70	23.43	60.96	59.14	62.51	73.82	69.90
m4	18.57	21.34	63.72	58.83	63.70	37.56	42.00	51.69	59.12	86.86	86.34	72.57
m2	24.08	15.98	60.43	66.02	43.60	41.83	38.03	60.42	59.53	82.01	88.55	72.39
Original	18.07	25.89	59.50	53.31	55.67	30.46	49.67	45.59	65.62	68.81	78.01	78.83
p2	29.28	14.72	45.92	51.35	40.17	54.95	46.31	56.85	54.48	73.35	60.19	66.55
p4	28.99	15.82	49.27	57.24	34.16	63.30	30.61	48.59	57.94	74.31	71.23	72.35
p6	23.10	15.31	45.72	67.24	34.94	43.14	39.26	55.19	53.63	62.19	42.59	59.94
p8	41.37	20.08	51.01	55.79	27.37	35.40	45.34	54.27	53.37	69.08	53.65	65.55
p10	22.00	21.91	57.63	51.77	41.51	51.82	44.38	59.41	62.43	69.62	58.77	68.54
p12	29.62	25.10	47.66	57.20	32.13	50.04	34.57	60.98	54.81	58.69	50.35	60.20

Table 4.7 Results of the five-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions X 2 listener gender) Mixed Model ANOVA conducted on all listeners' breathiness ratings of **voice patients'** voice samples.

Effect	F	Hypothesis df	Error df	p	η_p^2
Speaker Gender (SG)	17.20	1	8	0.003*	0.68
H1-H2 Level (L)	24.61	2	16	< 0.001*	0.76
Signal Condition (C)	0.85	12	96	0.598	0.10
Vowel Type (V)	0.01	1	8	0.913	0.00
Listener Gender (LG)	0.01	1	8	0.925	0.00
SG x L	3.03	2	16	0.077	0.28
SG x C	1.95	12	96	0.038*	0.20
SG x V	17.93	1	8	0.003*	0.69
SG x LG	0.27	1	8	0.617	0.03
L x C	2.35	24	192	0.001*	0.23
L x V	2.63	2	16	0.103	0.25
L x LG	0.21	2	16	0.816	0.03
C x V	2.72	12	96	0.003*	0.25
C x LG	0.49	12	96	0.915	0.06
V x LG	3.68	1	8	0.091	0.32
SG x L x C	1.47	24	192	0.080	0.16
SG x L x V	1.03	2	16	0.381	0.11
SG x L x LG	7.59	2	16	0.005*	0.49
SG x C x V	1.62	12	96	0.099	0.17
SG x C x LG	1.20	12	96	0.297	0.13
SG x V x LG	0.23	1	8	0.644	0.03
L x C x V	2.17	24	192	0.002*	0.21
L x C x LG	1.25	24	192	0.202	0.14
L x V x LG	0.03	2	16	0.975	0.00
C x V x LG	0.55	12	96	0.875	0.07
SG x L x C x V	1.94	24	192	0.008*	0.20
SG x L x C x LG	0.68	24	192	0.869	0.08
SG x L x V x LG	1.31	2	16	0.298	0.14
SG x C x V x LG	0.56	12	96	0.871	0.07
L x C x V x LG	0.89	24	192	0.616	0.10
SG x L x C x V x LG	0.73	24	192	0.820	0.08

* Significant at the 0.05 level.

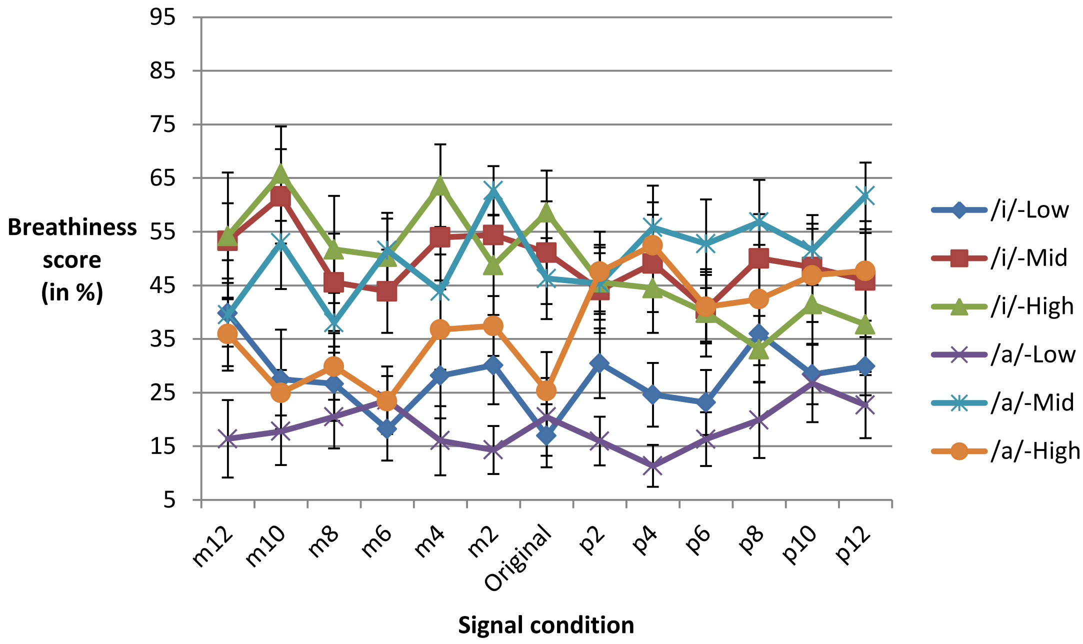


Figure 4.6 Means and standard errors of breathiness scores across three H1-H2 levels (Low, Mid, and High) for vowels /i/ and /a/ obtained from **female voice patients**, with both listener genders combined.

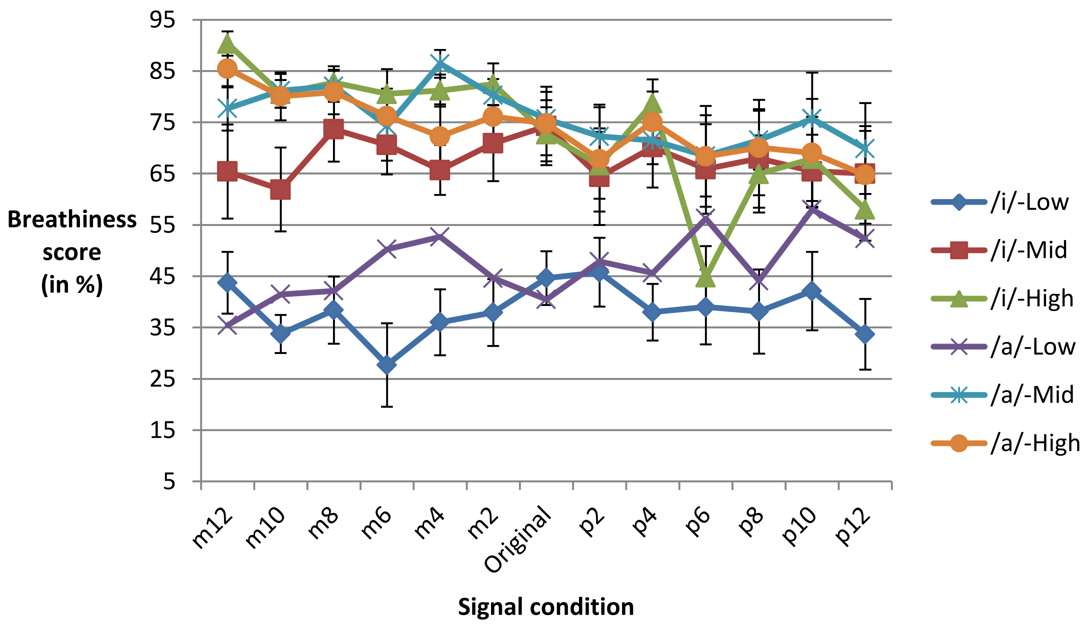


Figure 4.7 Means and standard errors of breathiness scores across three H1-H2 levels (Low, Mid, and High) for vowels /i/ and /a/ obtained from **male voice patients**, with both listener genders combined.

Signal Condition

With both listener genders combined, there was a significant linear trend of breathiness scores decreasing as the signal condition rose from “m12” (lowest H1-H2) to “p12” (highest H1-H2) at the “High” H1-H2 level for the vowel /i/ in both female [$F(1, 9) = 3.905$, $p = 0.08$; see Figure 4.6] and male voice patients [$F(1, 9) = 23.146$, $p = 0.001$; see Figure 4.7].

H1-H2 Level

With both listener genders combined, there was a significant linear trend of breathiness scores increasing with an increase of H1-H2 level for the vowel /i/ obtained from female (see Figure 4.6) and male voice patients (see Figure 4.7) except for the “m12”, “m2”, “p2”, “p6”, “p8”, “p10”, and “p12” signal conditions in female voice (see Figure 4.6) and the “m6” signal condition in male voice. For the vowel /a/, there was also a significant linear trend of breathiness scores increasing with an increase of H1-H2 level for the “m2”, “p2”, “p4”, “p6” signal conditions in female pathological voice (see Figure 4.6) and all the signal conditions in male pathological voice except for the “m4”, “p2”, “p6”, “p10”, and “p12” signal conditions (see Figure 4.7).

Vowel

With both listener genders combined, the vowel /i/ showed a significantly higher mean breathiness score than /a/ for female pathological voice in the “m12”, “m2”, “p2”, “p4”, and “p8” signal conditions at the “Low” H1-H2 level and in the “m10”, “m8”, “m6”, “m4”, and original signal conditions at the “High” H1-H2 level (see Figure 4.6). In contrast, the vowel /i/ showed a significantly lower mean breathiness score than /a/ for male pathological voice in the “m6”, “p6”, and “p10” signal conditions at the “Low” H1-H2 level and the “m4” and “p10” signal conditions at the “Mid” H1-H2 level (see Figure 4.7).

Speaker Gender

With both listener genders combined, male pathological voice received a higher mean breathiness score than female pathological voice across all vowels by signal condition groups. This speaker gender difference reached the significance level ($p < 0.05$) in the original signal condition at the "Low" H1-H2 level, the "m8" and "m6" signal conditions at the "Mid" H1-H2 level, and the "m12", "m8", "m6", "m4", "m2", "p2", "p4", and "p8" signal conditions at the "High" H1-H2 level. For the vowel /a/, the speaker gender difference was significant across all signal conditions ($p < 0.05$) and H1-H2 levels except for the "m8" and original signal conditions at the "Low" H1-H2 level, the "p4", p6", "p8", and "p12" signal conditions at the "Mid" H1-H2 level, the "p2", "p4", "p6", "p8", "p10", and "p12" signal conditions at the "High" H1-H2 level.

4.1.2.2 Listeners Showing Reliable Breathiness Scoring Only

A four-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions) Mixed Model ANOVA was conducted on the five reliable raters' (i.e., PF1, PF2, PF5, PM1, and PM2) breathiness ratings for the voice obtained from voice patients. As shown in Table 4.8, there were significant H1-H2 level, speaker gender by vowel interaction, and H1-H2 level by signal condition interaction effects. Figure 4.8 shows the means and standard errors of breathiness scores across signal conditions and H1-H2 levels for voice patients' voice with all speaker genders and vowels combined. Figure 4.9 shows the means and standard errors of breathiness scores for the vowels /i/ and /a/ produced by female and male voice patients with all signal conditions and H1-H2 levels combined.

Signal Condition

With all speaker genders and vowels combined, there was a significant trend of breathiness scores decreasing as signal condition rose from "m12" (lowest H1-H2) to "p12" (highest H1-H2) for pathological voice at the "High" H1-H2 level (see Figure 4.8). This finding suggests that an increase of H1-H2 in pathological voice through signal manipulation did not result in the expected increase of perceived breathiness. On the contrary, with pathological voice showing the highest H1-H2 level (i.e., "High" H1-H2 level) among the chosen samples, an increase of H1-H2 through signal manipulation actually resulted in a decrease of perceived breathiness.

H1-H2 Level

With all speaker genders and vowels combined, there was a significant trend of breathiness scores increasing with an increase of H1-H2 levels for pathological voice at the signal conditions with a H1-H2 value smaller that of the "p6" signal condition, including "m12" [$F(1, 19) = 22.882$, $p < 0.001$], "m10" [$F(1, 19) = 28.432$, $p < 0.001$], "m8" [$F(1, 19) = 29.969$, $p < 0.001$], "m6" [$F(1, 19) = 15.357$, $p = 0.001$], "m4" [$F(1, 19) = 15.346$, $p < 0.001$], "m2" [$F(1, 19) = 38.155$, $p < 0.001$], original [$F(1, 19) = 17.774$, $p < 0.001$], "p2" [$F(1, 19) = 17.271$, $p = 0.001$], and "p4" signal conditions [$F(1, 19) = 38.061$, $p < 0.001$]. For all of these signal conditions, the "Low" H1-H2 level showed a significantly lower mean breathiness score than both "Mid" and "High" H1-H1 levels (see Figure 4.8).

For all the other signal conditions, there was also a significant H1-H2 level effect, including the "p6" [$F(2, 38) = 5.248$, $p = 0.01$], "p8" [$F(2, 38) = 7.236$, $p = 0.002$], "p10" [$F(2, 38) = 3.339$, $p = 0.046$], and "p12" signal conditions [$F(2, 38) = 5.667$, $p = 0.007$], where the "Low" signal condition showing a significantly lower breathiness score than the "Mid" H1-H2 level only (see Figure 4.8).

These findings indicate that 1) the original pathological voice with a lower H1-H2 value was perceived to be less breathy than that with a higher H1-H2 value and that 2) signal manipulation did not affect this relationship except that the difference in perceived breathiness between the "Low" and "High"H1-H2 levels was reduced when the H1-H2 value was increased from that of the original signal by more than 4 dB. For signals at the “Mid” and “High” H1-H2 levels, there was a tendency for breathiness scores to decrease when the H1-H2 value was increased, through signal manipulation, to be higher than that of the original signal.

Vowel

Figure 4.9 illustrates the speaker gender by vowel interaction effect on the breathiness scores. With all H1-H2 levels and signal conditions combined, the mean breathiness score for the vowel /i/, compared to the vowel /a/, was significantly higher in female pathological voice ($t = 4.25$, $df = 194$, $p < 0.001$) but significantly lower in male pathological voice ($t = -2.489$, $df = 194$, $p = 0.014$).

Speaker Gender

With all H1-H2 levels and signal conditions combined, male pathological voice showed a significantly higher mean breathiness score than female pathological voice in both /i/ ($t = 5.564$, $df = 194$, $p < 0.001$) and /a/ ($t = 11.675$, $df = 194$, $p < 0.001$).

Table 4.8 Results of the four-way (3 vowels X 2 speaker genders X 3 H1-H2 levels X 13 signal conditions) Mixed Model ANOVA conducted on the five reliable raters' breathiness ratings of **voice patients'** voice samples.

Effect	F	Hypothesis df	Error df	p	η_p^2
Speaker Gender (SG)	3.963	1	4	0.117	0.498
H1-H2 Level (L)	8.219	2	8	0.011*	0.673
Signal Condition (C)	1.681	12	48	0.101	0.296
Vowel Type (V)	0.226	1	4	0.659	0.054
SG x L	1.969	2	8	0.202	0.330
SG x C	0.749	12	48	0.697	0.158
SG x V	25.768	1	4	0.007*	0.866
L x C	1.786	24	96	0.025*	0.309
L x V	2.635	2	8	0.132	0.397
C x V	0.602	12	48	0.830	0.131
SG x L x C	0.926	24	96	0.567	0.188
SG x L x V	0.222	2	8	0.806	0.053
SG x C x V	1.717	12	48	0.092	0.300
L x C x V	0.530	24	96	0.961	0.117
SG x L x C x V	1.353	24	96	0.153	0.253

* Significant at the 0.05 level.

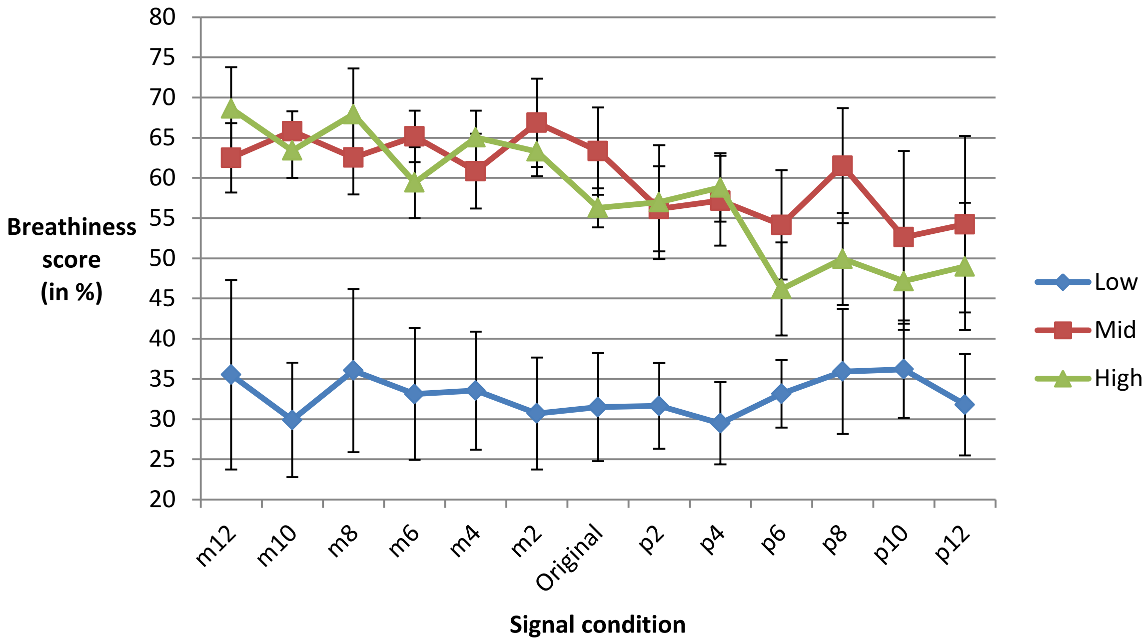


Figure 4.8 Means and standard errors of the five reliable raters' breathiness ratings across 13 signal conditions for signals obtained from **voice patients**, with all speaker genders and vowels combined.

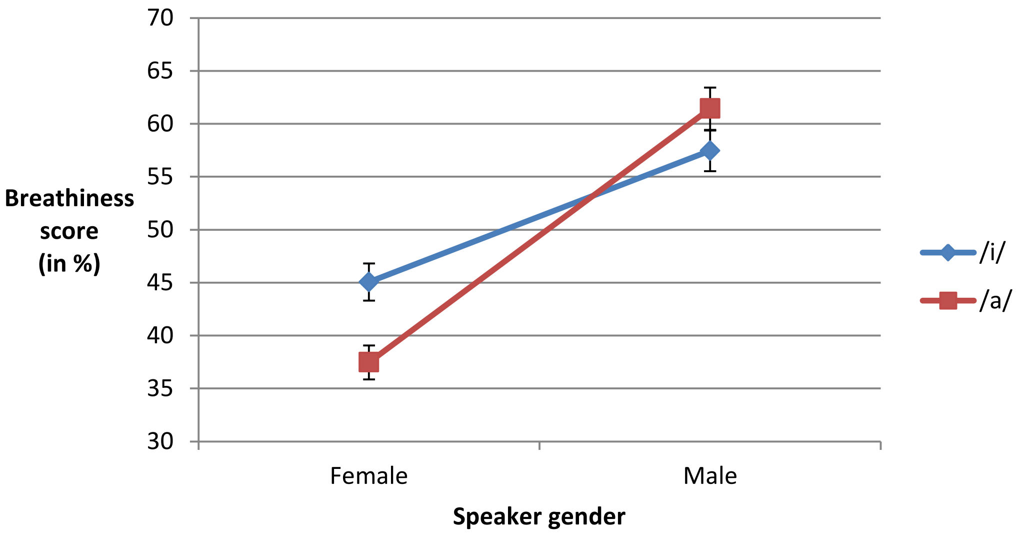


Figure 4.9 Means and standard errors of the five reliable raters' breathiness ratings for the vowels /i/ and /a/ obtained from female and male **voice patients** respectively, with all H1-H2 levels and signal conditions combined.

4.1.2.3 Summary

As for the vowel effect on the perception of breathiness, the vowel /a/ was consistently perceived to be more breathy than /i/ only at the “Low” H1-H2 level. Male voice was generally perceived as more breathy than female voice. Male listeners tended to perceive higher breathiness than female listeners.

For pathological voice with a H1-H2 value at the higher end, an increase of H1-H2 through signal manipulation appeared to result in a reduction of perceived breathiness for the vowel /i/. Perceived breathiness generally increased with an increase of H1-H2 level but this relationship could be disrupted by an increase of H1-H2 through signal manipulation, especially in pathological voice already at a higher H1-H2 level. The vowel /a/ was perceived to be more breathy than /i/ in male voice. However, the vowel /a/ was perceived to be less breathy than /i/ in female voice. Male voice was generally perceived as more breathy than female voice.

4.1.3 Summary

In both normal and pathological voices, male voice tended to receive a higher breathiness score than female voice. An increase of H1-H2 through signal manipulation led to an increase of perceived breathiness only when performed on vowel /i/ produced by female normal speakers. The classification of “low”, “mid”, and “high” H1-H2 level in pathological voice corresponded to the perception of breathiness in the same direction while the three categories in normal voice did not. The vowel /a/ was perceived to be more breathy than /i/ in normal voice generally at the “Low” H1-H2 level and in male pathological voice. Difference between female and male listeners on breathiness ratings was found when rating normal voice but not when rating pathological voice, with male listeners giving higher breathiness scores than female listeners.

4.2 Percentage of Correct Vowel Identification

The counts of vowel misidentification across signal conditions were analyzed for female and male speakers as a whole and separately. To promote ease of visual interpretation, the findings have been represented on the line graph-y axes as ‘vowel misidentification’ rather than correct vowel identification.

4.2.1 Female and Male Voice

Table 4.9 shows the percentages of incorrect vowel identification across speaker genders, H1-H2 levels (Low, Mid, and High), and vowel types (/i/ and /a/). With all vowel types, speaker genders, and H1-H2 levels combined, a significant signal condition effect on the number of incorrect vowel identification was found ($\chi^2 = 188.585$, df = 10, $p < 0.001$).

Table 4.10 shows the correlations between signal condition levels and the count of vowel misidentification across speaker genders and H1-H2 levels, with all signal conditions included (“Overall”), signals conditions above the original signal only, and signals conditions below the original signal only.

Table 4.9 Percentages of incorrect vowel identification (in %) across speaker genders, H1-H2 levels (Low, Mid, and High), and vowel types (/i/ and /a/).

	Female									Male									All H1-H2 levels and vowels
	Low			Mid			High			Low			Mid			High			
	/i/	/a/	/o/	/i/	/a/	/o/	/i/	/a/	/o/	/i/	/a/	/o/	/i/	/a/	/o/	/i/	/a/	/o/	
m12	50	50	90	60	70	50	60	30	60	30	70	50	30	60	30	20	20	40	48.3
m10	50	60	100	80	70	50	50	50	60	40	30	60	60	30	30	40	30	90	54.4
m8	90	20	90	50	60	40	60	90	80	90	20	60	30	100	90	80	50	80	65.6
m6	50	20	100	40	70	70	30	80	60	80	40	50	20	30	40	40	80	60	53.3
m4	60	20	70	40	90	60	40	60	70	60	40	20	10	80	90	90	90	90	60.0
m2	60	80	40	40	90	40	90	60	60	60	70	80	90	10	50	60	50	40	59.5
Original	60	20	30	80	40	60	50	50	50	30	60	30	90	20	60	90	50	90	53.3
p2	70	70	30	50	70	60	70	40	80	70	60	10	50	40	0	70	100	90	57.2
p4	80	80	90	50	60	60	20	30	50	20	40	70	60	30	80	40	100	50	56.1
p6	20	30	60	90	60	50	60	80	80	20	50	90	50	70	80	20	70	50	57.2
p8	70	60	80	40	90	80	60	80	30	80	70	80	70	80	50	40	40	30	62.8
p10	70	60	70	70	60	20	40	70	80	70	50	40	80	90	70	80	50	60	62.8
p12	50	50	20	70	40	40	80	80	40	40	30	30	70	60	80	70	40	10	50.0

Table 4.10 Correlations (Spearman rho) between signal conditions and percentage of correct vowel identification (in %), across speaker genders and H1-H2 levels (Low, Mid, and High), with all signal conditions included ("Overall"), signals conditions associated with H1-H2 amplitude difference greater than the original signals ("Above") only, and signals conditions associated with H1-H2 amplitude difference smaller than the original signals ("Below") only.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	r	p	r	p	r	p	r	p	r	p	r	p
Overall-all	-0.091	0.588	-0.059	0.723	0.090	0.588	0.010	0.951	0.330	0.038*	-0.037	0.822
Overall-/i/	0.147	0.633	0.152	0.621	0.131	0.669	-0.078	0.801	0.520	0.068	0.137	0.655
Overall-/a/	0.292	0.333	-0.356	0.233	0.269	0.375	0.095	0.758	0.205	0.502	0.173	0.571
Overall-/o/	-0.605	0.029*	-0.082	0.791	-0.198	0.516	0.008	0.979	0.294	0.329	-0.408	0.166
Above-all	-0.020	0.932	-0.124	0.592	0.268	0.240	0.103	0.656	0.420	0.058	-0.502	0.020*
Above-/i/	-0.185	0.691	-0.073	0.877	0.270	0.558	0.309	0.500	0.109	0.816	-0.164	0.726
Above -/a/	0.018	0.969	-0.037	0.937	0.667	0.102	-0.455	0.305	0.750	0.052	-0.624	0.134
Above -/o/	-0.018	0.969	-0.556	0.195	-0.262	0.570	0.252	0.585	0.408	0.364	-0.764	0.046*
Below-all	0.268	0.239	-0.084	0.717	-0.010	0.965	0.030	0.898	0.064	0.783	0.510	0.018*
Below-/i/	0.501	0.252	-0.243	0.599	-0.091	0.846	0.018	0.969	0.346	0.448	0.764	0.046*
Below-/a/	-0.217	0.641	0.019	0.968	0.218	0.638	0.255	0.582	-0.541	0.210	0.580	0.172
Below-/o/	-0.764	0.046*	0.220	0.635	-0.335	0.463	-0.218	0.638	0.564	0.187	0.243	0.599

*Showing a significant correlation at the 0.05 level

With all speaker genders and H1-H2 levels combined, the number of incorrect vowel identifications across signal conditions significantly varied by vowel type (Lambda coefficient = 0.032, p = 0.014). Specifically, there were significantly more vowel misidentifications in male voice for vowel /i/ compared to vowel /o/ in the "p12" signal condition and compared to vowel /a/ in the original signal condition (see Figure 4.3).

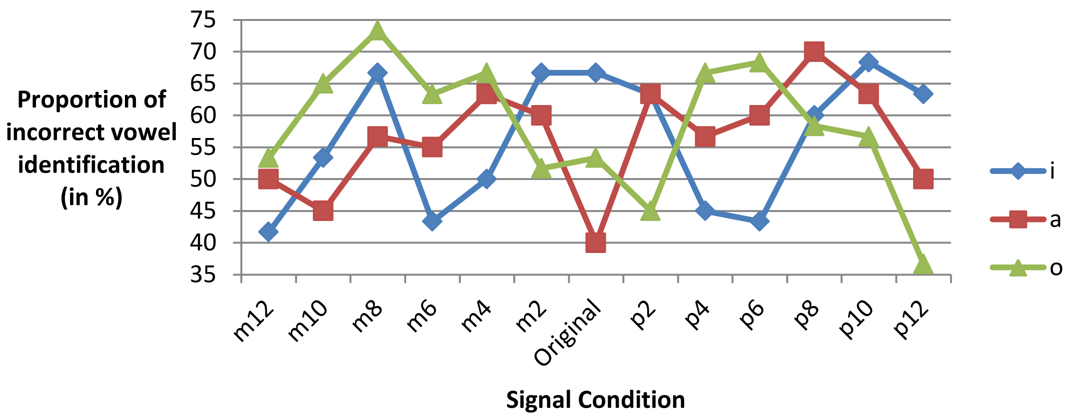


Figure 4.10 Percentages of incorrect vowel identification across signal conditions, with all **female and male** voice combined.

4.2.2 Female Voice

Female voice data were analyzed with data for all three H1-H2 levels combined and for each of the three H1-H2 levels separately.

4.2.2.1 All H1-H2 Levels Combined

With all H1-H2 levels combined, the number of incorrect vowel identifications for female voice across signal conditions was significantly affected by vowel type (Lambda coefficient = 2.015, p = 0.044). For female /i/, the "m6" signal condition showed a significantly fewer vowel misidentifications than the original and "p12" signal conditions (see

Figure 4.11), suggesting that an increase of H1-H2 resulted in poorer vowel identification. For the vowel /o/, however, this relationship was reversed, with the "m6" signal condition showing significantly more vowel misidentification than the "m2" and "p12" signal conditions ($p < 0.005$).

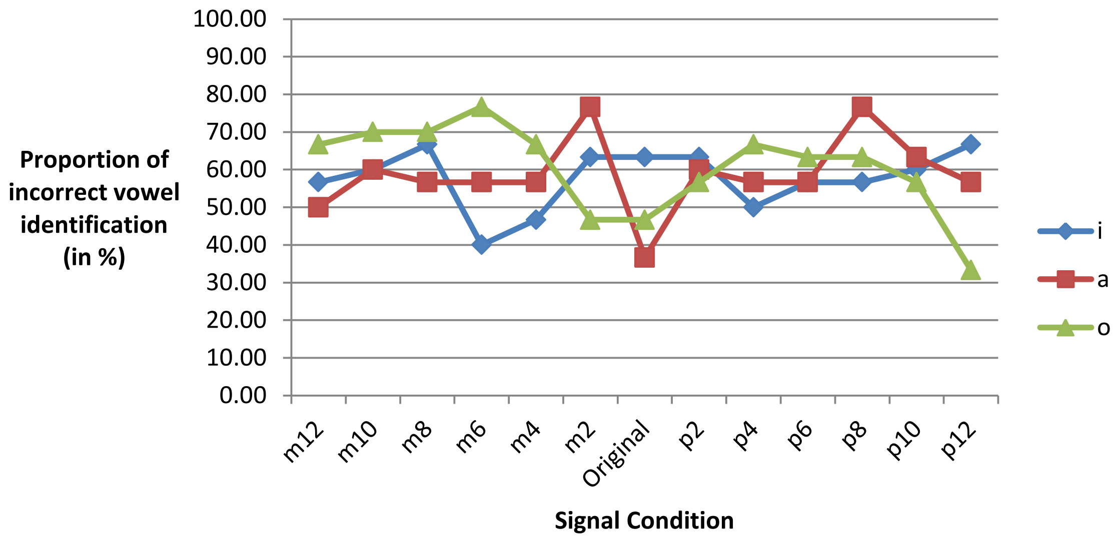


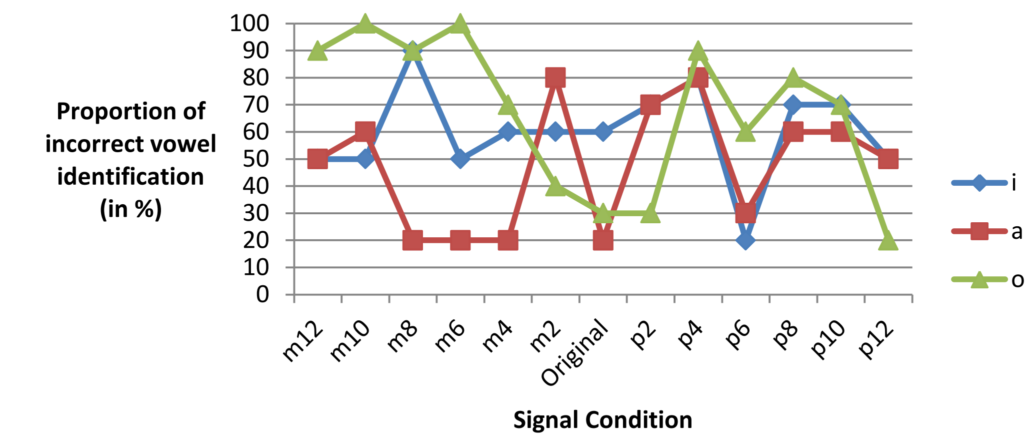
Figure 4.11 Percentage of incorrect vowel identification for **female** voice across signal conditions, with all H1-H2 levels combined.

4.2.2.2 Individual H1-H2 Levels

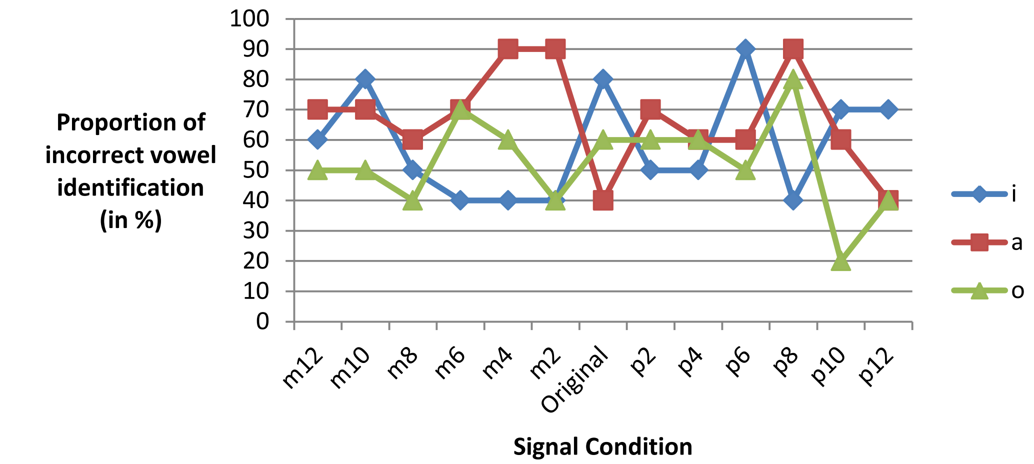
Figure 4.12 shows the proportion of incorrect vowel identification for female voice across signal conditions in each of the three H1-H2 levels. The count of incorrect vowel identification across different signal conditions was significantly affected by vowel type only at the "Mid" H1-H2 level (Lambda coefficient = 0.066, $p = 0.033$).

For signal conditions below the original signal, the vowel /o/ at the "Low" H1-H2 level was associated with a lower count of vowel misidentification as the signal condition level increased (see Table 4.10 and Figure 4.12). This finding suggests that the vowel /o/ at the "Low" H1-H2 level would be harder to identify if H1-H2 was decreased from that of the original signal.

Low



Mid



High

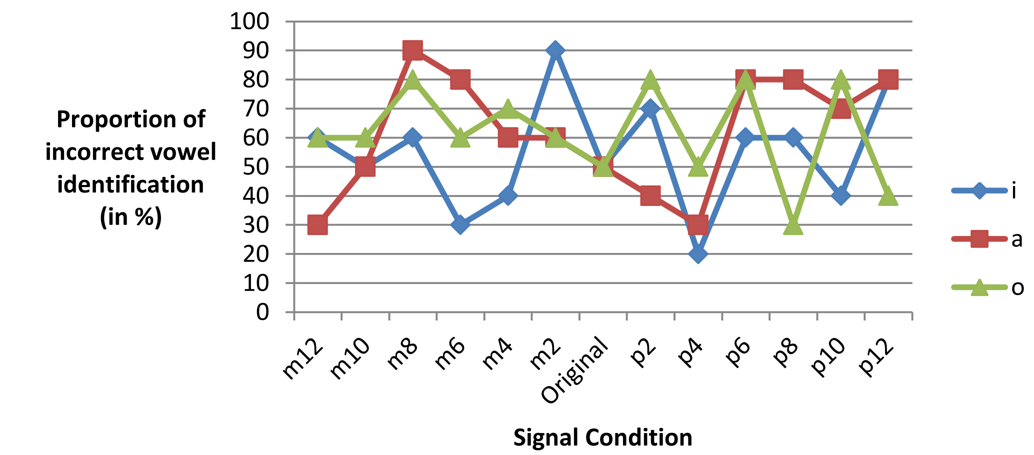


Figure 4.12 Percentages of incorrect vowel identification for **female** voice across signal conditions in each of the three H1-H2 levels.

4.2.3 Male Voice Only

Male voice samples in the “vowel identification” task were analyzed with data for all three H1-H2 levels combined and for each of the three H1-H2 levels separately.

4.2.3.1 All H1-H2 Levels Combined

With all H1-H2 levels combined, the number of incorrect vowel identification for male voice across signal conditions significantly varied by vowel type (Lambda coefficient = 0.044, $p = 0.044$). For the vowel /i/ in the signal conditions below the original signal, there was a significant, positive correlation between signal condition level and the count of vowel misidentification ($r = 0.855$, $p = 0.014$). In other words, when the H1-H2 value was decreased through signal manipulation from that of the original signal for male /i/, vowel identification was improved (see Figure 4.13).

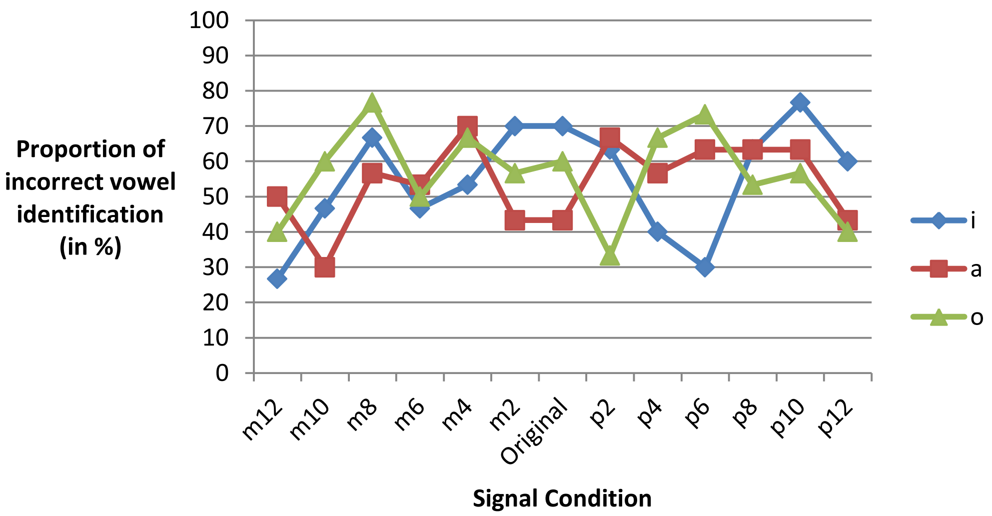
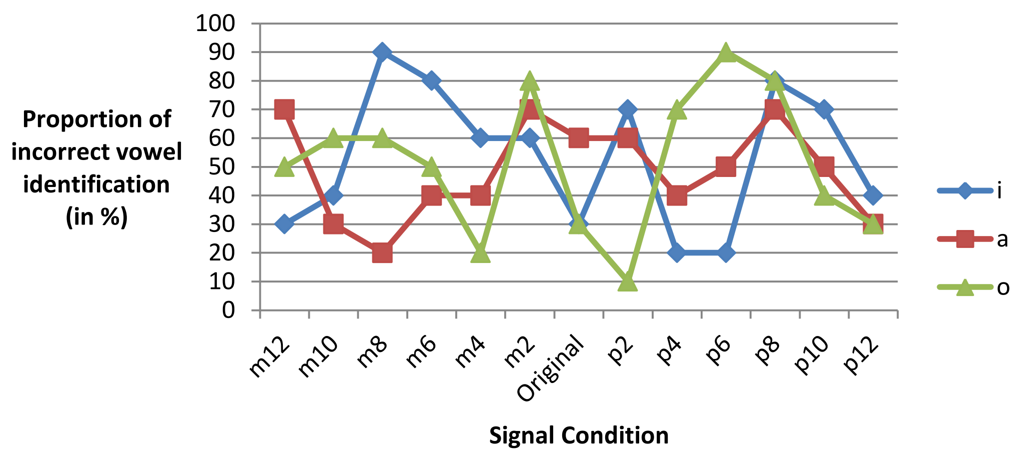


Figure 4.13 Percentages of incorrect vowel identification for **male** voice across signal conditions, with all H1-H2 levels combined.

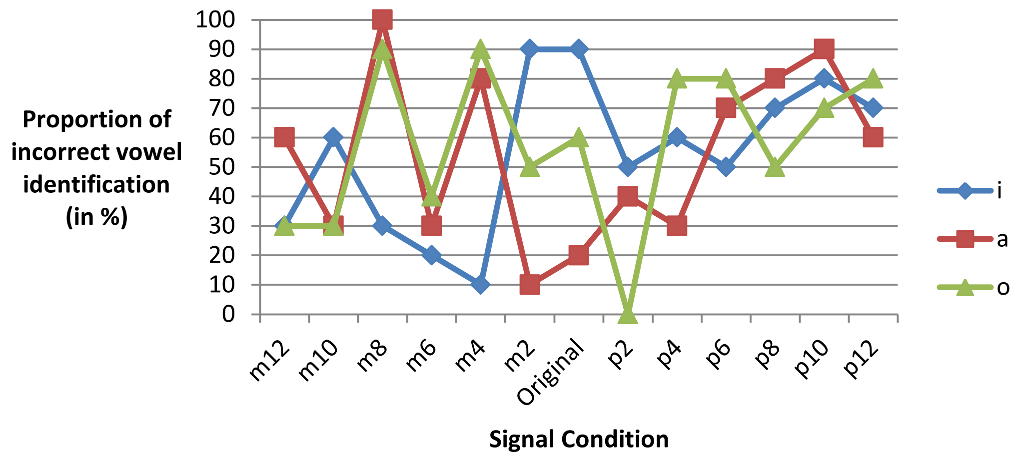
4.2.3.2 Individual H1-H2 Levels

The number of incorrect vowel identification for male voice across signal conditions was significantly affected by vowel type only at the "Mid" H1-H2 level (Lambda coefficient = 0.333, p = 0.038). With all vowels combined, there was a significant, positive correlation between signal condition levels and counts of vowel misidentification at the "Mid" H1-H2 level for male voice when only the original signal condition and all the signal conditions below it were considered (n = 18, r = 0.565, p = 0.015). As shown in Figure 4.14, the count of vowel misidentification increased as the signal condition level increased (i.e., H1-H2 value increased) for signals in the "p2" signal condition and up. An inspection of Table 4.9 revealed that the count of vowel misidentification for male /i/ at the "High" H1-H2 level increased as signal condition level increased for signal conditions below the original signal. For male /o/, the direction was reversed when observing signal conditions above the original signal. These findings suggest that decreasing H1-H2 for vowel /i/ and increasing H1-H2 for vowel /o/ from that of the original signal through signal manipulation would improve vowel identification.

Low



Mid



High

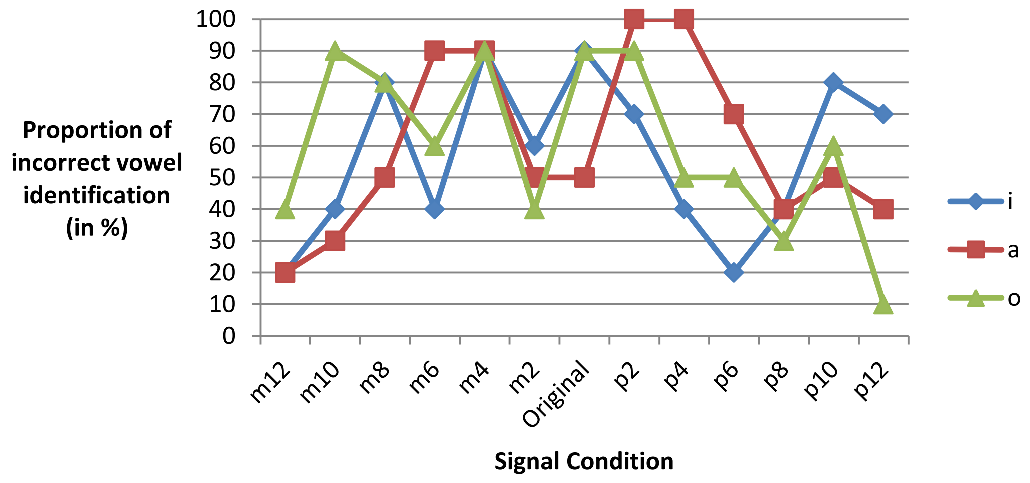


Figure 4.14 Percentages of incorrect vowel identification for **male** voice across signal conditions in each of the three H1-H2 levels (Low, Mid, and High).

4.2.4 Summary

A relationship between H1-H2 increment and vowel intelligibility was found but the relationship was affected by vowel type, speaker gender, and H1-H2 level. Generally, it appeared that an increase of H1-H2 would worsen the identification of /i/ but enhance that of /o/.

5. DISCUSSION

This chapter presents a discussion of the findings in relation to the proposed hypotheses and previous studies, clinical implications, limitations of the study, and future directions.

5.1 Findings in Relation to Hypotheses

This study aimed to determine whether excessiveness in the H1 dominance, which has been associated with breathy voice, may compromise vowel intelligibility and affect the perception of breathiness. There were two main hypotheses to be tested in this study. Firstly, it was hypothesized that changes to the magnitude of the amplitude difference between H1 and H2 (i.e., H1-H2) would lead to changes of the perception of breathiness and vowel identity. Specifically, an increase in the relative energy of the first harmonic would be expected to lead to an increase in perceived breathiness and an increase in the number of incorrectly identified vowels. Secondly, it was hypothesized that the relationship between voice quality and vowel intelligibility might be affected by vowel type and speaker gender as well as the speaker's vocal health status (i.e., normal vs. pathological).

With regard to the first hypothesis, the general finding was that acoustic manipulation of H1 affected vowel intelligibility and perceived breathiness. However, there were several key findings which only partially supported the hypothesis that an increase in the prominence of H1 would result in a higher level of perceived breathiness. Firstly, it was indeed observed that an increase of H1-H2 through signal manipulation led to an increase of perceived breathiness when the manipulation was performed on vowel /i/ produced by female normal speakers. Furthermore, the classification of the three H1-H2 levels in the pathological voice samples corresponded to the levels of perceived breathiness in the same direction as predicted in the first hypothesis. However, this pattern was not apparent in the results from the "breathiness rating" task for the same three H1-H2 categories of normal voice. The difference in findings between the normal speaker and voice patient groups suggests that the perception of different grades of breathiness in response to the change of H1-H2 is more readily detectable or consistent in pathological voice than in normal voice. In normal voice, the relationship between H1-H2 and perceived breathiness is only evident for the vowel /i/ produced by females.

In regard to gender-related factors that might affect the perception of breathiness, it was found that when rating normal voice, male listeners assigned higher breathiness scores than female listeners. However, there was no significant listener gender effect on the perception of breathiness when rating pathological voice. These findings reinforce the aforementioned observation that breathiness was more consistently rated in pathological voice than in normal voice. Nevertheless, regardless of the speaker's vocal health status or listener gender, there was a significant speaker gender effect on the extent of perceived breathiness. Overall, male voice was perceived as more breathy than female voice. In addition to gender-related factors, vowel type was also found to play role in affecting the perception of breathiness. Specifically, the vowel /a/ was perceived to be more breathy than /i/ in normal voice at the "Low" H1-H2 level and in male pathological voice in general. In summary, the pattern of the change of breathiness ratings and the accuracy of vowel identification as a function of H1-H2 varied by speaker gender, vowel type, the speaker's vocal health status (normal voice or pathological voice), and the classification of breathiness level ("Low", "Mid" and "High"). These findings agree with our second hypothesis that the effect of voice quality on vowel intelligibility would be affected by speaker gender, vowel type, and the speaker's vocal health status.

For the vowel intelligibility task, a relationship between H1-H2 increment and vowel intelligibility was found but this relationship also varied by vowel type, speaker gender, and

H1-H2 level. With all H1-H2 levels combined, the instances of incorrect vowel identification for male voice across signal conditions significantly varied by vowel type. Generally, an increase of H1-H2 would worsen the identification of /i/ but enhance that of /o/. Moreover, when the H1-H2 value was decreased through signal manipulation from that of the original signal for the vowel /i/ produced by males, vowel identification was improved. These results provide support for the two hypotheses demonstrating the effect of H1-H2 as well as speaker gender and vowel type on the performance in the vowel identification task.

The instances of incorrect vowel identification for male voice across signal conditions was significantly affected by vowel type only at the "Mid" H1-H2 level. With all vowels combined, there was a significant, positive correlation between signal condition levels and counts of vowel misidentification at the "Mid" H1-H2 level for male voice when only the original signal condition and all the signal conditions below it were considered. The count of vowel misidentification increased as the signal condition level increased (i.e., H1-H2 value increased) for signals in the "p2" signal condition and up. The count of vowel misidentification for male /i/ at the "High" H1-H2 level increased as signal condition level increased for signal conditions below the original signal. For male /o/, the direction was reversed when observing signal conditions above the original signal. These findings suggest that decreasing H1-H2 for vowel /i/ and increasing H1-H2 for vowel /o/ from that of the original signal through signal manipulation would improve vowel identification.

With respect to the second hypothesis, our findings suggest that the pattern of an increase in H1 energy leading to reduced vowel intelligibility varied by speaker gender, vowel type, and speaker's vocal health status (normal or pathological). With all H1-H2 levels combined, the number of incorrect vowel identifications for female voice across signal conditions was significantly affected by vowel type. For example, for female /i/, signal conditions below the original showed a significantly fewer vowel misidentifications than the original and the highest H1 signal condition , suggesting that an increase of H1-H2 resulted in poorer vowel identification and thereby supporting our first hypothesis. For the vowel /o/, however, this relationship was reversed, which supports the second hypothesis which says vowel type affects intelligibility.

The count of incorrect vowel identifications across different signal conditions was significantly affected by vowel type only at the “Mid” H1-H2 level. For signal conditions below the original signal, the vowel /o/ at the "Low” H1-H2 level was associated with a lower count of vowel misidentification as the signal condition level increased. This finding suggests that the vowel /o/ at the “Low” H1-H2 level would be harder to identify if H1-H2 was decreased from that of the original signal.

5.2 Findings in Relation to Previous Studies

In this study, male voice was perceived to be more breathy than female voice. This finding deviates from the common finding that females exhibit greater breathiness (Hanson & Chuang, 1999; Klatt & Klatt, 1990; Simpson, 2012). As described in the literature review (see Section 2.2), the anatomical differences between men and women may be a contributing factor to this difference in voice quality. Women have larger posterior cartilaginous spaces than men and enlarged glottal space, creating a more breathy voice quality (Sapienza & Ruddy, 2009). Females are also predisposed to breathiness due to a longer open phase of the glottal pulse which allows greater airflow into the vocal tract. However, Hillenbrand et al. (1994) yielded perceptual ratings indicating higher breathiness ratings for men in comparison to women in the very breathy condition. They suggested that it is quite likely that this was due to listeners making more of an allowance for breathiness in females and assigning relatively lower scores accordingly. This tendency may have been inherent in our study.

The increase of H1-H2 via signal manipulation resulted in a corresponding increase in perceived breathiness but only for the vowel /i/ produced by females. The increase of the perceptual salience of breathiness resulting from a greater H1-H2 relative amplitude is a common finding in the literature (Abramson et al., 2004; Bickley, 1982; Fischer-Jørgensen,1967; Garellek &Keating,2011; Hanson, 1995; Hanson & Chuang,1999; Henton &Bladon,1985; Huffman, 1987; Klatt &Klatt,1990; Ladefoged & Antonanzas-Barroso, 1985). In this study, this trend was also found but it was affected by both vowel type and speaker gender. With regard to the vowel type effect, using only open vowels (/a/ and /o/) for H1-H2 measurement may prevent interference from F1 (Hanson, 1997; Henton & Bladon, 1985). The inclusion of low (or open) vowels for the study of H-H2 measure ensures that H1 and F1 are well separated (Hanson, 1997). One plausible explanation for the vowel effect on the relationship between H1-H2 and vowel identification was that only an open vowel had a F1 at a frequency high enough not to increase the amplitudes of the lower harmonics (Henton & Bladon, 1985). There is an inverse relationship between the frequency of F1 and vowel height in that a high tongue position will result in a lower F1 frequency (Thompson, Lin, & Robb, 2011). The high vowel used in this study is /i/ while /a/ and /o/ are the low vowels. It is possible that H1-H2 amplitude difference is most usefully measured from low vowels (Thompson et al., 2011) and that the linear result found with the female /i/ is attributable to the exaggerated H1 prominence brought about by the combination of H1 and F1 energy.

Although the use of low vowels as stimuli in the “breathiness rating” task may minimize the impact of F1 on the lower harmonics, it is highly likely that the degree of nasality present in low vowels may also be a confounding factor, especially when making comparisons between male and female voices (Simpson, 2012). In Simpson (2012), the speculation is that the H2 measure derived from the female voice in particular was vulnerable to co-occurring with the nasal zero, which may attenuate the amplitude of H2 resulting in a larger difference between H1 and H2. The amplitude difference is widened more by the "enhancement of H1 by the first nasal pole" (p480). The finding of a linear trend of the perceived breathiness increasing with an increase of H1-H2 in the vowel /i/ but not in /a/ for female voice is indicative of the interaction between formant frequencies and H1 in affecting the perception of breathiness. It is unclear, however, whether the effect of F1 energy on H1 energy in the high vowel is greater than the effect of nasality on the energy of H1 and H2 in the low vowel. The finding that listeners in our study judged male voice to be breathier than female suggests that the impact of nasality on the perception of breathiness might be less than the tendency for listeners to consider high breathiness levels in males more abnormal than in females (Hillenbrand et al., 1994).

Hillenbrand et al. (1994), in a study of acoustic correlates of breathiness, concluded that the amplitude of H1 correlated moderately with breathiness perception. Statistical analysis of variance was performed to determine whether there was a vowel, phonation-type, or gender effect. They employed both high (/i/) and low (/a/, /o/) vowels and reported no significant vowel effect. It is noteworthy that the sustained phonation vowel tokens used in their study were produced by normal speakers who were simulating different degrees of breathy voice-normal, moderately breathy and very breathy while the present study employed natural vowels that were digitally manipulated. Despite this methodological difference, both Hillenbrand et al.'s (1994) and the present study showed a significant phonation-type effect on the perception of breathiness. In Hillenbrand et al.'s (1994) study, the analysis of variance investigating the effects of normal, moderately breathy and very breathy phonation on breathiness ratings revealed a phonation-type effect ($F(2,156=330.94$, $p=0.0001$) In the present study, the linear relationship between perceived breathiness and H1-H2 classification (Low, Mid, and High) was found in pathological voice but not in normal voice.

In addition, Hillenbrand et al. (1994) reported a significant speaker gender effect. They found that H1-H2 produced higher breathiness ratings for female speakers over male. However, as explained previously, perceptual ratings indicated higher breathiness ratings for men in comparison to women in the very breathy condition only. To an extent this pattern was observed in the current study as well. There is some evidence in figure 4.4 where as H1-H2 level increases, the perception of breathiness decreases considerably more in female relative to male voice, especially where the listener is also female. Further indications can be seen in Figure 4.6 which shows the means breathiness scores across the Low, Mid, and High H1-H2 levels for vowels /i/ and /a/ obtained from female voice patients, with both listener genders combined. Female /i/ has a considerable drop from m12 to p12 in the high H1-H2 level in pathological voice where the low and mid /i/ does not fluctuate greatly as the signal condition increases. This trend is seen in /i/ abut not for /a/ which, at the high H1-H2 level , has a more linear trend from m12 to p12. Based on this, it is very tentatively speculated that there is a higher tolerance for H1-H2/vocal fold opening for men over women. That is, it is only in females that a critical opening phase (see reference to Samlan & Story, 2011, below) initiates a pattern of decreasing breathiness perception in response to a rising in the H1-H2 amplitude difference.

The trend of numerous quadratic relationships between breathiness perception and H1-H2 is largely supported by Samlan et al.,(2013). Samlan and Story (2011) acknowledged the benefits of H1-H2 as an acoustic cue pertinent in perceiving breathy voice quality (Esposito et al., 2010; Kreiman & Gerratt, 2010) but pointed to the near-inability of H1-H2 to explain differing levels of breathiness in both normal and pathological speakers (Shrivastav, 2003) and also cited the influence of nasality on the perception of breathiness (Simpson, 2012). Samlan and Story (2011) showed that greater vocal fold opening resulted in an increase in H1-H2 until "a critical separation value" (Samlan et al., 2013) was reached, which initiated a decrease in H1-H2 despite the continued vocal fold opening. Samlan and Story (2011) concluded that the aerodynamic interaction of the vocal tract was responsible for the falling H1-H2 amplitude. A subsequent perceptual study by Samlan et al., (2013) found that H1-H2 levels were lower in the breathiest conditions as expected. They concluded that H1-H2 may be effective in predicting breathiness only in low to moderate levels. In the current study, we found a similar trend. For example, in normal speakers the "Low" H1-H2 level appeared to result in higher breathiness scores than the other two higher H1-H2 levels, suggesting that within normal voice, an increase of H1-H2 amplitude difference did not lead to an increase in perceived breathiness. Moreover, normal voice with an H1-H2 value at the lower end of the normal range of H1-H2 value may actually be perceived as more breathy than those at the higher end of the range.

Studies have shown variable patterns of breathiness perception in response to changes in H1 prominence. Kreiman and Gerratt (2010) found H1-H2 to play a role in the breathy and non-breathy voice discrimination. They carried out a perceptual experiment in both English and Mandarin, using natural samples the vowel /a/ with H1 levels increased and decreased in 15 0.5 dB steps relative to the original token. The task was to compare an unmanipulated token with a token which had H1 altered, and judge which token had been acoustically adjusted. They found that it was easy for listeners to detect differences based on changes in H1-H2.

Results from the "vowel identification" task in the present study failed to reveal a consistent pattern with regard to the relationship between H1-H2 level and the performance in vowel identification. However, it was found that the number of misidentified /i/ tokens increased with a corresponding increase in H1-H2. Conversely, /o/ tokens misidentification decreased as H1-H2 level increased. This pattern was observed in male voice only. Henton and Bladon (1985) found that breathiness resulted in reduced speech intelligibility.

Hillenbrand et al., (1994) found higher breathiness ratings for men in comparison to women in the very breathy condition. As noted previously, however, there was a high likelihood that listeners allowed for more breathiness in female voice while the same level of breathiness in males sounded abnormal. Therefore, it can be speculated that the present finding that there was a relationship between H1-H2 and vowel intelligibility in male voice but not in female voice may be attributable to the lower tolerance of a breathy component in male voice.

5.3 Clinical Implications

It is hoped that the relative amplitude difference between the first and second harmonics can be used clinically to quantify a normal range of breathiness. That is, via acoustic analysis of a given voice patient's speech sample, a differentiation can be made between pathological and non-pathological breathiness based on the magnitude of this acoustic correlate. The clinical relevance of our study could possibly materialize by a) providing support for the treatment of vocal pathologies involving breathy voice b) evaluation of progress over the course of the treatment process c) a basis to create a normal range of breathiness which can help to identify where breathy voice begins to sound pathological. In addition, as stated previously, a better understanding of the relationship between voice quality and speech intelligibility would help in improving the signal processing strategies used in technologies related to speech communication.

5.4 Limitations and Future Directions

Despite and number of results supporting our original hypotheses, it needs to be acknowledged that under the majority of conditions, a non-linear trend was observed in both breathiness perception and vowel intelligibility. This was contradictory to our first hypotheses.

For example, in the vowel intelligibility task, with all vowels combined, there was a significant, positive correlation between signal condition levels and counts of vowel misidentification at the “Mid” H1-H2 level for male voice when only the original signal condition and all the signal conditions below it were considered (as shown in Figure 4.14) This means that for conditions outside this description, (e.g. at “Low” and “High” H1-H2 levels, for female voice, for signal conditions greater than the original signal) the hypothesis was not supported. Furthermore, the instances of incorrect vowel identification for male **(**see Figure 4.14**)** and female voice (see Figure 4.12) across signal conditions were also significantly affected by vowel type only at the “Mid” H1-H2 level. There are numerous examples such as these. However, it is not incorrect to say that our second hypothesis accounts for this by predicting that certain vowel, gender and H1-H2 level conditions need to present before the first hypothesis is supported. Further investigation into the effect of the variable factors used in this study (vowel, gender and H1-H2 level conditions) on vowel intelligibility and breathiness perception is warranted.

One participant (GV) commented on the obvious ‘unnatural and artificial quality’ of tokens which had high levels of H1 prominence. Similar studies have used a scale to measure the perceived naturalness of speech samples (Klatt & Klatt, 1990) and considering the signal manipulation performed in this study, it would have been a worthwhile scale to include. It has been stated that an increase in H1 alone does not elicit a sense of breathiness for most listeners (Klatt & Klatt, 1990, p851). The reason for this is the acoustic effect (specifically with female speakers) is to increase the spectrum around 200Hz which produces a nasal pole consistent with nasalization. They add that a nasal voice quality is common in instances where only one fundamental component is increased and often results in somewhat unnatural sounding speech samples leading to perceptual ambiguity (Klatt & Klatt, 1990). Stimuli which included multiple acoustics cues (most importantly increased aspiration noise) were perceived to be more natural. Perhaps adding to the perceived unnatural quality employed in this study was the use of sustained phonation for tokens used in the “breathiness rating” task. The non-use of sentence embedded vowels (as used in the vowel intelligibility task) for the “breathiness rating” task in this study, may have made our findings arguably more difficult to be generalised to the clinical cases of vocal pathologies involving breathiness. As stated in the literature review (see Section 2.4), Klingholtz (1990) questioned the validity of using sustained phonation for judging voice quality in speech likening sustained phonation to more of a singing voice than a speaking voice. The question of whether the advantage of using simplified sustained vowel tokens for ease of measurement is more valuable than more complex but realistic speech signals remains.

In terms of future studies, there are several other experimental factors which should be researched to clarify the relationship between H1-H2 and breathiness perception. Considering we assert that listener fatigue is an argument against the use of pair-wise comparison, determining whether results yielded from direct magnitude estimation versus pair-wise comparison would significantly affect perceptual results would represent a viable research topic. The question of using sustained phonation for ease of analysis or sentence embedded vowels for a realistic and natural speech sample requires continued investigation also.

H1-H2 is a prime correlate for breathy voice quality but several studies (Fischer-Jorgensen, 1967; Klatt & Klatt, 1990; Samlan et al.,2013) have questioned the validity of using/adjusting a single acoustic parameter (as used in the current study) rather than a combination for perceptual studies. As Samlan et al., (2013) suggested, as separate single correlates, perhaps H1-H2 and spectral tilt could be used to measure a greater range of breathiness more accurately. The former could measure breathiness at low to mid levels, where the main acoustic changes involve decreased energy in the fundamental component and the latter, mid to high levels, which are characterized by greater noise in high frequency regions (Samlan et al., 2013).

5.5 Conclusion

As stated in the literature review, (2.1.1) as a consequence of not all breathy voice being considered abnormal and not all speech with breathy quality being unintelligible, the relationship between the acoustic correlate of breathiness and the perception of breathiness and speech intelligibility may well be non-linear or categorical and even language-specific or gender-dependent instead of linear or universal. The current study aimed to determine the effect of the signal manipulation of the acoustic correlate, namely the prominence of the first harmonic relative to the second harmonic (H1-H2), on both breathiness perception and vowel intelligibility. It also hypothesized that the relationship between them would vary depending on speaker gender, vowel and speaker's vocal health status. The relationship between H1 dominance and perceived breathiness was non-linear. Factors found to disrupt the linear relationship included speaker gender, vowel type, and the extent of H1 dominance. In addition, there was evidence that acoustic manipulation of the H1 amplitude would affect vowel intelligibility and the relationship between vowel intelligibility and H1-H2 values also vary by speaker genders and vowel types.

References

Abramson, A.S., Luangthongkum, T., & Nye, P.W. (2004). Voice register in Suai (Kuai): Analysis of perceptual and acoustic data. *Phonetica*, 61, 147–171.

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

Castillo-Guerra, E., & Ruiz, A. (2009). Automatic modeling of acoustic perception of breathiness in pathological voices. *Biomedical Engineering, IEEE Transactions on, 56*(4), 932-940.

DiCanio, C.T.(2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, 39(2), 162–188.

Esposito, C.M. (2010) The effects of linguistic experience on the perception of phonation *Journal of Phonetics* 38(2) 306 – 316

Esposito, C. M. (2011). The perception of pathologically-disordered phonation by Guajarati, English, and Spanish listeners. *Language and Speech, 54*, 415-30.

Esposito, C. M., & Khan, S. u. D. (2012). Contrastive breathiness across consonants and vowels: A comparative study of Guajarati and white Hmong. *Journal of the International Phonetic Association, 42*(2), 123-143.

Fischer-Jorgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics*, *28*, 71-139.

Fukazawa, T., el-Assuooty, A., & Honjo, I. (1988). A new index for evaluation of the turbulent noise in pathological voice. *Journal of the Acoustical Society of America, 83*(3), 1189-1193.

Garellek, M.,&Keating,P.A.(2011).The acoustic consequences of tone and phonation interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, 41(2), 185–205.

Gerratt, B. R., & Kreiman, J. (2004). Voice quality, perceptual evaluation of. In R. D. Kent (Ed.), *The MIT encyclopedia of communication disorders* (pp. 78-80). Cambridge, Massachusetts: The MIT Press.

Gordon, M. and P. Ladefoged (2001). "Phonation types: a cross-linguistic overview." *Journal of Phonetics* 29(4): 383-406.

Hanson, H.M.(1995). Glottal characteristics of female speakers. Ph.D.Thesis. Harvard: Harvard University.

Hanson, H.M.(1997).Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America,* 101, 466–481.

Hanson, H.M., & Chuang, E.(1999).Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America,* 106, 1064.

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

Hillenbrand, J., & Houde R.A., (1996) Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*,39, 311-321.

Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America*, *81*(2), 495-504.

Jing, Jack (2009) Vocal fold dynamics : phonation mechanism, aerodynamics, impact stress, and mucosal wave. San Diego, California: Plural Publishing.

Keating, P.A.,&Esposito,C.M.(2007).Linguistic voice quality. *UCLA Working Papers in Phonetics,* 105, 85–91.

Keating, P.A., Esposito, C.M., Garellek, M., Khan, S., & Kuang, J. (2010).Phonation contrasts across languages. *UCLA Working Papers in Phonetics,* 108, 188–202.

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.

Klingholtz, F. (1990). Acoustic recognition of voice disorders: A comparative study of running speech versus sustained vowels. The Journal of the Acoustical Society of America, 87(5)

Kreiman, J & Gerratt BR (2010) Perceptual sensitivity to first harmonic amplitude in the voice source. *Journal of the Acoustical Society of America* 128, 2085.

Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *Journal of the Acoustical Society of America, 96*(3), 1291-1302.

Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2007). Measures of the glottal source spectrum. *Journal of Speech, Language and Hearing Research, 50*(3), 595-610.

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

Krom, G. d. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research,, 38*(4), 794-811.

Ladefoged, P.,&Antonanzas Barroso,N.(1985).Computer measures of breathy phonation. *Working Papers in Phonetics,UCLA,* 61, 79–86.

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

Law, T., Kim, J.H., Lee, K.Y., Tang, E.C., Lam, .H. Hasselt, A.C (2012) Comparison of rater's reliability on perceptual evaluation of different types of voice sample. *Journal of Voice.* 26 (5) 666e13-666e21

Pandit, P.B.(1957).Nasalization, aspiration and murmur in Gujarati. *Indian Linguistics*, 17, 165–172.

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. *Journal of Speech and Language Hearing Research, 44*(2), 327-339.

Patel, S., Shrivastav, R., & Eddins, D. A. (2012). Developing a single comparison stimulus for matching breathy voice quality.*Journal of Speech, Language and Hearing Research (Online), 55*(2), 639-647.

Peterson G, . E., and BarneyH, . L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175-184.

Radish Kumar, B., Bhat, J. S., & Prasad, N. (2009). Cepstral analysis of voice in persons with vocal nodules. [Article in Press]. *Journal of Voice*, 1-3.

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

Samlan, R. A., & Story, B. H. (2011). Relation of structural and vibratory kinematics of the vocal folds to two acoustic measures of breathy voice based on computational modeling. *Journal of Speech, Language and Hearing Research (Online), 54*(5), 1267-1283.

Samlan, R. A., Story, B. H., & Bunton, K. (2013). Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *Journal of Speech, Language and Hearing Research (Online), 56*(4), 1209-1223.

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

Shrivastav, R., Camacho, A.,Patel, S & Eddins, DA. (2011) A model for the prediction of breathiness in vowels. *The Journal of the Acoustical Society of America*, 129, 1605-1615

Shrivastav, R.,&Sapienza,C.M.(2003).Objective measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustical Society of America,* 114, 2217–2224.

Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voice. *Journal of Phonetics* 40(3): 477-490.

Södersten, M., Hertegård, S., & Hammarberg, B. (1995). Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *Journal of Voice, 9*(2), 182-197.

Thompson, L., Lin, E., Robb, M.P. (2011) The impact of breathiness on the intelligibility of speech. Christchurch, New Zealand: 8th Asia-Pacific Conference on Speech, Language and Hearing, 11-14 Jan 2011.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, N.J.: Prentice Hall.

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

Appendices

Appendix 1. Standard deviations of **female listeners'** (n = 5) breathiness ratings (in %) on the **normal speakers'** voice samples as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	23.70	35.91	25.55	2.18	22.14	12.37	21.12	16.87	27.90	23.28	32.46	29.81
m10	34.24	13.49	22.53	3.36	11.55	20.13	28.64	38.36	22.78	18.84	22.24	35.82
m8	21.44	18.38	27.29	5.13	15.30	26.25	8.38	21.87	28.03	21.93	19.70	40.05
m6	24.25	15.59	32.86	15.56	24.46	16.00	34.26	16.41	27.78	27.92	20.83	28.68
m4	22.11	29.35	7.64	7.29	10.15	16.69	14.67	31.15	32.65	28.17	20.80	27.55
m2	20.57	19.79	12.41	28.82	12.34	12.38	12.32	42.35	23.38	16.28	31.26	25.40
Original	24.48	18.05	14.41	9.58	19.87	4.28	26.70	14.97	24.34	22.00	19.88	26.25
p2	24.06	20.00	28.63	14.12	14.49	10.47	22.10	23.41	27.41	32.28	24.15	26.58
p4	24.32	14.90	19.15	5.47	16.85	17.06	25.72	15.01	25.25	33.24	25.50	15.09
p6	25.31	16.16	31.31	8.60	29.27	26.59	22.45	15.51	31.80	24.09	20.62	29.50
p8	21.74	13.71	25.83	10.45	12.09	22.25	12.57	24.62	29.15	27.52	30.04	19.93
p10	14.91	19.31	25.69	5.04	24.29	25.29	23.62	13.86	18.34	30.80	26.79	27.01
p12	19.69	35.92	30.25	21.73	26.08	16.60	12.76	31.46	25.92	28.86	15.70	22.24

Appendix 2. Standard deviations of **male listeners'** (n = 5) breathiness ratings (in %) on the **normal speakers'** voice sample as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	13.07	22.91	12.69	23.11	21.72	26.99	7.50	14.01	18.61	6.74	30.40	22.57
m10	14.77	26.59	18.11	12.51	26.43	21.37	17.67	8.89	18.81	16.49	21.89	28.17
m8	17.43	27.76	17.31	16.90	22.90	18.27	16.78	9.10	27.75	13.55	24.77	28.67
m6	10.15	28.78	15.49	19.61	9.49	10.66	5.88	16.11	22.66	12.19	16.52	26.70
m4	20.67	20.54	21.25	16.96	18.08	16.28	27.06	9.10	14.59	20.63	27.44	16.40
m2	19.84	25.47	23.08	17.16	16.61	15.22	14.75	11.16	30.99	14.84	32.92	14.85
Original	11.30	22.63	29.55	19.83	7.68	19.52	16.93	16.02	13.98	19.70	13.81	20.88
p2	14.31	20.72	30.07	12.33	15.53	16.17	16.70	10.40	10.68	21.25	21.16	21.27
p4	16.85	13.98	17.58	12.41	33.03	24.01	26.48	25.26	25.36	15.53	24.75	21.04
p6	13.45	15.33	30.32	22.67	29.76	18.51	21.48	10.74	14.51	19.18	10.66	26.42
p8	32.37	32.16	30.49	8.48	16.93	19.79	27.00	24.58	17.24	20.24	19.88	28.41
p10	31.85	28.31	28.21	20.45	18.26	32.81	23.87	6.90	9.01	21.92	14.50	16.64
p12	28.36	29.18	29.25	11.61	34.88	33.66	23.83	26.86	20.88	5.75	24.32	21.09

Appendix 3. Standard deviations of **female listeners'** (n = 5) breathiness ratings (in %) on the **voice patients'** voice samples as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	34.31	23.00	16.96	14.18	33.54	20.13	23.70	17.46	11.75	13.74	9.14	8.19
m10	33.52	27.36	27.26	26.60	25.76	14.66	11.11	28.15	25.34	9.72	10.66	15.15
m8	25.60	19.23	25.97	17.90	31.13	24.02	25.75	22.50	15.45	5.93	6.10	14.33
m6	10.10	21.23	19.31	26.06	21.01	8.18	20.94	28.24	9.35	8.85	6.25	11.07
m4	23.96	12.80	28.22	10.94	19.62	27.13	20.28	26.77	11.55	7.85	5.06	26.47
m2	29.54	3.96	26.73	14.27	27.45	18.04	25.55	19.72	8.46	5.30	13.75	13.96
Original	11.93	14.59	30.18	24.64	22.67	10.43	19.02	27.19	9.61	5.97	20.13	26.02
p2	19.04	8.06	25.86	22.42	18.58	18.29	24.21	21.60	25.16	13.80	18.01	35.58
p4	11.10	4.69	29.86	21.48	20.59	21.29	21.93	22.94	13.89	18.39	10.04	21.77
p6	13.48	7.10	21.65	26.61	30.80	24.67	16.55	28.91	20.28	22.26	18.20	30.59
p8	17.29	20.60	35.35	29.25	15.21	29.40	21.84	25.33	11.49	5.75	9.45	35.73
p10	17.60	22.89	21.47	12.95	25.85	29.37	28.66	28.60	18.37	13.64	14.91	31.17
p12	18.65	20.77	32.58	16.97	30.71	33.33	20.61	28.64	15.42	7.39	13.52	27.03

Appendix 4. Standard deviations of **male listeners'** (n = 5) breathiness ratings (in %) on the **voice patients'** voice samples as grouped by speaker genders, H1-H2 levels (Low, Mid, and High), vowel types (/i/ and /a/), and signal conditions.

	Female						Male					
	Low		Mid		High		Low		Mid		High	
	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/
m12	28.13	25.36	28.16	21.92	44.78	25.11	15.72	15.24	35.90	14.63	6.27	15.14
m10	26.01	6.60	20.36	10.90	32.48	13.21	12.70	34.35	23.60	12.13	6.98	15.88
m8	15.95	18.96	33.65	16.16	35.36	14.60	17.06	23.33	25.59	12.98	11.84	14.42
m6	25.80	19.98	28.57	17.27	25.04	20.06	31.71	25.19	18.04	24.49	19.33	20.41
m4	25.00	26.40	20.23	19.74	30.79	23.10	20.74	7.47	17.11	10.20	11.33	13.77
m2	15.10	20.68	17.96	15.23	33.38	18.02	17.55	24.85	28.89	13.98	9.26	20.46
Original	25.16	29.85	31.26	23.55	28.94	31.76	13.64	20.39	32.12	12.25	12.18	11.78
p2	24.24	19.97	27.39	11.57	25.91	27.77	20.51	15.86	33.31	25.76	23.02	32.57
p4	24.90	16.50	30.92	30.15	29.49	24.04	8.64	17.12	28.82	11.14	15.43	17.54
p6	25.39	22.60	19.56	17.35	21.96	18.02	30.36	25.31	30.53	28.51	22.12	32.60
p8	20.44	26.66	16.52	23.18	23.42	36.20	30.05	8.05	37.57	26.26	29.34	24.85
p10	16.62	24.31	28.16	27.98	23.28	27.74	21.84	37.30	27.84	38.97	32.94	38.91
p12	17.72	20.29	26.06	22.98	31.12	28.47	25.45	32.31	32.39	37.33	22.69	35.54

Appendix 5. Individual female listener's breathiness ratings on the vowels /i/ and /a/ produced by female and male normal speakers across three H1-H2 levels

NF1: "Low" H1-H2 level	NF1: "Mid" H1-H2 level	NF1: "High" H1-H2 level
NF2: "Low" H1-H2 level	NF2: "Mid" H1-H2 level	NF2: "High" H1-H2 level
NF3: "Low" H1-H2 level	NF3: "Mid" H1-H2 level	NF3: "High" H1-H2 level
NF4: "Low" H1-H2 level	NF4: "Mid" H1-H2 level	NF4: "High" H1-H2 level
NF5: "Low" H1-H2 level	NF5: "Mid" H1-H2 level	NF5: "High" H1-H2 level

Appendix 6. Individual male listener's breathiness ratings on the vowels /i/ and /a/ produced by female and male normal speakers across three H1-H2 levels

NM1: "Low" H1-H2 level	NM1: "Mid" H1-H2 level	NM1: "High" H1-H2 level
NM2: "Low" H1-H2 level	NM2: "Mid" H1-H2 level	NM2: "High" H1-H2 level
NM3: "Low" H1-H2 level	NM3: "Mid" H1-H2 level	NM3: "High" H1-H2 level
NM4: "Low" H1-H2 level	NM4: "Mid" H1-H2 level	NM4: "High" H1-H2 level
NM5: "Low" H1-H2 level	NM5: "Mid" H1-H2 level	NM5: "High" H1-H2 level

Appendix 7. Individual female listener's breathiness ratings on the vowels /i/ and /a/ produced by female and male voice patients across three H1-H2 levels

PF1: "Low" H1-H2 level	PF1: "Mid" H1-H2 level	PF1: "High" H1-H2 level
PF2: "Low" H1-H2 level	PF2: "Mid" H1-H2 level	PF2: "High" H1-H2 level
PF3: "Low" H1-H2 level	PF3: "Mid" H1-H2 level	PF3: "High" H1-H2 level
PF4: "Low" H1-H2 level	PF4: "Mid" H1-H2 level	PF4: "High" H1-H2 level
PF5: "Low" H1-H2 level	PF5: "Mid" H1-H2 level	PF5: "High" H1-H2 level

Appendix 8. Individual male listener's breathiness ratings on the vowels /i/ and /a/ produced by female and male voice patients across three H1-H2 levels

PM1: "Low" H1-H2 level	PM 1: "Mid" H1-H2 level	PM 1: "High" H1-H2 level
PM2: "Low" H1-H2 level	PM 2: "Mid" H1-H2 level	PM 2: "High" H1-H2 level
PM 3: "Low" H1-H2 level	PM 3: "Mid" H1-H2 level	PM 3: "High" H1-H2 level
PM4: "Low" H1-H2 level	PM 4: "Mid" H1-H2 level	PM 4: "High" H1-H2 level
PM 5: "Low" H1-H2 level	PM5: "Mid" H1-H2 level	PM5: "High" H1-H2 level