Audio - Aero tactile integration in speech perception using an open choice paradigm

A thesis submitted in partial fulfilment of the
requirements for the Degree of
Master of Science

Jilcy George Madapallimattam
Department of Communication Disorders
The University of Canterbury, Christchurch, New Zealand
2017 - 2018

“Education is the kindling of a flame, not the filling of a vessel.”

(Socrates)

Acknowledgement

The path towards this thesis has been circuitous and intense. As I complete this, it is time to write a note of thanks to the special people who challenged, motivated, supported and stuck with me along the way.

Firstly, I would like to extend my deep sense of gratitude to my thesis mentor, Dr. Catherine Theys whose expertise, patience, understanding, generous guidance and constant support helped me to accomplish my research and writings.

Next, I owe my co-guide, Dr. Donald Derrick, for his unwavering support, guidance, and insight throughout this research project.

Besides my supervisors, I would like to thank everyone in the NZILLB speech lab, especially Dr. Doreen Hansmann for her invaluable help and encouragement during the course of this research work.

Finally, I am deeply thankful to my beloved family and friends who put faith in me and urged me to do better.

Most of all, I am deeply indebted to God, the Almighty for giving his endless blessings, knowledge and strength without which this work won`t have been possible.

Abstract

Purpose: The purpose of this study was to investigate the influence of AT integration during speech perception measured using an open choice response elicitation.

Methods: 34 untrained native English speakers received puffs of air (aero tactile stimuli) on their neck while simultaneously hearing aspirated or unaspirated English plosives (i.e., /pa/, /ba/, /ka/ and /ga/). These monosyllables were presented in congruent and incongruent stimulus conditions at different SNRs, following an adaptive staircase to get 80% accuracy. Participant responses for the monosyllabic identification task was recorded by asking them to type down the syllable they heard.

Results: Air puffs did not have a statistically significant influence on SNR 80% accuracy thresholds. However, there was a significant effect of both place of articulation and the interaction of place and manner of articulation.

Conclusion: The study found no evidence of benefit from aero tactile stimuli on speech perception in open choice tasks. However, the results suggest a place and manner dependency in speech perception, as demonstrated in previous studies.

Table of Contents

List of Tables

List of Figures

Abbreviations

AT	Auditory – tactile or Auditory – aero tactile
AV	Auditory – visual
VT	Visual – tactile
CV	Consonant – vowel
SNR	Signal – to – noise ratio
GLMM	Generalised linear mixed effects model
AIC	Akaike information criterion
IC	Incongruent stimuli
C	Congruent stimuli
AAC	Alternative augmentative communication

Chapter I: Background

1.1 Introduction

Humans are equipped with a broad category of senses, with the classic five being vision, hearing, smell, taste and touch. Together, these make up the sensory system. Experiences, whether enjoyable or miserable, can be perceived because of these sensory modalities. Our sensory organs provide the interface to the brain, our coordinating centre for sensation and intellect, enabling us to interpret and understand our surrounding environment. There are obvious benefits associated with having multiple senses. Each sense s of optimal use in various circumstances and collectively they increase the likelihood of identifying and understanding the events and objects in our everyday life. For example, when lying down in a grassy field, we could smell the flowers, feel the grass, and see the sky. The whole combine to make the experience. Risberg and Lubker (1978) found that integrating information from different sensory channels had an additive effect on comprehension of speech as well. This interaction among the senses and the fusion of their information is described by the phrase multisensory or multimodal integration. So, multisensory integration refers to the influence of one sensory modality over another in the form of enhancement or suppression relative to the strongest unimodal response (Stein & Meredith, 1993).

Initially, multisensory studies in speech perception focused primarily on the integration of auditory and visual information (Green & Kuhl, 1989; Green, Kuhl, Meltzoff, & Stevens, 1991; McGurk & MacDonald, 1976). Research along this line has shown that visual information does not only enhance our perception of speech but can also alter it. Recently, researchers started to focus on the effect of tactile sensation beside the auditory signal, starting to reveal the impact of tactile information on speech perception (Derrick, Anderson, Gick, & Green, 2009; Reed, Rabinowitz, Durlach, & Braida, 1983). Moreover, different modes of response elicitation (open choice and forced choice, which will be discussed later) have been used to study these multisensory interactions (Colin, Radeau, & Deltenre, 2005; Sekiyama & Tohkura, 1991; Van Wassenhove, Grant, & Poeppel, 2005).

This thesis is based on multisensory integration, more specifically auditory – aero tactile (AT) integration. So, this study will focus on investigating the influence of AT integration during speech perception measured using an open choice response elicitation task. When an individual speaks, apart from lip movements, they release tiny air puffs as well, so the knowledge of the characteristics of articulation would be beneficial to understand how these cues can help us understand speech better. In the next section of the chapter, first and foremost, the production of speech in relation to the speech characteristics that distinguishes perception of speech sounds will be discussed. The effect of stimulus congruency in multisensory integration will be then briefly described in the section. Then, review of speech perception as a multisensory event including auditory – visual, auditory – tactile and visuo – tactile interaction with evidences from behavioural studies will be focused. Next, methodological aspects that can influence speech perception such as response type and stimulus type will be elaborated. The section will be concluded with the description of the statement of problem. Chapter II will describe the methodological design of the study. In Chapter III, the results provided by the current study will be presented. Finally, in the Discussion (Chapter IV), the empirical findings will be summarized, and further directions of research will be suggested.

1.2 Theoretical Background

1.2.1 Role of Articulatory Features in Speech Perception

A series of visual and tactile cues are generated along with auditory information when an individual speaks, so an understanding of speech production mechanism and how different sensory modalities contribute to speech perception using these cues need to be understood. Speech production is underpinned by a series of complex interactions of numerous individual processes beginning at the level of the brain with phonetic and motoric planning, followed by expelling of air from lungs that leads to vibration of vocal cords, termed as phonation. Air flow, including puffs of air released after phonation, are then routed to oral or nasal cavities, which gives the resonance quality needed. To become speech sounds, this air flow undergoes further shaping by various oral structures called articulators. The basic unit of a speech sound is called a phoneme, which may consist of vowels and consonants. Importantly, phonemes are classified based on features of articulation process that are used to generate them (Liberman, 1957).

Table 1: *Description of the summary of articulatory features for consonants - /p/, /b/, /k/, /g/ (phonetically represented).*

Consonants	**[p]**	**[b]**	**[k]**	**[g]**
Voicing	Voiceless	Voiced	Voiceless	Voiced
Place	Labial	Labial	Velar	Velar
Manner	Stop	Stop	Stop	Stop
Nasality	Oral	Oral	Oral	Oral

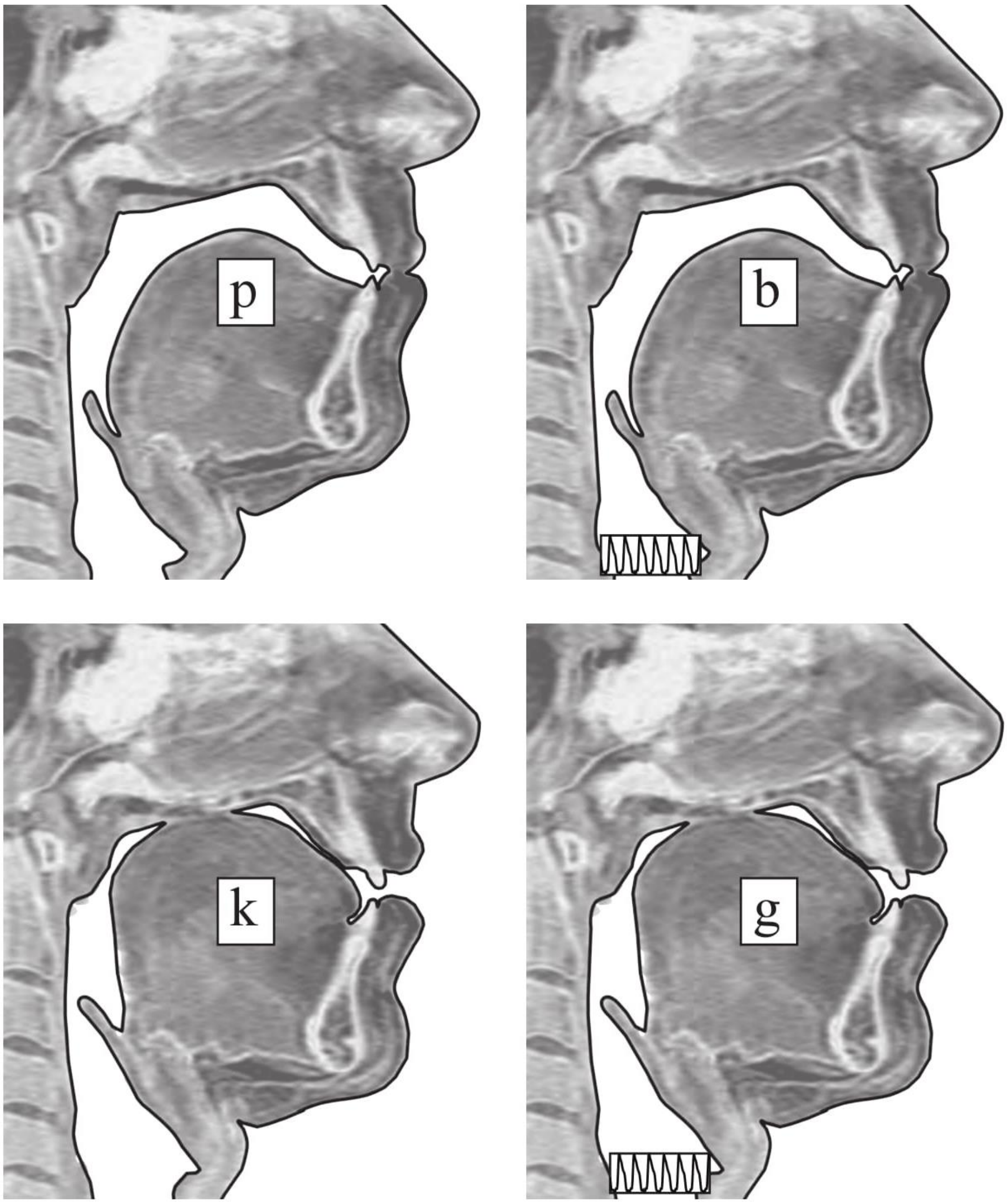


Figure 1. *MRI images traced out to demonstrate articulation of the consonants - /p/, /b/, /k/ and /g/.* Image courtesy by Dr. Donald Derrick.

The articulatory features that describe a consonant are its place and manner of articulation.

Figure 1 illustrates these articulatory features for consonants - /p/, /b/, /g/ and/k/ and the Table 1 summarizes the features for each of these consonants. For example, /b/ (phonetically [b]) made at the lips by stopping the airstream, is voiced, and is oral. These features of speech production are reflected in certain acoustic characteristics that can be presumably discriminated by the listener. Among them, voicing and place of articulation are two primary features that provide perceptually apparent acoustic cues, thus enabling easier distinction of the consonants. Place of articulation refers to the point of constriction in the vocal tract, especially in oral cavity, where closure occurs. Formant transitions and formant frequencies are the acoustic cues that underlie the place of articulation in consonant – vowel (CV) syllables (Liberman et al., 1967). On the other hand, presence or absence of periodic vocal cord vibration is the voicing feature. The acoustic cue that underlies the voicing feature is voice onset time (Lisker & Abramson, 1964) and it corresponds to the time interval between release from stop closure and the onset of laryngeal pulsing.

In English, six stops of three cognate pairs share place of articulation but differ in voicing feature. These consonants are the labial /p/ and /b/, alveolar /t/ and /d/, and velar /k/ and /g/, where the first of each pair is voiceless and the second is voiced. This is one class of consonants which are of particular interest in the field of multisensory integration. The stop consonants are a set of speech sounds that share the same manner of articulation. Their production begins with the build – up of pressure behind some point in the vocal tract, followed by a sudden release of that pressure. Stops have well defined acoustic properties, provided by place and voicing feature, with minimal difference in production.

Thus, it can be said that there is a clear link between how speech is produced and how it is perceived. In face – to – face communication between normally hearing people, manner of articulation of consonantal utterances is detected by ear (e.g., whether the utterance is voiced or voiceless, oral or nasal, stopped or continuant, etc.) (Miller and Nicely (1955); place of articulation, on the other hand, is detected by eye (Binnie, Montgomery, & Jackson, 1974).

When you stand close to a person, you can also feel the difference between voiced and voiceless stops. Recently, Derrick et al. (2009) demonstrated that the tiny puffs of air you can feel, released when we normally speak, can enable us to listen with our skin as well. The research in this domain has been actively in progress since, to study to what extent this phenomenon can be made possible to facilitate communication in everyday life.

Provided literature already indicated that speech perception is a multisensory process involving simultaneous syncing of visual, acoustic and tactile cues generated during phonetic production. When information from multiple senses are processed, misconceptions and confusions can occur as demonstrated previously in literature (Stroop, 1935) which will be discussed in the following section.

1.2.2 Role of Stimulus Congruency in Multisensory Integration

When multisensory information conflicts, the efficiency of the perceptual system, the superiority of one sensory modality over the other in various circumstances, and the essence of multisensory integration can be studied more accurately. In an experimental setting, this can be achieved by using congruent and incongruent stimulus conditions. Stimulus congruency is defined in terms of the properties of the stimulus. In stimulus congruency, at least a feature or dimension of the stimulus is same when it is presented through a single modality. The Stroop effect (Stroop, 1935) is a classic example to understand stimulus congruency through a single modality. In the *Stroop* task, participants respond to the ink colour of printed coloured words. Typically, participants performed better in terms of the reaction times and accuracy if the target word's meaning is congruent with its colour (e.g., the word "BLUE" in blue ink) than if colour and meaning are incongruent (e.g., the word "BLUE" in green ink). Importantly, in stimulus congruency, stimulus features could also be presented through different modalities. Sumby and Pollack (1954) evidenced that in a noisy background, watching congruent articulatory gestures improves the perception of degraded acoustic speech signal. Similarly, when stimulus becomes incongruent, it can also lead to perceptual confusions. Ventriloquism effect is a typical example of such a perceptual effect or illusion that arises in stimulus localisation as a result of multisensory integration (Howard & Templeton, 1966). Here, when auditory stimulus is presented simultaneously, but in a different location to a visual stimulus, the participants perceived sound to be in the location of the visual stimulus.

As discussed so far, articulatory features and congruency of stimulus was found to have an effect on multisensory integration and could influence speech perception. The subsequent section will be aimed to continue reviewing speech perception as a multisensory event, focusing on behavioural studies.

1.3 Multisensory Speech Perception

1.3.1 Speech Perception as a Multisensory Event

Speech perception is the process by which the sounds of language are heard, interpreted and understood. Research in speech perception seeks to understand how human listeners recognize speech sounds and use this information to understand spoken language. Just like the speech production process, the perception of speech also is a very complicated, multi-faced process that is not yet fully understood.

As discussed, speech can be perceived by the functional collaboration of the sensory modalities. In ideal conditions, hearing a speaker`s words is sufficient to identify auditory information. But when a sensory signal is degraded, information from other sensory sources were found to compensate and help in better understanding of speech (Macleod & Summerfield, 1990). So, recently researchers have viewed speech perception as a specialized aspect of general human ability, the ability to seek and recognize patterns. These patterns can be acoustic, visual, tactile or a combination of these. Speech perception, as a unified phenomenon involving association of various multisensory modalities (e.g. auditory – visual, auditory – tactile, visuo – tactile) will be discussed in detail now, reviewing the significant behavioural studies of multisensory literature.

1.3.2 Auditory – Visual (AV) Integration in Speech Perception

The pioneering work on multisensory integration showed that while audition remains vital in perceiving speech, our understanding of auditory speech is supported by visual cues in everyday life, that is, seeing the articulatory mouth movements of the speaker. This is especially true when the auditory signal is degraded or distorted (e.g., due to hearing loss, environmental noise or reverberation). For example, the individuals with moderate to severe hearing impairment can achieve higher levels of oral communication skills when focusing on the auditory – visual cues (Grant, Walden, & Seitz, 1998; Thornton & Erber, 1979). Previous researches (Macleod & Summerfield, 1990; Sumby & Pollack, 1954) of speech perception in noise confirms that speech is better understood in noise with visual cues.

On the other hand, even with fully intact acoustic speech signal, visual cues can have an impact on speech recognition. Based on a general observation from a film, McGurk and MacDonald (1976) designed an AV study that lead to a remarkable break – through in speech perception studies. They presented incongruent AV combinations of 4 monosyllables (/pa/, /ba/, /ga/, /ka/) to school children and adults in auditory only and AV conditions. Subject`s responses were elicited by asking them to repeat the utterances they heard. Findings from the study demonstrated that when the auditory production of a syllable is synchronised with visual production of an incongruent syllable, a few subjects perceived a third syllable that is not represented by either auditory or visual modality. For instance, when the visual stimulus /ga/ is presented with the auditory stimulus /ba/, many subjects reported perceiving /da/. In contrast, reverse combinations of such incongruent monosyllabic stimuli usually elicit response combinations like /baga/ or /gaba/ (Hardison, 1996; Massaro, 1987; McGurk & MacDonald, 1976). This unified integrated illusionary percept, formed as a result of either fusion or combination of stimulus information, is termed as "McGurk effect". This phenomenon stands out to be the basis of speech perception literature substantiating multisensory integration.

Conversely, studies suggest that the McGurk effect cannot be easily induced in other languages (e.g. Japanese) as in English. Sekiyama and Tohkura (1991) evidenced this in their investigation to evaluate how Japanese perceivers respond to the McGurk effect. Ten Japanese monosyllables were presented in AV and auditory only condition in a noisy and noise-free environment. Perceivers had to write down the syllable they perceived. When the AV and auditory only conditions were compared, the results suggested that the McGurk effect depended on auditory intelligibility and was induced when auditory intelligibility was lower than 100%, else it was absent or weak.

The ability to combine auditory – visual stimuli was found to be existing in the earlier years of life when researches evidenced AV integration in pre – linguistic children (Burnham & Dodd, 2004; Kuhl & Meltzoff, 1982; Pons et al., 2009) and in non – human primates (Ghazanfar & Logothetis, 2003).

Kuhl and Meltzoff (1982) investigated AV integration of vowels (/a/ and /i/) in 18 to 20 – week – old typically developing infants by scoring their visual fixation. Results showed a significant effect of auditory – visual correspondence and vocal imitation by some infants, which is suggestive of multimodal representation of speech. Similarly, the McGurk effect was replicated in pre – linguistic infants aged four months employing a visual fixation paradigm by Burnham and Dodd (2004). Infant studies indicate that new – borns also possess sophisticated speech perception abilities, as demonstrated by their multisensory syncing capacity which lay the foundations for their subsequent language learning.

In addition, it is quite interesting to know that other non – human primates also exhibit a similar AV synchronisation, as humans, in their vocal communication system. Rhesus monkeys were assessed to see if they could recognize AV correspondence between their ‘coo’ and ‘threat’ calls by Ghazanfar and Logothetis (2003) using preferential – looking technique to elicit responses generated during the AV task. Their findings were suggestive of an inherent ability in rhesus monkeys to match their species – typical vocalizations presented acoustically with the appropriate facial articulatory posture. The presence of multimodal perception in an animal`s communication signals may represent an evolutionary precursor of human`s ability to make the multimodal associations necessary for speech perception.

Complementarity of visual signal to acoustic signal and how it is an advantage to speech perception has been reviewed so far. Moreover, the fact that AV association is an early existing ability as evidenced by animal and infant studies was also considered. Recently research progressed a step ahead beyond AV perception and extended findings by examining how tactile modality can influence auditory perception which is discussed in the following section.

1.3.3 Auditory – Tactile (AT) Integration in Speech Perception

Less intuitive than the auditory and visual modalities, the tactile modality also has an influence on speech perception. Remarkably, the literature suggests that speech can be perceived not only by eyes and ears but also by hand.

Robust evidence for manual – tactile speech perception, mainly derives from studies on the Tadoma method (Alcorn, 1932; Reed et al., 1978). The Tadoma method is a technique where oro – facial speech gestures are felt and monitored from hand contact (also haptic or manual – tactile) with the speaker's face. Previous research on the effect of tactile information on speech perception has focused primarily on enhancing the communication abilities of deaf – blind individuals (Chomsky, 1986; Reed et al., 1985).

However, numerous behavioural studies have also shown the influence of tactile cues on speech recognition in healthy individuals. Treille and colleagues (2014) showed that reaction time for speech perception is altered when haptic (manual) information is provided additionally to the auditory signal in untrained healthy adult participants. Fowler and Dekle (1991), on the other hand, described a subject who perceived /va/ responses when feeling tactile (mouthed) /ba/, presented simultaneously with acoustic /ga/. So, when the manual – tactile contact with a speaker's face was coupled with incongruous auditory input, integration of both auditory and tactile information evoked a fused perception of /va/, thereby representing an auditory – tactile McGurk effect.

Manual – tactile contact is not the only form of tactile information that has been shown to influence speech perception. Work conducted by Derrick and colleagues demonstrated that puffs of air on the skin, when combined with an auditory signal, can enhance speech perception. In a recent study conducted by this group (Gick & Derrick, 2009), untrained and uninformed healthy adult perceivers received puffs of air (aero tactile stimuli) on their neck or hand while simultaneously hearing aspirated or unaspirated English plosives (i.e., /pa/ or /ba/). Participant responses for the monosyllabic identification task was recorded by pressing keys corresponding to the syllable they heard. Participants reported that they perceived more /pa/ in the aero tactile condition indicating that listeners integrate this tactile and auditory speech information in much the same way as they synchronous visual and auditory information. A similar effect was replicated using puffs of air at the ankle (Derrick & Gick, 2013) demonstrating that the effect does not depend on spacial ecological validity.

In addition, in the supplementary methods of their original article, Derrick and Gick (2009b) demonstrated the validity of tactile integration in speech perception by replicating their original experiment (Gick & Derrick, 2009), this time not by presenting puffs on the hand but by replacing them with taps from a metallic solenoid plunger. No significant effect of tap stimulation on speech perception was observed. This confirmed that participants were not merely responding to generalized tactile information nor it was the result of increased attention. This also indicates that listeners were shown to respond to aero tactile cues normally produced during speech, which evidences multisensory ecological validity.

Moreover, to illustrate how airflow would help in distinguishing minor differences in speech sounds, Derrick and colleagues (2014) designed a battery of eight experiments consisting of comparisons with different combinations of voiced and voiceless (stops, fricatives and affricatives) English monosyllables (/pa/, /ba/, /ta/, /da/, /fa/, /ʃa/, /va/, /t͡ʃa/, /d͡ʒa/). The study was run on 24 healthy participants where auditory stimuli was presented simultaneously with air puffs. Participants were asked to choose which of two syllables they heard. Perception of stops and fricatives were found to be increased in this study with the presentation of puff. The obtained results show that AT integration enhance speech perception for a large class of speech sounds found in many languages across the world.

Extending Gick and Derrick (2009) findings on air puffs, Goldenberg, Tiede, and Whalen (2015) investigated the effect of air puffs during identification of syllables on a voicing continuum rather than using voiced and voiceless exemplars as in the original work (Gick & Derrick, 2009). English consonants (/pa/, /ba/, /ka/ and /ga/) were the syllables used for the study. The auditory signal was presented simultaneously with and without an air puff received on the participant`s hand. Eighteen healthy adults took part in the study and had to press a key corresponding to their response from a choice of two syllables. Their findings showed an increase in voiceless responses when co – occurring puffs of air were presented on the skin. In addition, this effect became less pronounced at the endpoints of the continuum. This suggests that the tactile stimuli exert greater influence in cases where auditory voicing cues are ambiguous, and that the perception system weighs auditory and aero tactile inputs differently.

Temporal asynchrony during speech perception was evaluated by Gick, Ikegami, and Derrick (2010) to establish the temporal ecological validity of AT integration. They assessed whether asynchronous cross – modal information can be integrated in a similar way by presenting auditory (aspirated “pa” and unaspirated “ba” stops) and tactile (slight, inaudible, cutaneous air puffs) signals synchronously and asynchronously. The experiment was conducted in 13 healthy participants who chose the speech sound they heard out of the two alternatives using a button box. Conclusions from the study suggested that subjects integrate auditory – aero tactile speech over a wide range of asynchronies, as in cases of auditory – visual speech events. Furthermore, findings also suggest that perceivers accommodate variability in the transmission speed of signal in multimodalities.

As AT integration was found to be effective for the perception of syllables, as evidenced in the studies above, researchers extended their evaluation to see if a similar effect could be elicited for complex stimulus types like words and sentences.

Derrick and colleageus (2016) studied the effect of speech air flow on syllable and word identification task with onset fricatives and affricates in two languages – Mandarin and English. All 24 participants in their study had to choose a response from an alternative of two expected responses when an auditory stimulus was presented with and without air puff. The results showed that air flow helps to distinguish syllables. The greater the distinction between the two choices, the more useful air flow is in helping to distinguish syllables. Though this effect was noticed to be stronger in English than in Mandarin, it was significantly present in both languages.

Recently, the benefit of AT integration in continuous speech perception was examined with highly complex stimuli and with an increased task complexity by Derrick et al. (2016). The study was conducted on hearing – typical and hearing – impaired adult who received air puffs on their temple while simultaneously hearing five – word English sentences (e.g. Amy bought eight big bikes). Data were recorded for puff and no puff conditions and the participants were asked to say aloud the words after perceiving each sentence. The outcome of the study suggested that air flow does not enhance recognition of continuous speech as it could not be demonstrated in the hearing – typical and hearing – impaired populations. Thus, in continuous speech the beneficial effect of air flow in speech perception could not be replicated.

The line of speech perception literature discussed above clearly illustrated the effect of AT integration in speech perception for several stimulus types (syllables or words), in different languages and for a wide range of temporal asynchronies. But a similar benefit could not be noticed in continuous speech. It remains unclear why this is the case and needs further investigation. Just as AT and AV integration were found to be beneficial in enhancing understanding of speech, investigations were carried out to see if visuo – tactile integration could lead to better perception of speech, which will be discussed in the upcoming section.

1.3.4 Visuo – Tactile (VT) Integration in Speech Perception

It is quite interesting that speech, once believed to be an aural only phenomenon, can be perceived without the actual presence of an auditory signal. Gick et al. (2008) examined the influence of tactile information on visual speech perception using the Tadoma method. They found that syllable perception of untrained adults improved by around 10% when they felt the speaker's face whilst watching them silently speak, when compared to visual speech information alone.

Recently, the effect of aero tactile information on visual speech perception of English labials (/pa/ and/ba/) in the absence of an audible speech signal has been investigated by Bicevskis, Derrick, and Gick (2016). Participants received visual signals of the syllables alone or synchronously with air puffs on neck at various timings. Participants had to identify the syllables perceived from a choice of two. Even with temporal asynchrony between air flow and video signal, perceivers were more likely to respond that they perceived /pa/ when air puffs were present. Findings of the study showed that perceivers have shown the ability to utilise aero tactile information to distinguish speech sounds when they are presented with an ambiguous visual speech signal which in turn confirms that VT integration occurs in the same way as AV and AT integration.

Research into multimodal speech perception thus shows that perceptual integration can occur with AV, AT, and VT modality combinations. These findings support the assumption that speech perception is the sub – total of all the information from different modalities, rather than being primarily an auditory signal that is merely supplemented by information from other modalities. However, the most recent AT study on continuous speech (Derrick et al., 2016) did not show the effect of multisensory integration even though majority of them support AT integration. So, a thorough understanding of factors that may have led to this contrasting outcome is needed, which is elaborately explained in the next section.

1.4 Methodological Factors that Affect Speech Perception

In the previous sections of this chapter, we discussed the wide range of behavioural studies on different sensory modality combinations (AV, VT and AT) that supported multisensory integration. However, a similar result could not be replicated in continuous speech perception study described in 1.3.3 (Derrick et al., 2016). The next step would be to determine factors that potentially resulted in this variability. Based on a careful literature review, the following methodological differences could have contributed to this contrasting finding: the response type used to collect response from the participants; and specific stimulus differences (e.g., stimulus type changing from syllable level to a five – word sentence identification task).

1.4.1 Effect of Response Type on Speech Perception

Speech perception studies have predominantly adopted two different ways in which individuals are asked to report what they perceived. These response types include open choice and forced choice responses. Open choice responses allow individuals to independently produce a response to a question by saying it aloud or by writing it down. In contrast, forced choice responses give preselected response choices out of which individuals are asked to choose the best (or correct) alternative. Moreover, forced choice task delivers cues that may not be spontaneously considered, while in open choice responses that are generated by the participants, no cues are made available. However, open choice responses can sometimes be difficult to code, unlike forced choice responses that are easier to code and work with (Cassels & Birch, 2014).

Therefore, different approaches may be used in these two response types. For example, participants might adopt an eliminative strategy for forced choice tasks, whereby participants continue to refine alternatives by eliminating the least likely alternative until they arrive at the expected option (Cassels & Birch, 2014). A task that exhibits demand characteristics or experimenter expectations (“did the stimulus sound like pa?”) might give different results than one that does not (“what did the stimulus sound like?”)(Orne, 1962). Forced choice tasks are quite often influenced by demand characteristics. However, non – directed questions can be used to avoid this in open choice tasks.

Importantly, the literature evidences that the experimental paradigm can affect the behavioural response. Colin et al. (2005) examined how sensory and cognitive factors regulate mechanisms of speech perception using the McGurk effect. They calculated McGurk response percentage by manipulating the auditory intensity of speech, face size of the speaker, and the participant instructions to make responses aligned with a forced choice or an open choice format. However, like many other studies, they instructed their participants to report what they heard. They found significant effect of instruction manipulation, with higher percentage of McGurk response for forced choice task. However, in the open choice task, the participant responses were diverse as they were not provided with any response alternatives. The reduction in the number of McGurk responses in the open choice task may be attributed to the participants being more conservative about their responses so that they could report exactly what they perceived. They also found an interaction between instructions used and the intensity of auditory speech. Likewise, Massaro (1998) found that stimuli were correctly identified and elicited more frequently in a forced choice task than in an open choice task.

Recently, Mallick, Magnotti, and Beauchamp (2015) replicated findings of Colin et al. (2005) when they examined the McGurk effect by modifying parameters like population, stimuli, time, and response type. They demonstrated that the frequency of the McGurk effect can be significantly altered by response type manipulation, with forced choice response increasing the frequency of McGurk perception by 18% approximately, when compared with open choice for identical stimuli.

In the literature review provided, it is clearly visible that there is a variation in the strength of the multisensory integration effect across the studies where the multisensory integration effect was found to be null in the most recent study on continuous speech. Table 2 given below summarizes the described studies from 1.3.3, providing an overview with respect to used stimuli, paradigm and effect of multisensory integration.

Table 2: *Summary of the AT studies from section 1.3.3 based on response type, stimulus type and study results.*

Article/ Study	**Response type/ Paradigm used**	**Stimuli used**	**Multisensory integration present or not**
Treille, Cordeboeuf, Vilain, & Sato (2014)	Forced choice	Monosyllables	Significant auditory-haptic (tactile) benefit
Fowler & Dekle (1991)	Forced choice	Monosyllables	Significant auditory-haptic (tactile) benefit

Gick & Derrick (2009)	Forced choice	Monosyllables	Significantly strong AT benefit
Derrick & Gick (2013)	Forced choice	Monosyllables	Significantly strong AT benefit
Derrick & Gick (2009b)	Forced choice	Monosyllables presented with metallic taps instead of air puff	No AT benefit
Derrick, O'Beirne, Rybel, & Hay (2014)	Forced choice	Monosyllables	Significantly strong AT benefit
Goldenberg, Tiede, & Whalen, (2015)	Forced choice	Monosyllables	Significant AT benefit
Gick, Ikegami, & Derrick (2010)	Forced choice	Monosyllables	Significantly strong AT benefit
Derrick, Heyne, O’Beirne, Rybel, Hay, & Fiasson (2016)	Forced choice	Syllables and words	AT integration present in Mandarin but not very strong as in English
Derrick, O'beirne, De Rybel, Hay & Fiasson (2016)	Open choice (say aloud)	5 - word sentences	No significant AT benefit

The conclusions of Colin et al. (2005) and the literature review suggest that response choice may be a significant contributor for variability in speech perception studies. In case of forced choice responses, participants can compare their percept with available response alternatives.

However, in case of open choice responses, participants attempt to retrieve which syllable closely matches their percept from an unrestricted number of possible syllables. In the continuous speech study, Derrick et al. (2016) used an open choice paradigm, which would have been a very complex task due to the unpredictability of the possible outcomes, and this might have led to a null result.

1.4.2 Effect of Stimulus Type on Speech Perception

As outlined in section 1.3 and illustrated in Table 2, the type of stimuli may also affect the behavioural outcome. Continuous speech stimuli like phrases or sentences, can be considered as complex stimuli type because they contain confounding factors (e.g., semantic information, context information, utterance length etc.). This additional information undergoes complex and higher language processing, unlike syllable, which is a simple unit of language whose recognition does not require complex language processing. Hence another possible cause of insignificant AT effect on continuous speech could be because of the use of sentences stimuli.

Similarly, Liu and Kewley-Port (2004) measured vowel formant discrimination in syllables, phrases, and sentences for high – fidelity speech and found that the thresholds of formant discrimination were poorest for the sentences context, the best for the syllable context, and the isolated vowel in between them. Thus, indicating that complexity of task increased with the stimulus type, where sentences stimuli was the most difficult.

In summary, Table 2 suggests that stimulus type represents context effect. It can be assumed that when an information in a sentence or phrase is decoded by brain, an additional effect of language processing (e.g., phonetic, phonological, semantic and syntactic processing) is added along with general auditory processing. And the by – product of this multiple level processing is context effect. Hence, type of stimuli (e.g., sentences) in a study can contribute to an increase in the task complexity.

So, methodological factors like response type and stimulus type, discussed above can be the possible features that interfered in multisensory integration of the continuous AT perception study (Derrick et al., 2016). So, in this study, these aspects will be taken into consideration to go a step closer towards more real – life stimuli than the original AT experiment (Gick & Derrick, 2009), to allow us to gradually test which of these factors was responsible for not being able to reproduce the AT effect in continuous speech study.

1.5 Statement of the Problem

Studies investigating AT integration using syllables as stimuli and a forced choice paradigm as response type, demonstrated AT integration. However, AT integration could not be replicated when investigated using sentences as stimuli and an open choice paradigm as response type. But the approach of the study by Derrick et al. (2016) did not follow a continuous hierarchical pattern, i.e., stimulus type was suddenly upgraded to five – word long sentences from monosyllables, which was a radical change. In addition to that, a comparatively sophisticated response type, the open choice task, was also chosen for the study. Thus, the unanswered question is to determine which of the following factors lead to null result in continuous speech perception paper; use of sentences or open choice task (Derrick et al., 2016). To determine this, it is important to take a step back and investigate multisensory AT integration by using syllables in an open choice paradigm. In order to make the task easier for participants and to make the experiment shorter, the accuracy level for task completion was set to 80% (Watson & Pelli, 1983). Furthermore, in the present study design, the congruency of the stimulus, presented in different modalities, were altered to obtain and study the efficacy of AT effect in perception of speech through skin.

1.5.1 Study Aim

The present study aims to identify whether the benefits from auditory – aero tactile integration uphold for a monosyllable identification task in varying signal – to – noise ratios during congruent and incongruent stimulus conditions when the participants do not have to make a forced choice between two alternatives but are presented with a more ecologically valid open choice condition.

1.5.2 Hypothesis

The research question is whether aero tactile information influences auditory syllable perception using an open choice identification task. This will be investigated by testing the following 2 hypotheses:

The signal – to – noise ratio (SNR) at 80% accuracy levels will interact with phoneme and air flow such that:

Hypothesis 1: - SNR at the 80% accuracy level will be decreased when listening to congruent AT stimuli compared to audio only stimuli.

Hypothesis 2: - SNR at 80% accuracy level will be increased when listening to incongruent AT stimuli than audio only stimuli.

1.5.3 Justification

The proposed study is an extended version of the previous work of Gick and Derrick (2009) but with the mode of response being an open choice design instead of a 2 – way forced choice paradigm task. This response format was chosen because it has been shown to provide a more conservative estimate of the participant`s percept in previous studies (Colin et al., 2005; Massaro, 1998). Moreover, an open choice design allows for a better assessment of the precision of AT integration in speech perception as the possibility of subject guessing can be minimized. Hence, the outcome of this study will extend our knowledge on the effectiveness of integration of tactile information in the enhancement of auditory speech perception, in a more natural setting with minimal cues.

In addition to this, simplifying the stimulus type to monosyllables would allow us to identify the conditions under which AT integration occurs, without the confounding factors needing higher cognitive and linguistic processing (semantic information, context information, utterance length etc.) that were present during the continuous speech studies (Derrick et al., 2016). Congruent and incongruent stimulus conditions used in the study enables us to confirm the effectiveness and validity of multisensory integration.

1.5.4 Significance

This study will be a valuable contribution to the multisensory speech perception literature as it adds on to the fundamental scientific knowledge of how information from various sources are unified in body. Moreover, insights from the study could be used for evidence – based clinical practise by professionals developing strategies and communication aids for training communication skills in individuals with sensory deficits.

Chapter II: Methodology

The current study builds on the methodology of the original aero tactile integration paper (Gick & Derrick, 2009), coupling an acoustic speech signal with small puffs of air on the skin. The difference with the present study is that this time the participants are free to choose their response, without any constraints, based on their own perceptual judgement rather than having to choose between two response alternatives.

2.1 Participants

Forty – four healthy participants (40 females), with a mean age of 23.34 years were recruited for the study. The University of Canterbury Human Ethics Committee has reviewed and approved this study on 15 May 2017 (Approval number 2017-21 LR).

Participants (n = 34) were primarily undergraduate speech – language therapy students and the remaining (n = 10) were recruited via email, Facebook, advertisement on the New Zealand Institute of Language, Brain and Behaviour (NZILBB) website and around the university. Undergraduate speech – language therapy students received course credits for their research participation while other volunteers were given a NZ$10 gift voucher as compensation for their time. As part of the recruitment process, participants received an information sheet (Appendix A), which was discussed with them before beginning any of the procedures. Following this discussion, if they chose to participate, they were asked to sign a written consent form (Appendix B).

Inclusion criteria set for the recruitment process were: (1) Native English speaker, (2) Aged between 18 to 45, and (3) No current/history of speech, language or hearing issues.

Participants not meeting the inclusion criteria could choose to still complete the study to gain experience participating in research. Of the 44 participants, seven participants did not meet language criteria (native English speaker) and three participants did not meet the hearing criteria (i.e., they had a pure tone average threshold of >25db in at least one ear). This resulted in a total of 34 participants. In addition, five participants were not too good at the experiment (ceiling effect) particularly for some of the conditions, hence their data could not be included fully. A ceiling effect was reached as they were unable to correctly identify some of the stimuli at a +10 dB SNR level, suggesting they had difficulty doing task in an effectively noiseless environment. None had to be completely excluded for this reason, but participant 2's /ka/ and /ba/, participant 6's /pa/, participant 8's /ka/, participant 14's /ka/, and participant 37's /ga/, /da/ and /ba/ data had to be excluded.

All the participants were asked to complete a questionnaire (Appendix C) detailing demographic information on age, native language and speech, language and hearing difficulties. As part of the initial protocol, participants underwent an audiological screening. Pure tone audiometry testing was carried out for frequencies 500Hz, 1KHz, 2KHz and 4KHz using an Interacoustics AS608 screening audiometer. Average pure tone thresholds were calculated and if the threshold was less than or equal to 25dB HL, hearing sensitivity was considered to be within normal range.

2.2 Recording Procedure and Stimulus Generation

The speech stimuli were recorded in a sound – attenuated booth using a Sennheiser MKH-416 microphone attached to a Sound Devices USB – Pre2 microphone amplifier fed into a PC. Video recordings of the English syllables, with labials (/pa/ and /ba/) and velars (/ka/ and /ga/), spoken by a female native New Zealand English speaker were recorded using a video camera (Panasonic Lumix DMC-LX100). The speaker produced twenty repetitions of each stimulus, and the stimuli were presented in randomized order to be read aloud off a screen.

To produce the air puff, an 80ms long 12kHz sine wave was used to drive the pump action of Aerotak (Derrick & De Rybel, 2015). This system stores the auditory signal and the air flow signal in the left and right channel of a stereo audio output respectively. The stored audio is used to drive a conversion unit that splits the audio into a headphone out (to both ears) and the right channel air pump drives the signal to a piezoelectric pump, mounted on the tripod, to release air puff to the participant.

2.2.1 Auditory Stimuli

The English syllables were matched for duration (390 – 450 ms each), fundamental frequency (falling pitch from 90 Hz to 70 Hz) and intensity (70 decibels). Using an automated process, speech token recordings were randomly superimposed 10,000 times within a 10 second looped sound file to generate speech noise for the speaker. According to Jansen and colleagues (2010) and Smits and colleagues (2004), this method of noise generation results in a noise spectrum virtually identical to the long – term spectrum of the speech tokens of the speaker and thus ensuring accurate SNRs for each speaker and token. Speech tokens and the noise samples were adjusted to the same A – weighted sound level prior to mixing at different SNRs.

Recordings of the four underlying tokens of /ba/, /pa/, /ga/, and /ka/ were overlaid with speech in noise generated using the same techniques described in Derrick et al. (2016) with exception that the software used was custom R and FFMPEG. The SNR of the stimuli ranged from -20 to 10 SNR with 0.1 SNR increments. From -20 to 0 SNR, the volume of the signal was decreased, and volume of the noise was kept stable. From 0 to 10 SNR, the signal was kept the same volume and noise was decreased. Thus, the overall amplitude was maintained stable throughout the experiment.

2.2.2 Tactile Stimuli

In order to create the airflow produced by the airflow generation system (Aerotak), with the dynamics of it produced in speech that is measured during recordings, the air flow outputs were generated by a Murata MZB1001T02 piezoelectric device (Tokyo, Japan). This is controlled through the Aerotak system, as described in Derrick, De Rybel, and Fiasson (2015). The Aerotak system uses the air flow signal to activate the Murata pump, which delivers air flow to the skin of the participant.

2.3 Procedure

The study was designed to examine the influence of AT integration on speech perception using an open choice task. Each participant`s perception was assessed using randomized presentation of 8 possible combinations of auditory only and congruent and incongruent auditory and AT stimuli of English syllables - /pa/, /ba/, /ga/ and /ka/ represented in table 3. Each participant heard these 8 tokens 4 times each and there were 32 repetitions for each syllable.

The experiment was run individually for each participant. The entire procedure lasted approximately 40 minutes. Data was collected using an Apple MacBook Air laptop in a sound attenuated room using a sound attenuating (closed) headphone (Panasonic Stereo Headphones RP-HT265).

Table 3: *Congruent and incongruent stimulus conditions for syllables - /pa/, /ba/, /ga/ and /ka/.*

Congruent stimulus conditions	**Incongruent stimulus conditions**
Voiced no puff	Voiced puff
• /ba/ no puff	• /ba/ puff
• /ga/ no puff	• /ga/ puff
Voiceless puff	Voiceless no puff
• /pa/ puff	• /pa/ no puff
• /ka/ puff	• /ka/ no puff

Once the initial screening protocol was done, participants were presented with the auditory stimuli via headphones at a comfortable loudness level through an experiment designed in PsychoPy software (Peirce, 2007, 2009). The aero tactile stimuli were delivered at the suprasternal notch via the piezoelectric pump, described above, placed aiming at the subject`s neck fixed at ~2.2cm from the skin surface. It has the following specifications: the 5-95% rise time takes 30 ms (Derrick et al., 2015), with a maximum pressure of 1.5 kPa (15.29 cm H_2O, normal speech volume ~7 cm H_20) at the source, and a maximum flow rate of 0.8 l/m (normal speech volume ~10 l/m), which corresponds to about a twelfth of that of actual speech. The suprasternal notch at the neck was chosen to present the tactile stimuli because it is a location where participants do not normally receive a direct airflow during their own speech production.

Participant`s perception was estimated by asking them to type out the perceived syllable into the experiment control program. The length of time taken to identify one token was 6.5 seconds on average. The correct responses were /pa/, /ba/, /ga/, and /ka/ or/ca/ based on the underlying auditory signal. Whenever the participant responded accurately, the SNR decreased, thereby increasing the task complexity. Similarly, for every incorrect response, SNR increased, making the signal clearer and simplifying task for the participants. Thus, results of each trial allowed for a re – tuning of the SNR`s for each syllable to compensate for how easy the individual recording was for perceivers to detect in noise. This method of assigning stimulus values based on preceding response follows that of an adaptive staircase. In the current study, an adaptive transformed up – down staircase was used to vary the SNRs and obtain a psychometric curve of perception based on the 80% accuracy response in noise, which is achieved by the 32th repetition where the SNR becomes almost stable with minimal fluctuations. The transformed up – down method (Quest staircase) has been adopted as it is a reasonably fast and typical method (Watson & Pelli, 1983).

Participants were told that they might experience some noise and unexpected puffs of air along with syllables, consisting of a consonant and a vowel, during the task. Participants were asked to type the syllables that they heard and push the enter key to record their responses. As the experimental part of the study, requiring active listening, lasted about 20 minutes, participants could take short listening breaks if they required one. The researcher stayed inside the experiment room with the participant during the experiment to make sure that pump placement was not disturbed and to ensure that participants were comfortable.

2.4 Data Analysis

Of the forty – four participants who took part in the study, data of thirty – four participants who fit the inclusion criteria were analysed to answer the research questions. Initially, data was entered and sorted in a Microsoft Excel 2016 spreadsheet. The participant`s SNR level for the last repetition (32th), where 80% accuracy was achieved, was extracted and statistics were run on it. Previous literature suggested that articulatory features like place and manner of articulation were beneficial cues, that contribute to perception of syllables (Eimas et al., 1978; Lisker & Abramson, 1970; Miller & Eimas, 1977; Sawusch & Pisoni, 1974). Hence while analysing data, effect of these articulatory features in the perception of speech were also studied.

Descriptive statistics were run, and box plots were used to visualize variation of SNR with place of articulation. Variation of SNR with the audio only and AT condition for each target stimulus was also plotted using a boxplot.

Generalized linear mixed – effects models (GLMM), analysed using R statistical software (R Development Core Team, 2016), were run on the interaction between aspiration [aspirated (/pa/ and /ka/) vs. unaspirated (/ba/ and /ga/) stops], place [labial (/pa/ and /ba/) vs. velar (/ga/ and /ka/) stops], and artificial air puff (present vs. absent).

Model fitting was then performed in a stepwise backwards iterative fashion, and models were back – fit along the Akaike information criterion (AIC), to measure quality of fit. This technique isolates the statistical model that provides the best fit for the data, allowing elimination of interactions in a statistically appropriate manner. The final model was:

SNR ~ place * manner + (1 + (place * manner) | participant)

In this model, the SNR at 80% accuracy was compared to the fixed effects. These included: 1) Place of articulation (labial vs. velar), 2) manner of articulation (voiced vs. voiceless), 3) the interaction of place and manner, and 4) The full – factorial random effect covering place and manner of articulation by participant.

Chapter III: Results

The aim of the current study was to evaluate the effect of air puffs on perception of speech by employing open choice response elicitation. Specifically, SNR levels where 80% accuracy was obtained were of interest to measure the influence of the puff stimuli. Other factors like place and manner of articulation were also taken into consideration. In the following section of this chapter, initially results of the performance of native English speakers in the open choice speech perception task will be presented. Finally, the chapter will be concluded by evaluating the individual performance of the seven non – native English speakers to the speech recognition task.

3.1 Native English Speakers

3.1.1 Descriptive Statistics

The results report the SNR values (in dB) at 80% accuracy level for each target phoneme (/pa/, /ba/, /ka/ and /ga/), averaged over the 34 native English participants in various stimulus conditions. As stated in the hypothesis, the influence of stimulus congruency determined by the presentation of puff stimuli was studied based on the SNR values of the participants and the results are illustrated in Table 4 and using Figure 2. Voiced (/ba/ and/ga/) and voiceless (/pa/ and /ka/) syllables were presented with and without puff, generating 4 congruent (C) and 4 incongruent (IC) conditions as described earlier (Table 3, section 2.3). It was hypothesised that stimulus congruency would improve perception and that for incongruent conditions, the performance of participants might be poorer.

Table 4: *Mean and standard deviation of native English speaker for AT and audio only condition for different syllables.*

Place	Labials				Velars			
Manner	Voiced (/ba/)		Voiceless(/pa/)		Voiced (/ga/)		Voiceless(/ka/)	
Conditions	A only	AT	A only	AT	A only	AT	A only	AT
Congruency of stimuli	C	I C	I C	C	C	I C	I C	C
Mean	**-4.51**	**-5.39**	**-5.32**	**-5.13**	**-8.72**	**-8.24**	**-4.14**	**-3.35**
(SD)	(1.74)	(1.90)	(2.13)	**(**3.40)	(3.52)	(2.61)	(4.28)	(3.68)

Note: A only = Auditory only, AT = Auditory – aero tactile, C = Congruent stimulus,

IC = Incongruent stimulus and SD = Standard Deviation.

Contradicting observations were seen in the results. When voiceless syllables, /pa/ and /ka/ were presented with puff (congruent stimulus condition), the results (/pa/: C = - 5.13 dB, IC = - 5.32 dB and /ka/: C = - 3.35 dB, IC = - 4.14 dB) does not suggest any significant difference regarding improvement of perception with the introduction of puff stimuli. Similarly, the voiced syllables in the absence of puff (congruent stimulus condition), does not demonstrate an ease to perceive either, when compared with presence of puff (/ba/: C = - 4.51 dB, IC = - 5.39 dB and /ga/: C = - 8.72 dB, IC = - 8.24 dB).

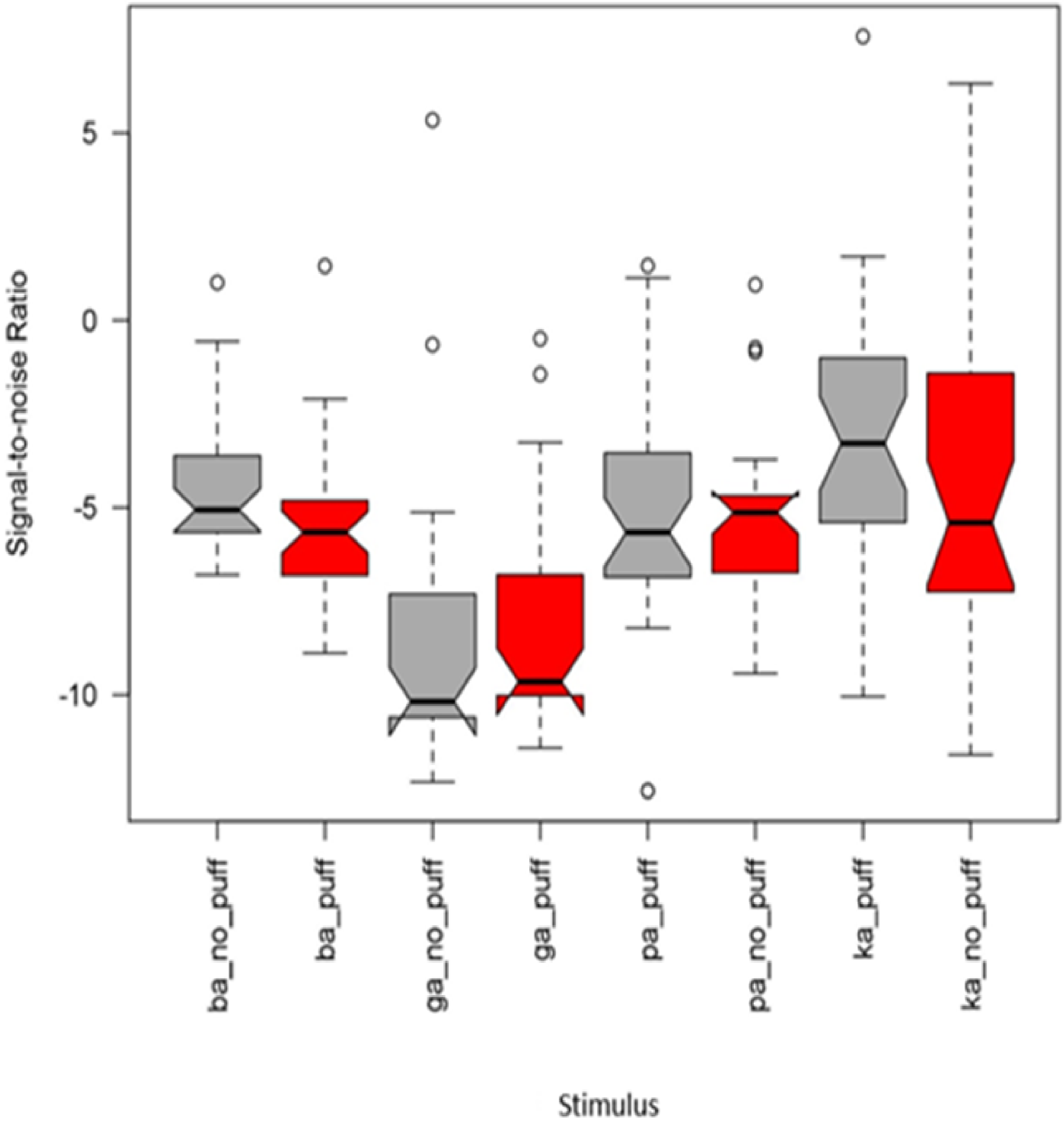


Figure 2. *Boxplot demonstrating variability of SNR as a function of different stimulus conditions. Red for incongruent conditions while grey is for congruent stimulus conditions.*

The observations from the Table 4 are supported by the visualisation of the participant`s performance, marked by the SNR levels, using a boxplot for the syllables studied in puff and no puff conditions. In the boxplot (Figure 2), the incongruent stimulus conditions are marked by red colour, while grey is used for congruent stimulus conditions. The average (median) of the participant responses were obtained at higher SNR levels for congruent stimuli than incongruent for all other syllables except /ga/ and /pa/. In case of /ga/ and /pa/, the median of the congruent stimuli (/ga/ no puff and /pa/ puff) was at a slightly lower SNR level than incongruent stimuli (/ga/ puff and /pa/ no puff). But this difference also does not appear to be a significant one, being negligibly low.

Thus, based on the information from the Table 4 and boxplot (Figure 2) it can be understood that the SNR values for none of the syllables in congruent AT conditions were significantly greater, with lower SNR averages, than incongruent conditions to advocate for the influence of air puff on perception of these syllables.

Furthermore, in the boxplot (Figure 2), outliers can be seen for all the syllable conditions. These outliers were not excluded from objective evaluation of data using GLMM, as the undesired outliers (data of participants that demonstrated ceiling effect, discussed in Section 2.1) were eliminated before applying different descriptive statistical tools.

Interestingly, among the syllables studied, /ga/ was the easiest to perceive in the puff and no puff conditions for participants, with the lowest SNR scores -8.24 dB and -8.72 dB respectively. This finding is in line with that of the observation in the plot as well, where SNR level of /ga/ was found to be the lowest of all. On the other hand, the voiceless velar (/ka/) was the hardest to perceive in both stimulus conditions. While perception of voiced (/ba/) and voiceless (/pa/) labials were more identical in both cases.

As discussed in previous chapter (Section 2.5), place of articulation could be an influential aspect in speech perception and hence it was included in the statistical model used in the study. Also following the observations made from the Table 4, the influence of place of articulation was also visualised using the following boxplot (Figure 3).

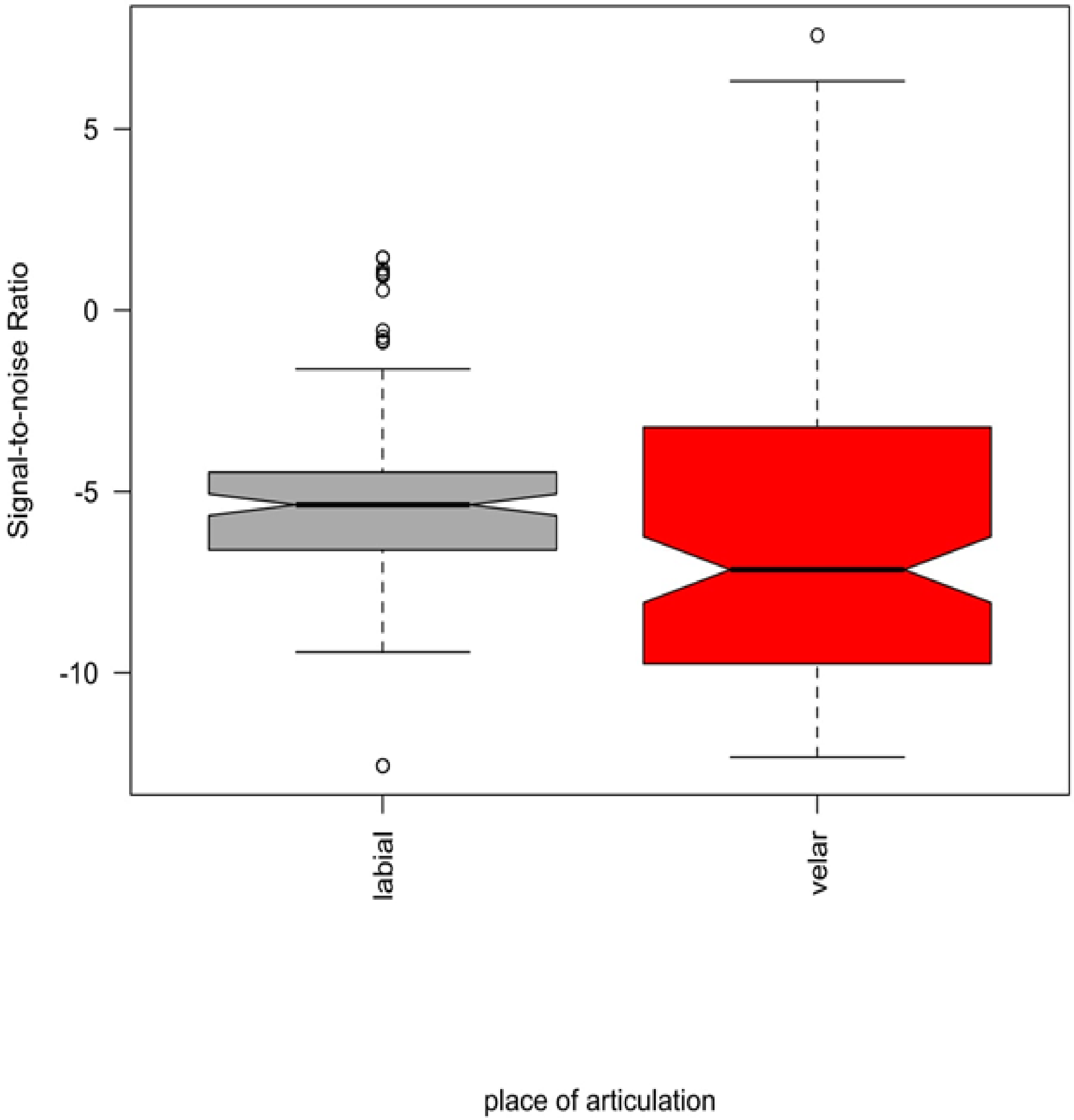


Figure 3. *Boxplot of variability of SNR as a function of place of articulation.*

As shown in Figure 3, labials (/pa/ and /ba/) have a comparatively shorter boxplot than velars (/ka/ and /ga/). This suggests that their variability is less across the participants, when compared with velars where the participants have high variability; hence data has a wider range. Even though majority of the participants demonstrated less variability in labial perception, there were a few outliers beyond the upper limit of the plot, who required a higher SNR level to perceive labials than others in the group. Conversely, one participant demonstrated ease to perceive labials with the lowest SNR level than others, which is represented by the outlier below the lower limit of the plot.

Even when participants demonstrated variability in velar perception, as a group they found velars to be easier to perceive than labials. The average SNR (median) was significantly lower for velars than the labial syllables. There is an outlier in velar group as well, whose SNR level is higher than the upper limit of the plot. These outliers were not eliminated while applying various statistical tools (GLMM) as unnecessary outliers were removed during initial evaluation of data. And we assume that these results were due to normal variation between participants.

3.1.2 Statistical Analysis using GLMM

From the descriptive statistics results, the effect of puff and AT congruency in AT integration was found to be negligibly low to show an effect, while place and manner of articulation showed an effect on SNRs. As stated in the methods section 2.4, the behavioural data were further tested by fitting them with Generalised linear mixed – effects model (GLMM), with the influence of puff, manner, place as fixed effects and individual subject variability across participants as random effects. The results are depicted in Table 5 below.

Table 5: *Summary of the GLMM model.*

	Estimate	**Standard Error**	**T-value**
Intercept	-4.951	0.297	-16.646
place (velar)	-3.564	0.490	-7.272
manner (voiceless)	-0.291	0.471	-0.619
place x manner	5.058	0.888	5.698

Statistical analysis using GLMM model confirmed the findings of descriptive statistics that the effect of air puff in perception of these syllables are statistically insignificant. As puff did not have any significant effect on these results, it is not included in the best-fit model. Thus, this finding answers the research question that aero tactile information does not influence auditory syllable perception during an open choice syllable identification task, leading to rejection of both the hypotheses stated in this study.

The mean SNR value of the sample is -4.95 dB. Velars had lower SNR than labials (estimate = -3.564, t-value = -7.272). SNR levels of the voiceless syllables were non – significantly lower than voiced syllables (estimate = - 0.291, t-value = - 0.619). Place and manner of articulation showed an interaction effect. The absolute t-value (5.698) is relatively far away from zero and is larger than the standard error (0.888) indicating that relationship exists between these articulatory features. Thus, the results show that participants had an easier time recognizing velar as opposed to labials. The interaction effect shows that it was also harder to understand voiceless velars compared to voiced velars.

3.2 Non – native English Speakers

Seven participants were not included in the above analysis as they were non – native English speakers and therefore did not meet our inclusion criteria. Their data were evaluated on a case – by – case basis. The table 6 below represents the origin and the native language of these participants.

Table 6: *Descriptive summary of non – native participants.*

Participant	Country of origin	First /Native Language
1	India	Telugu
2	Philippines	Tagalog
3	Hong Kong	Cantonese
4	Nepal	Nepali
5	Korea	Korean
6	Malaysia	Mandarin
7	Singapore	Chinese

Results of these 7 non – native speakers are illustrated in Table 7. The different languages spoken by the non – native English participants are listed vertically in the first column. The syllables in puff and no puff conditions are indicted horizontally across the top of the table. The value in each cell is the SNR level where 80% accuracy was obtained by the participant, specifically 32th repetition where the SNR level is mostly stabilised for the participants.

Table 7: *Representation of the SNR levels of non-native speakers in the study at 80% accuracy level.*

Languages	/ba/ puff	/ba/ no puff	/pa/ puff	/pa/ no puff	/ka/ puff	/ka/ no puff	/ga/ puff	/ga/ no puff
Stimulus congruency	IC	C	C	IC	C	IC	IC	C
Telugu	-6.23	-2.59	-6.79	-4.33	10	10	-9.72	-11.11
Tagalog	10	10	4.25	-4.25	-5.21	-2.78	10	10
Cantonese	-4.34	-5.45	-3.98	-4.65	-2.75	-7.73	-11.19	-12.21
Nepali	10	10	10	10	10	10	2.20	-6.46
Korean	-6.79	-6.96	-6.31	-5.37	-1.03	-9.25	-9.97	-10.73
Chinese	-3.69	-2.88	0.54	-1.05	1.88	-2.59	-7.11	-6.79
Mandarin	-2.90	0.65	-5.15	-6.79	2.59	2.43	-8.31	-10.71

Results of these non – native English speakers largely fell into 2 categories. The first group consisting of Mandarin, Chinese, Cantonese and Korean speakers, whose performance in the task had close resemblance to native English speakers. Conversely, Telugu, Tagalog and Nepali speakers constituted the second group who exhibited a pattern of perception that differs from the one we observed in the English speakers. Hence results of these non – native English speakers were grouped and were discussed accordingly.

Mandarin, Chinese, Cantonese and Korean speakers could perceive the syllables at similar SNR levels as the native English speakers. These participants did not demonstrate an improvement in perception of target syllables with the introduction of air puff along with auditory signal in congruent stimulus conditions. This trend was found to be similar to that of native English speakers (/pa/: C= - 5.13 dB, IC= - 5.32 dB, /ka/: C= - 3.35 dB, IC= - 4.14 dB, /ba/: C= - 4.51 dB, IC= - 5.39 dB and /ga/: C= - 8.72 dB, IC = - 8.24 dB) but variability exists between different non – native English participants. To exemplify, the Cantonese speaker demonstrated a slight improvement in the performance of /ba/ (C = - 5.45 dB, IC= - 4.34 dB) and /ga/ when stimulus condition was congruent, while Korean and Mandarin speakers showed mild effect of stimulus congruency in syllables /pa/ (C = - 6.31 dB, IC= - 5.37 dB), /ga/ (C = -10.73 dB, IC = - 9.97 dB) and /ga/ (C = - 10.71 dB, IC = - 8.31 dB) respectively. Though slight improvements were seen, which is indicative of a trend to demonstrate the effect of stimulus congruency and air puff, these differences were low and hence the effect of puff on speech perception could not be claimed. For all the four participants from different languages, puff and no puff conditions of /ga/ syllable were easiest to perceive with the lowest SNR levels. This is also similar to the behaviour of the native English speakers in the study who had the lowest SNR for /ga/ (i.e., - 8.24 dB for puff condition and - 8.72 dB for no puff condition). Perception of velars was found to be comparatively better than labials for these non – native speakers, again similar to that of native English speakers.

On the other hand, speakers of Group 2 (Telugu, Tagalog and Nepali speakers) illustrated similar pattern in the task, they all found the task to be extremely difficult because they reached ceiling effect (i.e. maximum SNR level 10 dB, discussed in section 2.1) for one or more stimulus conditions. Among them, Nepali speakers demonstrated the highest task difficulty, they could only perceive puff (2.20 dB) and no puff condition (-6.46 dB) of /ga/ with ease and for every other syllable conditions they found it extremely hard to perceive the syllable. These speakers also could not demonstrate the benefit of AT effect in their perception of syllables as SNR values of puff condition was poorer than no puff condition.

Moreover, stimulus congruency also did not seem to have an influenced the syllable perception for these non – native English speakers. Finally, for Telugu and Nepali speakers, perception of velars, specifically voiced /ga/, was observed to better than labials. Unlike them, voiceless labials and velars were easier for Tagalog speakers.

Since the data of non – native English speakers were analysed individually, none of the statistical tools were applied. The findings from their data evaluation does not suggest an influence of air puff and hence does not support the hypothesis stated in the study.

Thus, interaction between SNR and air puff for the syllables studied, in congruent and incongruent stimulus conditions, does not suggest any significant effect of air puff in speech recognition for the native and non – native English speaking participants in the current study, however place and manner demonstrated interaction which is suggestive of their influence in speech perception. The possible justifications for these findings will be discussed in the following chapter.

Chapter IV: Discussion

The array of multisensory speech perception literature discussed earlier suggests that perceptual integration occurs for various bimodal combinations like auditory – visual, visuo – tactile and auditory – tactile. Thus, speech perception can be considered as the integrated resultant of the information from various sources. In the landmark study on auditory – aero tactile perception by Gick and Derrick (2009), the presentation of aero tactile stimuli that are congruent with the auditory signal, have been shown to aid speech perception in a 2 – way forced choice experiment. This effect has been replicated many times using similar 2 – way forced choice syllable identification tasks. As an extended version of these aero tactile studies, Derrick et al. (2016) designed an aero tactile continuous speech study using an open choice paradigm. In this study, they did not find any benefit of tactile stimuli on the identification of auditory signals in noise. As the 2009 and 2016 studies differ on numerous variables (syllables vs. sentences, forced choice vs. open choice), the present study was designed to investigate the effect of aero tactile stimuli on speech perception by only changing one variable from the original study, the response task. Thus, the focus of present study was to examine the effect of AT integration on speech perception using an open choice paradigm. In order to evaluate this effect, four monosyllables (/pa/, /ba/, /ka/ and /ga/) were presented in congruent and incongruent conditions at different SNRs, following an adaptive staircase to get 80% accuracy. It was hypothesized that SNR at 80% accuracy level would interact with phoneme and air flow in such a way that it would be decreased for congruent listening conditions and increased for incongruent listening conditions in AT stimuli compared to audio only stimuli.

The results of the current study do not confirm the hypotheses stated. The effect of air puff did not significantly influence SNR in congruent and incongruent listening conditions. Based on our descriptive statistics, the influence of place and manner of articulation on speech perception was also considered and added to the statistical model to analyse the relationship between these parameters in speech perception. The result shows a statistically significant effect of place, and an interaction effect of place and manner in speech perception.

As described above, results of the 7 non – native English participants in the study were also analysed independently from that of the native English participants. Some possible reasons for the difference in their perceptual results from those of native English speakers will be discussed briefly towards the end of this chapter.

This chapter will begin by discussing the prominent findings on the grounds of the effect of air puff, AT congruency, and the effect of the two articulatory features; place and manner of articulation in native English speakers.

4.1 Native English Speakers

4.1.1 Effect of Air puff

A strong multisensory association has been previously shown in bimodal speech perception studies between auditory and tactile stimuli in recognition of speech (Derrick & Gick, 2013; Derrick et al., 2014; Gick & Derrick, 2009). The primary finding of the present study showed that air puff does not have a significant effect on perception of speech. This is in contrast with the AT studies (discussed in section 1.3.3) carried out using forced choice response elicitation. When the data was visualised and analysed using descriptive statistics, the effect of puff was not consistent for syllables or at least for a particular class of syllables (either labials, velars, voiced or voiceless) during congruent and incongruent stimulus conditions and the differences in the SNR levels with puff presentation was negligibly low or minimal. Thus, the overall result is statistically insignificant. Next, using the GLMM model, the result, which appeared to be insignificant on initial investigation with descriptive statistical analysis, was further evaluated. A three – way interaction between the effect of artificial airflow and other parameters like place of articulation (labial vs velar), manner (voiced vs voiceless) was calculated using the GLMM model, where a back – fit using the Akaike information criterion (AIC) was used to obtain the best fit model. Interaction could not be obtained with the introduction of puff stimuli, confirming that no effect of air puff in the perception of these syllables. But when puff was eliminated, the other two parameters showed interaction, which is suggestive of the influence of place and manner in speech perception. Thus, air puff had to be eliminated from the best fit model used in the study with the conclusion that it may not be important in the perception of speech in any but the most constrained of conditions.

So, the findings of the present open choice study on AT integration, demonstrated no obvious interaction between auditory stimuli and air puff. Hence, we could not confirm the findings of the original AT paper (Gick & Derrick, 2009) that advocates AT integration. However, this result is in line with the finding of Derrick and colleagues (Derrick et al., 2016), which showed that the effect of air puff was insignificant in the perception of continuous speech. It was assumed that methodological modifications like response type and stimulus type might have influenced the findings and varying them might have led to responses suggestive of AT integration in participants. However, even with considerable methodological differences in both these open choice studies (e. g., stimuli used, congruency of stimuli, type of open choice task, and participant population) AT benefit could not be obtained in the current study. Since these methodological modifications did not confirm our hypotheses of aero tactile influence on auditory speech perception, one of the most obvious plausible causes of the null result could be because air puff may not be essential stimuli to perceive speech in a natural setting. Interestingly, Blamey and colleagues (1989) found that even though in tactile only condition, valuable information can be conveyed through electro – tactile stimulation during a vowel recognition task, significant effect of tactile information could not be seen when combined with other modalities (auditory and visual) during open – choice speech recognition tasks. The results of the present study are supportive for these findings by showing insignificant effect of air puff on perception of monosyllables during an open choice task. So, based on these findings it can be concluded that in an open choice response setting, where the listener has no or minimal access to cues regarding what they hear, the influence of tactile stimuli including electro tactile and air puff in recognition of speech would be negligibly low. This is suggestive of a weaker role of tactile modality during multisensory integration for understanding speech.

Thus, another possible cause of the absence of an AT effect in the current study could be the dominancy of the auditory modality over the tactile modality in understanding the speech signal. Blamey et al. (1989) evidenced this dominancy in his study with auditory, visual and tactile modality, where he found that performance was poorest for the tactile modality (electro – tactile stimuli) for all tasks expect vowel recognition. Even when he integrated this tactile information with auditory and visual information, he could not observe a significantly beneficial effect of tactile cues. Enhancement of speech recognition with the addition of aero tactile information to auditory signal could not be evidenced in the current study as well. Hence, the findings of the current study are in line with Blamey et al. (1989) `s observations that in speech recognition auditory modality is far more superior than tactile modality. Goldenberg et al. (2015) also confirmed this finding in his forced choice study. They found that the perceptual system weighs auditory and aero tactile inputs differently based on their observation that the tactile stimuli (aero tactile) exert greater influence in cases where auditory voicing cues are ambiguous. The possible justifications for this recessive nature of tactile information (whether electro – tactile or aero tactile) in speech recognition for participants could be because of the lack of experience or stimulation through the modality. The lack of experience or training in tactile modality for understanding speech would have led to reduced neural networks and connections to process this information quickly and effectively (Blamey et al., 1989). The listeners in the present study were also untrained and unexperienced in tactile modality.

Finally, a methodological drawback that was described earlier (Derrick et al., 2016), which is valid in this study as well is that the tactile stimuli (air puff) used in the present study had reduced airflow when compared to actual speech. The air flow system used for the present study is the same as Derrick et al. (2016) which uses same pressure (max 1.5kPa) as speech but only one twelfth of the air flow (0.8 l/m) normally produced in the speech (11.1 l/m). So, it could be a hypothetical assumption that the lack of adequate air flow might have reduced the impact of tactile stimuli, leading to a nullified effect of aero tactile stimuli in perception of speech in this study.

4.1.2 Effect of Congruency

English voiceless stops (such as /pa/) have puff of air called aspiration when articulated, hence air puff would help in distinction of voiced (/ba/) syllable from voiceless syllable (/pa/). Gick and Derrick (2009) found that voiceless judgements were made for stop consonants when they were presented with air puffs. The congruency of voicing feature and air puff was chosen for the study as their effect is perceptually evident as evidenced by Gick and Derrick (2009). Stimulus congruency was found to have certain influence in multisensory integration as discussed in section 1.2.2. Voiceless syllables (/pa/ and /ka/) did not show statistically significant difference in perception with the presentation of puff when compared to no puff condition. Similarly, significant congruency effect could not be obtained for voiced syllables (/ba/ and /ga/) in the absence of puff. Thus, congruency of stimulus did not demonstrate any improvement in the AT integration in this study. Since, the result being statistically insignificant, the effect of stimulus congruency could not be evidenced in multisensory integration. Thus, it could be concluded that integration of auditory and tactile information for speech perception is not disrupted by the mismatch of puff and voicing feature, even though it appears perceptually apparent. Green et al. (1991) also made a similar observation when he studied the magnitude of McGurk effect with gender incongruent (male voice dubbed for female faces) and gender congruent videos where he found that the magnitude of the McGurk effect was not statistically different for congruent conditions when compared to incongruent conditions. Based on the findings from these studies it can concluded that even when congruency of stimulus influence results, it may not always be an essential factor of multisensory integration in perception of speech.

4.1.3 Effect of Place and Manner of Articulation

Dependency of the place and manner of articulation on the recognition of speech has been demonstrated earlier in speech perception literature (Eimas et al., 1978; Miller & Eimas, 1977; Sawusch & Pisoni, 1974). In this AT study, designed to investigate the effect of air puff, SNR values were analysed to estimate how place and manner features contribute to easier understanding of the target stimuli. Findings of the study advocate for some influence of articulatory features in speech processing.

Interestingly, velar stops demonstrated to be easier to identify than labial stops in perception with low SNR values. Specifically, the velar voiced syllable (/ga/) was found to be the easiest to perceive among all the other syllable conditions. For labial stops, a significant difference between voiced and voiceless condition could not be obtained. Hence, dominance of voiced over voiceless was not found to be uniform and statistically significant across all the syllables studied. Furthermore, the results of the study also suggest interaction between place and manner of articulation.

Voicing and nasality features were found to be least affected by noise when compared to other articulatory features (Miller & Nicely, 1955). They found that when SNR drops below 6 dB, place feature becomes hard to discriminate. While voicing and nasality is distinguishable even at SNR as low as -12 dB. This could also be reason why voiced were slightly better for some syllable conditions.

As discussed in section 1.2.1, an acoustic cue for perception of voicing feature is VOT (Lisker & Abramson, 1970) while formant transitions and F1 frequency act as the acoustic cues for place of articulation (Liberman et al., 1967). Benkí (2001) studied English stops by varying their formant transition duration and F1 frequency. From his study, he concluded that bilabials and alveolars, with high F1 frequencies and short F1 transition, are more likely to be categorized as voiceless more frequently than velars. So, it could be a hypothetical assumption that the speaker of the current study might have had a similar formant transition duration and F1 frequency for bilabials as described by Benkí (2001). These acoustic characteristics would have resulted in perception of /ba/ (voiced bilabial) as voiceless. Thus, voiced was not significantly better than voiceless for bilabials. Since velars are least affected by these cues according to Benkí (2001), voiced was better for velar consonant than voiceless.

Another plausible reason for /ga/ perception in noise to be easier than /ka/ could be because of segmental bias described by Benkí (2001). He evidenced that bilabials and alveolars have shorter VOT boundaries than velars. Due to segmental bias, voiceless bilabial(/pa/) and alveolar(/ta/) is dominant over voiced bilabial(/ba/) and alveolar(/da/). On the other hand, velar voiced (/ga/) dominant over velar voiceless (/ka/).

Based on above discussions, it is evident that place is more influenced by manner feature (Eimas et al., 1978) which confirms the mutual dependency of these articulatory features. Moreover, both place and manner feature have acoustic cues delivered through same modality, hence undergo similar type of processing when though they are defined by different parameters (Miller & Eimas, 1977). Thus, the findings of the present study are suggestive of dependency between place and manner of articulation in perception of speech which is in agreement with earlier literature.

4.2 Non – native English Speakers

In the present study, 7 participants did not fit our native language criteria and were hence eliminated from the main study results. The results for non – native English speakers are harder to understand than those for native English speakers. Processing differences for non – native English speakers are strongly influenced by the phonological systems of their native languages (Abramson & Lisker, 1970).

When their data was evaluated, it could be grouped into two categories. Group 1, consisting of a Mandarin, a Chinese (likely Mandarin), a Cantonese and a Korean speaker, replicated responses similar to that of the native English speakers. Conversely, the Telugu, Tagalog and Nepali speakers (one each) were part of group 2, all of whom had considerable difficulty with the task.

4.2.1 Potential Cause of Differences between Group 1 and 2

For Mandarin, the distinction between voiced and voiceless stops, including the /p/ vs. /b/ (phonetically [p^h] vs [p]) and /k/ vs /g/ (phonetically [k^h] vs [k]) is structurally identical to that in English. This holds true for Cantonese as well.(Li & Thompson, 1989) In Korean, there is a three – way distinction between plain, tense, and aspirated. However, the differences between plain and tense have little impact on either voicing or air flow and are instead related to the size of the pharyngeal cavity (Brown & Yeon, 2015; Shin, 2015). So Korean also has an English – like distinction between voiced and voiceless stops. The similarities between English, Cantonese, Mandarin, and Korean on this form of stop distinction could make it easier for listeners to distinguish them in English. This ease allows the results for these non – native speakers to be grouped together as group 1.

In contrast, in Telugu, a language of South – central East India, there is a four – way distinction between stop consonants, phonetically [p], [b], [p^h], and [b^h] (Krishnamurti, 1998). These contrasts contain the English version of the voiced/voiceless distinction, which is actually an aspirated/unaspirated distinction. However, they also contain a literal voiced/voiceless distinction, with pre – voicing before the stop burst contrasted with no pre – voicing before the stop burst. This means that the listener can expect four distinctions, only two of which are addressed by air flow changes. As a result, the task remains confusing, and not enough information is provided to resolve that confusion. The same four – way distinction also exists in Nepali (Khatiwada, 2009). For Tagalog, they don't use aspiration as a mechanism for distinguishing stops at all. They just use a literal (pre) – voiced vs voiceless distinction. Therefore, nothing in their native language provides them either the auditory or air flow cues needed for this experiment (Guevarra, 2015; Llamzon, 1966).

The results did not show universal confusion, but rather a pattern of task confusion for some tokens, and not for others. This is likely due to the fact that they have learned English as a second language, and the patterns of language learning for each speaker are unknown to us.

4.3 Limitations and Future Directions

Though the findings of the study were interesting, it also had some limitations too. One of the potential drawbacks was that the task was not functional for the participants. They did not have to use any cognitive or linguistic skills to perform the task and after a couple of trials participants could possibly guess that they had to choose between four syllables (constricted open choice task). Since the task was not cognitively challenging, participants might have ignored the tactile cues. Moreover, the air flow system used for the present study had same pressure (max 1.5kPa) as speech but only one twelfth of the air flow (0.8 l/m) normally produced in the speech (11.1 l/m) (Derrick et al., 2016). Finally, gender representation and population of native English speakers were not uniform, which was also a weakness of the study.

The current study is one of the studies in the series of AT studies to investigate the AT effect in open choice task which could not be replicated in the continuous speech study (Derrick et al., 2016). Since the findings of this study also does not advocate for AT integration as in continuous paper, other parameters like consonants types chosen, usage of adaptive staircases that were different in present study from that of the original AT paper, might have influenced the results of the study. Thus, the future research directions need to be focused on investigating the relationship between these aspects and AT integration. Moreover, a better version of airflow system that can replicate the air flow equivalent to actual speech should be devised in the future aero tactile studies. In addition, measuring the effect of aero tactile stimuli in speech recognition using open choice design after training participants, to feel speech through tactile modality, could also be explored in future research. As the possibility that participants might have ignored the tactile cues considering it to be less important to understand to speech still remains, future research could also aim at evaluating effect of aero tactile information in infants and pre – linguistic children to examine if AT integration is an innate ability like AV integration.

4.4 Implications

Despite these limitations, there are some implications for this study. Based on the observations from the current study it could be understood that inefficiency of tactile cues (air puffs) in understanding speech is due to the lack of appropriate and adequate training. When an individual is trained to perceive speech by stimulating tactile modality, neural networking necessary to process tactile cues for speech occurs at brain level. As a result of this, easier and quick processing of tactile cues for speech like auditory or visual cues could be made possible. Thus, this study contributes to the fundamental scientific knowledge on the ways humans perceive multi – modal speech. Importantly, it adds to the line of multisensory literature investigating influence of puff in speech perception. Importantly, rehabilitation of individuals with sensory deficits specifically, individuals with visual and hearing impairment should be focused on using total communication approach, where all possible means of communication like cued speech, Tadoma method, tactile sign language, Braille, Alternative and Augmentative Communication (AAC) etc. are used (Chomsky, 1986; Vernon & Andrews, 1990). Though the insights from the present study does not support beneficial effects of air puffs in everyday speech perception, previous aero tactile literatures have proven aero tactile information to be useful when choices are provided for participants (Derrick et al., 2014; Gick & Derrick, 2009). So, rehabilitation professionals in the field of communication disorders could use forced choice tasks to train tactile speech perception for individuals with sensory deficits rather than open choice tasks. Hence this study is beneficial for evidence – based practice as it confirms Derrick et al. (2016)`s findings that tactile information may not be beneficial as auditory in a natural communication environment.

4.5 Conclusion

The findings of the current study could not prove the beneficial influence of air puff information for syllable perception when examined using a constricted open choice design in native English speakers. In addition to this, mismatch between aspiration (an aspect of voicing) and puff was found to have insignificant effect on AT integration for speech perception, even though they are perceptually evident. However other articulatory features like place and manner of articulation was found to contribute significantly on perception of speech. Interestingly, voiced velar (/ga/) was observed to be easier to perceive for majority of the participants in the study. Though the insignificant effect of AT integration in the current study confirms Derrick et al. (2016)`s findings on continuous speech with open choice design, it is not in line with the array of aero tactile literature that advocated AT integration using forced choice design. So, other parameters that were different in these studies needs to be examined in future to rule out why aero tactile information which was proven to be useful, could not be a beneficial cue in open choice tasks. And these aero tactile research should be done with a better air flow system that can replicate air flow equivalent to speech. Moreover, this study improved our understanding of how individuals perceive speech using different sensory systems. And from clinical perspective, it provided insights to rehabilitation professionals, working with individuals having multiple sensory impairments, to be careful when using and recommending tactile stimuli with open choice tasks. It also suggested the need of using total communication, including training in tactile modality with forced choice design, to improve communication skills for people with visual and hearing impairments.

REFERENCES

Abramson, A. S., & Lisker, L. (1970). *Discriminability along the voicing continuum: Cross-language tests.* Paper presented at the Proceedings of the sixth international congress of phonetic sciences.

Alcorn, S. (1932). The Tadoma method. *Volta Review, 34,* 195-198.

Benkí, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, *29*(1), 1-22.

Bicevskis, K., Derrick, D., & Gick, B. (2016). Visual-tactile integration in speech perception: Evidence for modality neutral speech primitives. *The Journal of the Acoustical Society of America*, *140*(5), 3531-3539.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech, Language, and Hearing Research, 17*(4), 619-630.

Blamey, P. J., Cowan, R. S., Alcantara, J. I., Whitford, L. A., & Clark, G. M. (1989). Speech perception using combinations of auditory, visual, and tactile information. *Scientific publications, vol. 5, 1989-1990, no. 268*.

Brown, L., & Yeon, J. (2015). *The handbook of Korean linguistics*: John Wiley & Sons.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental psychobiology, 45*(4), 204-220.

Cassels, T., & Birch, S. (2014). Comparisons of an Open-Ended vs. Forced-Choice ‘Mind Reading’Task: Implications for Measuring Perspective-Taking and Emotion Recognition. *PLoS ONE, 9*(12), e93653.

Chomsky, C. (1986). Analytic study of the Tadoma method: Language abilities of three deaf-blind subjects. *Journal of Speech, Language, and Hearing Research, 29*(3), 332-347.

Colin, C., Radeau, M., & Deltenre, P. (2005). Top-down and bottom-up modulation of audiovisual integration in speech. *European Journal of Cognitive Psychology, 17*(4), 541-560.

Derrick, D., Anderson, P., Gick, B., & Green, S. (2009). Characteristics of air puffs produced in English “pa”: Experiments and simulations. *The Journal of the Acoustical Society of America, 125*(4), 2272-2281.

Derrick, D., & Gick, B. (2009b). Aero-tactile integration in speech perception (supplementary methods). *Nature, 463*(7272).

Derrick, D., & Gick, B. (2013). Aerotactile integration from distal skin stimuli. *Multisensory research, 26*(5), 405-416.

Derrick, D., Heyne, M., O’Beirne, G. A., Rybel, T. d., Hay, J. & Fiasson, R. (2016). Effects of speech air flow on 2AFC word identification in Mandarin and English

Derrick, D., O'beirne, G. A., De Rybel, T., Hay, J., & Fiasson, R. (2016). Effects of aero-tactile stimuli on continuous speech perception. *The Journal of the Acoustical Society of America, 140*(4), 3225-3225.

Derrick, D., O'Beirne, G. A., Rybel, T. d., & Hay, J. (2014). *Aero-tactile integration in fricatives: Converting audio to air flow information for speech perception enhancement.* Paper presented at the Fifteenth Annual Conference of the International Speech Communication Association.

Derrick, D. J., De Rybel, T., & Fiasson, R. (2015). Recording and reproducing speech airflow outside the mouth. *Canadian Acoustics, 43*(3).

Eimas, P. D., Tartter, V. C., Miller, J. L., & Keuthen, N. J. (1978). Asymmetric dependencies in processing phonetic features. *Attention, Perception, & Psychophysics, 23*(1), 12-20.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: cross-modal contributions to speech perception. *Journal of experimental psychology: Human perception and performance, 17*(3), 816.

Ghazanfar, A. A., & Logothetis, N. K. (2003). Neuroperception: Facial expressions linked to monkey calls. *Nature, 423*(6943), 937-938.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 462*(7272), 502-504.

Gick, B., Ikegami, Y., & Derrick, D. (2010). The temporal window of audio-tactile integration in speech perception. *The Journal of the Acoustical Society of America, 128*(5), EL342-EL346.

Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., & Mühlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *The Journal of the Acoustical Society of America, 123*(4), EL72-EL76.

Goldenberg, D., Tiede, M. K., & Whalen, D. (2015). Aero-tactile on speech perception of voicing continua.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America, 103*(5), 2677-2690.

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of. *Attention, Perception, & Psychophysics, 45*(1), 34-42.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50*(6), 524-536.

Guevarra, K. (2015). Phonological Analysis of Tagalog.

Hardison, D. M. (1996). Bimodal speech perception by native and nonnative speakers of English: factors influencing the McGurk effect. *Language Learning, 46*(1), 3-73.

Howard, I. P., & Templeton, W. B. (1966). Human spatial orientation.

Khatiwada, R. (2009). Nepali. *Journal of the International Phonetic Association, 39*(3), 373-380.

Krishnamurti, B. (1998). Telugu. *The Dravidian languages*, 202-240.

Kuhl, P. K., & Meltzoff, A. N. (1982). *The bimodal perception of speech in infancy*.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*: Univ of California Press.

Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America, 29*(1), 117-123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review, 74*(6), 431.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

Lisker, L., & Abramson, A. S. (1970). *The voicing dimension: Some experiments in comparative phonetics.* Paper presented at the Proceedings of the 6th international congress of phonetic sciences.

Liu, C., & Kewley-Port, D. (2004). Vowel formant discrimination for high-fidelity speech. *The Journal of the Acoustical Society of America*, *116*(2), 1224-1233.

Llamzon, T. A. (1966). Tagalog phonology. *Anthropological Linguistics*, 30-39.

Macleod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British journal of audiology, 24*(1), 29-43.

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic bulletin & review, 22*(5), 1299-1307.

Massaro, D. W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale: NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle* (Vol. 1): Mit Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338-352.

Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *The Journal of the Acoustical Society of America, 61*(3), 835-845.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist, 17*(11), 776.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods, 162*(1-2), 8-13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics, 2,* 10.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences, 106*(26), 10598-10602.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., & Braida, L. D. (1983). Research on the tadoma method of speech communication. *The Journal of the Acoustical Society of America, 73*(S1), S26-S26.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., & Schultz, M. C. (1985). Research on the Tadoma method of speech communication. *The Journal of the Acoustical Society of America, 77*(1), 247-257.

Reed, C. M., Rubin, S. I., Braida, L. D., & Durlach, N. I. (1978). Analytic study of the Tadoma method: Discrimination ability of untrained observers. *Journal of speech and hearing research, 21*(4), 625-637.

Risberg, A., & Lubker, J. (1978). Prosody and speechreading. *Speech Transmission Laboratory Quarterly Progress Report and Status Report, 4*, 1-16.

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics, 2*, 181-194.

Sekiyama, K., & Tohkura, Y. i. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America, 90*(4), 1797-1805.

Shin, J. (2015). Vowels and consonants. *The handbook of Korean linguistics*, 1-21.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*: The MIT Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Thornton, N. E., & Erber, N. (1979). Auditory-visual speech perception by hearing impaired children. *Hearing Aid Journal, 32*(6), 32.

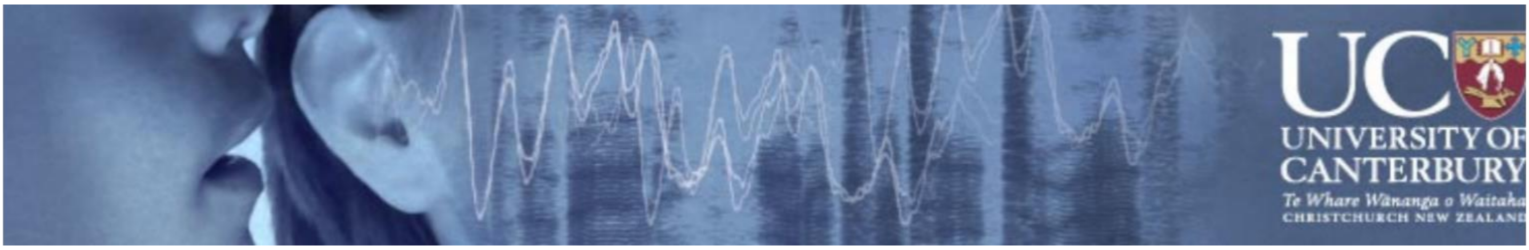
Treille, A., Cordeboeuf, C., Vilain, C., & Sato, M. (2014). Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia, 57*, 71-77.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181-1186.

Vernon, M., & Andrews, J. F. (1990). *The psychology of deafness: Understanding deaf and hard-of-hearing people*: Longman Publishing Group.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Attention, Perception, & Psychophysics*, *33*(2), 113-120.

APPENDIX A: INFORMATION SHEET



Speech perception experiment

INFORMATION FORM

Dear,

You are invited to participate in an experiment on speech perception.

Your involvement in this project will require you to come to the lab at the Child Language Centre of the University of Canterbury (7 Creyke Road, Christchurch) for up to 1 hour.

We will ask some background questions and test your hearing levels. This is to make sure that the results we obtain are not confounded by any potential hearing, vision, speech and language or health issue.

You will be asked to sit in a sound-attenuated booth while wearing headphones. You will then be presented with different speech sounds and asked to indicate what you heard.

You will receive a $10 gift voucher as compensation for your time. Your participation in this study is entirely voluntary. Up until publication of the results you have the right to withdraw from the project at any time, including withdrawal of any information provided without penalty. Please note that this project forms part of master's thesis, which will be publically available via the University of Canterbury database. The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: the identity of participants will not be made public without their consent.

The project is being carried out by Zoë Haws (Master of Audiology) and Jilcy Madappallimattam (Master of Science Speech and Language, Department of Communication Disorders) and supervised by Dr. Catherine Theys. Catherine can be contacted at 03 369 4516 ext. 94516 or via Catherine.Theys@canterbury.ac.nz. She will be pleased to discuss any questions or concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (humanethics@canterbury.ac.nz).

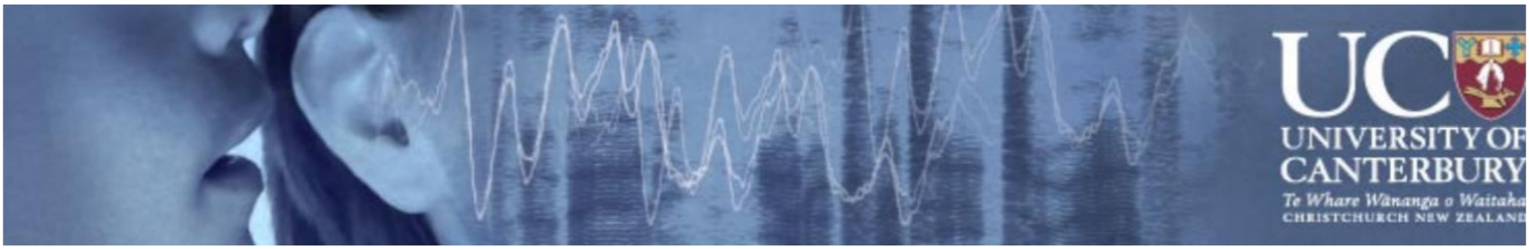
Thank you very much for you participation,

Zoë Haws
Master of Audiology Student
zoe.haws@pg.canterbury.ac.nz

Jilcy Madappallimattam
Master of Science Speech and Language Student
jilcy.madappallimattam@pg.canterbury.ac.nz

Dr. C. Theys
Lecturer, Department of Communication Disorders & New Zealand Institute of Language, Brain & Behaviour
Private Bag 4800 – Christchurch 8140
Catherine.Theys@canterbury.ac.nz - Phone: 03 369 4516 ext. 94516

APPENDIX B: CONSENT FORM



Dr. C. Theys – Dr. D. Derrick -Prof. M. McAuliffe
Project Leader: Dr Catherine Theys Catherine.Theys@canterbury.ac.nz
Phone: 03 369 4516 ext. 94516
Department of Communication Disorders
University of Canterbury

Speech perception experiment

CONSENT FORM

I have read and understood the description of the above-named project. On this basis I agree to participate as a participant in the project, and I consent to publication of the results of the project with the understanding that confidentiality will be preserved.

I understand also that I may at any time withdraw from the project, including withdrawal of any information I have provided. Withdrawal is possible up until publication of the study results.

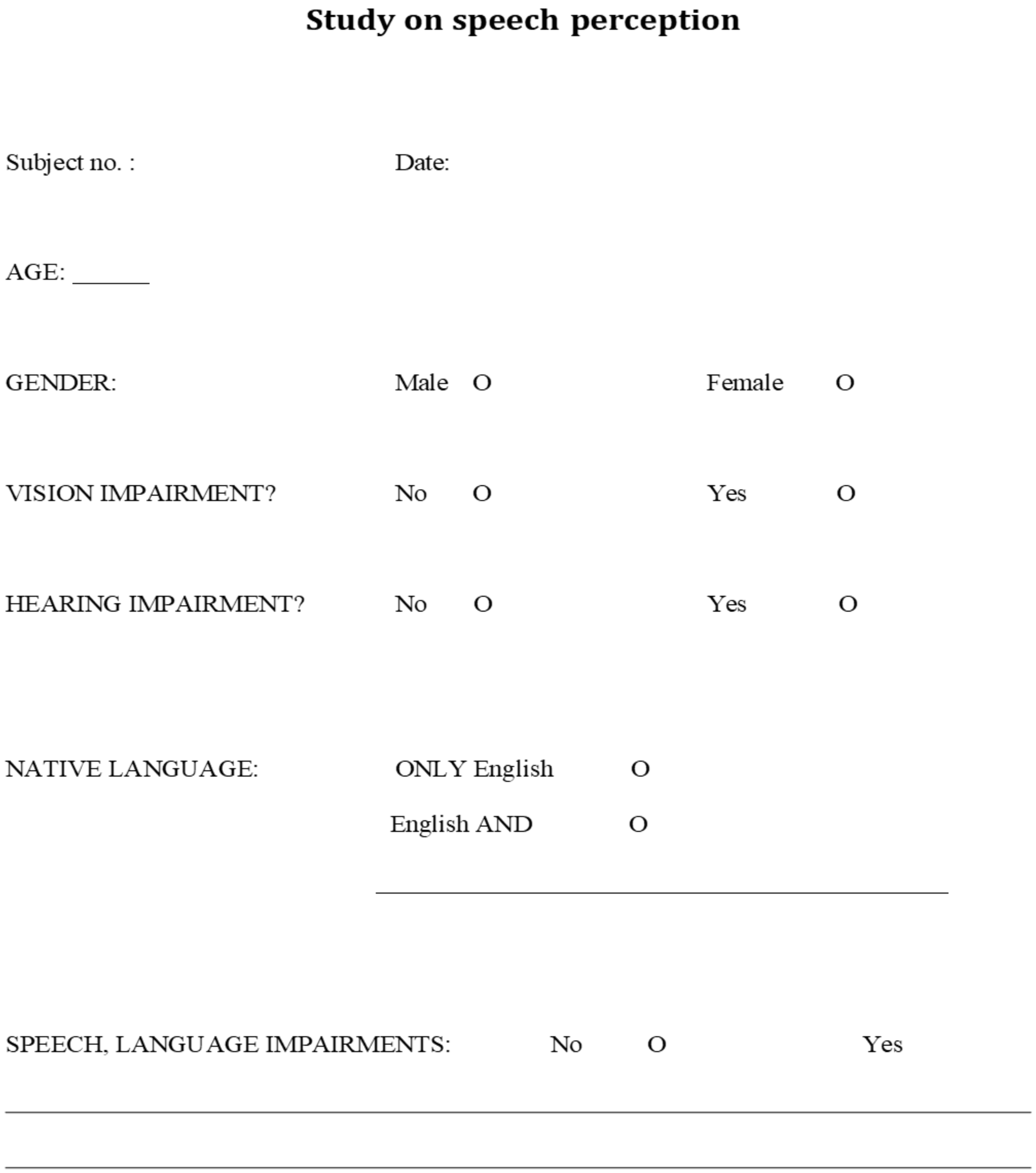
I note that the project has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Name (please print): …………………………………………………………..

I would like to receive a summary of the results upon completion of the study: yes / no

Signature:

APPENDIX C: DEMOGRAPHIC QUESTIONNAIRE



Study on speech perception

Subject no. : Date:

AGE: ______

GENDER: Male O Female O

VISION IMPAIRMENT? No O Yes O

HEARING IMPAIRMENT? No O Yes O

NATIVE LANGUAGE: ONLY English O

English AND O

____

SPEECH, LANGUAGE IMPAIRMENTS: No O Yes

____

____