TONES AND VOWELS IN CANTONESE INFANT DIRECTED SPEECH – HYPERARTICULATION DURING THE FIRST 12 MONTHS OF INFANCY

Nan Xu

BA (Psychology) Honour Class 1, University of Western Sydney

BA Applied Science, University of Sydney

A thesis submitted for the degree of Doctor of Philosophy

MARCS Auditory Laboratories

University of Western Sydney

29th August 2008

I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by any other person, nor material which has been accepted for the award of any other degree or diploma at the University of Western Sydney, or at any other educational institution, except where due acknowledgement is made in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception is acknowledged.

<div style="text-align: right;">_____</div>

Acknowledgements

Firstly, I would like to acknowledge my primary supervisor Prof. Denis Burnham not only for his guidance, support, and enthusiasm, but also for always having faith in my abilities and convincing me to take on the challenge of a PhD. It was Denis who first roused my interest in infant and language research when I was an undergraduate student completely naïve to the world of research. I want to thank you Denis for the time you always manage to find and the seamless effort in guiding and reviewing my work, and above all I want to thank you for your dependability and calmness which has given me confidence and motivation throughout my candidature.

I acknowledge my co-supervisors Dr Christine Kitamura and Professor Cathi Best for our many conversations together and their wisdom on infant-directed speech and speech perception, and I thank them for the invaluable advice they had provided especially during the development of my experiments.

I want to acknowledge and thank most warm-heartedly the Chinese welfare workers Quinnie at Cabramatta Anglicare and Sue Loh at Parramatta Early Childhood Centre for their help with the recruitment of Cantonese mothers. Without their kindness and help this PhD would never have taken place. To all my Cantonese mothers thank you for your participation, time and patience.

I want to thank all of the MARCS support staff: Colin Schoknecht, our lab manager, not only for his organisational astuteness but also for his perceptiveness in anticipating unforeseeable technical and non-technical issues; Gail Charlton for her kindness and willingness to help, and for organising Denis; Arman Abrahamyan and Johnson Chen for their technical assistance and Johnson especially for installing, removing and installing many programs on many laptops whenever I needed them; Mel Gallagher, Darlene Williams, and Karen McConachie for their behind the scenes contributions that makes the turning of the clockwork at MARCS appear seamless.

A big thank you to all my proof readers – Mark Antoniou (for putting up with the longest and most tedious chapter), Christa Lam, Damien Smith, Benjawan Kasisopa, Tim Byron and Rikke Bundgaard-Nielsen – whose handprints together with Denis's are all over these pages. I want to thank especially Benjawan, Rikke, and Tim for setting aside their own work in the final moments of my PhD to honour my timeline and I am more appreciative then I can express in words for their kindness.

I thank all of my friends both at and outside of the MARCS lunch table, firstly for being patient with my PhD, secondly for being patient with my PhD-self during my PhD, also for make my life interesting and relevant, and finally for challenging and supporting me intellectually, mentally, emotionally and physically.

To Lidija, I want to thank you for your friendship and for your support throughout our PhD together. It has been a great privilege and a great honour to know you as a person, and to have shared this PhD experience with you.

To my Mama (妈妈) and Baba (爸爸), whose support has enabled me to pursue a PhD in relative ease and comfort. I am so grateful to you both for allowing me to follow my heart as well as my head. To my Sam, always a loving, giving, and forbearing partner, I thank you for your support and dependability, you are truly my solace.

For everyone who has, in passing, helped me in making this PhD happen, I thank you.

Table of Contents

List of Tables

List of Figures

Appendices

One

Chapter 1

Speech Production – Segmental & Suprasegmental Speech

1.1   How is the Speech Signal Produced? Articulatory Phonetics

One way of looking at the human speech production system is to divide it into three units: (1) the pressure system often referred to as the respiratory system (2) the voicing system of the larynx and its surrounding structures, and (3) the articulators and resonating cavities above the larynx. This structural division illustrates the functional difference between the subglottal system that provides an energy source for speech production, and the articulators above the larynx that determine the phonetic quality of speech sounds.

1.1.1   *The Subglottal System*

Speech sounds are made by constricting, partially constricting, or not constricting the passage of air from the lungs while it is passing through the various sub-laryngeal structures.

The production of speech sounds requires an energy source – flow of air from the lungs. Immediately below the glottis (a valve-like structure consisting of the vocal folds and cartilages of the laryngeal system) is an airway consisting of the trachea (a single tube), which is divided into two bronchi that are further divided into smaller bronchioles terminating in alveolar sacs which nourish the lungs with oxygen from air moving in through the upper respiratory tract (see figure 1). The volume of air inside the lungs changes during inspiration and expiration. During expiration, lung pressure rises beyond its resting volume and the elastic coils of the lungs, together with the intercostal (or abdominal) muscles act in concert pushing back on the expanded lung space (Clark & Yallop, 1995). Expiration for humans has the subsidiary, but socially important, function of assisting in the production of speech. Air flow from expiration is constricted, partially constricted, or not constricted at the glottis and manipulated by the articulators situated higher in the laryngeal system and the resonators (in the oral or nasal cavities) to produce speech.
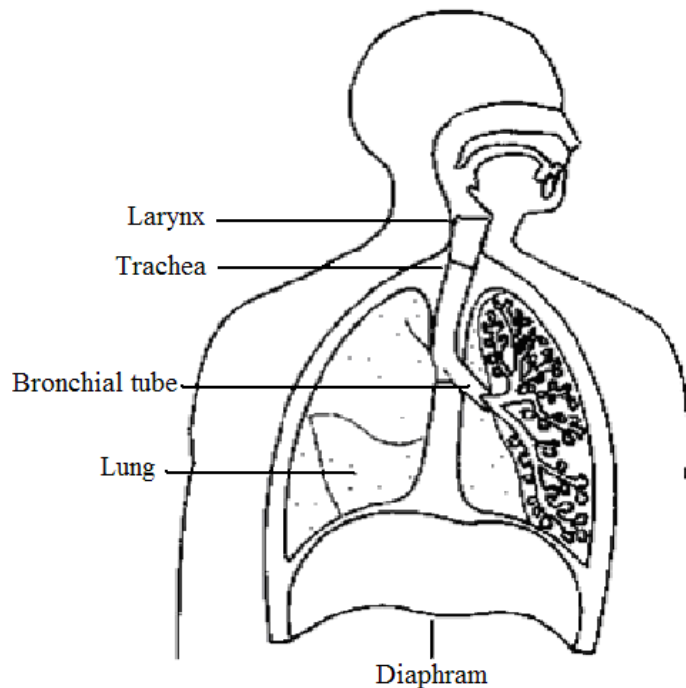
Figure 1. Structures of the Respiratory System (Clark & Yallop, 1995)

During production of an utterance consisting of several words, the pressure in the lungs is maintained by the respiratory musculature at a value that is proportional to the overall level of vocal effort, which rises and falls with different types of sounds (Stevens, 1999).   Lung volume can contribute to such speech characteristics as the duration of speech sounds (perceived as vowel length), amplitude (perceived as loudness), and changes in the fundamental frequency (which is perceived as pitch).

1.1.2   *The Larynx and Surrounding Structures*

Higher up the trachea are the principal structures responsible for speech production collectively called the larynx.

1.1.2.1   Laryngeal Cartilages and Muscles

The larynx consists of four major cartilages (see Figure 2).   Immediately above the trachea, the ring-shaped cricoid cartilage is connected at the superior end to the wing-shaped thyroid cartilage, and the posterior end (towards the back of the neck) to the tetrahedron-shaped arytenoid cartilages.
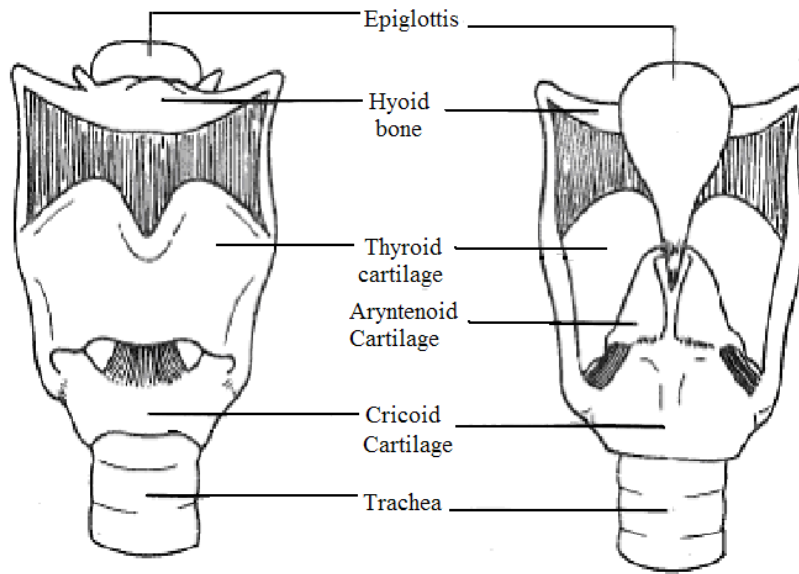
Figure 2. Laryngeal Structures (Clark & Yallop, 1995)

The vocal folds, commonly called the vocal cords, are two ligaments that are connected to the thyroid and arytenoid cartilages. See Figure 3 for an illustration of the vocal folds and surrounding structures. The cartilages and muscles of the larynx work in concert to manipulate the vocal folds so that they move together (adduct) or apart (abduct). This difference in the state of the vocal folds allows for the manipulation of air expired from the lungs and the articulators higher up in the oral cavity to produce actions that are related to perceptual distinctions among individual speech sounds, as well as syllable, word, and sentence level linguistic meaning to be made.
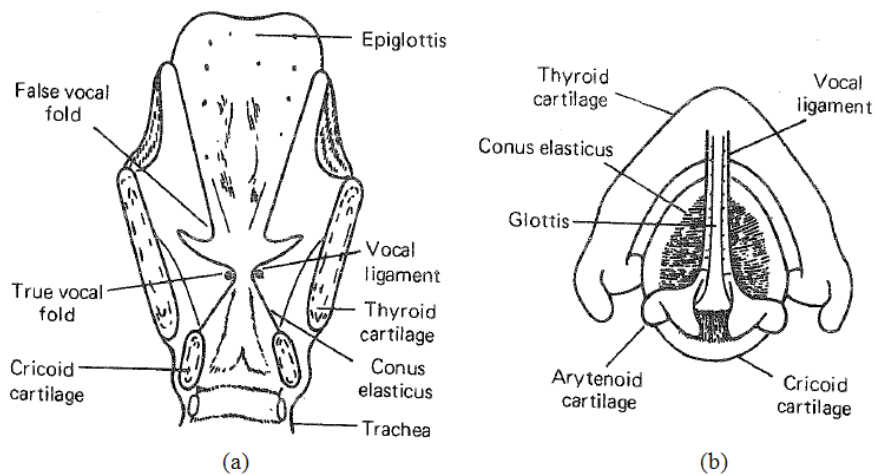
Figure 3. Vocal fold structure: (a) anterior view; (b) coronal section (Clark & Yallop, 1995)

Adduction of the vocal folds involves contraction of the posterior cricoarytenoid muscles – a pair of muscles that connect the cricoid cartilage to the arytenoid cartilages – which causes the arytenoid cartilages to tilt back lengthening and increasing the stiffness of the vocal folds. The interarytenoid muscle connecting the two arytenoid cartilages, with the assistance of the lateral cricoarytenoid muscles that connect the arytenoid cartilages and the cricoid cartilage brings the arytenoid cartilages together and causes the adduction of the posterior end of the vocal folds. In addition, the cricothyroid muscles that connect the cricoid and thyroid cartilages can rock the thyroid cartilage back and forth thus stretching the vocal folds. The thyroarytenoid muscle is connected to the entire length of the vocal folds and with contraction serves to shorten and thicken the vocal folds as well as adducting them. These muscles are implicated in pitch modifications that are important in intonational and tonal variation, which will be described further in Section 1.4.1.

1.1.2.2 *Phonation*

The larynx and surrounding structures are responsible for phonation, puffs of air released periodically during voicing via movements of the muscles surrounding the vocal folds. All speech sounds are produced in this way and are further modified by the resonators above the larynx (the oral and nasal cavities). Maintenance of vocal fold vibration, which is perceived as pitch during speech and singing, is dependent on the independent movement of the different layers of the vocal folds (superior and inferior levels).

Speech sounds also differ on voicing, with some sounds voiced and others voiceless. In many languages, this difference contributes to phonemic (lexical-relevant) distinctions. When the vocal folds are open, voicing always occurs. However, when the vocal folds are closed, voiced sounds like [z] can be produced, as well as voiceless sounds such as [s] produced with air coming from the subglottal system but not vibrating the vocal folds.

1.1.3 *The Articulators above the Larynx*

The articulators above the larynx include the pharynx and the structures inside the oral cavity, as well as the nasal cavity (see Figure 4). These are responsible for articulating specific speech sounds.
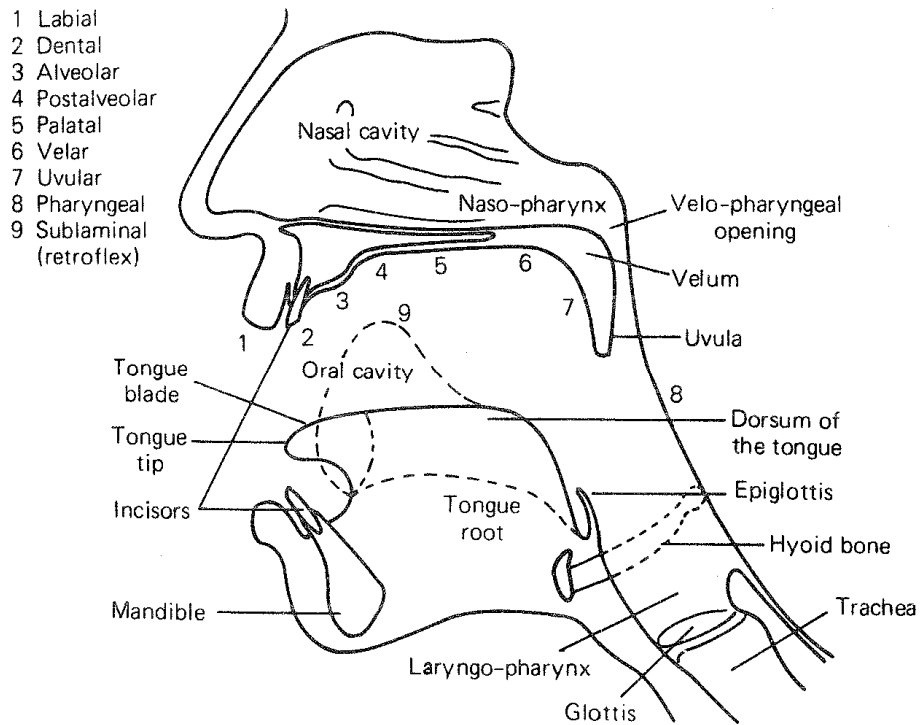
Figure 4. Articulators above the Larynx (Clark & Yallop, 1995)

1.1.3.1 The Pharynx

The pharynx is located just in front of the spinal cord extending from the laryngeal region up to the hyoid bone, then to the soft palate (back roof of the mouth), and finally to the beginning of the nasal passages. The pharynx is surrounded by a set of constrictor muscles and the genioglossus muscle, which contract or relax to narrow or widen the pharyngeal airways respectively.

The pharynx functions as a food passage, an air passage, and a passage for the drainage of nasal cavities. As an air passage, it also assists in speech production during expiration.

1.1.3.2 The Soft Palate (Velum)

The soft palate or velum can be thought of as a continuation of the roof of the mouth and extension from the hard palate, and is covered by a muscular tissue ending at the

uvula (a small flexible tissue hanging from the velum). The velum is a flap of tissue about 4cm long, 2cm wide, and 0.5cm thick, and can be raised or lowered by a set of muscles (Stevens, 1999).

When the velum is raised it forms the floor of the nasal cavity, and when lowered it allows for airflow between the vocal tract and the nasal cavity. The velum is lowered to allow this opening when a speaker is producing sounds that require nasal cavity participation (e.g., [n]). The uvula, in conjunction with the body of the tongue, can also function as an articulator.

1.1.3.3 The Nasal Cavity

The nasal cavity (typically 10cm long from the pharynx to the nostrils) is divided into two passages separated by the septum, which begins as a bony tissue connected to the skull and ends as a cartilaginous tissue at the nostrils (Clark & Yallop, 1995). Along the divided sections of the passage are undulations called conchae that increase the area of resonating space within the nasal cavity several times over a plain circular passage.

The nasal cavity is predominantly responsible for warming and humidifying air during inspiration. As it lacks muscular structure, it cannot be voluntarily manipulated, and acts as a passive resonating chamber in speech production. The quality of nasalised sounds can be somewhat altered by position and movement of the velum, and involuntarily by changes in the volume of the mucus inside the nasal cavity from externally caused events such as during influenza which causes inflammation of the sinuses.

Nasal sounds in English and Cantonese include [m], [n], and [ŋ] as in the final sound in si**ng**. In English as in Cantonese the nasal [ŋ] cannot occur word initially.

1.1.3.4 The Oral Cavity

The oral cavity is the most important part of the vocal tract because of the relative freedom of movement of its parts to change the shape of this resonating chamber during articulation. It consists of the hard palate (roof of the mouth); alveolar ridge from which the upper teeth project; the mandible, from which the lower teeth project, is

23

connected to the skull via the temporomandibular joint, which allows for vertical and some horizontal movement that change the position and height of the tongue and teeth.

Tongue movement is controlled by both external and internal muscles, and it is the internal muscles that allow the tongue to move independently. The tongue deserves special attention as it can be moved to make contact with other parts of the oral cavity to articulate different speech sounds, and is important for both consonant and vowel production.

The lips are the anterior opening of the vocal tract, and can move both with the mandible and independent of the mandible to change lip shape and mouth aperture, both important for articulating consonants and vowels.

1.2 Segmental Speech Sounds: Organisation of sounds across languages

Speech segments are often defined as discretely produced units of sound that contribute directly to lexical meaning, and include both consonants and vowels. In order to appreciate the repertoires of speech sounds in different languages, it is important to understand the meaning of the terms phones, phonemes, and allophones.

Phones are the most basic units of speech sounds that can be produced and perceived as distinctly different from other sound units across languages, where as phonemes are the smallest units of sound that distinguish lexical meaning of words in a given language. For example, the English words rock and lock are differentiated only by the initial phonemes /r/ and /l/. However, some languages (e.g., Japanese) do not make this phonemic distinction, and Japanese learners of English have difficulty discriminating /r/ and /l/ as separate phonemes (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Yamada, & Tohkura, 1997). Allophones are systematic phonetic variants of a phoneme that are realised in specific contexts. The allophones of the English phoneme /p/ in the words pat [pʰ] and spat [p] are produced differently and so are different phones, but in English they are not perceived differently where as in Thai the phones [pʰ] and [p] are different phonemes. On the other hand, the

24

phonetic variations of the voiceless unaspirated [p] and the voiced [b] within the English phoneme /b/, are separate phonemes, /p/ and /b/, in Thai[1].

While it is convenient to describe spoken language as individual consonants and vowels, phones are rarely uttered in isolation; they generally occur in combinations called syllables. While syllables usually contain consonant(s) and vowel(s), a few contain only vowels ("owe"), and there are even some cases of single consonant syllables such as the nasal [ŋ] in Cantonese. In general, there is usually consensus across speakers on the number of syllables a word contains, but there are instances of dissension when dialect and personal factors intrude. For example, the word predatory can be produced with either three or four syllables. There is also great debate about a specific definition of syllable, whether it is based on articulatory information acoustic information or higher mental organisation of perceived units of speech. For a discussion of this topic see (Ladefoged, 2001a).

Armed with this knowledge we now turn to a description of the articulation of consonants and vowels with especial reference to Cantonese where appropriate.

1.2.1 *Consonants*

Consonants can be classified by their place and manner of articulation. Each language has its own set of consonants, and these for English and Cantonese are shown in Figures 5 and 6 respectively for comparison purposes. In the following subsections place and manner of articulation are explained with appropriate examples.

---

[1] Phones are written within square brackets e.g., [pʰ] whereas phonemes are written between forward slashes /p/.

Consonant phonemes of Australian English

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Nasal	m			n			ŋ	
Plosive	p　b			t　d			k　ɡ	
Affricate					tʃ　dʒ			
Fricative		f　v	θ　ð	s　z	ʃ　ʒ			h
Approximant				ɹ		j	w	
Lateral				l				

Figure 5. Australian English Consonants Chart by Place and Type of Articulation (Cox & Palethorpe, 2007)

Consonant phonemes of Chinese Cantonese

	Labial	Coronal		Palatal	Velar		Glottal
		plain	sibilant		plain	labialized	
Nasal	m	n			ŋ		
Stop — plain	p	t	ts		k	(kʷ)	(ʔ)
Stop — aspirated	pʰ	tʰ	tsʰ		kʰ	(kʷʰ)	
Fricative	f		s				h
Approximant		l		(j)		(w)	

Figure 6. Cantonese Consonants Chart by Place and Type of Articulation

1.2.1.1   Place of Articulation

Places of articulation are points in the oral cavity at which the airflow is stopped or otherwise impeded to produce particular consonant sounds (for locations of these articulators in the oral cavity see Figure 4).

Bilabial sounds are made with participation of both the lower and upper lips as in the sounds [p], [b], and [m]. Labiodental sounds are made with the lower lips touching the upper teeth, as in the consonants [f] and [v]. Dental sounds are made with the tongue touching the upper teeth, e.g., [ð] as in the initial sound in the, which is also articulated by some speakers as an interdental – with the tongue protruding between the teeth. Such sounds do not exist in Cantonese but dental sounds in Cantonese include [z] and [s].

26

Alveolar sounds like [l] are made with the tongue touching the alveolar ridge. Retroflex sounds are made with the tongue tip touching the back of the alveolar ridge and some [r] sounds in some American English dialects are produced in this way.

Palato-alveolar sounds such as [ʃ], as in the initial sound in "sheep" are made with the tongue blade touching the back of the alveolar ridge. While Mandarin Chinese has a substantial repertoire of retroflexed dental-alveolar sounds like [ʤ] and [tʃ] as in jug and chug respectively, and retroflexed palatal sounds like [ʂ], they are rare in Cantonese, and due to language drift are no longer present in modern Cantonese.

Produced further back in the oral cavity are velar sounds [h], [g], and [k], made with the back of the tongue and the soft palate. There are sounds that are articulated even further back, such as the glottal [ɦ], which occur in some languages but are rare.

1.2.1.2 *Manner of Articulation*

Manner of articulation describes the ways in which sounds may be articulated at particular places along the vocal track.

Stops occur when articulators close off the airstreams so that air cannot escape. Oral stops occur when pressure builds up in the mouth due to complete closure of the mouth at a particular place of articulation, e.g., pie, die, and guy, involving bilabial, alveolar, and velar closures respectively. When the air is completely obstructed in the oral cavity but the velum is down and air is allowed to pass through the nose, the sound is called a nasal stop. Nasal stops occur in the beginning of the words my, nigh, and at the end of sang with bilabial, alveolar, and velar closures respectively. Egressive (breathing out) pulmonary stops are the most common type of stop consonants and are often referred to as plosives (Clark & Yallop, 1995). In Cantonese, for example, the dental-alveolar unaspirated voiceless [t] and aspirated voiceless [tʰ] are both plosives.

Fricatives are produced with close approximation of two articulators so that the airstream is partially obstructed and turbulent airflow is produced. Voiceless fricatives [f], [θ], [s] as in fie, thigh, and sigh respectively begin with the vocal cords being held

apart so that they do not vibrate (hence voiceless) and air is forced through a narrow gap producing relatively high-pitch frication noise (Ladefoged, 2001a). Voiced fricatives [ð], [z], [v], and [ʒ] as in the initial consonants of the words thy, zeal, vie and the medial consonant in measure respectively are produced with the vocal cords vibrating (Ladefoged, 2001a). Other variations of fricatives include sibilants with higher pitched sounds and a more obvious hissing quality such as [ʃ] as in shy.

Approximants [w], [j], [l], and [r] involve a narrowing at some point in the oral cavity but not to the extent of producing turbulence (Ladefoged, 2001a). For example, [l] as in lie, is produced with obstruction at a point along the centre of the oral cavity, with incomplete closure between one or both sides of the tongue and the roof of the mouth.

Other consonantal articulations in English include trills or tongue rolls (Scottish pronunciations of [r] in words like rye), and taps [t], as in the medial sound of the word letter for some English speakers, produced with the tongue making a single tap against the alveolar ridge.

1.2.2 *Vowels*

In vowel production, the air stream is relatively unobstructed and it is tongue position that is most important for variations in articulation. Vowel articulation can be described in terms of three factors: (1) the height of the tongue (high/low), (2) the horizontal position of the tongue within the oral cavity (front/back), and (3) the degree of lip rounding (rounded/unrounded). This is not an exhaustive description of vowel sounds, but is sufficient for our purposes.

For example, the vowel in heed [i] is a high front vowel, meaning that the tongue is in the front part of the mouth and the tongue body raised with respect to the hard palate. The vowel [a] in father as spoken in most English dialects is a low back vowel produced with the tongue drawn back toward the throat and the tongue body low

against the mandible.   In producing the low front [u] vowel as in who the lips are rounded.

The chart in Figure 7 is based on the cardinal vowels, vowels mapped by relative tongue position with fixed coordinates in the horizontal and vertical planes. This does not represent an exhaustive repertoire of vowels in all languages, nor the actual coordinates for all vowels across languages, but rather provides vowels with fixed coordinates so that descriptions and comparisons among different languages can be made. The horizontal plane maps frontness and backness of the tongue, and the vertical plane maps tongue height represented here by openness (close vs. open).
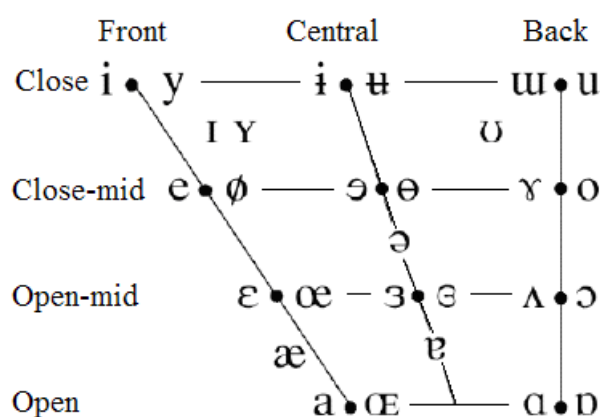
Figure 7.   Vowels charted by tongue position along dimensions of height and backness (Ladefoged, 2001a)

Each language has its own set of vowels. According to (Harrington, Cox, & Evans, 1997), Australian English contains 13 monophthongs (vowels with a single vowel quality) and seven diphthongs (vowels which change from one vowel quality to another at a certain point in time). In Cantonese, the exact size of the vowel inventory is controversial, with the number of monophthongs and diphthongs (not universally recognised) varying from 8 to 19 depending on the source (see Barrie, 2003; Bauer & Benedict, 1997; Chao, 1947; Hashimoto, 1972; Mathews & Yip, 1994; Mok & Hawkins, 2004; Zee, 2003 for descriptions of Cantonese vowel inventories). In order to illustrate the differences between English and Cantonese vowels some further information regarding the fundamental frequency and formants is required. We will return to English and Cantonese vowels after this introduction.

1.2.3 *How do we hear the speech signal? Acoustic Phonetics*

Following the above articulatory information and the mechanics of speech production, we now turn to acoustic phonetics, the study of sound properties and how we hear them.

Vibration of the vocal folds during expiration produce modified streams of air propagating sound waves which are further modified by the shape and position of the oral articulators in the oral and nasal cavities. When sound waves arrive at our ears, two basic properties are important for distinguishing sounds: frequency and amplitude.

Frequency is number of complete wave cycles per second (Hertz, Hz) and is related to the perception of pitch. Higher frequency waveforms are heard as higher pitch and lower frequency waveforms as lower pitch. Amplitude is the height of each waveform, and is correlated with acoustic intensity. Intensity is proportional to the average size of amplitude of the variation in air pressure (Ladefoged, 2001a), and is usually measured in decibels (dB) and is a relative measure in which the target is compared to some other sound. Intensity is perceived by the listener as loudness.
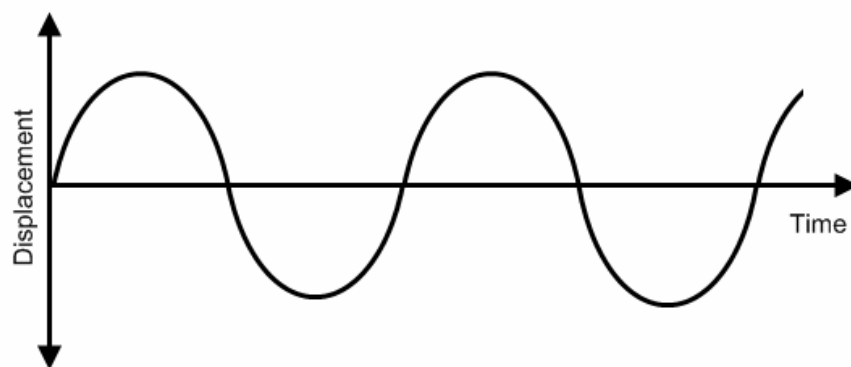
Figure 8. Simple waveform

There are two main types of sounds and these can be specified by their waveforms. Aperiodic sounds are noise-like vibrations that are random without regular repeated patterns. Consonants are, in general, aperiodic. Because there is no repeated waveform in aperiodic sounds, there is no cycling, no frequency and this no pitch as such. Periodic sounds have vibrations that follow a regular pattern repeated over a period of time. The perceived pitch of periodic sounds depends on the frequency of the

waveform.   In speech, vowels are periodic sounds the central non-transient part of a single vowel (a monophthong) is a steady state part of a speech production.

Figure 9 shows a complex waveform of the Cantonese spoken word [daːn], meaning egg, with the segmental information labelled.
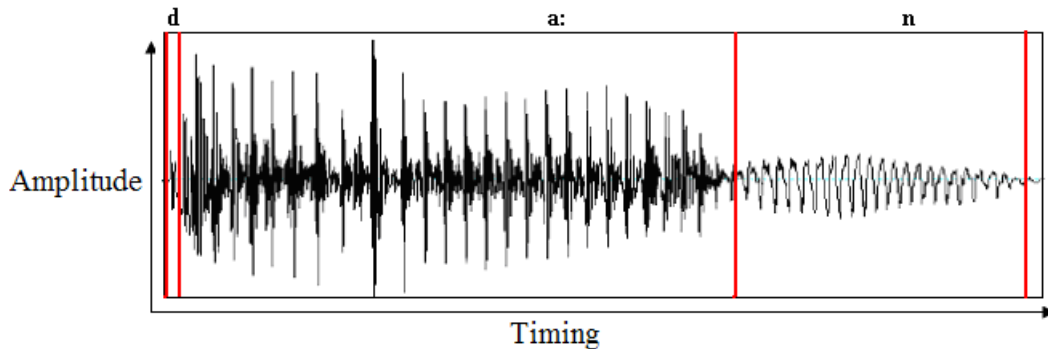
Figure 9. Segmentation of the Waveform [daːn]

The above waveform in Figure 8 is complex and contains many different components, which reflect the physical structure of the larynx, the oral and the nasal resonating cavities. In speech the fundamental frequency (F0) refers to frequency of vibration of the larynx during phonation, and differs as a function of the length and musculature of the larynx.   Children, with the smallest sized vocal folds, produce speech with high F0 in the range of 200-500Hz, women at around 150-300Hz, and men, having the longest vocal folds, produce lowest frequency speech at around 80-200Hz (Clark & Yallop, 1995).

In terms of perception, the human ear can hear frequency ranges roughly between 20 and 20,000 vibrations per second. The auditory system is sensitive to changes differentially across the frequency spectrum; for frequencies below 1000Hz, changes of 4 to 5Hz can be perceived, but for frequencies over 8000Hz, changes less than 40Hz to 50Hz are very difficult to perceive. The relationship between physical frequency changes and the perception of pitch is logarithmic (Clark & Yallop, 1995).

Apart from F0, the speech signal also contains formants, which reflect the structure of the resonating cavities and position of the articulators in the oral cavity, especially during articulation of vowels. Figure 10 is a spectrogram of the same Cantonese word [daːn]. The vowel (periodic) part of the word contains distinct dark

31

bands of energy across the frequency domain and extending over time known as formants.
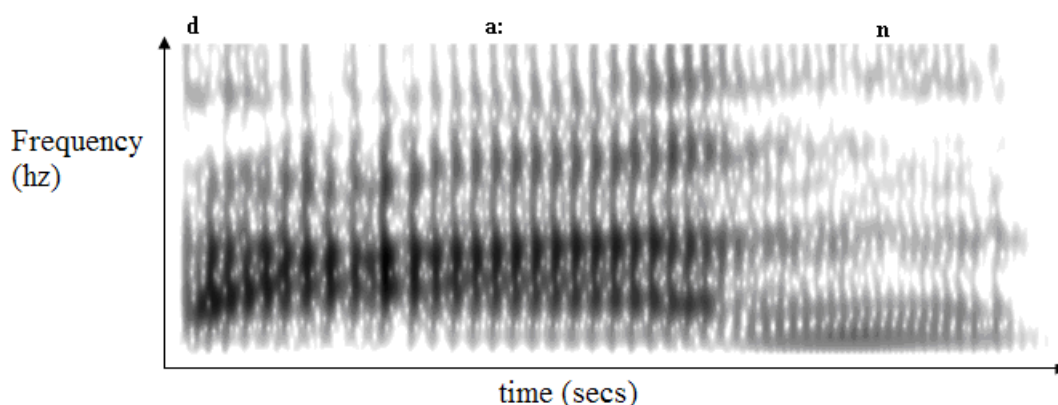
Figure 10. Spectrogram segmented for the Cantonese spoken word /daːn/

Vowels are often described and discriminated based on the first two formants (F1 and F2 the two lowest bands of energy Figure 10). While there are many more formants, (Shepard, 1978) has shown that confusions related to judgment of vowels are mostly related to the first three formants, and (Bernard & Mannell, 1986) have found similar results with Australian English.

The first formant (F1) correlates with the vertical plane (high or low/open or closed), and the second formant (F2) correlates roughly to tongue position along the horizontal plane (front or back). While vowel height is inversely related to F1 (i.e., lower frequencies on the first formant correlate to a higher vowels), F2 is not very well related to backness of the vowel, as lip rounding also influences F2. So the degree of backness is best related to the difference between F1 and F2; the closer they are to each other, the further back the vowels are produced (Ladefoged, 2001b). Since the relationship between frequency changes and pitch perception is logarithmic, formants are sometimes transformed to reflect this relationship by using either the Mel or the Bark scales (Fant, 1968).

In Figure 11 Australian English vowels are plotted along F1 and F2 dimensions in Hz (Cox, 1996). Figure 12 is a chart of Cantonese long and short vowels by F1 and F2 (Zee, 2003).

Australian English Monophthongs
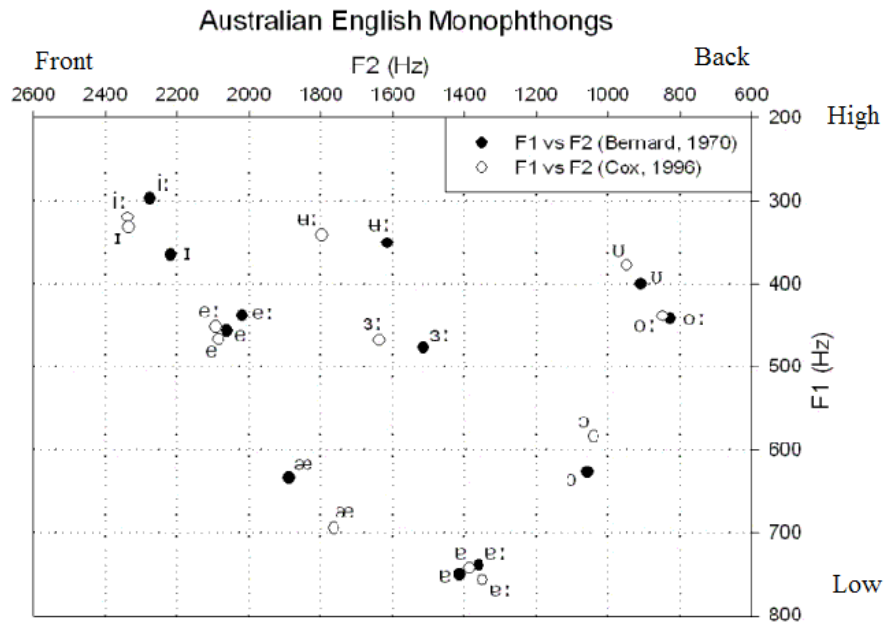
Figure 11. Australian English vowels from (Cox, 1996) with vowels represented in black versus white dots indicate findings from two different studies in 1970 and 1996.

Hong Kong Cantonese Monophthongs
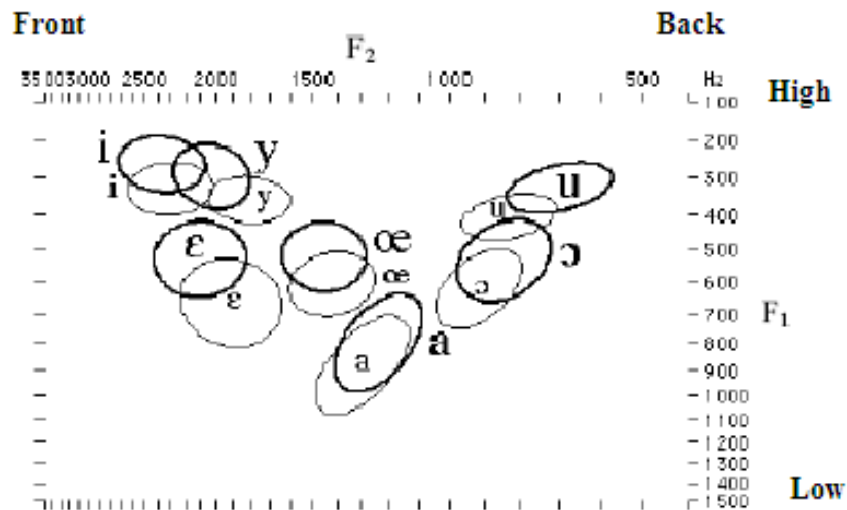
Figure 12. Hong Kong Cantonese long (dark lines) and short (light lines) vowels from (Zee, 2003)

So far we have only examined monophthongs, however, in Australian English and Cantonese there are many diphthongs. Figure 13 illustrates diphthongs again from the (Bernard, 1970) and (Cox, 1996) studies and the transition from one vowel quality

to another along F1 and F2 dimensions. Figure 13 is an illustration of Hong Kong Cantonese diphthongs and transition shown for two diphthongs.

While this is not an exhaustive description of vowels or segmental information in general, it is sufficient for the purposes of this thesis. Apart from segmentals that carry lexical information are suprasegmentals that carry important linguistic and paralinguistic information.
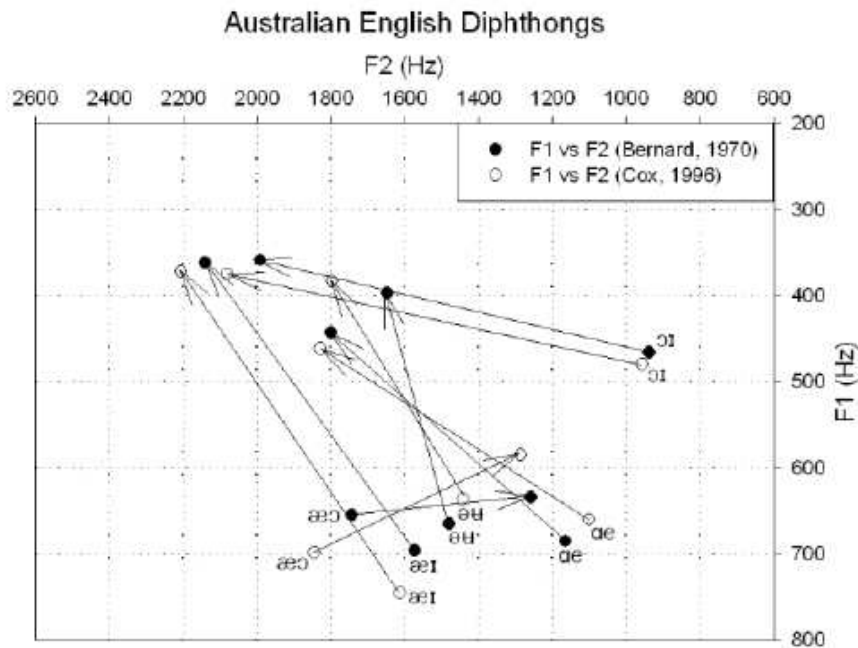
Figure 13. Australian English vowels from (Cox, 1996)
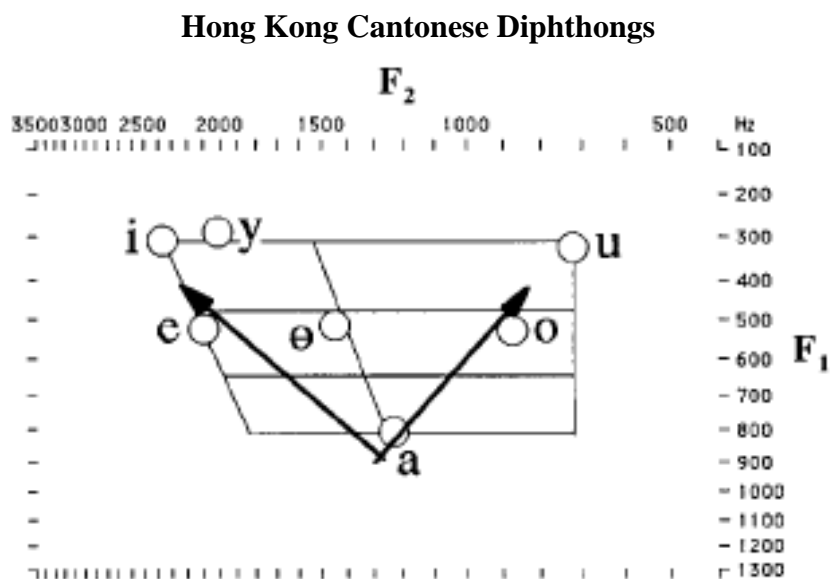
Figure 14. Cantonese Diphthongs from (Zee, 1999)

34

1.2.4 *Suprasegmental Speech Sounds*

Supra-segmental features are characteristics of speech sounds considered to be imposed on top of segments, transient, and not discretely produced in speech like consonants and vowels. However, in reality suprasegmental information is not as clear-cut as this description may suggest. Suprasegmental features, often collectively known as prosody, can be organised in different ways. One way of looking at prosody is by organising it in terms of perceptual factors of loudness, duration, and pitch, on which stress, intonation and rhythm are based.

1.2.5 *Stress*

Stress is the prominence or emphasis of certain units of speech over others. To produce stressed sounds, speakers expend more muscular energy which usually involves pushing out more air from the lungs and this extra activity may result in the sound having greater intensity, pitch, and duration (Ladefoged, 2001a).

In some languages there are specific rhythmic stress patterns for words, for example, most English bisyllabic words e.g., over, supper, and broken have a strong-weak pattern while a minority (e.g., ahead, suppose, and before) have a weak-strong pattern. Another way in which stress is used in English is to change some words from nouns to verbs, for example the word suspect can be produced with emphasis on the first syllable SUSpect or by emphasising the last syllable susPECT. The use of specific stress patterns in words is called word-stress. Some languages such as Czech and Polish have fixed word stress, whereas English, as demonstrated in the examples above, has variable word stress (Ladefoged, 2001a). Stress has other linguistic functions such as signifying the focus of the sentence, e.g., "SAM started running" versus "Sam STARTed running" with the subject as the focus in the first sentence and the verb as the focus in the second. While word stress does not usually alter lexical meaning, it does convey important language-specific information, and failure to follow such stress patterns can give the impression of foreignness.

1.2.6 *Intonation*

Intonation is the use of relative pitch variations (represented acoustically by F0 contours) along with duration and intensity, and can be used to convey syntactic information. For example, in most languages rising intonation pattern at the end of a sentence usually indicates a question, while a lowering of intonation can signal a confirmation. There is also an almost universal tendency across languages for a moderate progressive fall in pitch from the beginning to the end of a speech sequence called declination (also known as downdrift or downstep). This occurs for biomechanical reasons to do with airflow from the lungs and is not of great interest here. Like stress, failing to produce intonation will result in the speaker sounding foreign.

Besides its purely linguistic function, intonation also conveys affective information. Prosodic modification such as exaggeration in intonation is evident in special speech registers for instance infant- and pet-directed speech and appears to be used to express heightened affect (Burnham, Kitamura, & Vollmer-Conna, 2002; Grieser & Kuhl, 1988; Papousek & Hwang, 1992; N. Xu, Burnham, Kitamura, & Vollmer-Conna, 2004).

While intonation and stress can affect the meaning of the message but not specific lexical items, tones (changes of perceived pitch height and contour) do affect the lexical meaning of words.

1.3 Tones

Tone is a linguistic term referring to variations mainly of pitch height and contour, but may also include voice quality and durational variations that characterise particular tones. Tone is produced discretely and is in general carried on the steady state portion of a word and this is often reflected in written form with tone markers over vowels or in the case of Thai, over consonants. While all languages use changes in pitch height (related to stress and intonation), in tonal languages, pitch modifications affects the lexical meaning of the word "not just its nuances, but its core meaning" (Yip, 2002, p.1). In this way, it may appear that tone is not different to consonants and vowels, and could

be considered as segmental (see Burnham & Mattock, 2007; Mattock, Molnar, Polka, & Burnham, 2008).

Cantonese has six tones and an example of how tones operate in Cantonese is given in Table with the syllable /fan/, which can be produced with the six different tones to induce six different lexical items.

Table 1. Examples of Cantonese words on 6 Cantonese tone.

Tone	Meaning	Chinese Logographs
High level	to divide	分
High rising	powder	粉
Mid level	lecture	吩
Low-level	portion	份
Very low-level	grave	坟
Low rising	angry	愤

Although in tone languages, tones precede other suprasegmental information, it does not follow that tone languages do not utilise intonation. However, it is the case that intonation patterns tend to be flatter in tone than that in non-tone languages (Clark & Yallop, 1995). This is not necessarily true for all speech registers, as will become clear in the consideration of infant-directed speech in Chapter 3.

1.3.1   *How is Tone Articulated?*

Ahead of a more detailed discussion of tone in Chapter 2, this chapter will conclude with a brief discussion of how tone is produced. Tone production is influenced somewhat by the activity of the pulmonary system below the laryngeal system, but not at all by the articulators and resonance chambers above the larynx. Early studies on speech production involved freshly excised human larynges with weights attached to parts of the larynx to simulate the laryngeal muscles, and blowing into the trachea to produce sounds. Such studies suggest that the pulmonary system plays a part in pitch variation (Ohala, 1978). However, later studies using more modern methods, which include acoustic analysis to monitor F0 changes, and electromyography to monitor various laryngeal muscle, as well as observation of the pulmonary system activity, confirmed that perceived pitch changes based on F0 are best related to changes in

activity of the cricothyroid muscle in the larynx (Ohala, 1978). Moreover in tone languages the activity of the cricothyroid muscle precedes each pitch peak by just a few milliseconds suggesting that it is the cricothyroid muscle that is primarily responsible for raising pitch (Yip, 2002).

In rising pitch, the cricothyroid muscle elongates the vocal folds, which decreases their effective mass and increases their stiffness, and thereby increases the frequency of vibrations resulting in increased frequency and this increased perceived pitch (Ohala, 1978). Pitch lowering is more complex. The activity of the cricothyroid muscle increases while the thyroarytenoid muscle contracts to thicken the vocal folds; however, other external laryngeal muscles also work to lower the larynx and thus lower the frequency and perceived pitch (Ohala, 1978; Yip, 2002).

Two

Chapter 2

Tonology, Tonetics, & Tones of Cantonese

2.1 Terminology and Definitions

Before embarking on an exposition of tone, it is important to make clear the distinction between fundamental frequency (F0), pitch, and tone, and intonation. The acoustic term F0 refers to the acoustic property of the signal or vibration frequency of the vocal folds in the larynx (Yip, 2002), and is measured in Hertz (cycles per second). Pitch is the psychological counterpart of F0 and perceptual dimension concerned with whether the sound high, low, mid, and also involves movement such as high to low, or high to low to high. Tone is a linguistic term that refers to a group of cues that are used to convey lexical information. The main cue for tone is F0 (changes in F0 height and/or movement); but tone variations can incorporate other factors such as duration, amplitude, and voice quality, although these cues are not as central as F0 to tone perception in languages generally and in Cantonese in particular. Intonation, another linguistic term, is a suprasegmental feature of speech that also makes use of F0 movement for grammatical or affective, but not lexical ends. The perception of tone and intonation are therefore dependent on F0. A further distinction that can be make is between tonetic (tone) and tonemic (toneme) in a similar fashion to phonetic (phone) and phonemic (phoneme), in which tones are tonetically distinct sounds, and tonemes are groups of tones used in a particular language to distinguish meaning.

Here it is useful briefly to make the distinction between stress, tone and pitch-accented languages. The discussion above has introduced stressed languages such as English (and also Polish and Czech), and tone languages such as Cantonese. However, there is one more a class of language to consider, pitch-accented languages, which fall somewhere between stressed and tone languages. Japanese is an example of a pitch-accent language where an accent is always realised as a high pitch, and words differ in the placement of the accent (Yip, 2002). For example, "Ka\ki ga" (with first syllable accented) means oyster, "kaki\ ga" (with second syllable accented) means fence, while "kaki ga" the unaccented combination means persimmon in Japanese, where capital letters denote accents (Clark & Yallop, 1995).

Yip (2002) argues that the majority of "accent languages" have lexical tones, but the number is small, they are sparsely distributed, are absent on some words, and frequently lexically-associated with specific tone-bearing units (e.g., particular morae in Japanese). Therefore, she argues there is no absolute division between accent languages and tone languages, just a continuum from accent to tone, as the number and density of tones increases, and becomes less constrained.

2.2   Tonology: What is Tone?

According to Goldsmith (1994) tone languages differ from other languages in three ways: firstly, the span of each tone melody is roughly the size of a word, whereas for non-tone languages the span of an intonation melody size ranges between that of a syntactic phrase and a sentence; secondly, in tone languages the tone melody of an utterance is comprised of the tone melodies that are directly contributed by the lexical items in the utterance and to a slightly lesser extent by the syntactic constructions in the sentence, but for non-tone languages melody is generally determined by the information structure of the sentence; and finally, tone languages generally have phonological rules that modify the tone melodies depending on the tones found around them as well as on the syntactic structure in which they occur (p.4626). However, not all tone languages are exactly alike as will become evident in the following section.

Although the concept of tone may be a novel one for non-tone language speakers, it is estimated that as many as 60-70 per cent of the world's languages are tonal, and the number of speakers of Mandarin Chinese alone is estimated at about 885,000,000 (Yip, 2002). Since such a large percentage of the world's languages are tonal, let us now turn to a continent by continent view of tone languages.

2.3   Tonal Languages

Beginning in Australasia, languages in Australia and New Zealand are not tonal, which is in stark contrast to Asian languages that are predominantly tonal, e.g., Thai, Burmese, Vietnamese, with Chinese languages comprising the largest population of tone speakers. Other Asian languages such as Japanese are pitch-accented languages. Moving west to

India, the languages are predominantly non-tonal. Across India to the Middle East, no tonal languages are reported; and then north-west in Europe, the same trend continues with most European languages being non-tonal except Serbo-Croatian, Lithuanian, Swedish, and some dialects of Dutch that are all pitch-accented languages. Moving south to Africa there is a high concentration of tone languages. Crossing the Atlantic to the Americas, Central America also has a large cluster of tone languages.

Thus there are high concentrations of tone languages in Africa, America, and Asia. Each of these language families is addressed in the following sections in terms of tones type, size of tone inventories, grammatical implementation, and notation system for representing tone in these families.

2.3.1   *African Tone Languages*

The Sub-Saharan Africa has the largest concentration of tonal languages in the world, with the Niger-Congo family of languages including Bantu being almost entirely tonal (Yip, 2002). The African tone languages do not in general have a large tone inventory and it is common to have only two to three tones. As these tones are usually level (register) tones, rather than contour tones these languages are referred to as register languages, with only a very few of these languages having contour tones in addition to register tones. The most distinct feature of African tones that differentiate them from American and Asian tones is tonal mobility. This refers to tone movement of the tone of a particular toneme to be manifested on another morpheme in some circumstances. Yip (2002, p. 133) describes tonal mobility as follows "…when morphemes are concatenated into words or even phrases, the tone of one morpheme may migrate some distance from their point of origin". The exact manner of operation for tonal mobility in African tone languages is rather complex and is beyond the scope of this thesis but see Yip (2002) for a thorough treatment.

2.3.2   *American Tone Languages*

Most of the Central American tone languages belong to the Otomanguean family (Yip, 2002). American tone languages can be considered to fall somewhere between the African and Asian tone languages; their tone inventories resemble those of the Asian

languages with many level and contour tones, but like the African languages, most of their contour tones are derived concatenations of level tones (Yip, 2002). What makes American languages unique among tone languages types is the interaction of stress and tone, in both directions (see Yip, 2002 for discussion on these interesting phenomena).

2.3.3   *Asian Tone Languages*

Perhaps the most well known and well studied tone languages are the Asian languages in which tones are almost exclusively used lexically, with no correlation with syntactic or morphological aspects of the languages Wang (1967).   Asian languages can be roughly divided into three major families: Sino-Tibetan languages, including the Chinese language family and Tibeto-Burman, are almost completely tonal, the Austro-Tai family including Tai-Kadai (both Thai and Lao belong to this language family) are also tonal; while the Austro-Asiatic family are mainly non-tonal with the exception of the Mon-Khmer branch which contains both tonal languages like Vietnamese and non-tonal languages like Cambodian (Yip, 2002). Other languages in Asia such as Japanese are pitch-accent languages.

The Asian languages have by far the largest inventory of tones, e.g., six in Cantonese and five in Thai, as well as having both static and contour tone and, within these, quite complex contour tones, e.g., the Mandarin fall-rise third tone. The large inventory of contour tones that are more the norm than the exception in Asian tone languages is what sets them apart from the African and American tone languages. Asian tone languages that have attracted the greatest amount of research include Chinese Mandarin (four tone system), and Cantonese (six tone system), as well as Thai (five tone system). Belonging to a different family of Asian languages, Thai differs from Chinese in many ways, one of which is its lack of tone sandhi rules. Now, following an exposition of tone notation, a more detailed description of Cantonese is given.

2.4   Tone Notation Systems

With variations of tone languages come variations of tone notation systems. Tone notations described in the following sections are accepted, linguistically-imposed prescriptions that aim to describe a given language's tone system and not necessarily a form employed in the indigenous orthography.

2.4.1   *Asian Tone Notation*

The Asian tone notation system was developed by Chao (1930) who proposed that tone languages require speakers to divide their natural pitch range into a maximum of five levels. Accordingly, he devised a notation system in which each tone is labelled using two integers from 1 to 5 one each to indicate the onset and offset pitch height of the tone, with 5 representing the highest pitch and 1 the lowest. Occasionally a third middle number is added for complex tones that change direction during tone production, as for the third tone in Mandarin (213). This assignment is relative rather than absolute, as it depends on the individual speaker's pitch range, not an absolute set of pitch targets that everyone need produce (which arguably makes learning a tone language difficult because simply imitating another speaker's absolute pitch will not result in the correct production of the tone). Table 2 shows a comparison of three authors' assignment of the Chao notation to Cantonese tones:

Table 2.   A Comparison of Cantonese Tone Notations from Chao (1947), Hashimoto (1972), Rose (2000), and Vance (1976).

		(Chao, 1947)	(Hashimoto, 1972)	(Vance, 1976)	(Rose, 2000)
Level Tones	High	53	53	55	55
	Mid	33	44	33	33
	Low	22	33	22	22
Contour Tones	High-rising	35	35	35	25
	Low-rising	23	24	13	23
	Low-falling	21	21	11	21

2.4.2   *American Tone Notations*

The Central American languages have their own tradition of labelling tones, not entirely unlike the Asian system. However, the representation of pitch range is reversed with respect to the Chao system, with 1 representing the highest pitch range and 5 the lowest. Level tones are represented with only one integer and contour tones have a hyphen between the onset and offset pitch values (Yip, 2002). An example of this system from Yip (2002, p19) is given in Table 3.

Table 3. Central American Tone Notation System (Yip, 2002)

Tone Type	Tone	Notation
Level:	High	si^1
	Low	si^4
Contour:	High-rising	si^{3-2}
	High-fallings	si^{2-3}

2.4.3   *African Tone Notations*

For African languages, the norm is to use accents over vowels to mark tone. A high tone is marked with a rising accent [á], low tones with a falling accent [à], mid tones a level accent [ā] or more commonly left unmarked, and in cases where there are contour tones, rising [ǎ] and falling [â] contour tones are also marked. In some cases, two rising or falling accents are used on the same vowel to indicate "extra-high" [ő] or "extra-low" [ȍ] tones (Yip, 2002, p.19).

2.5   The Chinese Tones in Detail

Chinese, a member of the Sino-Tibetan branch of the Asian language family can be divided into at least eight sub-divisions (Mandarin, Wu, Yue, Min, Xiang, Hakka, Gan, and Jin) which are widely known as Chinese dialects. However, as Yip (2002) and Ladefoged (2001) point out, from a linguistic point of view they should be more appropriately recognised as different languages as they are not mutually intelligible, have different phonologies and lexicons, and have some syntactic differences. The

picture complicates further when we consider that these eight Chinese languages are spoken by the Han majority while the 55 minority groups in China speak languages that are members of the Altaic, Austroasiatic, Tai-Kadai, Indo-European, and Austronesian families. For the purposes of this thesis, only Han Chinese will be considered.

Unlike African languages, Chinese languages in general tend to have simple syllable structures and morphology, and morphemes are almost entirely monosyllabic (Yip, 2002). The tone contrasts thus considerably enlarge the otherwise small syllable inventory. Of the eight languages belonging to the Chinese language family, Mandarin and Cantonese (which is actually a dialect of Yue) have attracted greatest research interest and so are well documented, and it is to these languages that the following sections will now turn.

2.5.1 *Mandarin Tones*

The largest language family within Chinese is Mandarin, with Beijing, Tianjin, and Taiwan as major dialects. Mandarin is often used both to represent the language and dialect. Mandarin is also known as Putonghua or "Common Speech". It is the official language in China, and has the largest population of speakers of any Chinese language.

Mandarin in particular and Chinese in general are monosyllabic languages and this is reflected in the logographs employed across all Chinese languages, with each logograph representing a single monosyllabic word. The logography of Chinese characters had been Romanised in Mandarin for ease of learning, via the Han-Yu Pin-Yin system, Pin-Yin for short, which carries diacritics or accents to represent the four tones. This has however not become a standard working orthography and was developed only with the purpose to aid learning Chinese logographs and tones in schools. This system also inadvertently increases the learners' awareness of the sound structure in Chinese that the logographs do not reflect (Read, Yun-Fei, Hong-Yin, & Bao-Qing, 1986).

A comparison between the Pin-yin system and Chao tone letters for the word [ma] is given in Table 4.

Table 4. A comparison of Mandarin Tones using Chao tone letter and Pin-yin

Tone	Logographs	Lexical meaning	Chao Tone letters	Pin-yin
High level	妈	mum	55	mā
Mid-rise	麻	hemp	35	má
Fall-rise	马	horse	214	mǎ
High-fall	骂	to scold	41	mà

Compared to other Asian languages, Mandarin does not have a large tone inventory; it consists of a static high level tone, two contour tones (a mid rising and a high falling tone), and a complex contour tone that is low, falling then rising. The four-tone system of Mandarin means that a given monosyllabic phonetic string [ma] may have up to four lexical meanings differentiated by tone. While the tone system allows an increase in the syllable inventory of Mandarin, there are still a considerable number of homophones for each syllable by tone combination.

Tone Sandhi: Tone sandhi occurs when tones are changed systematically depending on other tones in the environment; the most prominent rule that Mandarin follows which differs from other Chinese languages is its third tone sandhi rule (Ladefoged, 2001a). This refers to the practice of changing the third or fall-rise tone to a second or mid-rising tone when it is followed by another third tone in running speech (Chen, 2000). This complicates the perception of Mandarin tones in running speech, as one need to take into consideration both acoustic phonetics and lexical processes when dealing with tone sandhi. Interestingly, some linguists also believe that in addition to this there is another tone sandhi rule for tone 3; it is also changed to the high level tone (tone 1) when preceding high level, rising, or falling tones (Li & Thompson, 1978) . If one considers that there are only four tones in Mandarin, then if the third tone is modified by each tone following it then it could be argued that the third tone does not exist at all in its canonical form in running speech.

2.5.2   *Cantonese*

Cantonese is a dialect of Yue spoken in the state of Guang Dong, with two major standard dialects spoken in the city of Guang Zhuo and in Hong Kong.   Strictly speaking, the term "Yue" should be used to indicate the language and "Cantonese" the dialect, however, as both native and non-native speakers have come to treat "Cantonese" as synonymous to "Yue", Cantonese will be used here to indicate both the language and the dialect. Cantonese, like Mandarin, is also in general monosyllabic, but unlike Mandarin which only allows vowels or nasal consonants in final word position, Cantonese has final stop consonants /p/, /t/, and /k/ which also play a role in Cantonese tone.   In addition, unlike Mandarin which unquestionably has a four-tone system, the Cantonese tone system is more complex and the specific number of tones in the tone inventory is a contentious issue.

There has been a long debate about the exact number of tones in Cantonese with the possible number ranging from 6 to 10.   This lack of consensus may be traced to several factors.   In Cantonese there are high-level and low-entering tones which occur before and are always associated with the unreleased final consonants /p/, /t/, and /k/. Some early authors considered such checked syllables (with unreleased consonants) to belong to categories to be contrasted with other tonal categories, however Mathews and Yip (1994) argue that the distinct tones before unreleased consonants are simply abbreviated counterparts of the three level tones. Like Mathews and Yip, many modern writers have come to regard the traditional analysis as unnecessarily complicating the system, and believe that the three tones associated with checked syllables need not be included in phonetic transcriptions.

In addition, with dialect drift it has become increasingly the case that the high level (55) and high falling tones (53) are no longer phonemically distinctive in modern Cantonese, especially in the Hong Kong dialect, further reducing the size of the tone inventory (Yip, 2002).   Most modern linguists consider Modern Cantonese as having only six distinct tones.

Even with this reduced number of tones (six), Cantonese is still considered a complex tone language because it contains many level and contour tones with similar

pitch onsets (see Table 2). Comparison of the different notations used to describe the pitch height of the Cantonese tones in Section 2.3.1 illustrates the complexity of Cantonese tone system, especially the lack of consensus for the tone notations, thus reflecting conceptual inconsistencies about the phonetic properties of level tones and tone onsets. The high-rising tone (which some prefer to call the mid-rising tone because it does not begin high in the speaker's pitch range) and the low-rising tone, are sometimes not distinguished by onset, but do differ in offset and gradient, with the high-rising tone having a sharper gradient and rising to a higher pitch level than the low-rising tone which only rises to the mid-level pitch range. The low-falling tone (21), which starts low and drops further giving it a creaky voice quality, can also be produced with a less accentuated drop in pitch by some speakers and can sometimes be produced as a level pitch (22, 11) below the low level tone (22/33).

Recent research has assisted in the clarification of tone notation in Cantonese. A study by Li (2004) shows that of the original tone letter assignments Chao's (1947) and Vance's (1976) descriptions are more accurate than Hashimoto's (1972) system, which involves more crowding in the lower pitch ranges than high for Cantonese tones (Vance, 1977) since it is easier to distinguish between the high-level tone and mid-level tones than the mid-level and low-level tones. Recently using acoustic data from male and female native Hong Kong Cantonese speakers with F0 normalised by duration, Rose (2000) was able to plot pitch trajectories of modern Cantonese (see Table 2). Rose's (2000) descriptions of the Cantonese tones are consistent with Li's (2004) findings and is adopted for the description of Cantonese tones in this thesis.

2.6   Tone Perception

There are several issues related to studying tone perception which are addressed here. Firstly, it is important to consider whether tones are perceived in the same way as phones (segmental information including vowels and consonants), whether prosody modifies tone perception, and the degree to which tones and phones are coarticulated and perceptually interdependent. Secondly, the acoustic features that are important for tone perception are of interest, as is the role of linguistic experience with a particular

tone language and how context influences tone perception. Finally, the acquisition of tone perception and how this might differ in first and second language learning needs to be considered in order to develop an understanding of tone perception, and to feed into a discussion of the concept of a tone space in the final section of the chapter.

2.6.1   *Perception of Tones and Phones*

One way to address the issue of tone and phone perception is from categorical perception data. Categorical perception is measured by measuring listeners' responses to a continuum created along which some physical dimension is varied. In consonants, voice onset time (VOT, the time between complete obstruction at the articulators and vibration of the vocal folds) is often used as a dimension of variation and one for which categorical perception has been found consistently (Blumstein & Stevens, 1980; Liberman, Harris, Hoffman, & Griffith, 1957). Categorical perception refers to the phenomenon in which small physical changes are not detected better than chance within phonemic categories but are detected around category boundaries, where perception usually undergoes a sudden shift from perceiving one consonant, e.g., /ba/, to perceiving another, e.g., /pa/. This change in perception is also evident in identification curves, plots of response accuracy over the relevant dimension, by a sudden change from response as one to response as another category.

Unlike consonants, vowels are perceived more continuously (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973). This difference in perception may be attributed to the kinds of information carried in each type of segment, with consonants specifying more lexical information and vowels carrying more prosodic and speaker related information (Werker & Polka, 1993).

The data on categorical tone perception is relatively inconsistent, with some studies finding evidence for continuous perception of tones (Abramson, 1979b; Francis, Ciocca, & Ng, 2003; Li, 2004) and others for categorical perception of tones (Chan, Chuang, & Wang, 1975; Stagray & Downs, 1993; Wang, 1976). Abramson (1979b) examined the three level tones in Thai and found continuous perception, while Chan et al. (1975), Stagray and Downs (1993), and Wang (1976) found categorical perception

using level and contour tones in Mandarin. From these data it appears possible that the perception of level versus contour tones is categorical while perception among level tones is continuous. However, both Francis et al. (2003) and Li (2004), found the perception of the six contrastive tones in Cantonese to be continuous, with identification curves similar to those of vowel identification curves (with no steep abrupt boundaries). It appears then that continuous perception of tones is associated with Thai and Cantonese tones but categorical perception is associated with Mandarin tones (Chan et al., 1975; Stagray & Downs, 1993; Wang, 1976) This may be a reflection of the different tone inventories of these languages – note that Mandarin is the only language with a single level tone, while the others each have three level tones.

Although clearly more research is required into the continuous/categorical nature of tone perception, from the results so far it would appear that tone perception falls somewhere between vowels and consonants, and that the presence or absence of categorical segment perception may be more parsimoniously explained by the different tone systems and how they are used in each language rather than by the kinds of information they carry (e.g., lexical for consonants versus suprasegmental for vowel, Werker & Polka, 1993).

2.6.2 *Tone Coarticulation*

Coarticulation occurs in speech when phonetic properties of given speech sounds are modified by surrounding sounds, and can be either anticipatory where speech sounds are influenced by articulation of subsequent sounds, or preservaratory where speech sounds are influenced by the articulation of previous sounds. Tone coarticulation has been found in Thai (Abramson, 1979a), Vietnamese (Brunelle, 2003; Han & Kim, 1974), and Mandarin (Wang, 2000; Xu, 1994), and is, in general, preservaratory rather than anticipatory (Ladefoged, 1982). Brunelle (2003) found that in Northern Vietnamese the broken and rising tone appears to raise the pitch onset of following tones, but the falling and curve tones lowers them, while the level tone has minimal effects on following tones. In Mandarin, all tones following tones 1 and 2 (high level and rising) have a higher pitch than when following tones 3 and 4 (dipping and falling),

and tone 2 has a smaller rising slope following tones 1 and 2 than when following tones 3 and 4 (Wang & Seneff, 2000). However, there is also evidence of anticipatory tone coarticulation in Mandarin; over and above tone sandhi, all tones preceding tone 3 have a much higher pitch and all tones preceding tone 1 have a lower pitch (Wang & Seneff,2000). Abramson (1979a) found that compared to tones in citation form, Thai tones in running speech tend to have F0 that is more compressed, and contour tones disrupted at the beginning and ends showing both anticipatory and preservaratory coarticulation effects.

Khouw and Ciocca (2007)have shown show that the direction and magnitude of F0 change toward the later parts of the vocalic segment correlate more with tone identity than do the earlier F0 elements, suggesting that perception of tone may not be affected by preservaratory coarticulation (which has its effect mostly on the initial part of the next tone) and probably only minimally by anticipatory coarticulation.

Coarticulation of Tones with Phones: In general tone is more affected by consonants and vowels than vice versa (see Hombert, 1978) for a thorough treatment of consonant and vowel influence on tones). The notion that consonants play a part in the historical development of tone is a popular one in theories of tonogenesis. It is believed that lower and higher pitched vowels developed as a product of perservarative coarticulation from preceding voiced and voiceless stops, or anticipatory coarticulation from following voiced and voiceless stops. The latter is thought to be so especially in Chinese in which syllable final stops have almost completely disappeared from modern Chinese presumably as a product of their being replaced by pitch (tone) differences (Hombert, Ohala, & Ewan, 1979). A similar pattern occurs for vowels in that high vowels tend to be accompanied by high tones and low vowels by low tones (Hombert, 1978). The exact mechanisms behind such patterns of coupling of tones with phones is unclear, and so it is debatable whether this effect is due to psychoacoustic or articulatory influences.

As tone is carried mainly on the steady state vowel part of a syllable, it is unlikely that there will be perceptual interactions between consonants and tones. There

is also evidence that vowels and tones are statistically uncorrelated in perception (Mixdorff, Luksaneeyanawin, Fujisaki, & Charnvivit, 2002).

Coarticulation of Tones with Prosody: Ross, Edmondson and Seiburt (1986), found that compared to English, the presence of tone in Taiwanese, Mandarin, and Thai significantly inhibits the unrestricted manipulation of affective prosody via changes to pitch and F0 contour. However, there is evidence to suggest that rising intonations in questions and down-step or sentence final declination that is found in non-tones languages also occurs in tone languages. Ma, Ciocca, and Whitehill (2004) found tone level and contour changes with sentence intonation and rising intonation for all six tones at the final positions of questions in Cantonese just as in other languages such as English. In a similar vein, Vance (1976) found sentence final declination occurring in Cantonese, and Kitamura, Thanavishuth, Burnham, and Luksaneeyanawin (2002) have found the same trend in Thai and at the expense of intelligibility for tone recognition.

In summary, the results of coarticulation studies suggest that there is indeed between-tone, tone-phone, and tone-prosody coarticulation effects in the production of tones. However, tone perception is minimally affected by coarticulation with phones, and somewhat but not extensively affected by between-tone coarticulation. The general effect of coarticulation for tone perception occurs for the intonation of tones and prosodic information.

2.6.3 *Salient Features of Tone Perception: From Studies of Native & Non-native Tone Perception*

Recall that, despite the involvement of duration, amplitude and voice quality, the most important acoustic correlate of tone is the F0. Using multidimensional scaling Gandour and Harshman (1978) found for Thai, Yoruba, and American English speakers, five dimensions that are most used in the perception of tone: average pitch, direction, duration, extreme endpoint, and slope, with tonal speakers more likely to use F0 direction, length, and slope. Using a larger sample of tone speakers Gandour (1983) presented 19 different F0 trajectories (5 level, 4 rising, 4 falling, 3 falling-rising, and 3

rising falling) superimposed on a synthetic speech-like monosyllable, to native Cantonese, Mandarin, Taiwanese, Thai, and English speakers and found that two dimensions were most important for tone speakers, F0 level and direction of movement. Other studies have confirmed that F0 height and movement are both necessary and sufficient for identifying tone in various languages (Gandour, 1981; Massaro, Cohen, & Tseng, 1985).

While tone speakers may use a constrained set of acoustic cues to perceive tones, tones are not equal in terms of ease of detection. Li (2004) found Cantonese listeners had more difficulty discriminating the mid and low level tones (33 versus 22) which have similar F0 onset and offset, and the two rising tones (25 versus 23) which have the same F0 onset.

Looking at native tone speakers' ability to generalise use of tone information to another tone language, Lee, Vakoch, and Wurm (1996) found that tone speakers were generally able to discriminate tones of non-native languages, but there were some asymmetrical effects; Cantonese speakers discriminated Mandarin tones better than vice versa. These results may suggest that experience with a large tonal inventory as in Cantonese may facilitate the perception of tones in a language with a smaller tone inventory such as Mandarin but not vice versa.

The role of context: Using natural speech, low-pass filtered speech, and musical sounds that resemble the pitch contours of normal speech, Burnham et al. (1996) found that English speakers performed significantly better in the music condition than filtered speech condition and in turn better than the speech condition, whereas native speakers of Thai and Cantonese discriminated contrasts equally well in all three contexts. Their results suggest that tone and non-tone language speakers alike are familiar with pitch information in a non-speech context but tone perception requires familiarity with using pitch information in a linguistic (specifically lexical) context. In a similar vein, (Shen & Lin, 1991) using native Mandarin listeners and more recently (Schwanhäußer & Burnham, 2005) with native Mandarin and Vietnamese listeners, found native tone

language listeners rely less on and English listeners rely more on phonetic information, suggesting that phonology becomes more important with familiarity of tones.

2.6.4   *How is Tone learned?*

In addressing the issue of tone acquisition, two areas of research must be examined: adults learning of non-native tones and tone acquisition in children.

2.6.4.1   *Training*

Wang, Spence, Jongman, and Sereno (1999) found that training English speakers to identify Mandarin tone increased identification accuracy from pre-test to post-test even when novel stimuli (21%) and new speakers (25%) were presented at post-test, and even when tested later, six months post training (21%). In addition Wayland and Guion (2004) reported that both identification and discrimination training procedures were effective in increasing both Mandarin and English speakers' ability to identify and discriminate Thai tones.

Looking at production data, Wang, Jongman, and Sereno (2003) using only perceptual training found gains in native English speakers' productions of novel Mandarin tones as indexed by native Mandarin listeners' ratings of the tone productions with average ratings increasing by 18% in post- compared to pre-test. It is of interest that this improvement was uneven over tones, ratings of non-native productions of the fall-rise tone were the lowest compared to productions of the other tones. Therefore, it appears that pitch height and pitch contour may not be mastered in parallel, with pitch contour being more resistant to improvement than pitch height. This suggests that the relationship between perception and production of tones is not related in linear manner (Wang et al., 2003).

2.6.4.2   *Acquisition of Tones in Infant and Children*

In a survey of early tone acquisition literature, Li and Thompson (1978) concluded that the acquisition of tones productions appears to be completed early in childhood and before the completion of segmental production. More recently, Hua and Dodd (2000) similarly found that for Mandarin children tones were acquired first then syllable-final consonants and vowels, and then syllable-initial consonants. Given that tones and

phone appear to be equally well perceived in infancy (see Chapter 3) the reason for earlier completion of tone acquisition is unclear but one explanation may be that there are fewer tones than vowels and consonants, and learning is therefore completed earlier.

What is also unclear is the order of tone acquisition. Most studies suggest that level tones are acquired before contour tones, but this is also dependent on the tone context and how tones behave in given languages. For example, the Mandarin rising and dipping tones are acquired after the other low tones possibly due to tone sandhi creating some ambiguity about the third tone (Li & Thompson, 1978). It would be interesting to investigate whether a similar pattern occurs for others languages such as Cantonese in which the two rising tones, and the low rising and low level tones are very similar in acoustic properties.

First and Second Language Learning: Despite some similarities (both adults and children appear to have difficulties with producing the dipping tone in Mandarin) the relative ease of acquisition of native tones by children is not reflected in the training data, with most individuals only improving very modestly (approx. 20%) albeit sustained improvement even after some time post-training. Tones, like vowels and consonants appear to be acquired with ease but learned by second language learners with great difficulty, which may relate to possible similarities in the psycholinguistic organisation of phones and tones.

2.7   Tonal Mapping

According to Abramson (1986) tone space is "the set of articulatory and auditory dimensions by which the speaker is constrained in production and perception" (p. 105). In the same way formants, especially F1 and F2, are important for vowel perception, the perception of tone is very much dependent on articulatory determinants and acoustic properties related to F0. The task here is to find an appropriate measure that would map tones onto a tone space in a similar way that vowels are mapped by F1 and F2. From the above survey of literature, it is clear that F0 height and contour are particularly important for the perception of tones, including Cantonese tones. Any method of tone

mapping will need necessarily to account for these two basic dimensions of tone perception.

Early Studies: Early studies on tone perception relied on comparative descriptions of tone contours produced by different speakers (Gandour, Weinberg, Petty, & Dardarananda, 1988; Kent & Murray, 1982). Rose (1987; 1993) developed a method of plotting tones by F0 height with duration normalised which appears to capture both height and tone contours across time. Rose's method assumes that all tone speakers of a given language would produce the same variety of linguistic tones and a similar set of tone targets. To establish the pitch height and contours of a given language's tone inventory tone productions from multiple native speakers can be compared using Rose's normalisation method both for within and between participant productions (see Rose, 1987, 1993 on specific details of this method). While this method is very useful in providing a picture of a given language's tonal inventory, it cannot provide an index of the size of a given individual's tone space as can be obtained for vowels. Recall in Chapter 1 (see section 1.3), the plots of the Australian English and Hong Kong Cantonese vowel spaces. Taking the most extreme vowels, the corner vowels [i], [a], and [u] in most languages including Cantonese, a vowel triangle can be formed, the area of which provides an index of the size of a given speaker's vowel space for comparison across speakers and speech registers within speakers. Using such measures studies comparing productions of vowels across speech registers such as infant- and adult-directed speech have found vowel hyperarticulation or the expansion of vowel space in infant- compared to adult-directed speech (Burnham et al., 2002; Kuhl et al., 1997). In order to investigate the possibility of a parallel phenomenon for tones in infant-directed speech, the main purpose of this thesis, a comparable method for tone space mapping is required.

2.7.1 *New Tonal Mapping*

In order to make comparisons of tone productions between hearing impaired and normal hearing children Barry and Blamey (2004) developed a method of mapping Cantonese

tones without the need for normalisation across speakers, i.e., with "no requirement for a speaker to have the same tonal target as any other speaker" (p. 1741), making it also suitable for within-participant comparisons across speech registers. This method involves plotting tones by F0 onset and F0 offset. This method captures average pitch, direction, extreme end points and slope, all of the dimensions found by Gandour and Harshman (1978) to be important for tonal perception except length; although recall from sections 2.5.3 that F0 level and contour are necessary and together sufficient features for tone perception, which suggests that duration information may not be central to Cantonese tone perception (Gandour, 1981; Massaro et al., 1985).

By plotting multiple tokens of tone productions from the same speaker, tone ellipses can be created for each tone in a specific area within the F0 onset and offset tone space for that speaker. From this basic analysis (at least) two measures can be derived: tone differentiation within the tone space, and differentiation among tonemes (Barry & Blamey, 2004). The first represents the spread among the different tones within the tone space, while the second represents the spread of tone productions of each tone within the tonal space, that is how much space each individual tone occupies in the tone space. Note that this method has been developed for Cantonese tones and so these different measures will be illustrated with reference to Cantonese.

Differentiation among Tones in the Tone Space: The differentiation of tonal space is given by the following formula:

$$\text{Between-Tone Differentiation} = At\,/\,Ae_{1,2,4}$$

where At represents the total area of the Cantonese tone space and $Ae_{1,2,4}$ represents the spread of the three most separated tones in Cantonese (where 1, 2, and 4 represent tones 55, 25, and 21 respectively). The lower the ratio the greater the amount of overlap between ellipses and the less differentiated the tonal space.

Differentiation within Tones: The differentiation among each tone in the tone space is given by the following formula:

$$\text{Within-Tone Variability} = \text{Ave Dist.}/\text{Ave Ax}_{1+2}$$

where Ave Dist represents the distance between the six tone ellipses centres from every other tone centre, and Ave Ax1+2 the length of the axes of each ellipse. This measure indexes how much of the total tone space is occupied by each tone ellipse. The lower the ratio the greater the amount of tone space each tone occupies and the less differentiated is each individual.

Figure 16 shows plots borrowed from Barry and Blamey's (2004) study comparing the tone space of normal hearing adults and children, to children with cochlear implants. Plots Adult and Adult 2 are based on productions from two adult native Cantonese speakers showing tightly contained tone ellipses and the triangular shape of the tonal space in Cantonese as determined by the three most extreme tones (55, 25, and 21). Note that some crowding of the three tones (Tone 33, 23, and 22) is also apparent in the centre of the tone space.
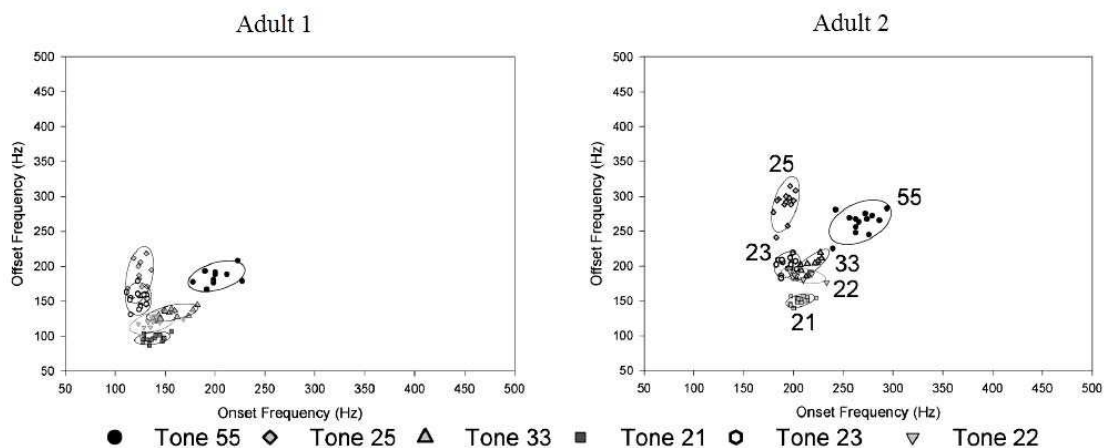
Figure 15. Tone Space of 2 Normal Hearing Adults

Plots Child 1 and Child 2 are from Children with normal hearing illustrating enlarged tone ellipses and reduced between-tone differentiation relative to the tone plots from adult speakers.
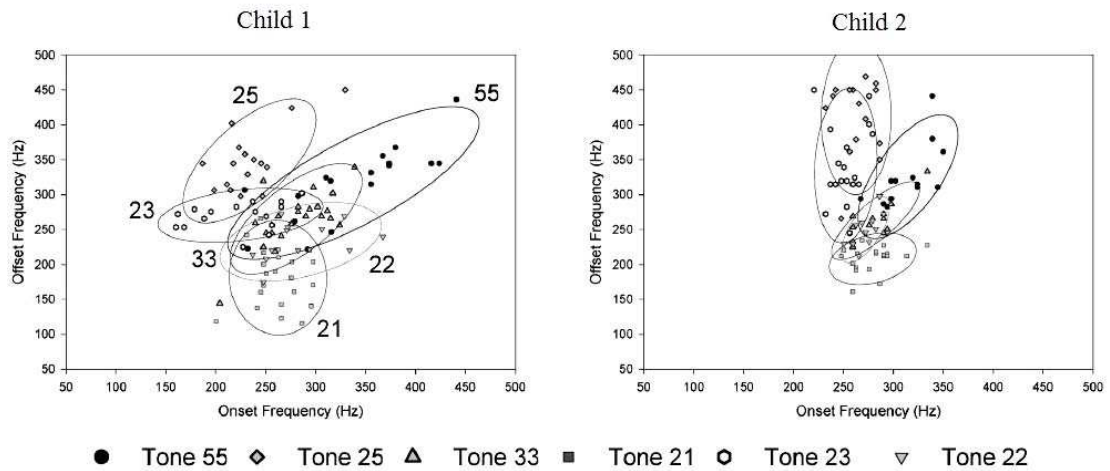
Figure 16. Tone Space of 2 Normal Hearing Children

Plots Cochlear Implant Child 1 and Cochlear Implant Child 2 are from two children with cochlear implants. The first shows centrally concentrated tone ellipses which reflect similar values of F0 onset and offset for most tone tokens regardless of tone type; the ellipses all cluster in one location indicating little tone differentiation. Plot Cochlear Implant Child 2 demonstrates better tone differentiation with more dispersed ellipses than Cochlear Implant Child 1, with the most clearly differentiated tone being tone 55, which appears as a distinct ellipse separated from the other five. This suggests a clear opposition between the features high and not-high for Cantonese tones by this child with cochlear implant (Barry & Blamey, 2004).
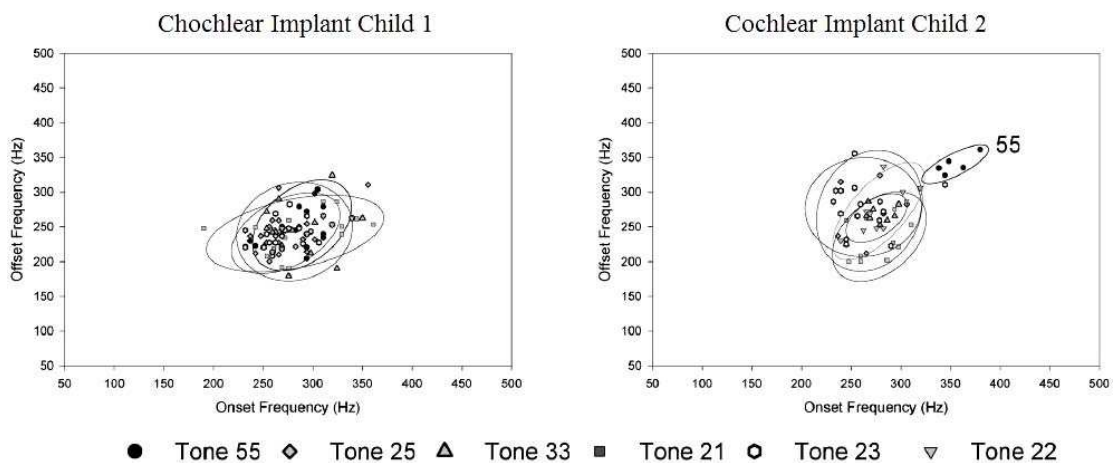
Figure 17. Tone Space of 2 Children with Normal Hearing Impairments

This method of mapping Cantonese tones clearly demonstrates variation of degree of differentiation between tones and the spread of tonal targets in the Cantonese

tone space for individuals with different linguistic experiences and abilities. The consistency of tonal targets and ellipse locations within the Cantonese tone space across speakers of similar experience and ability, suggests that this method is a valid and consistent way of mapping Cantonese tones. More importantly for the current purposes, by locating the centre coordinates of tones 55, 25, and 21, (the most dispersed tones in Cantonese) and calculating the area subtended by that triangle, a single index of Cantonese tone space may be obtained for each speaker. This method can be employed for within-speaker comparisons of tone productions in different speech registers such as infant- and adult-directed speech which are introduced in detail the next chapter.

Three

Chapter 3

Infant Directed Speech – Vowel Hyperarticulation & Exaggerated Pitch Modulations

3.1 What can Infants Hear?

Even as a foetus, the developing infant has a functional auditory system by the beginning of the third trimester (Bredberg, 1985; Querleu, Renard, Versyp, Paris-Delrue, & Crepin, 1988), and is able to hear a variety of sounds, in particular the mother's voice, which is low-pass filtered by the time it reaches the foetus through the womb and via bone conduction (Querleu, Renard, & Versyp, 1981; Querleu et al., 1986). Studies investigating neonates' ability to recognise a variety of speech-related properties have demonstrated that not only can foetuses hear they also have some memory for sounds. For example, using a non-nutritive sucking method, in which infants can control what they hear by their rate of sucking, neonates have been shown to prefer their own mother's voice (DeCasper & Fifer, 1980) and her native language (Mehler et al., 1988). In fact, at birth, infants' ability to discriminate sound properties is quite advanced. For example, Christophe, Dupoux, Bertoncini, and Mehler (1994) found neonates are able to hear word boundaries despite the lack of obvious cues such as silent pauses. However many of their earlier abilities appear to be based, to a large extent on the paralinguistic prosodic properties of language (Cutler & Mehler, 1993), rather than linguistic properties, since both native language preference and word segmentation abilities are lost when speech is played backwards thus disrupting the natural prosody (Ramus, Hauser, Miller, Morris, & Mehler, 2000).

Over and above infants' pre-tuning for paralinguistic suprasegmental information, they are also born with the ability to hear critical aspects of segmental information (consonants, vowels, and tones). During early development, infants can discriminate just about every phonetic contrast across the world's languages, and this includes contrasts that adults can no longer discriminate phonologically. For example, infants raised in an English language environment can discriminate phonologically irrelevant sounds [b] versus [p] (Aslin, Pisoni, Hennessy, & Perey, 1981; Burnham, Earnshaw, & Clark, 1991), the Indian voiceless, unaspirated, retroflex [ʈ] versus dental [t̪] contrast (Werker, Gilbert, Humphrey, & Tees, 1981), and the Native American

Nthlakampx glotallised velar [k̓] and uvular [q̓] contrasts. Their abilities are thought to be psychoacoustically- rather than psycholinguistically-based and the result of particular nature of the human auditory system (Kuhl, 1978; Pisoni, 1977). Speech perception becomes more linguistically based in the second half of the first year when specific language experience, more particularly the lack thereof affects infants' speech perception.   So the initially universal speech perception ability for consonants diminishes appreciably by 8 months onwards (Werker & Tees, 1984a) and most native English speaking adults have great difficulties discriminating among these phonetic contrasts. The same pattern of results is found for vowels, such that younger 4-mont-old infants raised in an English language environment are able to discriminate the German high front-rounded /U/ and /u/ (tense and lax respectively), and the high back-round /Y/ and /y/ (tense and lax) vowels, but by 6 months their ability to discriminate these non-native vowels has decreased significantly (Polka & Werker, 1994). Finally, with respect to tones in a recent study investigating tone discrimination ability in infants from an English language environment, Mattock and Burnham (2006) found that like consonant and vowel segments, non-tone language infants' ability to discriminate tone is also subject to attenuation sometime between 6 and 9 months, and certainly not earlier than 6 months (Mattock et al., 2008). This attenuated ability to discriminate non-native speech contrasts, which begins around the second half of the infant's first year, reflects an attentional effect rather than any sensorineural loss for it has been found that adults can be trained to hear non-native speech sounds (Werker & Tees, 1984b) and that when short inter-stimulus intervals (ISIs) between pairs of speech sounds are given, older infants can discriminate non-native speech sounds (Werker & Logan, 1985). This attentional shift appears to occur around the same time for consonants and tone (6-12 months) but slightly earlier for vowels (around 4-6 months). In addition, the timing of the onset of attenuation differs among consonants, vowels, and tones, with vowels occurring earlier at around 4-6 months, and consonants and tones occurring between 6 and 12 months.

3.2   Speech Input to Infants

3.2.1   *Historical Developments in Infant-Directed Speech Research*

The term termed "motherese", was originally the term usually used to describe that set of linguistic modifications that mothers use while speaking to their infants, however, following findings that fathers (Engle, 1979; Fernald et al., 1989; Golinkoff & Ames, 1979), single women (Ikeda & Masataka, 1999), child care givers (Nwokah, 1987), and even 4-year-old children (Sach & Devin, 1976; Weppelman, Bostow, Schiffer, Elbert-Perez, & Newman, 2003) modify their speech when speaking to infants, the term infant-directed speech (IDS) is a more appropriate term that reflects the pervasive use of this style of speech across the population.

IDS contain grammatical/syntactical, suprasegmental, and segmental characteristics that distinguish it from adult-directed speech (ADS). Early research on IDS was in the descriptive linguistics vein − if focused on aspects such as syntax, linguistic structure, semantics, and mean length of utterance. Later research has focused on the acoustic and phonetic characteristics of IDS, which allows for more basic investigations of IDS, and expositions of the functions these acoustic modifications may serve during early speech development.

3.2.1.1   *Early Descriptive Research*

Descriptive linguistic research has shown that IDS is syntactically less complex and contains more redundancies (Snow, 1972), is less varied and more concrete (Phillips, 1973), and more directed in content to the immediate situation than ADS (Snow & Ferguson, 1977). IDS also contains longer pauses and shorter utterances (Fernald & Simon, 1984), and word-final syllables are more consistently stressed than in ADS (Albin & Echols, 1996). Some investigators believe that taken together, these modified features of speech should allow infants to learn the ambient language more easily (Snow, 1972), but others believe that simply providing a reduction in the complexity of speech alone is insufficient to explain how infants gain linguistic competence (Gleitman, Newport, & Gleitman, 1984; Scarborough & Wyckoff, 1986).

There is also the more basic problem of how infants learn to produce speech sounds in their native language, even before they have mastered phonological, syntactical, and grammatical rules of that language. The next sections shows how later research focusing on the acoustic properties of IDS has addressed these basic bootstrapping issues associated with infant speech acquisition.

3.2.1.2 Later Acoustic Research

Studies examining the acoustic characteristics of IDS have found that it contains higher F0, wider pitch excursions, and more prosodic repetitions than ADS (Fernald, 1989; Fernald & Simon, 1984; Grieser & Kuhl, 1988); and that higher pitch is particularly associated with utterance-final positions and focus words (Fernald & Mazzie, 1991). In addition to higher pitch and more prosodic modifications, IDS also contains modifications at the segmental level namely vowel-hyperarticulation – the stretching of the psychoacoustic space so that vowel tokens are more separated from each other than they are in otherwise equivalent ADS (Burnham et al., 2002; Kuhl et al., 1997; Liu, Kuhl, & Tsao, 2003; Rvachew, Mattock, Polka, & Menard, 2006).

3.2.2 Pervasiveness of Infant –Directed Speech

3.2.2.1 Cross–Cultural Studies

A study examining mothers' and fathers' speech to infants across a range of languages including French, Italian, German, Japanese, British, and American English, found that irrespective of sex and across all the languages studied, IDS contained consistently higher F0, greater frequency variability, shorter utterances, and longer pauses than ADS (Fernald et al., 1989). The same pattern of increased pitch and extended pitch contours has been found in IDS of tone languages such as Mandarin Chinese and Cantonese, as in non-tonal English and German (Grieser & Kuhl, 1988; Papousek & Bethesda, 1987; Papousek, Bornstein, Nuzzo, Papousek, & Symmes, 1990; Papousek & Hwang, 1992; Papousek, Papousek, & Symmes, 1991; Werker, Pegg, & McLeod, 1994). Similarly, pitch contours that convey social meaning are consistently employed across tone and non-tone languages, such that rising melodies have been found to be associated with encouragement, and falling patterns with soothing (Papousek et al., 1991). The same

pattern of results has been found for vowel hyperarticulation, first found by Kuhl et al., (1997) for American English, Russian. Swedish, but not has also been found for Australian English (Burnham et al., 2002) and Mandarin (Liu et al., 2003). However, see Englund and Behne (2005; 2006) for a description of the lack of vowel hyperarticulation in Norwegian IDS during the first 6 months of infancy.

3.2.2.2   *Cross-modal IDS (Infant-Directed Signing):*

In addition to spoken language, there is also evidence for infant-directed (ID) sign in Japanese and American sign languages. Mothers who use Japanese sign were found to use signs at a significantly slower tempo, with more repetitions, and with more exaggerated movements associated with each sign when signing to infants than to adults (Masataka, 1992, 1996). In a similar vein, infant-directed signing in American sign language is characterised by slower sign speed, longer sign duration, and repetitiousness (Holzrichter, 2000).

3.3   Functions of IDS

To summarise thus far: infants are born with the ability to hear fine-grained acoustic detail of just about any speech contrast, and the input they receive is modified at the suprasegmental level with higher pitch and greater prosodic modulations, and at the segmental level with vowel hyperarticulation. These results beg the question: What, if any possible functions do these acoustic modifications in IDS serve in terms of infant development? The following sections will examine specific acoustic modifications of IDS and the possible functions they may serve (albeit unconsciously by the speaker) in infant development.

3.3.1   *Infants Prefer Listening to IDS*

It is one thing to have modified speech input in IDS, but examining whether infants are sensitive to such modifications is of great importance if we are to conclude that there is a relationship between IDS and infant speech development. The prevalence of IDS across languages and linguistic modes has prompted enquiries as to what extent infants are aware of the acoustic content of IDS. In this vein neonates have been found to be

able to discriminate not only between their mother's and a stranger's voice, but also between IDS and ADS (Cooper & Aslin, 1990; Hepper, Scott, & Shahidullah, 1993). While one-month-old infants do not show preference for IDS over ADS for their own mother's voice (Cooper, Abraham, Berman, & Staska, 1997), they do demonstrate, together with neonates, a clear preference for recordings of female strangers' IDS compared with ADS as indexed in an operant auditory preference procedure (Cooper & Aslin, 1990). However, older 4-month-old infants demonstrate preference for IDS, both in their mother's voice and female strangers' voices, as indexed by using the operant auditory preference head-turn procedure (Cooper et al., 1997; Fernald, 1985). The same preference for IDS has also been found for male speakers by 4- to 9-month-old (Werker & McLeod, 1989), and 7-week-old infants (Pegg, Werker, & McLeod, 1992). Cross-language research has also found that both Cantonese and English infants prefer IDS to ADS Cantonese (Werker et al., 1994). Similar preferences have been found across linguistic modes, in that 6-month-old deaf infants show a preference for ID over AD sign as indexed by longer visual fixation times on ID than AD sign videos (Masataka, 1996).

Taken together, these findings suggest that infants' preference for IDS is not simply a preference for the familiarity of their own mother's voice, but one that is consistent across gender of speakers, languages, mode of linguistic communication, and familiarity level of the speaker. However, simple preference is not sufficient to answer the question of whether and what functions IDS may serve in infant development. It is necessary to tease apart specific acoustic modifications in IDS and their possible separate functions in infant development.

3.3.2   *Attention Getting Through Higher Pitch*

Fernald and Kuhl (1987) investigated, by systematically manipulating via computer synthesis, three prosodic characteristics of IDS – F0, amplitude, and word duration (related to the perception of speech rhythm). They found that infants' preference for IDS is based on higher F0 and not on higher amplitude or longer duration. Other studies have found similar results with 4-month-olds (Papousek et al., 1990), while

(Trainor & Zacharias, 1998) found the same-aged infants also prefer higher pitched to lower pitched singing of otherwise identical songs. (Trainor & Desjardins, 2002) tested the impact of higher pitch on infants' vowel perception and found that pitch modification in IDS actually impaired infants' vowel perception. This implies that higher pitch in IDS serve a function other than enhancing linguistic discrimination, for example, gaining infants' attention. Using factor analysis so adults' ratings of low-pass filtered IDS, Kitamura and Burnham (2003; 2002) found greater attentional components of IDS than ADS. Moreover, comparisons of age trends for the attention component of IDS and mean pitch level over the first 12 months of mothers' IDS to their infants, suggested to Kitamura and Burnham that the increased pitch in IDS is related to gaining infants' attention, rather than any linguistic functions.

3.3.3 *Affect Through Prosody: Socio-Emotional Function of IDS Speech*

Other research has found that it is not simply pitch height that infants prefer but increased frequency modulations (Kaplan, Goldstein, Huckeby, Owren, & Cooper, 1995). However, Cooper and Aslin (1994) found that 1-month-old infants did not prefer low-pass filtered IDS – lacking linguistic information but with intact frequency modulation – over ADS, but did show preference for unfiltered IDS over ADS. In a similar study, Colombo and Horowitz (1986) found infants discriminated but did not show a preference for frequency-modulated sweeps that resemble IDS over those that resemble ADS. These results suggest that infants' preference for IDS may be much more complex than first imagined.

Kitamura and Burnham (1998) found that when pitch characteristics – mean pitch and pitch range – are held constant, infants prefer to listen to IDS that expresses higher rather than lower vocal affect, but when affect is held constant, infants do not display preference for high over low mean pitch in speech. This suggests that the critical aspect in capturing infants' attention is affect rather than pitch characteristics alone (Fernald & Kuhl, 1987). These results are not inconsistent with studies showing pitch-based preferences, but rather imply that heightened pitch may act as a vehicle for heightened affect. In a similar vein, Singh Morgan, and Best (2002) found 6-month-old

infants preferred "happy speech " to speech with less positive affect, regardless of whether it was IDS or ADS; and ratings of affective categories for emotional ADS are equivalent to those in IDS, suggesting that prosodic modifications in IDS are related to the communication of affect (Trainor, Austin, & Desjardins, 2000). Other studies with 6-month-old infants have found that they are able to distinguish between prosodic speech that represents comforting versus approving utterances, even in low-pass filtered speech (Moore, Spence, & Katz, 1997). Consistent with these findings, 5-month-old English, German, and Italian language environment infants were found better able to discriminate affective intent in IDS than in ADS, and preferred speech expressing approval over that expressing disapproval (Fernald, 1993); younger 4-month-old infants also prefer acoustic patterns typical of "approval" than "disapproval" (Papousek et al., 1990); and even 2-month-old infants prefer rising contours associated with approval, than falling contours associated with disapproval (Sullivan & Horowitz, 1983).

Taken together these findings suggest that for infants, at least the older 5- to 6-months-olds in the Kitamura and Burnham (1998) and Fernald (1993) studies, it is the affective intent conveyed by mothers' speech, and carried through intonation patterns of this special speech style that is most salient to them. This would seem to imply that infants may develop awareness for social implications of speech very early in infancy. In support of this notion Werker and McLeod (1989), found that adults rated 4- to 9-month-old infants in mute video recordings to be more friendly and cuddly when the infants were listening to IDS than to ADS. Moreover, adults' ratings and categorisations on scales of communicative intent of ADS and of IDS directed to 7- to 9-month-olds, and 12-month-olds, revealed that affect was significantly more consistently rated for IDS than ADS conditions (Fernald, 1989; Trainor et al., 2000), suggesting IDS more accurately communicates an emotional message to infants than does ADS. Additionally, it has been found that over and above acoustic cues in IDS, and sign cues in ID sign, there are accompanying gestural cues (Brand, Baldwin, & Ashburn, 2002; Iverson, Capirci, Longobardi, & Cristina Caselli, 1999) and facial expressions (Chong, Werker, Russell, & Carroll, 2003), that accompany and perhaps reinforce or augment the spoken message, not only in terms of the linguistic content, but

also in terms of the social and emotional message. There is also evidence to suggest a universality of the communication of affect in infant-directed communication across linguistic modes. In a study involving 45 six-month-old hearing infants watching ID and AD sign, hearing infants showed greater affective responsiveness to ID sign than to AD sign (Masataka, 1998).

3.3.4 *IDS and General Learning*

Aside from the attentional and socio-emotional function of IDS, many studies have also linked IDS to general learning in infancy. Four-month-old infants of clinically depressed mothers – who show less modulation of fundamental frequency in their IDS – failed to learn simple association tasks such as the operant auditory preference procedure when they listened to their own mother's IDS, but were able to learn the tasks when listening to the speech of a non-depressed mother (Kaplan, Bachorowski, Smoski, & Hudenko, 2002). However, older 5- to 13-month-olds could only learn such tasks by listening to a non-depressed father's speech (Kaplan, Dungan, & Zinser, 2004) suggesting that the inability to learn from speech of their depressed motheres had now generalised to all female voices. These results illustrate the importance of IDS in general learning as well as infant's ability to tune out irrelevant or non-salient input. While this is analogous to the tuning out of irrelevant non-native segmental information in this case with depressed mothers' speech the selective attention has negative rather than positive consequences.

3.3.5 *Linguistic Function of IDS Speech*

Studies examining linguistic features of IDS have found that infants learn to identify word boundaries when listening to IDS but not ADS (Thiessen, Hill, & Saffran, 2005); and it has been found that infants prefer IDS containing natural as opposed to unnatural pauses, and indeed when IDS and ADS are matched for natural pauses no IDS over ADS preference is found, suggesting that clear clausal boundaries in IDS are detectable by infants and may possibly assist them to learn the syntactic structure of speech (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). In a similar vein Chinese adults better recognised Chinese words when presented in IDS than ADS, presumably

because IDS is characterised by utterance-final placements with vowel lengthening, and frequency peaks that appear to aid word recognition (Golinkoff & Alioto, 1995).

Other studies examining the didactic nature of IDS found feedback from infants may influence mothers' speech to infants. Comparing mothers speaking to their actual infant and their imagined infant, the full range of speech modifications that characterise IDS were only evident in the infant-present condition: Fernald and Simon (1984), and Bohannon III and Marquis (1977) suggested that the presence of the infant is both necessary and sufficient to elicit IDS in carers. Early descriptive linguistic research showed that feedback from children changed both their mothers' and unfamiliar adults' mean length of utterance in IDS (Bohannon III & Marquis, 1977), suggesting that children may actually control the complexity of linguistic input from adults through the feedback they provide, and adults change (albeit unconsciously) their linguistic input as a function of infants' developing linguistic abilities. Following this developmental bent, Kitamura and Burnham (2003) compared pitch in IDS directed to 0-, 3-, 6-, 9-, and 12-month-old infants, and found significantly higher pitch in IDS to 3- and 6-month-olds, than 9- and 12-month-olds. This reaction in pitch at the latter being a time when infants' language perception is becoming more adult-like, due to their attunement to the native language environment (Tees & Werker, 1984; Werker & Tees, 1984a) would seem to herald a time of transition in which affective elements of IDS are being subjugated to more linguistic or didactic elements.

Not all studies have found positive linguistic benefits of IDS, for example see Furrow and Nelson (1986), and Gleitman et al., (1984) who proposed that a simple reduction in the complexity of speech input to infants would not explain how infants gain linguistic competence, involving the complexities typical of ADS. In addition Trainor and Desjardins (2002) show how higher pitch in IDS can hinder linguistic learning, and Ellen and Anderson (1983) demonstrated that word redundancy in IDS leads to poorer recognition of words by infants. While these studies and early reviews (see Newport, Gleitman, & Gleitman, 1977; Scarborough & Wyckoff, 1986) suggest that IDS may not positively influence linguistic development, more recently evidence

above, especially across different ages suggests that IDS has multiple functions and that the relative weights of these functions change over development.

3.3.5.1   Hyperarticulation: Mechanism for Infant Speech Development
Recent acoustic research identifies a phenomenon in IDS that may well be a basic mechanism that bootstraps infants' language acquisition. Kuhl et al., (1997) examined natural language input to infants in American English, Russian, and Swedish, and found that across all languages, when speaking to infants compared to adults, mothers produced vowels that were acoustically more extreme, resulting in an acoustically expanded vowel space. This effect is schematically represented in Figure 18 showing Kuhl et al.'s (1997) original findings of vowels hyperarticulation.
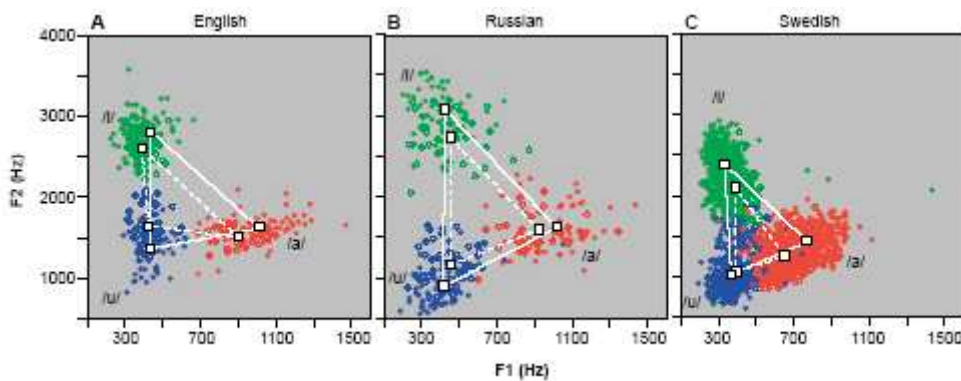
Figure 18.   Vowel space of IDS versus ADS in English, Russian, and Swedish plotted by F1 and F2 (Kuhl et al., 1997)

As can be seen in Figure 18 vowel triangles are created by joining the "point" (or corner) vowels /i/, /a/, and /u/, so called because they are considered to be the most extreme vowels in the acoustic vowel space of most languages. Each data point represents the coordinates of the first two formant frequencies of a vowel. A universal stretching of the vowel triangle can be seen in IDS (solid line) relative to ADS (dashed line), i.e., a larger triangle area for IDS than ADS, in all these languages, American English, Russian, and Swedish.

In vowel hyperarticulation, mothers do not simply raise formant frequencies when speaking to their infants, rather formant frequencies are selectively increased or decreased to achieve an expansion of the acoustic space encompassed by the three point vowels (/i/, /u/, and /a/). The degree of hyperarticulation is indicated by the size of the

area of the vowel triangle subtended by the corner vowels. Expanded vowel triangles reflect an increase in the acoustic distance between vowels, making them more distinct from one another with less acoustic overlap between vowel categories (Kuhl et al., 1997), which should allow vowels to be more easily discriminated and possibly more easily learned.

Since Kuhl's original study, further studies have found vowel hyperarticulation. Examining Australian English (Burnham et al., 2002), the pitch-accented language Japanese (Andruski, Kuhl, & Hayashi, 1999), and the tone language Mandarin (Liu et al., 2003) have also found vowel hyperarticulation in IDS compared with ADS. More recently, evidence have emerged that consonants are hyperarticulated in Norwegian IDS, in that voice onset times and fricative durations are longer for stop consonants and fricatives in IDS than in ADS (Englund, 2005; Englund & Behne, 2006). Even across linguistic modes there is evidence for sign hyperarticulation, in that mothers who use Japanese sign with their infants significantly exaggerated signs, as indexed by the angle of the hand and elbow subtended to the sagittal plane while signing, an effect akin to vowel hyperarticulation (Masataka, 1992, 1996).

Another line of research in vowel hyperarticulation investigates the long recognised uncanny similarity between IDS and pet-directed speech (PDS), and provides further support for the notion that vowel hyperarticulation is an agent of infant speech acquisition. Both IDS and PDS have syntactically simple structure and short sentences (Hirsh-Pasek & Treiman, 1982; Mitchell, 2000; Mitchell & Edmonson, 1999). Moreover, regardless of whether people are speaking to their own or someone else's infant or dog, the prosodic modifications of both IDS and dog-directed speech serve to gain attention, and communicate affection and conversational meaning (Mitchell, 2000; Mitchell & Edmonson, 1999). To ascertain whether vowel hyperarticulation is unique to IDS Burnham et al., (2002) compared IDS, ADS, and PDS. Their results indicate that although both IDS and PDS both contain higher pitch and affective information than ADS, only IDS, and not PDS, has vowel hyperarticulation compared with ADS. This suggests that not only is hyperarticulation specific to IDS, but that speakers consciously or unconsciously adjust their speech

depending on the ability of the recipient to learn to speak (Burnham et al., 2002). In a similar line of investigation, Uther, Knoll, & Burnham (2007) investigated IDS, ADS, and speech directed to foreigners, foreigner directed speech (FDS). They found that adult British English speakers showed increased pitch and affective content in IDS compared to ADS, but no such elevations in FDS. On the other hand, vowel hyperarticulation was produced in both IDS and FDS, both being styles of speech directed to recipients with the potential to learn the ambient language. Together, these results further strengthen the possibility that vowel hyperarticulation, in IDS, is a mechanism that promotes speech acquisition in infancy.

To further and more directly investigate the didactic role that vowel hyperarticulation may play in infant acquisition of speech, Liu et al. (2003) computed correlations between the degree of individual mothers' vowel hyperarticulation in IDS and their infants' native consonant discrimination. Liu et al. (2003) found a positive correlation between vowel hyperarticulation in the mothers' speech and their infants' discrimination ability. Their findings (although correlational) strongly suggest that hyperarticulation is implicated in the development of infants' vowel and consonant categories, and more generally language acquisition. Moreover, as vowel hyperarticulation appears to occur in a broad spectrum of languages, it is possible that vowel hyperarticulation is a universal linguistic mechanism that may promote language learning, and even that vowel hyperarticulation is a particular instance of a range of hyperartiuclations (of vowels, consonants, and even tones) that may be present in speech to infants.

3.4   Tone Hyperarticulation?

Although investigations revealed IDS hyperarticulation in tone and non-tone languages alike, it is not known if a similar phenomenon occurs for tones in tone language IDS. Does tone hyperarticulation occur, and if so, is it a similar phenomenon to vowel hyperarticulation? Grieser and Kuhl (1988) compared Mandarin IDS to English and German IDS, and found that in all cases IDS had significantly higher F0, exhibited larger F0 range over the entire sample, and larger F0 range per phrase than ADS.

Similarly Papousek and Hwang (1992) found that when native Mandarin speakers were asked to produce IDS, they increased peak and minimum F0, reduced the rate of F0 fluctuations and, in foreign language instruction, also expanded F0 patterns and F0 range in comparison to ADS. Kitamura et al., (2002) found in Thai that pitch modifications occurred in IDS, even at the expense of speech integrity. Finally, a recent study found evidence for the exaggeration of pitch in the four Mandarin tones in IDS but not ADS, and for differences in pitch height, F0 range, and duration (Liu, Tsao, & Kuhl, 2007). While there is still no research that examines the expansion of tone space for IDS, studies thus far on IDS suggest that, despite lexical tone restrictions, tone language speakers are able to modify the pitch quality of their speech directed to infants, and such F0 changes may transfer to tone space expansion in IDS compared to ADS.

With the advent of the tone space mapping method developed by Barry and Blamey (2004), it is now possible to investigate the possibility of tone hyperarticulation in IDS, and in parallel with the Liu et al., (2003) study using vowel hyperarticulation, the exposition of any relationship between any tone hyperarticulation, if present, and infants' ability to discriminate native language segmental speech sounds will also be examined in a preliminary investigation.

Four

Chapter 4

Vowel & Tone Hyperarticulation via IDS

4.1 Introduction: Investigating Tone and Vowel hyperarticulation in Cantonese IDS

There is now quite extensive research on infant-directed speech (IDS) and vowel hyperarticulation in IDS both for tone languages such as Mandarin (Liu et al., 2003), and non-tone languages such as American English, Russian, Swedish (Kuhl et al., 1997), and Australian English (Burnham et al., 2002). Vowel hyperarticulation in mothers' IDS has been found to be related to their infants' speech perception, in that mothers who hyperarticulate their vowels more have infants who are better able to discriminate native consonant contrasts (Liu et al., 2003). Studies comparing IDS with other speech registers have found that despite many parallels between IDS and pet-directed speech (PDS), only IDS has hyperarticulated vowels (Burnham et al., 2002); whereas, there is vowel hyperarticulation in both IDS and foreigner-directed speech (FDS) (Uther et al., 2007). Together these results suggest that vowel hyperarticulation may be related to the speaker's expectation of the target audience's ability or need to acquire the ambient language, or at least its phonology.

While vowel hyperarticulation in IDS has been investigated across a variety of languages, both tonal and non-tonal, the same cannot be said for tones. Considering that the majority of the world's languages are tonal (Yip, 2002), this lack of research on tone in IDS is unfortunate. Over and above the, until recently, relative Anglocentricity of this research, part of the reason for this may be the lack of a method to measure tone hyperarticulation that parallels vowel hyperarticulation analysis. Vowel hyperarticulation is measured by plotting the F1 and F2 values of the three corner vowels of a given language, joining the centroids of these vowels, and calculating the resultant triangle area as a measure of vowel space. If the vowel triangle area is larger for a particular special speech register than a referent or baseline register (usually adult-directed speech – ADS) then that register is said to show vowel hyperarticulation.

A method developed recently by Barry and Blamey (2004) can be used for plotting tones, in an analogous fashion to that used for vowels: tones are plotted in terms of heir F0 onset and F0 offset and then the corner tones or tones that are most dispersed in a particular tone language (e.g., tone 55, 25, and 21 for Cantonese) can be

78

connected to form a triangle in the same way that the corner vowels form a vowel triangle. The area of such a tone triangle can be used as an index of a given speaker's tone space, and in turn along with measures of tone space in different speech registers, tone hyper- or hypo-articulation can be determined

The first aim of this study is to examine both vowel hyperarticulation and tone hyperarticulation in Cantonese IDS. A recent study investigating Mandarin IDS has found evidence for exaggeration of the four Mandarin tones in IDS compared to ADS in terms of pitch height, F0 range, and duration (Liu et al., 2007). Given this, plus evidence suggesting that tone hyperarticulation occurs in IDS in a tone language (Liu et al., 2002), along with an appropriate tone space measurement procedure (Barry & Blamey, 2004), the time has come for a study that examines tone space and tone hyperarticulation in IDS. If vowel hyperarticulation in IDS does indeed have a linguistic function, then tones should also be hyperarticulated in tone language IDS, and such hyperarticulations should be evident via similar means – by comparison of tone triangle areas in IDS and ADS.

The second aim of this study is to investigate both vowel and tone production in Cantonese mothers' IDS longitudinally across the first 12 months of their infants' development, and track patterns of change over that period. There is developmental evidence to suggest that infants move from language-general to language-specific modes of speech discrimination, such that they are able to discriminate most of the world's speech contrasts at birth, but between 4 and 12 months of age this ability is attenuated for speech contrasts not present in the ambient language (Mattock & Burnham, 2006; Polka & Werker, 1994; Rvachew et al., 2006; Werker & Tees, 1984a). Despite this evidence for developmental reorganisation of speech perception mainly across the second half of the infant's first year, most IDS research has focused on speech directed to infants at a single point in their development, usually at 6 months. Thus, there has been little investigation of changes, if any, in mothers' speech production as a function of their infants' language development. Research investigating such changes is necessary if a full understanding of infants' linguistic development and the role that speech input plays in this is to be gained.

With these aims in mind, in this study Cantonese mothers' speech to their infants when they are 3, 6, 9, and 12 months of age, along with a reference register of ADS is collected. The study examines several hypotheses. Firstly, it is hypothesised that there will be tone hyperarticulation in Cantonese IDS at some point during the infants' first 12 months since no previous results are available to make specific predictions about when tone hyperarticulation should occur. Secondly, it is hypothesised there will be vowel hyperarticulation in Cantonese IDS and this will be most likely to occur for speech to 6-month-olds (mo) and as vowel hyperarticulation had previously been found in 6moIDS (Burnham et al., 2002; Kuhl et al., 1997; & Liu et al., 2004). Finally, it is hypothesised that there will be differences in the degree of hyperarticulation in 3-, 6-, 9-, and 12-month-old IDS for both tones and vowels, in that tones should peak at 9 months and vowels at 6 months in parallel with language specific developmental changes in infant speech perception in which perceptual reorganisation for tones (between 6 and 9 months) appears to occur later than for vowels (between 4 and 6 months; Polka & Werker, 1994).

4.2   Method

The procedures and analyses for vowel and tone analyses adopted for this study closely follow those implemented by previous studies (Barry & Blamey, 2004; Liu et al., 2003) and are outlined in the following sections.

4.2.1   *Design and Participants*

A longitudinal cohort sequential design was used. A total of 25 native Cantonese-speaking mothers and their infants were recruited from ethnic newspaper advertisements, and by referrals from ethnic health workers working in South-Western Sydney Area Health (see Appendices A.1 & A.2). These mothers and their infants formed two cohort groups, and each cohort participated in three-recording sessions with three monthly intervals between each session. The first cohort group of mothers (N = 11) had infants who were 3 months old at the start of the study, while the second cohort group of mothers (N = 11) had infants who were 6 months old at the start of the study.

Not included in the final sample are three mothers and their infants who could not continue after the initial recording session.

The two groups will be referred to as the 3-6-9 cohort and the 6-9-12 cohort because in the 3-6-9 cohort mothers' IDS was recorded when their infant was 3 months, 6 months, and 9 months old, and in the 6-9-12 cohort at 6, 9, and 12 months old. In the 3-6-9 cohort there were 5 mothers of boys and 6 mothers of girls and in the 6-9-12 group, 7 mothers of boys and 4 mothers of girls. Further details of age at testing are giving in the procedure.

4.2.2 *Equipment and Materials*

The recordings were collected using the unidirectional lapel microphone from a Sony DAT digital recorder (TCD-D100) with sampling rate set to 48kHz and recording level fixed (level 7) to accommodate relatively loud female speakers. Once collected, recordings were transferred to computer and normalised to 90% of amplitude using the Cool Edit 2000 speech analysis software. Formant and pitch analyses were conducted using the speech analysis software, Praat (Version 4.1.2) (Boersma & Weenink, 2008), and Praat scripts adapted from original scripts by (Lennes, 2002) for excising formants and pitch files from long speech files, as well as for extracting formant and pitch information from (see Appendices B.1 & B.2). The statistical analyses were conducted in PSY (Version 5.1.2) (Bird, Hadzi-Pavlovic, & Isaac, 1990).

Infant Details (see Appendix C.3) and Language Exposure Questionnaire (Appendix C.4) forms were sent out to mothers to obtain details relating to their infant's age, general health including hearing, degree of exposure to different languages, the mother's language background, and dominant language(s) used in the home and to the infant. An information sheet (Appendix C.1) was given to mothers at the beginning of the initial recording session and consent for mothers' and infants' participation in the study was obtained (Appendix C.2).

Several toys were used as aids to elicit particular target words (see Table 5) during recording sessions. These included a lion, a plastic number four, a clock, a coaster of a picture of Sydney, various plastic toys made to look like food items in a

container, a butterfly, a plastic toy made to look like poached eggs, and a rattle snake (Appendix D).

4.2.3 *Stimuli*

A total of nine target words were incorporated into the study, six for eliciting the six Cantonese tones: 55, 21, 25, 23, 33, and 22 in similar phonetic contexts; and for each of the three corner vowels in similar phonetic contexts: /a/, /i/, and /u/. All 6 Cantonese tones were included in the study in order to investigate the acoustic properties of all Cantonese tones, since so little is known about them in IDS. In order to elicit these target words during recording sessions appropriate toys were given to the mother (see pictures of these in Appendix D). The target vowels and tones, the target words in IPA and the logographic form along with their meaning, and the toys used to elicit the words are shown in Table 5.

Table 5. Vowels and Tones in Target Words

Phone	Word	Logograph	Gloss	Toy
[a]	[daːn]	蛋	Egg	Plastic toy poached egg
[i]	[dip]	蝶	Butterfly	Toy butterfly
[ü]	[dük]	毒	Poison	Toy rattle snake
Tone 55	[si]	獅	Lion	Toy lion
Tone 25	[sei]	死	Dead	Toy rattle snake
Tone 33	[sei]	四	Four	Plastic number four
Tone 21	[si]	时	Time	Clock
Tone 23	[si]	市	City	Coaster of a picture of Sydney
Tone 22	[sʰik]	食	Food	Plastic you food items in a container

To ensure that the words carrying the six tones were appropriate for use with infants while maintaining consistency in spectral and formant structure, the diphthong [ei] was used instead of [i] for some of the tones (see Table 5). The F1 and F2 structures are similar for [ei] and [i] and the transition from [e] to [i] in [ei] is rapid with a relatively short duration of [e] compared to [i] (Zee, 1999). Different dVC ([d]-vowel-consonant) contexts were used for the vowel carrier word. The stop consonants [p] and [k] in [dip] and [dük] are less problematic than the nasal [n] in [daːn], which

tends to lower the preceding vowel formants. However as [a] is the preceding vowel and as this already has low F1 and F2 values, this is unlikely to have any impact in terms of the purpose of this study.

In order to facilitate mothers' use of the nine Cantonese words, each word's logograph was written in black ink on pieces of white paper and were attached to the toys used to elicit the target words during IDS recordings (see Appendix D).

4.2.4 *Procedure*

Mothers and their infants were recruited via brochures (see Appendix A.1) sent out to local Early Childhood centres, advertisements (see Appendix A.2) placed in Chinese newspapers and via the assistance of ethnic community nurses. Interested mothers could telephone a number with a recorded message by a native Cantonese speaker asking them to leave their details for a return phone call. This native speaker then made the first return call, but since all participants could also speak either competent Mandarin or English, the Mandarin/English bilingual principal investigator then made subsequent contacts. Prior to the initial recoding session, the study and relevant procedural details were explained to the participants via telephone, and questions regarding the study answered. Relevant information regarding the infant and the mother were obtained via an Infant Details and Language Background Questionnaire form (see Appendix C.4), which was posted to mothers and then collected at the initial IDS recording session.

All recordings were collected in the mother and infant's home. On the first visit mothers were provided with further details about the study including an information sheet (Appendix C.1), and were given another opportunity to ask questions about the study. A guardian consent form (Appendix C.2) was explained to the mothers, and signed by them for their infant's participation in the study. For the recordings, mothers were given toys labelled with the target words, and instructed to use these words while speaking to their babies as naturally as possible for about 10 mins.

For the 3-6-9 cohort, the first data were collected when the infants were 3 months (mean = 92 days), the second at 6 months (mean = 186 days) and the final at 9

months (mean = 284 days). For the 6-9-12 cohort, recordings were collected at 6 months (mean = 183 days), 9 months (mean = 281 days) and 12 months (mean = 359 days). At each recording age, multiple visits were required in some mother-infant dyads. Nevertheless, a strict 14-day recording window was adhered to at each IDS recording interval.

4.2.4.1   Cantonese Language Exposure

The Language Exposure Questionnaire (see 4.1.4 and Appendix C.4) asked participants to provide information on their native language, language use in everyday life (at home and at work), dominant language(s), number of languages used, relative proficiency in native and English language, and primary language input to their infant. The questionnaire also obtained detailed information on language(s) used with and proportion of time exposure of the infant to each language (Cantonese, English, Other) from carers, siblings, and other family members in the form of spoken language, and also in television programs and music. Table 6 lists the numbers of mothers speaking each language, and Table 7 lists language input to infants from mothers and other family members.

Table 6. Languages spoken and used by mothers with and without their infant with each cohort 3-6-9 (n=12) and 6-9-12 (n=12) reported separately

		Native		Preferred		Tone	
		3-6-9 Cohort	6-9-12 Cohort	3-6-9 Cohort	6-9-12 Cohort	3-6-9 Cohort	6-9-12 Cohort
Cantonese	Guang Dong	7	6	7	5		
	Hong Kong	2	3	2	3		
	Provincial	2	2	1	2		
Other Chinese	Kejia						1
	Min			1		1	
	Mandarin				1	9	8
Other Tone Languages	Vietnamese					1	1
Other Non-Tone Language	English					3	7

All 22 respondents were native Cantonese speakers. Of these 13 spoke the Mainland Guangdong dialect, four spoke a provincial form of this dialect, five spoke the Hong Kong dialect, and the remaining two were born in Vietnam but used Cantonese as their first language. All mothers spoke at least one tonal language with three mothers competent in three tonal languages – including Cantonese, Mandarin, Kejia, Hainan dialect of Min, and Vietnamese, 16 mothers competent in two tonal languages – Cantonese and Mandarin or Vietnamese, and the remaining three mothers competent in one tonal language – the Hong Kong dialect of Cantonese. All 22 respondents reported that English was not their dominant/preferred language of use, with most mothers nominating Cantonese as their dominant/preferred language, and two mothers nominating another tone language – Hainan dialect of Min or Mandarin. All mothers reported using only Cantonese with their infants. However, 10 mothers reported that their infants were regularly exposed to other languages by other family members, including grandparents and siblings. Of these, three reported English as the only other language used, three reported one other tonal language (Mandarin), three reported a combination of English and one other tone language (Mandarin), and one reported two additional tonal languages (Mandarin & Hainan dialect of Min).

Table 7. Language input to infant by their mothers and other family members with each cohort 3-6-9 and 6-9-12 reported separately.

		Mum to Infant		Others to Infant	
		3-6-9 Cohort	6-9-12 Cohort	3-6-9 Cohort	6-9-12 Cohort
Cantonese	Guang Dong	7	6		
	Hong Kong	2	3		
	Provincial	2	2		
Other Chinese	Kejia				
	Min			1	
	Mandarin			2	4
Other Non-Tone Language	English			2	3

4.3   Acoustic Analyses

The acoustic analysis for vowels and tones differ on the parameters that were extracted, as well as the methods of analysis.

4.3.1   *Tone Analysis*

Pitch information for tone analysis was also obtained using Praat, with F0 onset and offset marked for each tone using a text-grid. As in Barry and Blamey (2004), F0 onset was identified as the point at which the vowel amplitude reached a post-consonantal maximum, while F0 offset was identified as the point at which the vowel amplitude declined to 50% of the maximum. This was easier to determine for target words in sentence final positions and in citation forms, positions and forms which more frequently occurred in IDS than in ADS. However, on the other hand, measuring IDS was more difficult due to the greater interaction between tones and the exaggerated prosodic contours that occur in IDS compared with ADS. Similar to the steps used in vowel analysis, each tone was marked then excised and pitch information extracted using a Praat script (Appendix B.1) which exported F0 onset and offset frequencies into an Excel file.

Following the method used by Barry and Blamey (2004), the tone triangle area was found by first calculating the mean F0 onset and F0 offset for the most dispersed tones in Cantonese, tones 55, 25, and 21. A triangle based on the distance between each tone's mean F0 onset and offset coordinates was obtained, and the area of the resultant triangle was calculated using Heron's formula: $A= \sqrt{s(s-a)(s-b)(s-c)}$, where s=(a+b+c)/2, and a, b and c denote the distance between centroids (mean F0 onset and F0 offset) of tones 55 and 25, 55 and 21, and 25 and 21. This tone triangle area is used as a measure of a given speaker's Cantonese tone space in a particular speech register.
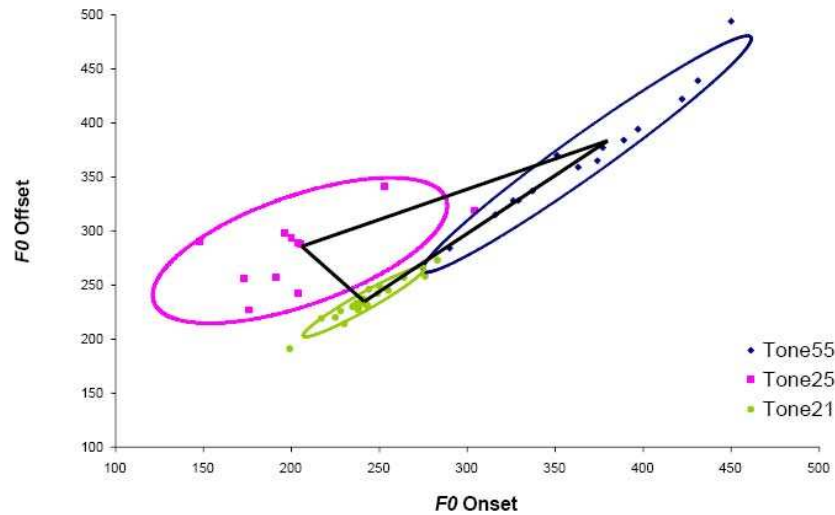
Figure 19. Example of a Cantonese speaker's Tone Triangle

4.3.2 *Vowel Analysis*

The vowel analysis was conducted using Praat software. Each vowel was labelled using a text-grid; the onset and offset of vowels were marked at the first and final periodic cycle respectively. Each vowel was then excised from the original long sound file by application of a Praat script (see Appendix B.1) and a second script (see Appendix B.2) was used to extract formant information including F1, F2 and F3 midpoints and means from the excised files and subsequent exportation into a Microsoft Excel file (Lennes, 2002). The F1 and F2 midpoint values are normally used for monophthongs as the midpoint is most likely to be the most steady-state part of the vowel. However, since some of the tokens used in this study contain diphthongs (which change from one vowel quality to the next at different points into the vowel for different diphthongs) the midpoint will not provide very representative F1 and F2 values. Therefore the mean F1 and F2 values were used for the diphthongs in this study. Each vowel's F1 and F2 values at each recording time interval (3, 6, 9, and 12 months, and ADS) were transformed into perceptual Mel units in Microsoft Excel using the formula: 3321.93*LOG10(1+X/1000)/LOG10(10). The vowel triangle areas were obtained in Excel using the following formula: ABS[F1i*(F2a-F2u)+F1a*(F2u-F2i)+F1u*(F2i-F2a)]/2 (where ABS denotes absolute values, F1 and F2 denote first and second formants, and a, i, and u, the /a/, /i/, and /u/ vowels), as in the Burnham et al., (2002), Kuhl et al., (1997), and Liu et al., (2003) studies. The triangle area created by the black

87

lines in Figure 20 below illustrates a vowel triangle formed by joining the centroids of /a/, /i/, and /u/ (mean F1 and F2) and are vowel productions in F1 by F2 vowel space for a given Cantonese speaker's 6- month-old IDS.
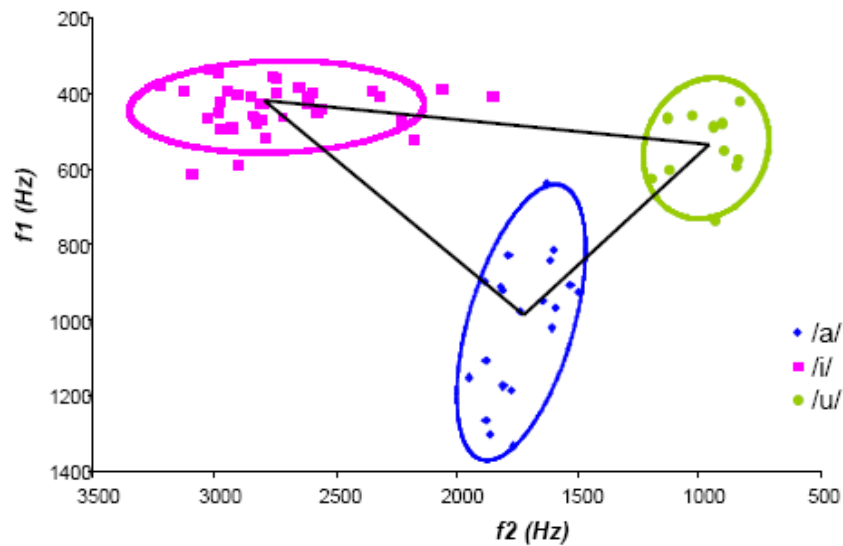
Figure 20. Example of a Cantonese speaker's Vowel Triangle

4.4 Results

The results for tone and vowel analyses will be presented separately. The cohort design did not allow full factorial analysis over the four infant ages – 3, 6, 9, and 12 months – as conducting a factorial design would entail the first, 3 months, and the last, 12 months, cell having half as many participants as the central two ages, 6 and 9 months, and the ADS. So data for the two cohorts (3-6-9 and 6-9-12) were analysed separately in the PSY program (Bird et al., 1990) developed for within subjects, repeated measures planned contrasts using confidence intervals (see Supplementary CD for results and coefficients). The raw data for individual tone and vowel triangle areas contained very large digits and were square-root transformed prior to running the analyses to reduce the effect of extreme scores, while still maintaining the relationship among scores (Field, 2005; Howell, 2002).

4.4.1 *Tone Analysis*

4.4.1.1 *Test of Tone Hyperarticulation Hypothesis*

The first analysis tests the first hypothesis that there will be tone hyperarticulation in Cantonese IDS at some point during the infants' first 12 months since no previous results are available to make specific predictions about when tone hyperarticulation should happen. Tone triangle areas over speech registers for each cohort group are shown in Figure 21.

The tone triangle area data from the two cohort groups 3-6-9 and 6-9-12 were analysed separately using planned comparisons to test the hypothesis. Mean tone triangle area data are shown in Figure 22. The tests of assumptions were satisfactory with no outliers identified and was alpha set at .05. The age effect was significant for both groups. For the 3-6-9 group the combined IDS versus ADS was significant ($F(1, 10) = 19.771$, Confidence Intervals (CIs) = 6.55 - 9.72) suggesting that tone triangle areas in IDS combined over the ages is larger than in ADS. The comparisons between 6-month-old (mo) IDS versus 3 and 9moIDS, and for 3 versus 9moIDS showed no significant results suggesting that despite the hypothesis tone triangle area was not larger in 6moIDS and that tone triangle areas are not significantly different in 3 and 9moIDS.

For the 6-9-12 group the combined IDS versus ADS was significant ($F(1, 10) = 5.23$, CIs = -13.18 - -.17) showing that tone triangle areas in combined IDS speech are larger than in ADS. The comparisons between 6mo IDS versus 9 and 12mo IDS, and for 9 versus 12mo IDS showed no significant results suggesting that tone triangle area was equivalent in 6, 9, and 12mo IDS.

3-6-9 Cohort   **6-9-12 Cohort**

3mo

6mo

9mo

12mo
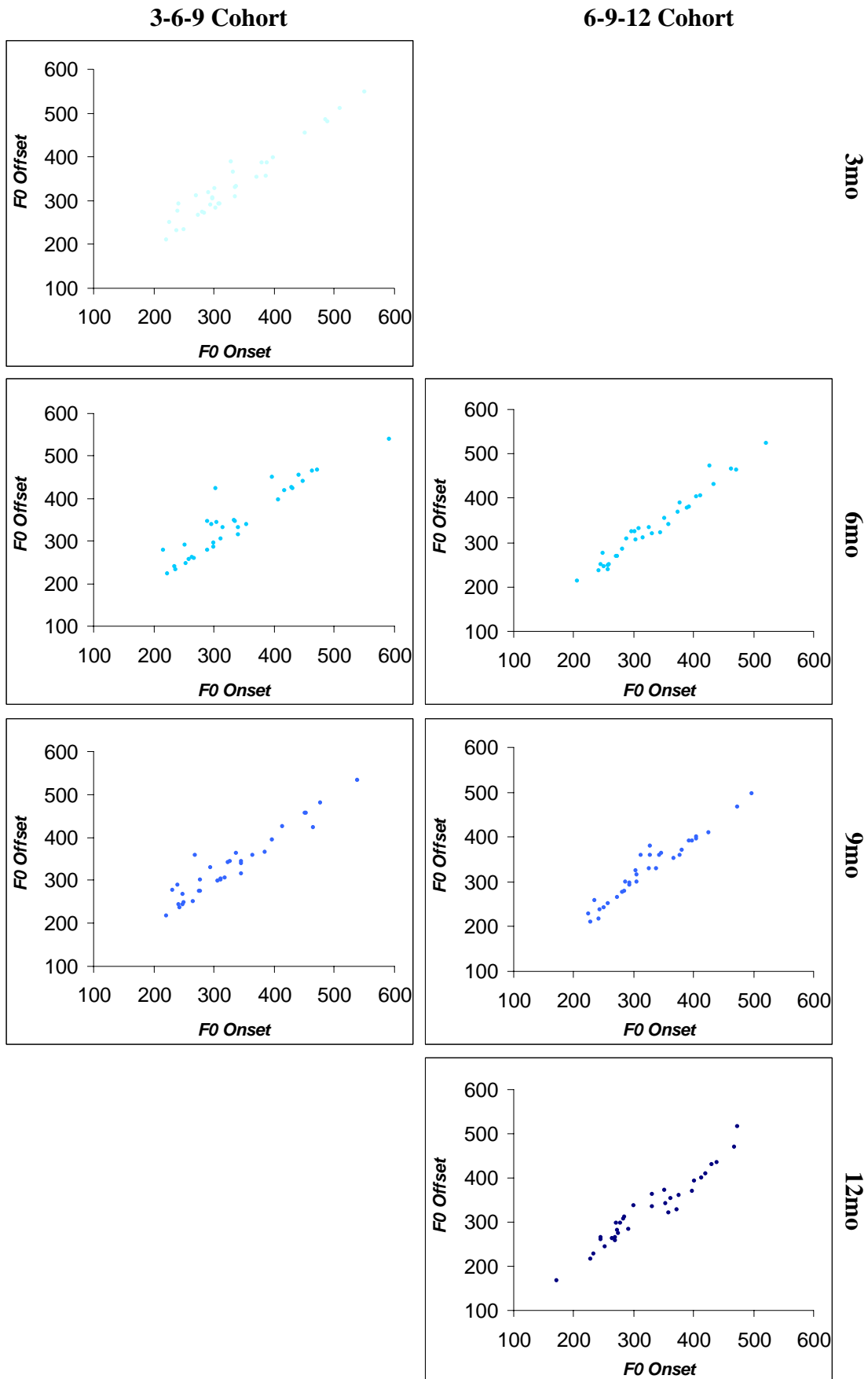
Figure 21 (a). Mean F0 Onset and Offset Data Points, Separated by Cohort and Age.

3-6-9 Cohort   **6-9-12 Cohort**

(i) ADS

(ii) Triangle Across Speech Registers

(iii) F0 Onset/Offsets plus Triangles
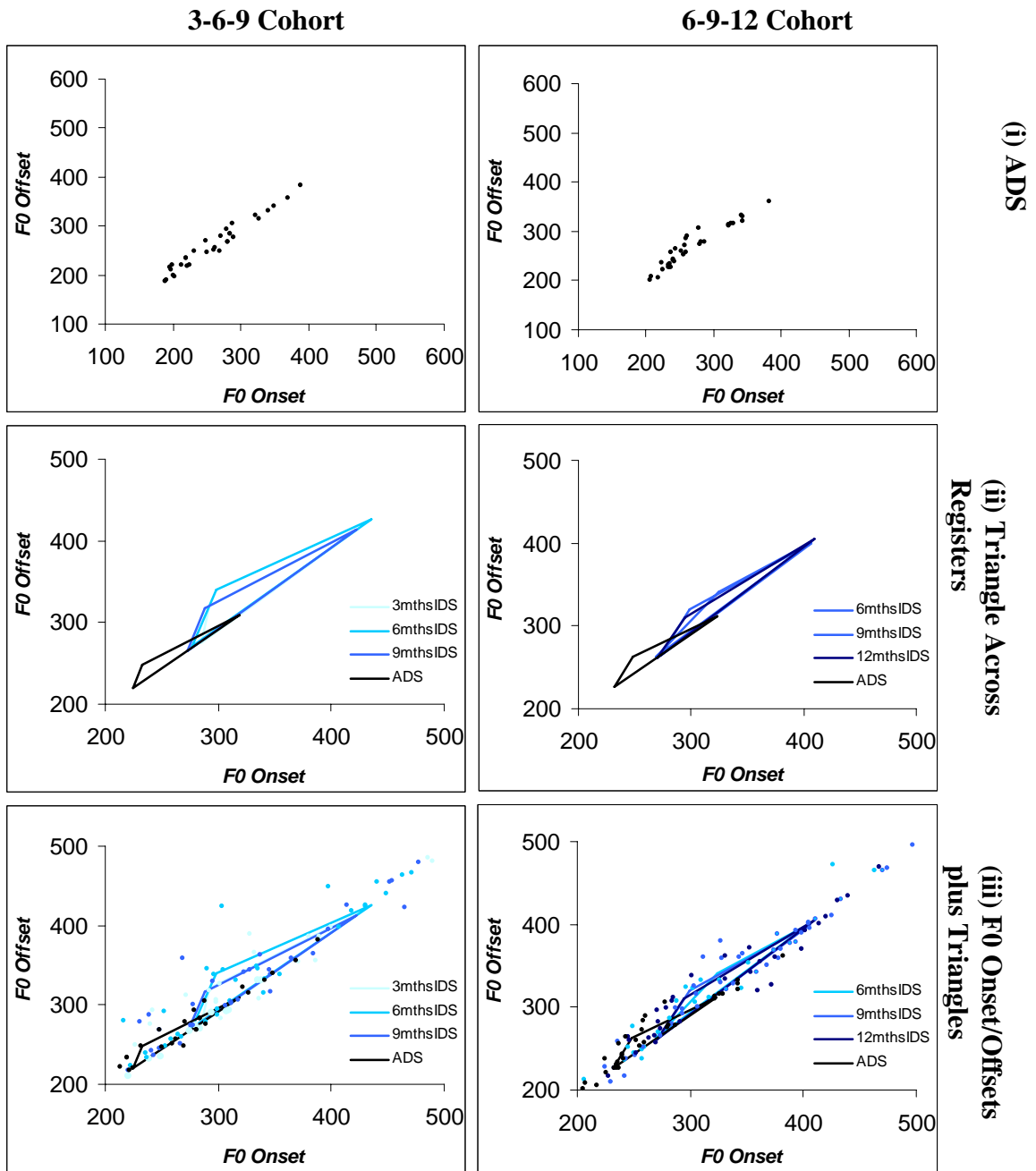
Figure 21 (b). Mean F0 Onsets and Offsets for (i) ADS, (ii) Tone Triangles across Speech Registers, and (iii) Mean F0 Onset and F0 Offset with Mean Tone Triangle Areas
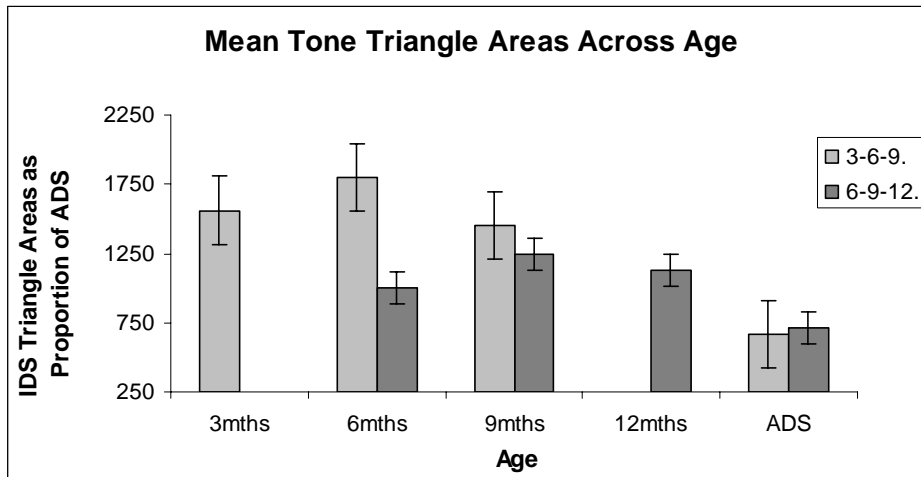
Figure 22. Mean IDS and ADS Tone Triangle Areas for 3-6-9 and 6-9-12 Cohort Groups

4.4.1.2 Investigating Test Order Effects for Tone

Due to the longitudinal cohort sequential design, each mother was recorded speaking to her own infant multiple occasions. Since no longitudinal data are available on vowel hyperarticulation, in addition to the hypothesis-driven analysis in 4.4.1.1, planned comparisons for polynomial trends were conducted to test whether order effects, unrelated to developmental factors, across IDS occurred as a result of conducting multiple recordings.

For the 3-6-9 cohort, planned trend comparisons showed significant linear ($F(1, 10) = 13.494$, CIs = -16.39 - -4.02), and quadratic trend($F(1, 10) = 15.56$, CIs = -11.51- -3.20). Together the linear and quadratic trends suggest that there is an overall increase in the size of tone triangle areas in IDS from the first recording session to the next session (i.e., from 3 to 6 months) and a decrease in tone triangle area after the second recording sessions to the last session (i.e., from 9 months and ADS) (see Figure 22).

For the 6-9-12 cohort, the planned comparisons revealed no significant linear, quadratic, or cubic trends suggesting no changes in tone triangle areas across recording sessions. The different results between the two cohorts suggest that the linear and quadratic trends observed in the 3-6-9 cohort are unlikely to be a result of any order effects of recording sessions. Moreover, together with the tone triangle areas planned comparisons in section 4.4.1.1 these results suggest that while tone triangle areas in IDS at 3, 6, 9 and 12 months are larger than those in ADS, there is a trend of initial

increasing of triangles areas within IDS from 3 to 6 months a level which is then maintained to 12 months, the maximum IDS age for this study.

4.4.2 *Vowel Analysis*

4.4.2.1 Test of Vowel Hyperarticulation Hypothesis

This second analysis tests the second hypothesis that there will be vowel hyperarticulation in Cantonese IDS and this will be most likely to occur for speech to 6mos, as vowel hyperarticulation had previously been found in 6moIDS (Burnham et al., 2002; Kuhl et al., 1997; & Liu et al., 2004).

The dispersion of productions in F1/F2 vowel space, along with triangles drawn between vowel centroids for each speech register are shown for each cohort in Figure 23. Mean vowel triangle area data are shown in Figure 24.
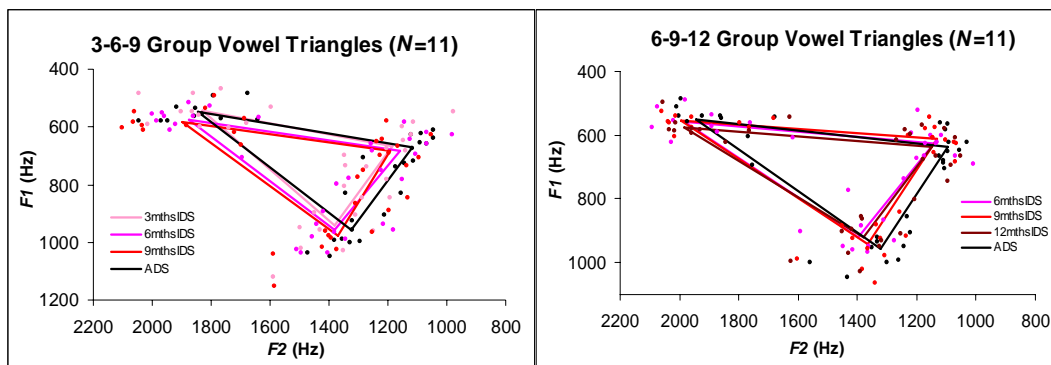
Figure 23.Vowel Triangles across Speech Registers for both Cohort Groups

The procedure for statistical analysis of the tone data was as for the tone analysis, with separate planned comparisons conducted for each cohort group, and alpha set at .05. The test of assumptions were met satisfactorily with no sphericity identified. Data screening using box plots identified four possible outliers of data points spread across three individuals, however these scores did not have corresponding z-scores in excess of 3.29, and following Tabachnik and Fidell (2001), these scores were not deleted or changed for this analysis. The three orthogonal comparisons were conducted and the results showed no significant effect of age (speech register) for either of the cohorts. Thus there were no differences between IDS and ADS, or within IDS, across ages for vowel hyperarticulation. These results suggest that all IDS vowel triangle

areas, across infant age and for both cohorts, are not significantly different to ADS vowel triangle areas.

Vowel Triangle Areas Across Age
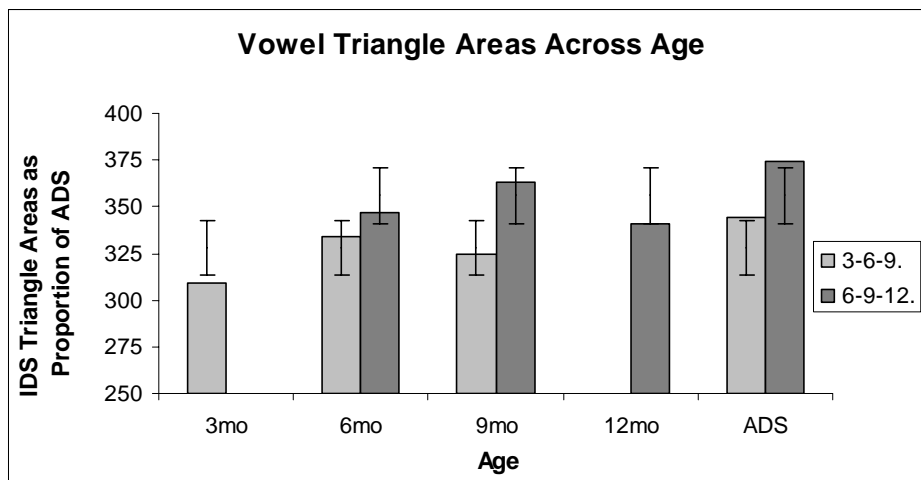
Figure 24. Mean IDS and ADS Vowel Triangle Areas for 3-6-9 and 6-9-12 Cohort Groups

4.4.2.2 Investigating Order Effects for Vowel

As for the tones analysis test of order effects were investigated. Planned trends comparisons revealed no significant linear, quadratic, or cubic trends for vowel triangle areas across test sessions for either cohort. These results are consistent with the previous vowel triangle areas analysis in section 4.4.2.1 in that the vowel triangle areas in IDS are not larger than those in ADS, nor do they differ across IDS age.

4.4.3 *Tone and Vowel Hyperarticulation Trends Across Age*

This analysis tests the final hypothesis that there will be differences in the degree of hyperarticulation in 3-, 6-, 9-, and 12-month-old IDS for both tones and vowels, in that tones should peak at 9 months and vowels at 6 months in parallel with language specific developmental changes in infant speech perception in which perceptual reorganisation for tones (between 6 and 9 months) appears to occur later than for vowels (between 4 and 6 months; Polka & Werker, 1994). Since tone and vowel triangle areas are in different units (F0 onset and offset versus F1 and F2) the triangle areas were standardised by expressing triangle areas in IDS as a proportion of triangles areas in ADS. Plots of IDS vowel and tone triangle areas indexed as a proportion of ADS vowel and tone triangle areas (IDS/ADS) for each cohort separately are shown in Figure 25

and collapsed across the two cohorts in Figure 26. As for the Tone and Vowel analyses, separate planned trend comparisons were conducted for the two cohorts.
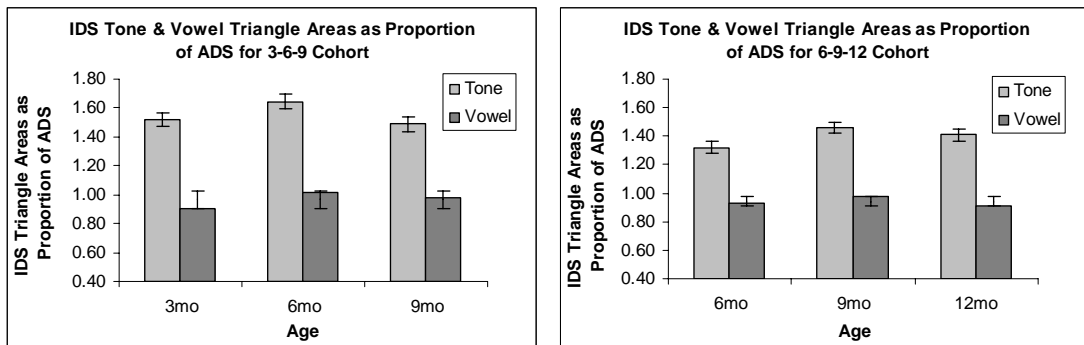
Figure 25. IDS Tone and Vowel Triangles as a Proportion of ADS for 3-6-9 Cohort (Left) and 6-9-12 Cohort (Right) with Consistent Tone and Vowel Trends across Cohorts

For the 3-6-9 cohort two comparisons were significant. First, the tone versus vowel comparison across IDS age was significant showing that standardised triangles areas were significantly larger for tones than for vowels ($F(1, 10) = 10.42$, CIs = .18 - .99). Second, there was a significant tone versus vowel quadratic trend ($F(1, 10) = 38.50$, CIs = .23 - .50), which is consistent with the results in sections 4.4.1.1 and 4.4.2.1 showing that while tone triangle areas increase from 3 to 6 months then decrease to 9 months, no changes in vowel triangle areas are apparent during the same period.

The same two comparisons were significant in the 6-9-12 cohort. Standardised triangle areas were significantly larger for tones than vowels across IDS age ($F(1, 10) = 7.37$, CIs = .08 - .83) and there was a significant quadratic trend for the tone versus vowel comparison ($F(1, 10) = 22.95$, CIs = .19 - .51), showing that for tones there is a increase in triangle area from 6 to 9 months then a decline to 12 months but no change for vowel triangle areas over the same period.

Together these results suggest that tone and vowel have different trends across age in IDS, namely they differ in that (i) tone triangle areas are consistently larger than vowel triangle areas across IDS ages for both cohorts and at all ages, (ii) tones show a quadratic trend across age in both cohorts, with initial increase followed by a decline, while vowels do not. Together these results suggest that tone and vowel productions are independently manipulated in Cantonese IDS. Moreover it should be noted that tones

are hyperarticulated in Cantonese IDS while vowels are not. and (iii) tone and vowel triangle areas are independently manipulated in Cantonese IDS.

IDS Tone and Vowel Triangle Areas (Mels) as a Proportion of ADS
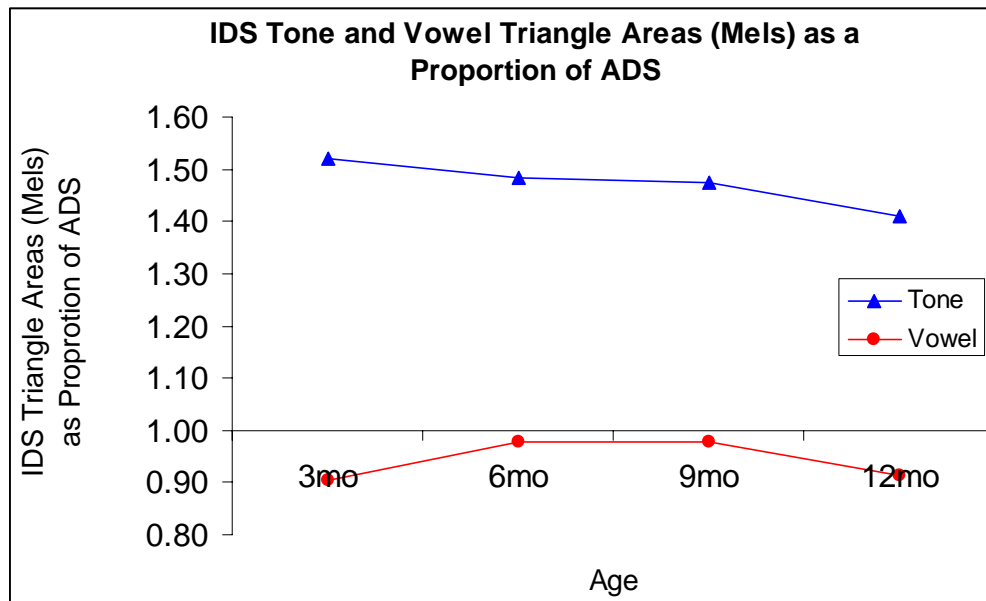
Figure 26. IDS Vowel and Tone Triangle Areas as proportion of ADS across both cohort groups (where 1 represents IDS=ADS)

4.5   Discussion

While the analyses in this chapter show tone hyperarticulation and the absence of vowel hyperarticulation in Cantonese IDS compared to ADS, the pattern of results and trends for tone and vowel triangle areas across IDS ages did not provide straight-forward support for the predictions set out in the introduction. Different aspects of the results are discussed below.

4.5.1   *Tone Hyperarticulation in Cantonese IDS: Early Emergence*

The tone triangle areas analysis provided support for the first hypothesis that there will be tone hyperarticulation in Cantonese IDS at some point during the infants' first 12 months. In fact, Cantonese mothers hyperarticulated tones in IDS compared to ADS at all ages (3, 6, 9, and 12 months) compared to ADS, so tone hyperarticulation in IDS is present early in infancy. While the degree of tone hyperarticulation did not vary considerably across infant ages, the quadratic trends found for standardised IDS on ADS triangle areas in both cohorts suggests that there is an increase in the degree of IDS tone hyperarticulation from 3 months to 6 months then a decrease from 9 to

12moIDS. This pattern of tone hyperarticulation fits well with infant development studies that suggest infants' speech perception is attenuated for non-native tone contrasts between 6 and 9 months (Mattock & Burnham, 2006; Mattock et al., 2008); if tone hyperarticulation is related to infant's language development, then it can be expected that a decline in the degree of tone hyperarticulation in IDS may occur after infants have tuned in to native tones.

4.5.2 *No Vowel Hyperarticulation in 6-month-old Cantonese IDS*

The second hypothesis that there will be vowel hyperarticulation in Cantonese IDS and this will be most likely to occur for speech to 6mos (Burnham et al., 2002; Kuhl et al., 1997; & Liu et al., 2004) was not supported. In fact, Cantonese mothers did not hyperarticulate vowels in IDS compared to IDS at any of the 3, 6, 9, or 12 months ages examined here. This is unexpected given that previous studies have found vowel hyperarticulation in other languages with 6-month-old infants (Burnham et al., 2002; Kuhl et al., 1997; Liu et al., 2003). However, it may be noted that this is the first direct acoustic analysis of IDS in Cantonese, and so it may indeed be the case that vowel productions are different in Cantonese IDS. Vowel hyperarticulation has so far only been found in one tone languages Mandarin, a tone language with a much less complicated tone inventory than Cantonese, so the possibility that vowel hyperarticulation may not occur in a tone language with a complex tone inventory must be seriously considered. Further studies on tone language IDS with complex tone inventories such as Thai must be conducted in order enrich comparisons of tone and vowel information in tone language IDS.

It is of note that a recent study comparing vowels in IDS and in ADS found the lack of vowel hyperarticulation in Norwegian IDS mainly due to the fronting of back vowels (Englund & Behne, 2005). Englund and Behne (2005) hypothesised that this fronting may be an artefact of the exaggerated facial gestures that mothers employ while speaking to their infants and may be related to the communication of affective information. Indeed, there is some cross-cultural evidence for ID facial gestures in Chinese (Cantonese and Mandarin) and English speaking mothers, with three distinct

97

facial gestures related to the communication of love, amazement and pleasure having been found (Chong et al., 2003). This issue of facial gesturing and fronting of back vowels will be considered in greater detail in the general discussion (Chapter 7).

It is also worth noting that while all three vowels /i/, /u/, and /a/, were recorded in similar consonant contexts (i.e., /d/VC), both the /i/ and /u/ vowels were followed by a stop consonant, while /a/ was followed by a nasal /n/. Nasalisation tends to lower the preceding vowel formants. However, since /a/ is a low vowel, it should be the least affected of all vowels by nasalisation (Ladefoged, 2003). So nasalisation is unlikely to have been the cause of the lack of vowel hyperarticulation observed in this study. There were also some consistent abnormalities observed in the /i/ vowel in IDS showing two very distinct frequency bands occurred for F2, the usual 2000-2800Hz band and a 1000-2000Hz band. The latter resulted in the lowering of the mean F2 frequencies which resulted in the reduction of the overall size of the IDS vowel triangle areas. There may be several reasons for this. One may be that the initial consonant /d/, with intrinsic formants at around 1000Hz, may have had a large effect on the vowel transition into /i/, which would be most affected since it has the highest F2 values of the three corner vowels, being a high vowel. An alternative explanation is that the lowering of F2 values resulted from excessive lip rounding during affective signalling. Since F2 lowering is only observed in IDS and not ADS, the second explanation appears to be the most likely. This issue together with ID facial gesturing in IDS will be taken up again in the general discussion.

Another possible reason for the lack of vowel hyperarticulation may be related to the fact that this is the first study to examine the hyperarticulation of two aspects of speech simultaneously (tone and vowels here), and so there is no accumulated experimental wisdom of how this (the demand characteristic that mothers may have been consciously or unconsciously subjected to) might affect mothers' behaviour. Moreover, one aspect of the experimental procedure, the fact that three vowels and six tones were studied meant that mothers were provided with three toys for eliciting the three vowels, but six toys for the six tones. This may have resulted in greater attention to tone than vowel items and/or a conscious or unconscious demand characteristic or

bias for tone more than vowel hyperarticulation in the experiment. While this is considered to be unlikely, further research is required on this issue. This and possible reasons for attenuated vowel and heightened tone hyperarticulation in Cantonese are considered in Chapter 7

4.5.3 *Developmental Trend for Tone Hyperarticulation but not for Vowels*

The third hypothesis that there will be differences in the degree of hyperarticulation in 3-, 6-, 9-, and 12-month-old IDS for both vowels and tones, in that tones should peak at 9 months and vowels at 6 months in parallel with language specific developmental changes in infant speech perception (Mattock & Burnham, 2006; Polka & Werker, 1994) was partially supported. The significant quadratic trend in the tone analysis suggest that tone hyperarticulation is present early at 3 months, and shows a decline from 6 to 9 to 12 months whereas there is no such trend, or any trend, or vowels. These results for tone hyperarticulation fit nicely with the known perceptual milestones for tone perception in infancy; early in infancy when infants are tuning in to native tones, mothers are hyperarticulating tone in IDS and this hyperarticulation peaks at 6 months and tends to decline just when infants' perceptual attention to non-native tones begins to attenuate between at 9 months (Burnham & Mattock, 2007; Mattock et al., 2008). Thee results are consistent with there being a correlation between tone hyperarticulation and infant speech perception for it appears that tone hyperarticulation decreases as infants' speech perception becomes more language specific. However, to test this possibility of a correlation, speech perception tests with these infants are required. Some preliminary evidence for this is provided in Chapter 6.

No trends are observed for vowel hyperarticulation in IDS across the first 12 months of infancy suggesting no vowel hyperarticulation in Cantonese IDS right across the first 12 months of infancy. One possible reason for this is that there are fewer tones (6) than vowels (8 to 19; see Sections 1.2.1 & 2.4.2 on Cantonese tones and vowels) in Cantonese and so acquisition of tones may be less difficult simply because there are less of them, on a numerical basis, to acquire. Such precedence in production of tones may

be evident even in young infants' perceptual responsivity, such that mothers respond to and build upon this predisposition in their IDS-borne implicit language instruction.

Finally, the lack of any consistent trend between speech register by tone and vowel triangle areas (IDS as proportion of ADS) comparison suggest the independence of tone and vowel information in IDS. While tone information is carried on vowels in Cantonese, tone and vowel information appear to be produced and manipulated independently in Cantonese IDS.

4.6   Conclusion

The results of the tone and vowel triangle areas analyses suggest that tone is hyperarticulated in Cantonese IDS consistently throughout the first 12 months of infancy. Tone hyperarticulation emerges early at 3 months, peaks at 6 and 9 months and decline somewhat at 12 months. This is consistent with known perceptual milestones for tone perception – tone hyperarticulation declines around 9 months when infants have already begun to tune in to native tones. Vowels were not hyperarticulated in IDS at any point over the first 12 months of infancy which is unexpected considering the existing body of research that has found vowel hyperarticulation in 6moIDS in other languages. However, it may be noted that this is the first study of Cantonese IDS (a complex tone language), and in this study tone and vowel information are examined together. Further implications of this and other possible reasons for these results as well as the implications of these results for future IDS and tone language research are taken up again the general discussion in Chapter 7.

Five

Chapter 5

IDS Tones and Vowels in Cantonese – A Closer Look

5.1 Tone Hyperarticulation in Cantonese IDS

An overall level of tone hyperarticulation in Cantonese IDS has now been established. Over and above this basic finding, there is a rich array of phonetic and acoustic information in the data that afford investigation, in particular the specific features of tone that Cantonese mothers may manipulate in their speech to their infants at different ages. In Chapter 3, a review of studies on tone perception and production suggested that the most salient features for Cantonese tone perception are F0 onsets and offsets. Given this and since Cantonese tones are complex – they include three level (high, mid, and low) and three contour (high-rising, mid-rising, and falling) tones and, tones with similar onsets (e.g., the rising tones 23 and 25) – it will be instructive to augment our knowledge of tone hyperarticulation by examining in more detail the F0 characteristic of mothers' tones in their IDS.

In this chapter, a detailed acoustic and phonetic analysis of tone differentiation in terms of F0 onsets and offsets is presented, as well as analysis of overall tone differentiation i.e., the degree of separation of each tone from every other tone. In addition to tone information, vowel information will also be subjected to a similar fine-grained examination using a vowel differentiation measure and similar to the investigation of the complexities of tones (contour versus level, high versus low, etc.) investigation of the differentiation of vowels in terms of the phonetic features (front versus back, and high versus low) will also be incorporated.

5.2 Method

As this is an extension of the analysis of tone and vowel hyperarticulation in Chapter 4, the design, participants, materials, procedure, and data collection and extraction here are the same. The planned contrasts analyses used here were conducted within the Psy program (Bird et al., 1990) developed for ANOVA with within-subjects, repeated measures planned contrasts using confidence intervals (CIs). As with the tone and vowel triangle analyses in Chapter 4, the data have been square-root transformed to reduce the effect of extreme scores.

5.2.1 *Data Analysis*

Tones: The analysis conducted here examines all six Cantonese tones; the three most dispersed tones 55, 25, and 21, used in the previous analyses to produce tone triangle areas, as well as the remaining three tones 23, 33, and 22. The three parameters examined here are F0 onset, F0 offset, and a differentiation index – the mean distance of each tone from every other tone in F0 onset by F0 offset space. This mean difference is calculated by first finding the centroid or mean of all data points for each tone based on the onset and offset values in a two-dimensional space, and then determining the distances between each tone and every other tone. Put simply, the tone differentiation index for each tone is the mean distance of each tone from each of the other five tones combined. As an extension of this, differentiation scores for level (55, 33, and 22) versus contour (21, 23, and 25) tone, and for relative height (55 versus 33 and 22; 33 versus 22) and contour type (21 versus 23 and 25; and 23 versus 25) are also derived.

Vowels: A vowel differentiation index, parallel to the tone index, was calculated by first determining the mean or centroids of each vowel along the F1 and F2 dimensions, from which mean distances between each vowel and the other remaining two vowels were calculated, resulting in a differentiation index for each of the three vowels. As for the tones, the centroids are the statistical mean of all data points for F1 and F2 for each vowel. In an extension of this method, differentiation scores for high versus low, the two high vowels /i/ and /u/ and the low vowel /a/; and for front versus back between the front vowel /i/ and back vowel /u/ were also derived.

Analyses of these scores were conducted with the same longitudinal sequential cohort design with two cohort groups 3-6-9 and 6-9-12 (with the numbers indicating infants' age at each recording session) as that used in Chapter 4, and due to the same limitations as in Chapter 4, data for the 3-6-9 and 6-9-12 cohort groups are analysed and reported separately for each tone and vowel index (see Supplementary CD for results and coefficients).

5.3   Results

5.3.1   *Tone Differentiation*

A mean tone differentiation score was established for each tone for each mother and planned contrasts conducted separately for each cohort group.

For the speech register factor, three planned comparisons were conducted for each cohort. For the 3-6-9 cohort, the three planned contrasts conducted were: ADS versus IDS (3, 6, 9 months combined), and then to determine differences over infant age (i) 3 months versus 6 and 9 months, and (ii) 6 versus 9 months IDS. For the 6-9-12 cohort group similar contrasts were conducted: ADS versus IDS; and then to determine differences over age (i) 6 versus 9 and 12 months; and (ii) 9 versus 12 months.

For the tones factor the same five planned contrasts were conducted for each cohort. These can be conceptualised in three groups of contrasts as follows (I) Level versus Contour: level tones (55, 33, 22) versus contour tones (21, 23, 25), (II) Within Level Tones (a) high-level tone (55) versus lower levels tones (33, 22), and (b) mid-level tone (33) versus low-level tone (22), and (III) Within Contour Tones: (a) falling (21) versus rising tones (23, 25), and (b) high-rising tone (25) versus the low-rising tone (23). These five planned contrasts for tones were investigated both alone and in interaction with each of the three speech register contrasts, separately for each cohort group (i.e., 3-6-9 and 6-9-12). For all contrasts, F-values and the 95% Confidence Intervals (CIs) are reported.

Figure 27 shows plots of mean tone difference scores for each tone across age separately for the two cohorts grouped by Level and Contour (all tone differentiation scores are actual F0 differences in Mels). The only significant contrast was in the 6-9-12 cohort for the main effect of age and the average F0 differentiation among tones is significantly greater in combined IDS than ADS ($F(1, 10) = 35.23$, CIs $= 1.23 – 2.70$).

3-6-9 Cohort   **6-9-12 Cohort**
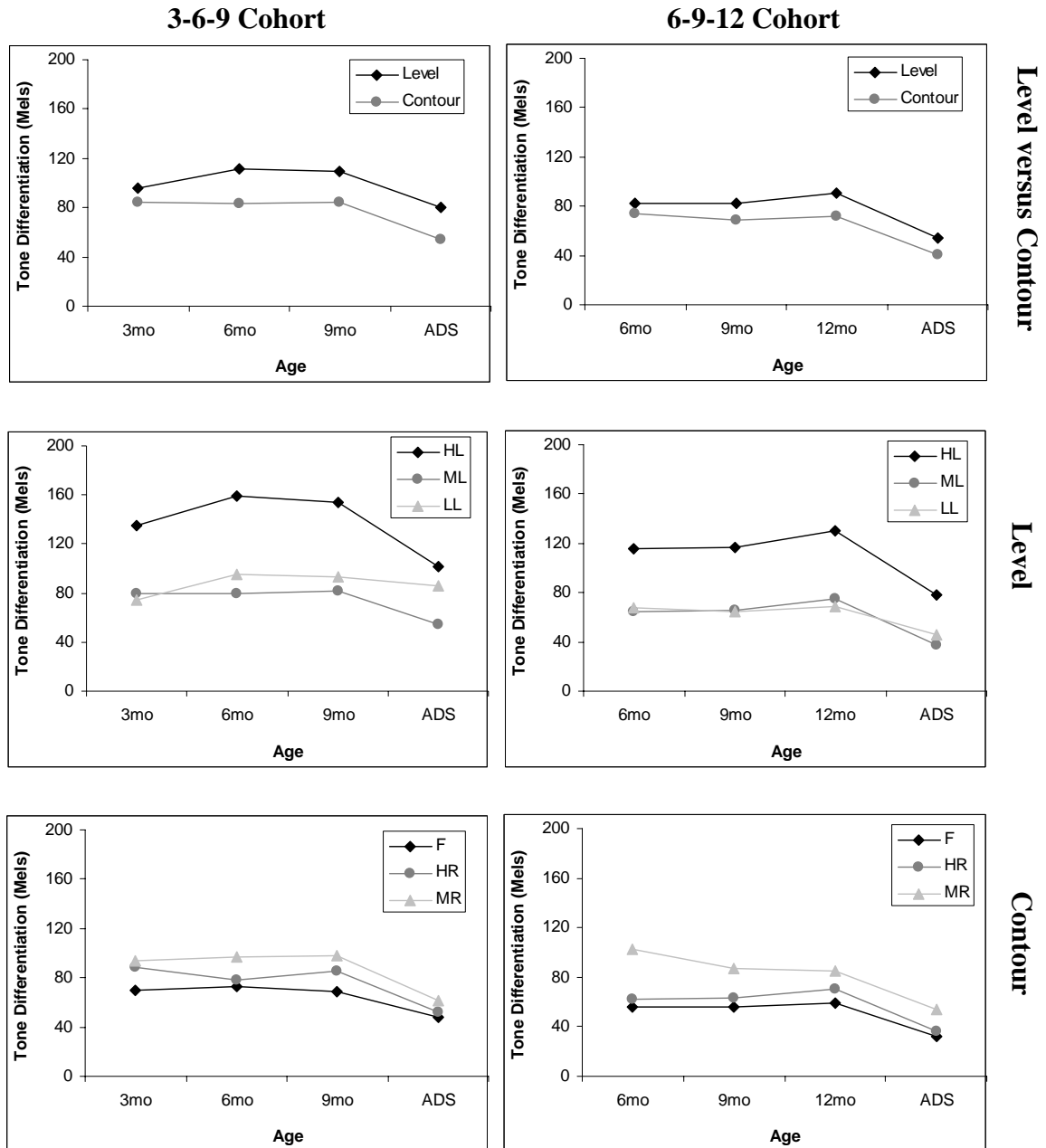
Figure 27.   **Tone Differentiation by Speech Register across cohort groups represented separately for Level only (Top) and Contour tones (Bottom)**

5.3.2 *F0 Onset*

The same planned contrasts as above were conducted for tone onset measurements. Figure 28 shows plots of F0 onset differentiation contrasts for level and contour tones, at each recording interval separately for each cohort.

3-6-9 Cohort **6-9-12 Cohort**
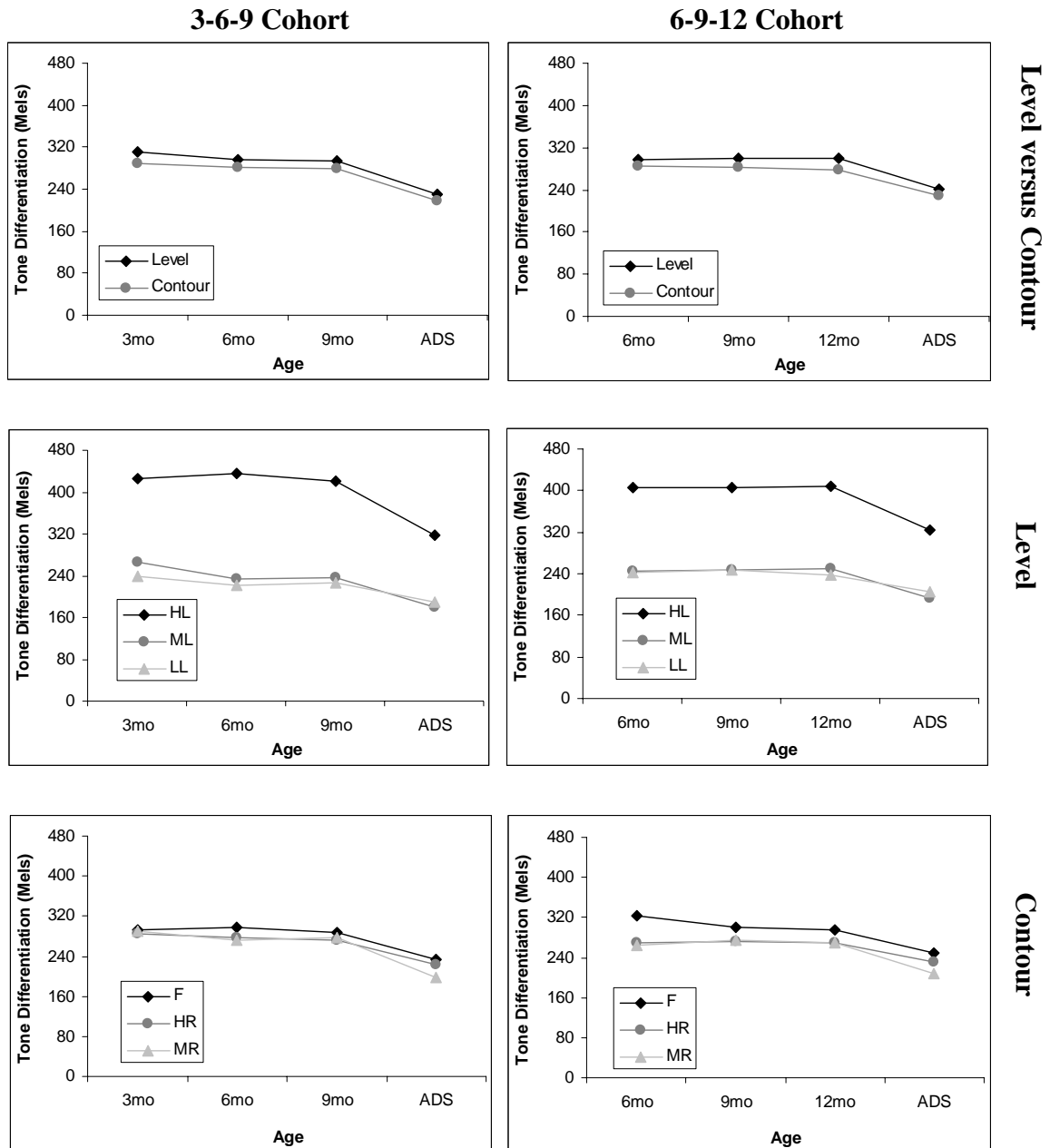
Figure 28. **Tone Onset Difference by Speech Register across cohort groups represented separately for Level only (Top) and Contour tones (Bottom)**

For the 3-6-9 cohort group, the mean tone onset differentiation was significantly greater for combined IDS than ADS ($F(1,10) = 39.83$, CI = 1.46 – 3.05), and there is a significant interaction with the high-rising, 25, versus mid rising, 23, tones in which the difference in F0 onset between high and mid rising tones is significantly less in IDS than ADS ($F(1,10) = 15.18$, CI = 0.23 – 0.86).

For the 6-9-12 cohort group, the contrast of IDS versus ADS was significant showing that, in general, tone onset was more differentiated in IDS than ADS ($F(3,12) =$

84.96, CI = 1.25 − 2.06). This main effect interacted significantly with (i) the high-level, 55, versus lower-level tones, 33 and 22, in which the difference between F0 onsets of the high- and lower-level tones was significantly greater for combined IDS than ADS (F(1,10) = 5.48, CIs = -.90 − -.02); and (ii) the high-rising, 25, versus mid-rising, 23, tones in which F0 onset differentiation was significantly less between high- and mid-rising tones for combined IDS than ADS (F(1,10) = 20.67, CI = . 26 − .77).

Together these results show that for both cohorts the mean F0 onset differentiation among all tones is significantly larger in combined IDS than ADS, but this is not true for the F0 onset differentiation between the high-rising, 25, and mid-rising, 23, tones which are significantly more reduced in combined IDS than ADS.

5.3.3   *F0 Offset*

Figure 29 shows plots F0 offset differentiation contrasts for level and contour tones at each recording interval separately for each cohort.

For the 3-6-9 cohort group, mean F0 offset differentiation is significantly greater in combined IDS than ADS (F(1,10) = 59.49, CIs = 1.49 − 2.70). This main effect interacted significantly with (i) high-level, 55, versus lower-level, 33 and 22, tones in which F0 offset was significantly more differentiated in combined IDS than in ADS (F(1,10) = 6.34, CIs = -.94 − -.06); and (ii) high-rising, 25, and mid-rising, 23, tones in which F0 offsets were significantly less differentiated in combined IDS than ADS (F(1,10) = 14.29, CIs =.19 − .74).

For the 6-9-12 cohort group, mean F0 offset differentiation among tones in combined IDS was greater than in ADS (F(1,10) = 90.15, CIs = 1.28 − 2.06). There was also a significant interaction between IDS versus ADS with (i) high-level, 55, and lower-level, 33 and 22, tones, in which the mean F0 differentiation between high- and lower-level tones is significantly greater in combined IDS than ADS, (F(1,10) = 5.90, CIs = -.68 − -.03); (ii) mid-level, 33, and low-level, 22, tones, in which the mean F0 offset is significantly more differentiated between mid- and low-level tones in combined IDS than ADS (F(1,10) = 8.50, CIs = .05 − .80); and (iii) high-rising, 25, and mid-

rising, 23, tones, in which the mean F0 offset is significantly less differentiated in combined IDS than ADS ($F(1,10) = 11.74$, CIs $= .14 - .66$).
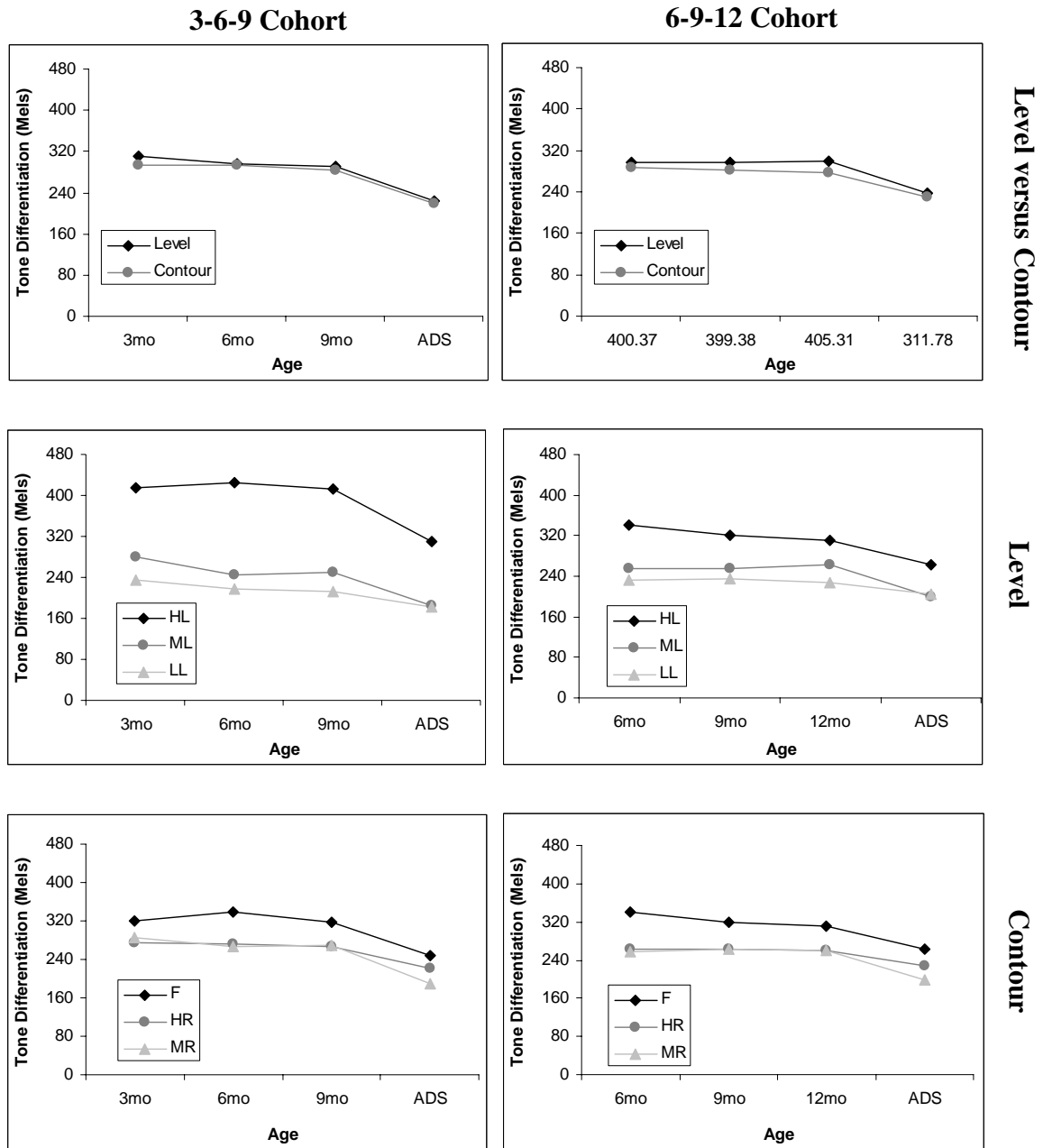
Figure 29. Tone Onset Difference by Speech Register across cohort groups represented separately for Level only (Top) and Contour tones only (Bottom)

Together these results suggest that for both cohorts overall F0 offset differentiation among tones is significantly greater in IDS combined than ADS, and for both cohorts high-level, 55, tones are significantly more differentiated from lower level, 33 and 22, tones, while the high-rising, 25, and mid-rising, 23, tones are significantly less differentiated in IDS combined than in ADS. For the older cohort 6-9-12, the mid-

level, 33, and low-level, 22, tones were significantly more differentiated in IDS combined than in ADS.

5.3.4 *Vowel Differentiation*

Analysis of speech register in terms of IDS versus ADS and within IDS levels were conducted as for the tones. The planned contrasts conducted for the vowels factor were: high vowels, /i/ and /u/, versus the low vowel, /a/, and within the high vowels, the front vowel, /i/, versus the back vowel, /u/. Table 8 shows mean vowel differentiation among the three corner vowels as well as for high versus low and front versus back dimensions. Figure 30 shows the mean differentiation of each vowel from all other vowels as well as differentiation along the high versus low and front versus back dimensions for each cohort.

Table 8. Mean Vowel Differentiation scores in Mels square-root Transformed for Vowel Contrasts across Speech Registers by Cohorts

Cohort 3-6-9 Planned Contrasts	3mo	6mo	9mo	ADS
/a/	22.74	22.08	22.11	22.89
/i/	26.93	25.56	25.53	27.02
/u/	23.68	22.65	22.67	23.75
High vs. Low	11.22	9.79	9.70	11.09
Front vs. Back	12.82	11.86	11.74	12.88
Column Mean	19.48	18.39	18.35	19.53
Cohort 6-9-12 Planned Contrasts	6mo	9mo	12mo	ADS
/a/	23.23	22.96	23.31	23.73
/i/	27.74	27.07	27.07	28.05
/u/	24.48	24.23	24.49	24.66
High vs. Low	12.03	11.51	11.09	11.60
Front vs. Back	13.06	12.06	11.54	13.36
Column Mean	20.11	19.57	19.50	20.28

Separate analyses for the two cohorts show that only in the 3-6-9 cohort were there significant results: mean vowel differentiation was significantly higher at 3 months than 6 and 9 months combined IDS ($F(1,10) = 6.28$, $CI = .11 - 1.82$).

109

Together these results show that there are no significant differences in vowel differentiation between IDS and ADS for either cohort, but that in the 3-6-9 cohort vowel differentiation is significantly greater in 3 months IDS than in 6 and 9 months IDS combined.
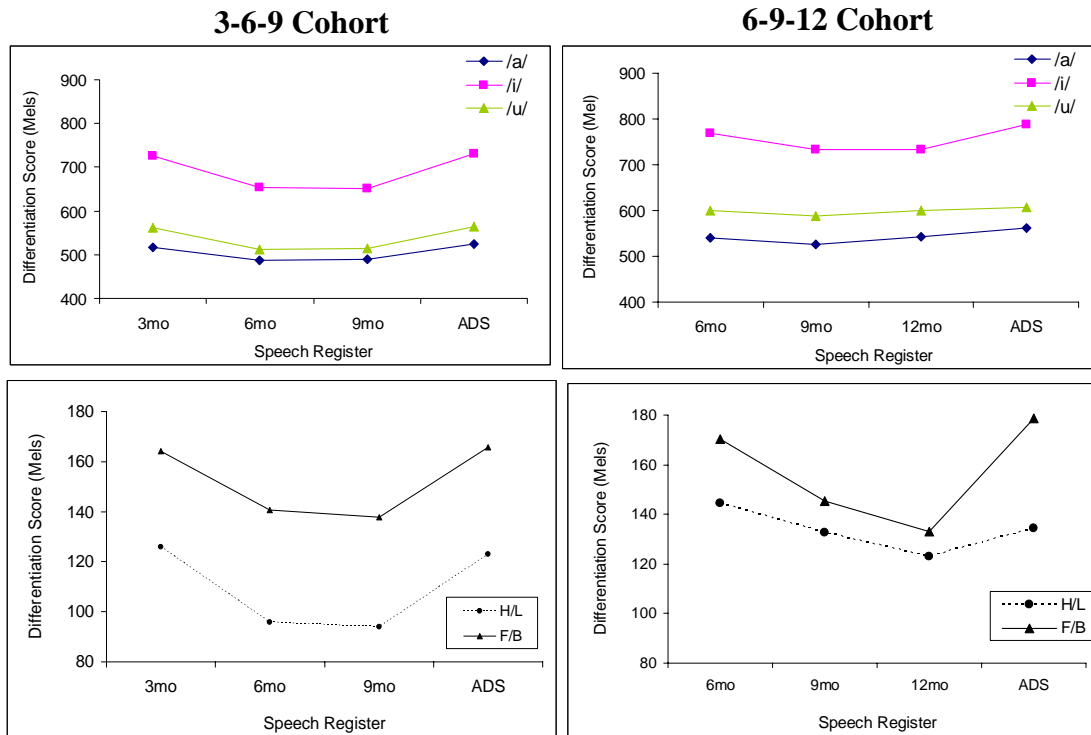
Figure 30. Vowel Differentiation scores by speech registers across cohort groups (Top) for the each of the three corner vowels and (Bottom) on dimensions of High versus Low (H/L) and Front versus Back (F/B)

5.4   Discussion

In general, the results here – for mean tone differentiation, onset and offset, and for mean vowel differentiation – are consistent with tone and vowel triangle areas analysis in Chapter 4. Both differentiation analyses (here) and triangle area analyses (in Chapter 4) show tone hyperarticulation across IDS ages compared to ADS and no vowel hyperarticulation at any infant age. However, further specific results regarding tone dimensions were found in the analyses here and these provide more detail about the dimensions that are manipulated in tone space expansion/hyperarticulation.

5.4.1 *Hyperarticulation of the Tone Space while Ignoring Detailed Onset and Offset Information*

In general, mothers exaggerated the mean F0 (in F0 onset/offset space, and onset and offset) across all tones in Cantonese IDS. This is especially the case for differentiation of F0 onset and offset for the high level tone 55 from the lower level tones 33 and 22, while the differentiation of F0 onset and offset between high rising 25 and the mid-rising 23 tones, having almost identical onsets, are consistently underspecified across IDS registers compared to ADS. The results suggest that the hyperarticulation of tones in Cantonese IDS is mainly reflected in the exaggerated differentiation of high and lower level tones. On the other hand, tones with similar onsets such as the two rising tones are hypoarticulated in IDS compared to ADS, suggesting that mothers do not over-specify tones with similar onsets and offsets in IDS. These findings suggest that mothers' specification of specifying pitch level alone in IDS may be sufficient for infants' initial learning of Cantonese tones. However, future studies are required to examine whether hyperarticulation between other tone qualities, e.g., the slope in the two rising tones 25 and 23 might occur as later developments in Cantonese, perhaps after the maximum age of 12 months studied here.

5.4.2 *No Vowel Hyperarticulation but 3 months IDS is More Differentiated than IDS during later Infancy*

Consistent with the results from the vowel triangle area analysis in Chapter 4, the vowel indices did not show vowel hyperarticulation across IDS registers compared to ADS. However, the results from the analysis in this chapter indicate that vowel differentiation overall is greater at 3 months compared to 6 and 9 months IDS, while vowel differentiation at 6, 9, and 12 months does not differ. This result could be taken to suggest that while vowels in Cantonese IDS are generally not hyperarticulated, vowels are better defined during early infancy at 3 months before infants have tuned in to native vowels between 4 to 6 months (Polka & Werker, 1994). Again, it must be borne in mind that this is the first study of IDS in Cantonese and vowel specification in IDS may be different for this complex tone language than in other less complex tone languages or in non-tonal languages (see Chapter 7 for further discussion).

5.4.3 *Applicability of Tone and Vowel Indices as Measures of Hyperarticulation*

The results of the tone analyses here parallel those of the tone triangle area analyses in Chapter 4, suggesting that both methods are appropriate measures of tone hyperarticulation in IDS. The differentiation-based analyses used here provide additional information about the types of tones that are hyperarticulated and hypoarticulated in IDS compared to ADS. Specifically, the overall tone differentiation scores tell a different story from tone triangles alone in that (i) the high- and mid-rising tones with similar F0 onset in Cantonese IDS are less differentiated in IDS than ADS and (ii) the high- and lower-level tones are more differentiated in IDS than ADS. It maybe the case that for tone languages with complex tone inventories such as Cantonese, the specification of tone space via the overall height of register tones is initially sufficient. This has implications for the study of tone acquisition in other languages, both for languages with simple and complex tone inventories such as Mandarin and Thai, respectively.

The vowel differentiation index, on the other hand, did not afford any additional clarification regarding vowel dimensions. However, this may be due to the lack of vowel hyperarticulation in this study. The assessment of whether the vowel differentiation analysis can provide additional information about the IDS vowel space awaits further studies.

5.4.4 *Conclusions*

The results from the vowel differentiation analysis demonstrate a lack of vowel hyperarticulation in Cantonese IDS. There may be several possible explanations for this, some of which have been discussed in Chapter 4 and will be taken up again in Chapter 7. Tone hyperarticulation in Cantonese IDS is mostly driven by the better differentiation of high from lower level tones, while contour tones with similar tone onsets and offsets are hypoarticulated in IDS compared to ADS. These results suggest that specification of overall height of tones in tone space may be initially sufficient for infants to learn to differentiate tones perceptually.

Six

Chapter 6

Infants' Native Speech Discrimination using the Conditioned Head-Turn Procedure

6.1   Introduction: A study on the relationship between Cantonese infant speech perception and tone hyperarticulation in Cantonese IDS

The results from Chapters 4 and 5 demonstrate that while vowels are not hyperarticulated in Cantonese IDS compared to ADS, tones are hyperarticulated at all the infant ages tested, 3, 6, 9 and 12 months. Moreover there is some evidence of a peak in tone hyperarticulation at 6 and 9 months. This pattern of results was found using both the tone triangle area method (Chapter 4) and tone space differentiations and F0 onset and offset differentiation methods (Chapter 5).   The peak in tone hyperarticulation at 6 and 9 months suggests that mothers' degree of tone hyperarticulation in IDS may be related to their infants' speech perception because attunement to tones is shown to occur between 6 and 9 months (Mattock & Burnham, 2006) but not earlier (Mattock et al., 2008).   Nevertheless, a direct test of the relationship between mothers tone hyperarticulation and infants' speech perception is yet to be conducted.

Such tests have been conducted with respect to vowel hyperarticulation. The degree of mother's vowel hyperarticulation in IDS has been found to be positively correlated with infants' native speech discrimination ability (Liu et al., 2003). This evidence plus the prevalence of vowel hyperarticulation in IDS in both tonal (Liu et al., 2003) and non-tonal languages (Burnham et al., 2002; Kuhl et al., 1997), as well as in speech directed to foreigners (Uther et al., 2007), suggests there is a strong possibility that vowel hyperarticulation is associated with infants' early language development.

Here, in this study we investigated the possibility of a relationship between mothers' tone hyperarticulation and infants' speech perception, specifically a correlation between tone hyperarticulation and infants' native consonant discrimination. If hyperarticulation is indeed associated with infant language development then tone hyperarticulation, like vowel hyperarticulation, should also be positively correlated with infants' native phonetic speech discrimination ability.

6.2 Method

A wide range of experimental paradigms for testing young infants' perception have been built up since Fantz (1963) first applied the visual preference method to young infants. Two particular paradigms are of interest here: the visual habituation paradigm and the conditioned head-turn (CHT) paradigm. Each has associated advantages and disadvantages, and they differ in terms of experimental set up, age of infants appropriate for testing with each method, experimenter error and types of data collected for analysis with differing purposes. Details of the relative merits of these two procedures are given in Appendix F. Here a Window version of the CHT method was used (Hayes & Slater, 1999). See procedure section 6.2.5.

In this CHT study we employed a one-way design with Cantonese environment infants who listened to a fricative Cantonese consonant contrast. The aim of this study is to yield graded CHT scores across infants to be correlated with the previously reported tone hyperarticulation scores from recordings of mothers' Cantonese IDS (see Chapter 4).

6.2.1 *Participants*

The CHT infant perception study involved the infants from the mother-infant dyads who participated in the IDS study in Chapters 4 and 5 (N = 22). Each infant participated in the CHT experiments after the second IDS recording session, i.e., when the infants were between 6 and 9 months (sample mean = 257 days; cohort means are 262 and 252 days for 3-6-9 and 6-9-12 cohorts respectively). For details of mother's language background and infant's language exposure see Table 6 and 7 in section 4.2.4.1.

6.2.2 *Stimuli*

The native Cantonese fricative consonants [ʥ] and [ʧ] were embedded in the target words [ʥa] and [ʧa] meaning to hold and fork respectively. This fricative contrast was chosen as the Liu et al. (2003) study had implemented a similar fricative pair which was evidently difficult enough to produce a spread of CHT scores. A female native

115

Cantonese speaker (age = 30) was recorded saying the words in the sentence frame: "I am going to say the word … for you" (我说…该你听), and the words then excised from the sentences. Many examples (10 tokens) of each word were recorded and measured in Praat. The final four exemplars of the stimuli were selected on the basis of matching the acoustic parameters of F0, F1, F2, duration, and intensity.

6.2.3 *Equipment and Materials*

A single experimenter Windows version of the CHT was used to control presentation of the stimuli. This Windows Conditioned Head-Turn procedure is based on the CHT procedure developed by Werker (Tees & Werker, 1984; Werker et al., 1981; Werker & Tees, 1983, 1984a). Following refinement of this method in Werker's laboratory, Hayes and. Slater at the University of Exeter worked with Werker to develop a single experimenter version of the procedure (Hayes & Slater, 1999). The basic concept is the same as the traditional CHT procedure with the main difference being that there is only one experimenter involved in testing (see Appendix E for a schematic setup of the experiment).

Two rooms were used for running the experiment. In the control room the CHT software was run on an IBM PC computer with Intel Pentium MMX (Microsoft Windows 98 operating system) with a Pine PCI digital wavetable sound card (Crystal 428, version 3.0). In the sound attenuated test room, the stimuli were played through an NAD C320 Stereo Integrated amplifier (40 watt output) connected to a single loudspeaker (Creative, Cambridge Sounds Inc SBS51). The computer screen, which changed colour to reflect differing stages of the experiment, in the control room was visible to the experimenter through a window, but behind the mother and infant.

The two visual reinforcers for rewarding infant head-turns were inside two wooden boxes (height 35cm, width 30cm, depth 25.5cm) on top of each other with clear plexiglass covering the front of each box. The top visual reinforcer when activated was a mechanical monkey playing symbols and the bottom a rabbit playing drums. It was not possible to see through the reinforcer box unless it was illuminated during the

delivery of reinforcement. The loudspeaker by which sounds were presented was situated on the bottom of the visual reinforcer box.

An enclosed box containing various toys was placed next to the experimenter's chair and these toys were used to engage infants' attention between sound change trials. The experimenter controlled two foot pedals (Pedal 1: start trial pedal and Pedal 2: reinforcer pedal) and two hand buttons (Button 1: start sound button and Button 2: reinforcement plus button). Pedal 1, Button 1, and Button 2 fed into a button junction box, connected to the games port of the PC. Pedal 2 connected to the printer port via a button box.

6.2.4 *Procedure*

The CHT study was conducted at the Baby Lab in MARCS Auditory Laboratories at the University of Western Sydney, Bankstown Campus. Infants were tested using a variation of the go/no-go conditioned heard-turn procedure. The experiment was conducted in a sound attenuated test room with plain walls, containing two chairs facing each other and 180cm apart – one for the experimenter and one for the mother. The infant was seated on the mother's lap during the experiment and behind them was the control room. The experimenter sat approximately 20 degrees to the right of the infant, while the visual reinforcer and loudspeaker were 135cm and approximately 45 to the left of the infant (see Appendix E for a schematic representation of the experimental setup).

Testing sessions averaged approximately 20 minutes in duration (range 8-40 minutes). Each session began with a brief explanation of the task to the mother. Both the mother and experimenter listened to concatenated stimulus sounds from the experiment and rock music on separate channels on a masking audiotape which were presented dichotically through closed ear headphones (Koss UR-20). The masking stimuli were concatenated in Cool Edit 2000 using the mix past function with no silence between each token and recorded on one track, and rock music recorded on the second track. Mother and experimenter listened to the masking stimuli to prevent them from inadvertently influencing the infant's response. Pinto, Fernald, McRoberts, and

Cole, (1999) reported that this method of dichotic listening of concatenated stimuli is the most effective means of masking the speech sounds presented to the infant.

6.2.5 *Windows Conditioned Head-Turn Procedure*

The version of the CHT described above, required a single experimenter to operate the computer and engage the infant with toys, use hand buttons and foot pedals to initiate trials when the infant's gaze is centred, and record infant head-turns. The program has various stages through which the infant progresses, which vary slightly to the traditional CHT. The entire experiment involved three stages and these stages are described in detail in Section 6.2.5.1, 6.2.5.2, and 6.2.5.3.

Each infant was presented with the Cantonese fricative consonant word pair ([ʥa] and [tʃa]), with half the infants hearing [ʥa] as the background stimulus, and the other half hearing [tʃa]. The background sound was repeated continuously at fixed intervals of 100msec from onset to offset. Whenever there was a change from the background to the target stimulus (i.e., a test trial), the visual reinforcer was activated and illuminated and infants were conditioned to turn their head away from the experimenter towards the sound and the visual reinforcer. The criterion definition for a head-turn was defined as a movement of the head to at least 45 degrees from the midline to the left of the infant towards the reinforcer. Incorrect head-turns or false-positives (when the infant performed a head-turn towards the reinforcer without a change in sound) were not reinforced. After a test trial was initiated, the experimenter was responsible for judging whether a criterion head-turn was produced within the test trial period as indicated by the colour of the screen on the computer in the control room, irrespective of whether the test trial was a change or a no-change control trial. During the initial Training Stage, most infants were immediately attracted to the now-illuminated animated toy and turned their head towards the visual reinforcer. However, if the infant did not show awareness of the visual reinforcer, the experimenter used a fairy wand to direct the infant's attention to the reinforcer at the appropriate time as part of the shaping process. In the training and conditioning stages of the experiment, after making a correct head-turn, the infant was given social reinforcement by the

experimenter saying "Clever … (child's name)", smiling, and clapping her hands together, in addition to the visual reinforcer. At the start of the experiment and during no-change control trials, the experimenter was required to maintain the infant's attention towards the experimenter for at least 2 seconds by manipulating toys to engage the infant's attention before initiating a trial. Toys were revealed from a box one by one so that the infant did not habituate to any one toy and lose interest. Toys were manipulated on a "stable table" positioned on the experimenter's lap.

In all stages each trial was activated by the start trial pedal which signalled the computer program to commence playing the repeating background stimulus. When the experimenter judged the infant to be in a state of readiness, that is, centred and not fussing, the stimulus change trials were activated also by the start trial pedal. As trial initiation was dependent on infants' readiness, the number of repeated presentations of the background stimulus before the start trial pedal was activated was variable, and this also prevented a response expectation based on temporal cues. During test trials, signalled by a change in colour of the computer screen, the reinforcer pedal was used to signal to the computer that a criterion head-turn toward the reinforcer was made regardless of whether the trial was a change or no-change trial. When a head-turn was made during a change trial the reinforcer pedal also served to activate the visual reinforcer and the infants saw at random one of the reinforcers illuminated, either the monkey playing cymbals or the rabbit playing a drum. Pressing the reinforcement plus button increased the duration of reinforcement from 4000ms to 5000ms, and was used if the experimenter decided the length of the reinforcement period was not salient enough to reward correct head-turns. This button was seldom used as the 4000 ms reinforcer was sufficient to engage the majority of infants.

6.2.5.1 Training Stage

In the Training Stage, all trials were change trials. During the first two trials of the Training Stage, the reinforcer was automatically activated for 4000ms by the computer program simultaneously with the change in speech stimulus. In addition, the infant was directed to turn their head towards the reinforcer within the activation period. The

interval between stimulus change and activation of the reinforcer was then gradually incremented by 1000ms per trial to a maximum of 4000ms. This gave the infant increasing opportunities to make anticipatory head-turns toward the visual reinforcer after a sound stimulus change but before the activation of the visual reinforcer. If the experimenter judged that a criterion head-turn was produced by the infant, she pressed the reinforcer pedal. Head-turns and reinforcer pedal presses that did not coincide with presentation of the target stimulus were not reinforced but were recorded by the computer as an index of between trial or random head-turns (false positives). Once the reinforcement period ended and the visual reinforcer was deactivated, the experimenter used toys to attract the infant's attention to the midline again. When the infant made three correct consecutive anticipatory head-turns within 15 trials, the Training stage was complete and the infant progressed to the Conditioning stage. Infants who failed to reach criterion for training (i.e., making three correct consecutive anticipatory head-turns within 15 trials) did not progress beyond the training stage and the session was terminated.

6.2.5.2 Conditioning Stage

In the Conditioning Stage, there was a 75% probability that a trial would be a change trial and a 25% probability that the trial would be a no-change control trial. Trials were presented in a quasi-random order with the constraint that no more than two change trials or two control trials occurred consecutively. Reinforcement was contingent upon an anticipatory head-turn within 4000ms of the onset of a change trial. During change trials, if the infant responded during this period a head-turn was signalled by the experimenter using the reinforcer foot pedal and a hit was recorded by the computer program, if no head-turn occurred during a change trial a miss would be recorded. In control trials, a correct-rejection was recorded in the absence of a head-turn, and a false alarm (incorrect response) was recorded if the infant produced an anticipatory head-turn. If the infant failed to produce anticipatory head-turns on three consecutive change trials, the program reverted to 'retraining', in which the change stimulus was paired with the activation of the reinforcer for three change trials or one hit, depending which came first.

The purpose of retraining was to remind the infant that producing a head-turn to a change trial would be positively reinforced. Performance in retraining trials was not used in calculating the infant's final performance score. The criterion for discrimination was set at 7 out of 8 correct consecutive responses (hits or correct rejections), as used by Werker and Tees (1984a), and infants who reached this preset criterion were deemed able to discriminate the contrast and progressed to the Experimental Stage. Infants who failed to reach criterion for discrimination within 25 conditioning trials did not progress beyond the Conditioning Stage and the session was terminated.

6.2.5.3   Experimental Stages 1 and 2

The general procedure for the experimental stages is similar to the previous conditioning stage with three variations. Firstly, the probability of a change trial occurring when the start trial pedal was pressed was 50% as opposed to 75% probability in the Conditioning stage. Secondly, there was no preset criterion for the number of consecutive anticipatory head-turns to change trials; rather infants were presented with 15 trials with the number of correct responses recorded using the reinforcement pedal. Finally, infants did not revert to retraining if they failed to produce anticipatory head-turns to three consecutive change trials.

Experimental Stage 1: The back ground and change stimulus are the same as those used previously in the training and conditioning stages. Infants who managed to complete Experimental stage 1 were invited to participate in Experimental stage 2.

Experimental Stage 2: In this stage of the experiment, the procedure was identical to that used in Experimental stage 1 but the auditory stimuli presented to the infants differed. Infants now listened to a new pair of Cantonese consonants, the contrast [da] versus [tʰa], used in a pilot study conducted with 10 native Cantonese and 10 monolingual Australian English speakers who were first-year undergraduate psychology students at the University of Western Sydney Milperra campus. The adult participants found the [da] versus [tʰa] much easier pair to discriminate than the [dʑa] versus [tʃa]

121

contrast. This final stage served to determine whether infants were able to apply their learning to a new sound pair.

6.3   Results

The results presented here begin with those of the entire sample ahead of detailed results of four infants who progressed through to Experimental stage 1, and three infants who completed Experimental stage 2 (the entire CHT test).

6.3.1   *Entire Sample*

Of the 22 infants who participated in the CHT experiment, 15 infants reached the training stage criterion of three consecutive anticipatory head-turns within 15 trials to progress into the Conditioning stage. The mean number of trials required to reach criterion was 5.80. Of these 15 infants, only 4 met the criterion (7 out of 8 consecutive correct responses over 30 trials) to progress through to Experimental Stage 1. The mean number of trials to criterion was 15 trials for these 4 participants. Only three infants completed Experimental stage 2 and all three infants required the maximum number of trials to criterion (i.e., 30).

The two sets of data for the correlation analysis were the CHT data to index infants' speech discrimination ability and IDS data to index mothers' degree of vowel or tone hyperarticulation. CHT data included Hits, Misses, Correct Rejections, False positives, Hits minus False positive, and Number of Trials to Criterion. Vowel and tone triangle areas at 6 and 9 months in both cohorts (only 6 and 9 months were used as they are the common ages sampled in both cohorts), along with the tone differentiation indices for high level 55 versus lower level tones 33 and 22, and high rising 25 versus mid rising tones 23 were used as these were the only significant comparisons found to be consistently hyper- and hypo-articulated respectively across IDS compared to ADS (see Chapter 5). Table 9 shows correlation coefficients for these items.

Table 9. Correlation for Conditioning stage CHT data (n= 15) with Tone and Vowel measures; numbers in bold indicate significant correlations (HL = high-level tone, LL = lower level tone, HR = high-rising tone, and MR = mid-rising tone)

	hits	misses	rejections	false positives	hits-false positives	No to Criterion	vowel_6mo	vowel_9mo	tone_6mo	tone_9mo	tone6_HL vs. LL	tone6_HR vs MR	tone9_HL vs. LL	tone9_HR vs MR
hits	1.00	**-1.00**	-0.13	0.13	**0.76**	0.02	0.20	0.12	0.09	0.08	0.32	0.19	0.45	-0.01
misses		1.00	0.13	-0.13	**-0.76**	-0.02	-0.20	-0.12	-0.09	-0.08	-0.32	-0.19	-0.45	0.01
rejections			1.00	**-1.00**	**0.55**	0.09	-0.22	-0.22	0.03	-0.07	-0.24	-0.34	-0.07	-0.08
false positives				1.00	**-0.55**	-0.09	0.22	0.22	-0.03	0.07	0.24	0.34	0.07	0.08
hits-false positives					1.00	0.08	0.03	-0.04	0.10	0.01	0.11	-0.06	0.33	-0.06
No to Criterion						1.00	-0.27	-0.18	-0.20	0.19	0.34	0.00	-0.05	-0.30
vowel_6mo							1.00	**0.91**	0.06	-0.04	0.18	-0.12	**0.77**	**0.86**
vowel_9mo								1.00	-0.15	-0.12	0.32	-0.04	**0.65**	**0.83**
tone_6mo									1.00	0.05	-0.12	-0.01	0.22	0.10
tone_9mo										1.00	-0.28	0.23	0.28	-0.23
tone6_HL vs. LL											1.00	-0.21	0.38	0.02
tone6_HR vs MR												1.00	-0.10	-0.08
tone9_HL vs. LL													1.00	0.50
tone9_HR vs MR														1.00

No informative significant correlations were found between the CHT measures and tone or vowel measures. Nevertheless it would be premature to conclude that tone hyperarticulation in Cantonese IDS is not correlated with infants' native speech perception. The results usually used in infant perception analyses generally include only those found in the experimental stages and not the conditioning stage of the experiment. This is partly because the conditioning stage contains a retraining phase and is not a good indicator of infants' discrimination abilities as infants are still being trained to learn the head-turn/reinforcer contingency. With only four infants completing Experimental stage 1, and three completing Experimental stage 2, the results from Experimental stages 1 and 2 here are limited, but may still provide some indication of whether there is a tone hyperarticulation-infant speech perception relationship as in the Liu et al. (2003) study.

6.4 Results for Infants Completing Experimental Stages 1 and 2

Of the four infants who met criterion and participated in Experimental stage 1, one infant subsequently scored 0 for both hits and false positives while scoring 100% for misses and correct rejections during experimental stage 1. Thus this infant simply ceased to make any head-turns, and her results will therefore not be reported.

Several measures of infant discrimination ability were obtained from the CHT procedure as described above. The measure of greatest importance here is the hits-false positives index, as this is the one that Kuhl et al. (1997) used in their study, and it is that one that is most informative about discriminative ability, as it takes into account the problem of the CHT procedure engendering head-turning responses irrespective of the stimuli, by subtracting false positive head-turns from the hits. The hits-false positives index can be thus be conceptualised as a discrimination index (DI). The number of trials to criterion was also used as it may be a sensitive indicator of task difficulty.

6.4.1 *Experimental Stages 1 and 2*

Table 10 shows DIs and number of trials to criterion, along with mothers' IDS vowel and tone triangle areas as a proportion of ADS for the three infants who completed Experimental Stages 1 and 2. Data for only 6 and 9 months vowel and tone triangle areas are shown as these are ages common to both cohort groups.

Table 10. Experimental Stages 1 & 2: DIs, Trials to Criterion, IDS Vowel and Tone Triangle Areas as proportion of ADS at 6 and 9 months IDS

Infant	DI (%)		Trials to Criterion		Tone Triangle Area				Vowel Triangle Area			
	Stage1	Stage2	Stage1	Stage2	6mths/ADS		9mths/ADS		6mths/ADS		9mths/ADS	
M1	0.38	0.51	19.00	30.00	5.05	5.05	1.44	1.44	3.96	3.96	2.49	2.49
M2	0.23	0.29	25.00	30.00	1.36	1.36	1.44	1.44	1.13	1.13	1.38	1.38
F1	0.13	0.10	8.00	30.00	1.81	1.81	1.87	1.87	0.82	0.82	0.92	0.92

All three infants had mothers who hyperarticulated tones, and two infants (M1 and M2) had mothers who also hyperarticulated vowels in IDS. The mother of infant M1 who had the highest DIs in both Experimental Stage 1 and 2 also had the largest tone and vowel triangle areas in IDS, particularly at 6 months. The mother of infant M2 with the next best DIs while she had reduced triangle areas she consistently hyperarticulated tones and vowels in IDS at both 6 and 9 months. It is also interesting to note that the mother of M1 hyperarticulated tones and vowels more at 6 than 9 months, while the other two mothers (of M2 and F1) hyperarticulated more at 9 months (although the mother of F2 did not hyperarticulate vowels at all).

The task appear to be a relatively difficult one based on the number of trials needed to reach criterion; all three infants required the maximum number of trials to reach criterion in experimental stage 2. Nevertheless, the DIs in both Experimental Stages 1 and 2 are comparable, and infants who performed well on one also performed well on the other, suggesting that both pairs of stimuli, the original pair used in Stage 1 and the new pair in Stage 2, are of comparable perceptual difficulty and in the discriminability/hyperarticulation relationship (such as it is) generalises across speech contrasts. Finally, the relative poor performance in Stage 2 may be due to fatigue or loss of interest in the task resulting in all three infants requiring the maximum number of trials to criterion. However, this seems unlikely as there is no obvious loss in accuracy of discrimination responses observed across the three infants in Stage 2.

6.5   Conclusions

Of the 22 infants who participated in the CHT study, three infants participated in Experimental Stages 1 and 2. For these three infants the DI measure of their speech perception ability was related to their mothers' degree of hyperarticulation; those with higher DI scores had mothers who had greater and/or more consistent tone and vowel hyperarticulation. These results, albeit for a very limited sample, appear to support the previous findings by Liu et al. (2003) that there is a positive correlation between mother's vowel hyperarticulation and infant's native consonant discrimination abilities. Thus it is possible to extend the results of the Liu et al. (2003) study to include tones. However, it is important to note that the preliminary speech discrimination results reported here must be replicated with a much larger sample, and that this is the first examination of mothers' tone hyperarticulation in Cantonese IDS and their infants' discrimination of native Cantonese sounds.

Seven

Chapter 7

General Discussion

With a growing body of research showing (1) that vowel hyperarticulation is present in infant- and foreigner-directed speech but not adult- or pet-directed speech (Burnham et al., 2002; Kuhl et al., 1997; Uther et al., 2007); (2) a positive correlation between vowel hyperarticulation and infants' ability to discriminate native speech sounds (Liu et al., 2003); and (3) tone hyperarticulation in tone language IDS (Liu et al., 2007); the studies in this thesis aimed to investigate several hypotheses regarding IDS in tone language Cantonese. The first aim was to investigate tone and vowel hyperarticulation in Cantonese IDS, and it was hypothesised that tones and vowels in Cantonese IDS will be hyperarticulated compared to ADS. The second aim was to document the relative development of tone and vowel hyperarticulation during the first 12 months of infancy. In this regard, developmental research on infant speech perception has shown that infants tune into native vowels between 4 and 6 months (Polka & Werker, 1994), and tones between 6 and 9 months (Mattock & Burnham, 2006). If mothers' productions are related to infant development then vowel hyperarticulation should occur particularly around 6 months and tones around 6 months but before 9 months. These two aims were investigated in Chapters 4 and 5. The third aim was to investigate the relationship between mothers' ton and vowel hyperarticulation on the one hand and their infants' speech discrimination on the other. To this end and infant speech discrimination study was conducted and reported in Chapter 6, another and the mother IDS/child speech perception relationship investigated.

7.1 Tone and Vowels in Cantonese IDS

The general findings from the tone and vowel triangle areas (Chapter 4) and tone and vowel space differentiation analyses (Chapter 5) is that Cantonese mothers hyperarticulate their tones and this emerges early at 3 months, peaks at 6 months before declining again from 9 to 12 months, but that , mothers did not hyperarticulate their vowels at any point over the infant ages that were studied. Further details of each of these results are discussed below.

7.1.1 *Tone Hyperarticulation in Cantonese IDS*

While previous studies have shown modifications of tones in tone language IDS compared with ADS (Kitamura et al., 2002; Liu et al., 2007), the results of the study reported here show the existence of tone hyperarticulation in IDS. Just as vowel hyperarticulation had been observed across various languages, we can now confirm that tones, like other segmental information (vowels and consonants), are better specified in IDS than ADS, at least for Cantonese. In addition, using the tone differentiation index, it has been shown that tone hyperarticulation/tone space expansions is mostly driven by better tone onset and offset differentiation between the high level and lower level tones, while contour tones with similar onsets or offsets are no better differentiated in IDS than ADS. These results have several implications for both infant studies and in tones language research.

7.1.1.1 *Implications for Research in IDS and Tone Languages*

The better specification of tone space information in IDs register tones may indicate that initially specifying the limits of the tone space is sufficient for infants to learn tones. So better specification of tones with similar onsets and offsets in IDS may be a later development beyond the first 12 months of infancy (the period for this study). Future research both in other tone languages such as Thai and Mandarin and in Cantonese IDS beyond the first 12 months of infancy are required to augment the results found here.

In addition to tone triangle areas, tone onset and offset differentiation scores were also used to provide additional information about aspects of the tone space that are manipulated in IDS. This measure was found to provide information about what aspects of tone, tone speakers over- and under-specify in their speech, so it has proved to be a useful tool in studying tone IDS. These results suggest that for other tone languages with a complex tone inventory and multiple level and contour tones such as Thai, tone onset and offset differentiation may be a better measure than overall tone space expansion as it conveys useful additional information. Admittedly, the results suggest that tone onset and offset information maybe less applicable for Mandarin which uses durational properties to signal tone, and this highlights the need for IDS

studies with different tone languages, including those with simple and complex tone inventories, and for the analyses to be tailored to the particular language.

It would be interesting to investigate whether similar changes in tone space and tone onset and offsets can be observed in foreigner-directed speech (FDS) in Cantonese and other tone languages. Now that tone hyperarticulation has been found in IDS, it is reasonable to expect that if tones are realised differently depending on the target audience's learning potential, then tones should also be hyperarticulated in FDS. In fact, it would be interesting to test this hypothesis by manipulating the FDS conditions e.g., whether and how tone speakers would manipulate tones to speakers of another tone language versus speakers of non-tone languages. For example, would tone language speakers specify the limits of tones for foreigners as they do for infants and would this be the same to tone versus non-tone languages foreigners?

7.1.2 *Vowel Hypoarticulation in Cantonese IDS*

The absence of vowel hyperarticulation in Cantonese IDS even at 6 months is unexpected given that previous studies have found vowel hyperarticulation in other languages with 6-month-old infants (Burnham et al., 2002; Kuhl et al., 1997). However the vowel space differentiation index did show a decreasing trend in vowel differentiation from 3 months to later stages of infancy within the first year so, since this is the first time Cantonese IDS has been investigated, and no prior accumulated wisdom is available, it is possible, though unlikely, that vowel hyperarticulation in Cantonese IDS occurs earlier than 3 months. While this is possible, it must also be seriously considered that in tone languages with complex tone inventories, like Cantonese and Thai, vowel information in IDS is not produced in the same manner as in non-tone languages or in tone languages with simple tone inventories such as Mandarin. Studies investigating IDS of other tone languages with complex tone inventories need to be conducted to test this hypothesis and Thai, with three level and two contour tones, would be a great candidate for such a study.

7.1.2.1   Why the lack of Vowel Hyperarticulation in Cantonese IDS?

Some possible reasons for the lack of vowel hyperarticulation in Cantonese IDS have been mentioned in this chapter and in Chapters 4 and 5, but considering that this result is so unexpected, more detailed consideration of this finding is required.

Phonetic Relationship between Tones and Vowels: It must be seriously considered that in Cantonese there is tone hyperarticulation but little if any vowel hyperarticulation in mothers' speech to their infants. This is not altogether out of the question for a variety of reasons. (i) Vowels are the primary vehicles on which tones are carried, so tone hyperarticulation may in some sense incorporate increased articulatory attention to vowels. (ii) Tones are a prominent part of Cantonese as it has six tones, a large number compared to other languages, so it is possible that this tendency to hyperarticulate tones but not vowels is present in Cantonese more than it is in other tone languages with less tones. (iii) As there are less tones than vowels (6 compared to 9 to 18, see sections 1.2.1 and 2.4.2 in Cantonese vowel and tone inventories) mothers may concentrate on tone hyperarticulation for this is less difficult for infant to acquire. Indeed, studies of tone production show that the acquisition of tones appears to be completed early in childhood and before the completion of segmental production (Hua & Dodd, 2000; Li & Thompson, 1978). (iv) Tones have more space to move than vowels, evident by the large tone triangle areas shown in IDS so, in a manner related to (iii) mothers may concentrate on that which is less difficult to teach (and learn). (iv) Tones must to be hyperarticulated in tone languages to set tones apart from pitch information relating to prosodic modifications in IDS that are related to the communication of affective information.

Whether any or all of these hypotheses are valid need to be tested by investigating tone and vowel in IDS of other tone languages with different number of tone, to examine Cantonese tone and vowel IDS in the second year of infancy, and to conduct studies in which the relationship between prosody and tone in IDS and in ADS is investigated in detail.

Infant Directed Facial Gestures take precedence over Vowel Hyperarticulation? At least one previous study, a study of Norwegian IDS, has failed to find vowel hyperarticulation, so it may be instructive to consider this study and the possible reasons for the lack of vowel hyperarticulation. Englund and Behne (2005) investigated Norwegian mothers' IDS to their infants over a period of 6 months beginning shortly after the infant's birth. They found a lack of vowel hyperarticulation in six Norwegian mothers' IDS and suggest the lack of vowel hyperarticulation to be largely due to the fronting of back vowels as part of mothers' affective signalling to her infant (Englund & Behne, 2005). In this regard, in another study examining ID facial gestures, three specific facial gestures were identified as part of affective signalling – "oochie", "wow", and "joy", communicating love, amazement and pleasure respectively (Chong et al., 2003). In the oochie gesture, many mothers were rated as making a "lip pucker" – with rounded and protruding lips. If mothers in this study were also making these lip movements systematically, then it could explain why two distinct bands of F2 frequencies in the IDS /i/ vowel; one could be due to /i/ vowels produced with the lips spread and one with lips rounded as in the "oochie" gesture. The /i/ vowel is the only vowel among the three Cantonese corner vowels that can be produced with lips spread as well as rounded, because the Cantonese /u/ vowel is a very front vowel produced with lip-rounding and the low back /a/ vowel is produced with spread lips they are not very likely to be affected by facial gestures. It would be profitable if future studies would analyse Cantonese vowels produced with different facial gestures to examine how vowel formants may be altered by ID facial gestures. It would also be informative to investigate other languages in which the /i/ vowel can be produced with spread and rounded lips to determine whether F2 values are influenced in the same way as they are in this study.

Methodological Reasons: Another possible explanation for the lack of vowel hyperarticulation is related to a methodological issue. This is the first study to examine the hyperarticulation of two aspects of speech simultaneously (tones and vowels here), and so there is no accumulated experimental wisdom of how this might affect mothers'

behaviour. Mothers may have consciously or unconsciously realised that tone and vowel contrasts were the issues of interest in the study and so have responded to this in some way. Moreover, one aspect of the experimental procedure, the fact that three vowels and six tones were studied meant that mothers were provided with three toys for eliciting the three vowels, but six toys for the six tones. Again, in either a conscious or unconscious manner, this may have resulted in greater attention to tone than vowel items, a conscious or unconscious demand characteristic resulting in tone but not vowel hyperarticulation. While this is considered to be unlikely, the problem can be easily overcome in future research by having the same number of tone and vowel target words.

7.1.3 *Tone and Vowel Changes in IDS during the First Year of Infancy*

The above findings for tone and vowel hyperarticulation entail the conclusion that the production of IDS tone and vowel information in the first 12 months of infancy are necessarily different. Two explanations for these results are considered below.

Tone hyperarticulation appears to be mainly driven by the better differentiation of tone onset and offset of the high level tone from the lower level tones and this is the case right across the first 12 months. The trends analysis of tone triangle areas suggest that tone hyperarticulation peaks at around 6 months before declining after 9 months. This trend is consistent with infant perception data in that mothers are hyperarticulating more between 6 and 9 months when infants show attunement to the tones of their native language (Mattock & Burnham, 2006). The vowel space differentiation measure used here suggests that vowels are overall more differentiated at 3 months compared to 6, 9, and 12 months. Even in the absence of vowel hyperarticulation these results suggest that vowels in IDS are produced differently before and after 3 months, and we know that between 4 and 6 months infants have already begun to tune into the vowel categories of the language they hear around them (Polka & Werker, 1994). Taken together, the vowel and especially the tone results over the first year of infancy are consistent with infant speech perception development – mothers appear to differentiate vowels and tones more when infants are tuning into their native language. When this attunement is sufficiently advanced, mothers appear to reduce the degree of exaggerated

differentiation of at least for tones in IDS, and maybe for vowels. While conclusions regarding mothers' vowel productions need to be treated with some caution here, the results lend support to the hypothesis that the hyperarticulation of segmental (including tonal) information in IDS may be a mechanism that contributes to the bootstrapping of infant language development. Indeed a study of speech perception here suggest this may be the case, as discussed in the next section.

7.1.4   *Is there a Relationship between Tone Hyperarticulation and Consonant Discrimination by Cantonese Infants?*

A Conditioned Head Turn (CHT) study investigating infants' ability to discriminate native Cantonese sounds was conducted with the aim of providing a native language speech discrimination sensitivity score for each participant similar to that obtained in the Liu et al. (2003) study. It was hypothesised that there would be a correlation of CHT scores with tone and vowel hyperarticulation indices, i.e., that there would be a relationship between mothers' degree of IDS and their infants' speech perception ability. However, due to the large attrition rate, only three infants were able to complete the study. While we are unable to comment on any statistical relationship between tone hyperarticulation and infants' native language discrimination abilities, the results for the three infants are informative. In general, infants who performed better on the Cantonese consonant discrimination task had mothers who had larger vowel and tone triangle areas. Future studies may consider using an alternative method to CHT, such as the stimulus alternating technique used in habituation paradigms which may be less susceptible to a large attrition rate (see Appendix E). To the extent that these results show a relationship between perception and production, they support the perception-production link, not only for vowels (Liu et al., 2003) but also for tones.

7.1.4.1   *Differentiation Index as a More Sensitive and Robust Measure of the Vowel and Tone Space*

Finally, it is worth noting that the tone onset and offset differentiation analysis is a useful tool in providing additional information about aspects of tones that tone speakers over- and under-specify in tone IDS. It appears that the tone onset and offset differentiation analysis would also be appropriate to use on other tone languages with

complex tone inventories, that is having multiple level and contour tones, such as Thai, and provides more detailed phonetic information about tones than triangle areas alone. The vowel differentiation index, on the other hand, did not afford any additional clarification regarding vowel dimensions, but this may be due to the lack of vowel hyperarticulation in this study.

7.2   Conclusion

In the longitudinal production study, tone hyperarticulation was found in Cantonese IDS while vowel hyperarticulation was not. Tone hyperarticulation in Cantonese emerges early in infancy at 3 months, peaks at 6 and 9 months and then declines at 12 months. Notably, this peak and decline occurs just as infants are tuning into native language tone categories (Mattock & Burnham, 2006). Moreover there is a strong indication from the results of the three infants who completed the speech perception study that mothers' tone hyperarticulation and infants' speech perception may well be related, just as mothers' vowel hyperarticulation and their infant' speech perception are related. For vowels, while no hyperarticulation in IDS was observed, vowel differentiation among IDS ages was best at 3 months and thus consistent with infant perception studies that have shown infants have tuned into the native vowels by 4 to 6 months (Polka & Werker, 1994). This interpretation needs to be treated with caution, however, as no vowel hyperarticulation was found in Cantonese IDS here.

The complexity of learning vowels, consonants, and tones in a tone language, such as Cantonese, may exert a different set of demands on both the infant and the way speech input is produced for the infant. The results in this study highlight the need for studies to be conducted that determine Cantonese language development milestones rather than relying on Anglocentric data. Cantonese, as a tone language, may exert a different set of demands on the first language learner such that infants' attunement to native vowels, consonants and tones, may follow a very different developmental trajectory to that in non-tone languages and even to that in tone languages with simpler tone systems (e.g., Mandarin). More light may be shed on the reasons behind some of the discrepancies between the results of this study and the existing body of knowledge

on IDS when further studies on tone IDS are conducted and infant language development milestones are established for Cantonese and other tone languages.

The research in this thesis is a first step in the study of IDS in Cantonese and will be fruitfully augmented by both perception and production research to chart Cantonese infants' developmental milestones, instead of simply applying Anglocentric data to tone language infants.

Eight

Chapter 8

Full Reference List

Abramson, A. S. (1979a). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie J. A. Henderson*. Bangkok: Indigenous Languages of Thailand Research Project.

Abramson, A. S. (1979b). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research* (pp. 127-134): Academic Press.

Abramson, A. S. (1986). The Thai tone space. *Haskin Laboratories: Status Report on Speech Research SR-85, 85*, 105-114.

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development, 19*(4), 401-418.

Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *The Journal of the Acoustical Society of America, 105*, 1095.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience. *Child Development, 52*, 1135-1145.

Barrie, M. (2003). Contrast in Cantonese vowels. *Toronto Working Papers in Linguistics, 20*, 1-19.

Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *Acoustical Society of America, 116*(3), 1739–1748.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter & Co.

Bernard, J. (1970). Towards the acoustic specification of Australian English. *Zeitschrift für Phonetik, 2*(3), 113–128.

Bernard, J., & Mannell, R. (1986). *A study of /h_d/ words in Australian English*: Working Papers of the Speech, Hearing and Language Research Centre.

Bird, K., Hadzi-Pavlovic, D., & Isaac, A. (1990). PSY: A program for contrast analysis (Version 5.1.2): © School of Psychology, University of New South Wales.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America, 67*, 648.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 4.5.13) [Computer Program].

Bohannon III, J. N., & Marquis, A. L. (1977). Children's Control of Adult Speech. *Child Development, 48*(3), 1002-1008.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101* (4), 2299-2310.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science, 5*(1), 72-83.

Bredberg, G. (1985). The anatomy of the developing ear. In S. E. Trehub & B. Schneider (Eds.), *Auditory development in infancy*. New York/London: Plenum Press.

Brunelle, M. (2003). *Tone Coarticulation in Northern Vietnamese.* Paper presented at the Proceedings of the 15th International Conference of phonetic Sciences.

Burnham, D., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language, 18*(2), 231-260.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., et al. (1996). *Perception of lexical tone across languages: Evidence for a linguistic mode of processing.* Paper presented at the ICSLP 96 Proceedings, Fourth International Conference on Spoken Language Philadelphia, PA, USA.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 259-280). Amsterdam: John Benjamins.

Chan, W. S., Chuang, C. K., & Wang, W. S.-Y. (1975). Crosslanguage study of categorical perception for lexical tone. *Journal of the Acoustical Society of America, 58*, 119.

Chao, Y. R. (1930). A system of "tone letters". *Le Maître Phonétique, 30*, 24-27.

Chao, Y. R. (1947). *Cantonese primer*. New York: Greenwood Press.

Chen, M. Y. (2000). *Tone Sandhi*. Cambridge, UK: Cambridge University Press.

Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three Facial Expressions Mothers Direct to Their Infants. *Infant and Child Development, 12*, 211–232.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America, 95*(3), 1570-1580.

Clark, J., & Yallop, C. (1995). *An introduction to phonetics and phonology (2nd Ed.)*. Oxford, UK: Blackwell Publishers.

Cohen, L. B. (2004). Uses and Misuses of Habituation and Related Preference Paradigms. *Infant and Child Development, 13*, 349-352.

Colombo, J., & Horowitz, F. D. (1986). Infants' Attentional Responses to Frequency Modulated Sweeps. *Child Development, 57*(2), 287.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development, 20*(4), 477-488.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584-1595.

Cooper, R. P., & Aslin, R. N. (1994). Developmental Differences in Infant Attention to the Spectral Properties of Infant-directed Speech. *Child Development, 65*(6), 1663-1677.

Cox, F. (1996). *An Acoustic Analysis of Vowel Variation in Australian English. Unpublished PhD Thesis.* Macquarie University, Sydney.

Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association, 37*(03), 341-350.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of phonetics, 21*(1-2), 103-108.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

Ellen, G. B., & Anderson, A. H. (1983). The Unintelligibility of Speech to Children. *Journal of Child Language, 10*(2), 265-292.

Engle, M. E. (1979). *Do fathers speak motherese? An analysis of the language environments of young children.* ProQuest Information & Learning, US.

Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language, 25*(2), 219-234.

Englund, K., & Behne, D. (2005). Infant Directed Speech in Natural Interaction-Norwegian Vowel Quantity and Quality. *Journal of Psycholinguistic Research, 34*, 259-280.

Englund, K., & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development, 15*(2), 139-160.

Fant, G. (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), *Manual of Phonetics* (pp. 173-276). Amsterdam: North-Holland Publication. Co.

Fantz, R. L. (1963). Pattern Vision in Newborn Infants. *Science, 140*(3564), 296-297.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181-195.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development, 60*(6), 1497-1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*(3), 657-674.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279-293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477-501.

Field, A. P. (2005). *Discovering Statistics Using SPSS*: Sage Publications Inc.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*, 1029-1044.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*(17), 1-189.

Furrow, D., & Nelson, K. (1986). A further look at the motherese hypothesis: A reply to Gleitman, Newport & Gleitman. *Journal of Child Language, 13*(1), 163-176.

Gandour, J. T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics, 9*, 20-36.

Gandour, J. T. (1983). Tone perception in far east languages. *Journal of Phonetics, 11*, 149-175.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception a multidimensional scaling investigation. *language and Speech, 21*(1), 1-33.

Gandour, J. T., Weinberg, B., Petty, S. H., & Dardarananda, R. (1988). Tone in Thai Laryngeal Speech. *Journal of Speech and Hearing Disorders, 53*, 23-29.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11*(1), 43-79.

Goldsmith, M. (1994). Tone languages. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics*. Oxford: Pergamon Press.

Golinkoff, R. M., & Alioto, A. (1995). Infant-Directed Speech Facilitates Lexical Learning in Adults Hearing Chinese: Implications for Language Acquisition. *Journal of Child Language, 22*(3), 703-726.

Golinkoff, R. M., & Ames, G. J. (1979). A Comparison of Fathers' and Mothers' Speech with Their Young Children. *Child Development, 50*(1), 28-32.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal Speech to Infants in a Tonal Language: Support for Universal Prosodic Features in Motherese. *Source: Developmental Psychology, 24*(1), 14-20.

Han, M. S., & Kim, K. O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics, 2*, 223-232.

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics, 17*(2), 155-184.

Hashimoto, O.-K. Y. (1972). *Studies in Yue Dialects 1: Phonology of Cantonese* (Vol. 1). Cambridge: Cambridge University Press.

Hepper, P. G., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive Infant Psychology, 11*(3), 147.

Hirsh-Pasek, K., & Treiman, R. (1982). Doggerel: Motherese in a new context. *Journal of Child Language, 9*(1), 229-237.

Holzrichter, A. S. (2000). Language acquisition by eye. In C. Chamberlain, J. P. Morford & R. I. Mayberry (Eds.), *Child-directed signing in American Sign Language* (pp. 25-40). Mahwah, NJ: Lawrence Erlbaum Associates.

Hombert, J. M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (Ed.), *Tone: A linguistic survey*. New York: Academic.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations of the development of tones. *Language, 55*(1), 37-58.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.): Duxbury Press Belmont, Calif.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*, 3-42.

Ikeda, Y., & Masataka, N. (1999). A variable that may affect individual differences in the child-directed speech of Japanese women. *Japanese Psychological Research, 41*, 203-208.

Iverson, J. M., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14*(1), 57-75.

Kaplan, P. S., Bachorowski, J.-A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Science, 13*(3), 268.

Kaplan, P. S., Dungan, J. K., & Zinser, M. C. (2004). Infants of Chronically Depressed Mothers Learn in Response to Male, But Not Female, Infant-Directed Speech. *Developmental Psychology, 40*(2), 140-148.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., & Cooper, R. P. (1995). Dishabituation of visual attention by infant- versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior and Development, 18*, 209-223.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*(1), 55-68.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America, 72*(2), 353-365.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104-117.

Kitamura, C., & Burnham, D. (1998). The infant's response to vocal affect in maternal speech. In C. Rovee-Collier (Ed.), *Advances in Infancy Research* (Vol. 12, pp. 221-236).

Kitamura, C., & Burnham, D. (2003). Pitch and Communicative Intent in Mother's Speech: Adjustments for Age and Sex in the First Year. *Infancy, 4*(1), 85-110.

143

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior & Development, 24*, 372-392.

Kuhl, P. K. (1978). Predispositions for the perception of speech-sound categories: A specifies-specific phonomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities: Early behavioral assessment*. Baltimore: University Park Press.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science, 277*(5326), 684-686.

Ladefoged, P. (1982). *A course in Phonetics (2nd ed.).* New York: Harcourt Brace Joanovich.

Ladefoged, P. (2001a). *A course in Phonetics (4th ed.)* (Fourth Edition. ed.). Boston: Heinle & Heinle, a division of Thomson Learning, Inc.

Ladefoged, P. (2001b). *Vowels and Consonants*. Oxford: Blackwell Publishers Ltd.

Ladefoged, P. (2003). *Phonetic Data Analysis: An introduction to fiedlwork and instrumental techniques*: Blackwell Publishing Ltd.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross linguistic comparison. *Journal of Psycholinguistic research, 25*(5), 527-542.

Lennes, M. (2002). Mietta's Praat scripts.

Li, C. M., & Thompson, S. A. (1978). The acquisition of Tone. In V. Fromkin (Ed.), *Tone: A linguistic survey*. New York: Academic.

Li, P.-Y. (2004). *Perceptual analysis of the six contrastive tones in Cantonese.* Paper presented at the Inter-nations Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1-F10.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43*(4), 912-917.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2004). *The effects of intonation patterns on lexical tone production in Cantonese*. Paper presented at the International symposium on tonal aspects of languages: With an emphasis on Tone Languages.

Masataka, N. (1992). Motherese in a signed language. *Infant Behavior and Development, 15*(4), 453-460.

Masataka, N. (1996). Perception of motherese in a signed language by 6-month-old deaf infants. *Developmental Psychology, 32*, 874-879.

Masataka, N. (1998). Perception of Motherese in Japanese Sign Language by 6-Month-Old Hearing Infants. *Developmental Psychology, 34*(2), 241-246.

Massaro, D. W., Cohen, M. M., & Tseng, C.-Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13*, 267-289.

Mathews, S., & Yip, V. (1994). *Cantonese: a comprehensive grammar*. NY: Routledge.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-264.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143-178.

Mitchell, R. W. (2000). Americans' Talk to Dogs: Similarities and Differences With Talk to Infants. *Research on Language and Social Interaction, 34*(2), 183-210.

Mitchell, R. W., & Edmonson, E. (1999). Functions of repetitive talk to dogs during play: Control, conversation, or planning. *Society and Animals, 7*, 55–81.

Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., & Charnvivit, P. (2002). *Perception of tones and vowel quantity in Thai.* Paper presented at the Proceedings of ICSLP, Denver, USA.

Mok, P. K. P., & Hawkins, S. (2004). *Effects of phonemic vs. allophonic density and stress on vowel-to-vowel coarticulation in Cantonese and Beijing Mandarin.* Paper presented at the 2004 International Symposium on Chinese Spoken Language Processing.

Moore, D. S., Spence, M. J., & Katz, G. S. (1997). Six-Month-Olds' Categorization of Natural Infant-Directed Utterances. *Developmental Psychology, 33*(6), 980-989.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition.* Cambridge: Cambridge University Press.

Nwokah, E. E. (1987). Maidese versus motherese--is the language input of child and adult caregivers similar? *Lang Speech, 30*(3), 213-237.

Ohala, J. J. (1978). Production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5-39). New York: Academic.

Papousek, M., & Bethesda, M. D. (1987). *Melodies in Motherese in Tonal and Nontonal Languages: Mandarin Chinese, Caucasian American, and German.*

Papousek, M., Bornstein, M. H., Nuzzo, C., Papousek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development, 13*(4), 539-545.

Papousek, M., & Hwang, S.-F. C. (1992). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics, 12*(4), 481-504.

Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development, 14*(4), 415-440.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15*(3), 325-345.

Phillips, J. R. (1973). Syntax and Vocabulary of Mothers' Speech to Young Children: Age and Sex Comparisons. *Child Development, 44*(1), 182-185.

Pisoni, D. B. (1973). Perceptual Processing Time for Consonants and Vowels. *The Journal of the Acoustical Society of America, 53*, 369.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustic Society of America, 61*(5), 1352-1361.

Polka, L., & Werker, J. F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435.

Querleu, D., Renard, X., & Versyp, F. (1981). Les perceptionsauditives du foetus humain. *Médecine et Hygiène, 39*(2101-2110).

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crepin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 28*(3), 191-212.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., Vervoort, P., & Crepin, G. (1986). Commentary. Can the fetus listen and learn. *BJOG: An International Journal of Obstetrics and Gynaecology, 93*(4), 411-412.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349.

Read, C., Yun-Fei, Z., Hong-Yin, N. I. E., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31-44.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication, 6*(4), 343-352.

Rose, P. (1993). A linguistic-phonetic acoustic analysis of Shanghai tones. *Australian Journal of Linguistics, 13*, 185-220.

Rose, P. (2000). *Hong Kong Cantonese citation tone acoustics: A linguistic tonetic study*. Paper presented at the Proceedings of 8th Australian International Conference of Speech Science and Technology, Canberra.

Ross, E. D., Edmondson, J. A., & Seiburt, C. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics, 14*, 283-302.

Rvachew, S., Mattock, K., Polka, L., & Menard, L. (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *Journal of the Acoustical Society of America, 120*, 2250-2259.

Sach, J., & Devin, J. (1976). Young children's use of age appropriate speech styles in social interaction and role-playing. *Journal of Child Language, 3*, 81-98.

Scarborough, H., & Wyckoff, J. (1986). Mother, I'd still rather do it myself: Some further non-effects of 'motherese'. *Journal of Child Language, 13*(2), 431-437.

Schwanhäußer, B., & Burnham, D. (2005). *Lexical tone and pitch perception in tone and non-tone language speakers*. Paper presented at the In Proceedings of the 9th European Conference on Speech Communication and Technology, Bonn, Germany.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*(2), 145-156.

Shepard, R. N. (1978). The circumplex and related topological manifolds in the study of perception. *Theory construction and data analysis in the behavioral sciences*, 29-80.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy, 3*(3), 365.

Snow, C. E. (1972). Mothers' Speech to Children Learning Language. *Child Development, 43*(2), 549-565.

Snow, C. E., & Ferguson, C. A. (Eds.). (1977). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and nontone language. *Journal of Chinese Linguistics, 21*(1), 143-163.

Stevens, K. N. (1999). *Acoustic phonetics*. Massachusetts: Massachusetts Institute of Technology.

Sullivan, J. W., & Horowitz, F. D. (1983). The Effects of Intonation on Infant Attention: The Role of the Rising Intonation Contour. *Journal of Child Language, 10*(3), 521-534.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics (4th Ed.)*. MA: USA: Allyn & Bacon: A Pearce Education Company.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579-590.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy, 7*(1), 53-71.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion? *Psychological Science, 11*(3), 188-195.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review, 9*, 335-340.

Trainor, L. J., & Zacharias, C. A. (1998). Infants prefer higher-pitched singing. *Infant Behavior and Development, 21*(4), 799-805.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49*(1), 2-7.

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica, 33*, 368-392.

Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica, 34*, 93-107.

Wang, C., & Seneff, S. (2000). *Improved Tone Recognition by Normalizing for Coarticulation and Intonation Effects.* Paper presented at the Sixth International Conference on Spoken Language Processing, Beijing, China.

Wang, W. S.-Y. (1967). Phonological Features of Tone *International Journal of American Linguistics, 33*(2), 93-105.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences, 280*(1), 61-72.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America, 113*(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*(6), 3649-3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning, 54*(4), 681–712.

Weppelman, T. L., Bostow, A., Schiffer, R., Elbert-Perez, E., & Newman, R. S. (2003). Children's use of the prosodic characteristics of infant-directed speech. *Language & Communication, 23*(1), 63-80.

Werker, J. F., Gilbert, J. H. V., Humphrey, G. K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*, 349-355.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology Revue Canadienne de Psychologie, 43*(2), 230-246.

Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development, 17*(3), 323-333.

Werker, J. F., & Polka, L. (1993). Developmental changes in speech perception. New challenges and new directions. *Journal of Phonetics, 21*(83-101).

Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedures as a method for assessing infant speech perception. *Early Development & Parenting, 6*(3-4), 171-178.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37*(2), 278-286.

Werker, J. F., & Tees, R. C. (1984a). cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*, 49-63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*(6), 1866-1878.

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2004). *Polly wanna cracker? On talking to infants and parrots. Refereed poster presentation.* Paper presented at the ICIS.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*(2240-2253).

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

Zee, E. (1999). *An acoustical analysis of the diphthongs in Cantonese.* Paper presented at the Conference proceedings from ICPhS, San Francisco, USA.

Zee, E. (2003). *Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers*. Paper presented at the 15th ICPhS.

Nine

Appendices

Appendix A: Recruitment Materials

Appendix A.1 : Recruitment Brochure

OUR SPONSORS

PENLINE Mon Ami • HUGGIES • PHOTO CORPORATION GROUP • PIGEON • REAL FOODS • Johnson & Johnson • MACMILLAN Pan Macmillan Australia • KRAFT • Safety 1st • practical parenting • Baby Wonder • Bébelle • HEINZ • CUBASH • Amcal • the first years • Happy Baby • golden circle

RESEARCH PARTNERS

Karitane

ST JOHN OF GOD HEALTH SERVICES

MARCS BABYLAB
UNIVERSITY OF WESTERN SYDNEY

BANKSTOWN CAMPUS
Building 1
Bullecourt Avenue
Milperra 2560
Phone: 9772 6448
www.uws.edu.au/marcs/

THE TEAM:
Prof. Denis Burnham
Dr Christine Kitamura
Ms Nan Xu (徐囡/Xu Nan)

您的嬰兒2至9個月大嗎?

媽咪,我可以上大學嗎??

對有意幫助者,我們贈送
✓ 贊助商提供的嬰兒精品
✓ 特制 T 恤上衣
✓ 嬰兒科學家科研獎狀
✓ 您的交通費 ($20)

University of Western Sydney
MARCS BABYLAB

馬克斯聰覺實驗室(MARCS)是西雪梨大學其中一個研究中心,致力於研究嬰兒、孩子和成人對語言、音樂和其他有聲現象的反應。在MARCS進行的研究重點是語言和交流技能的早期發展。

MARCS的主任丹尼斯本威教授說:"我們正在進行一項關於語言和音樂對嬰兒的影響、作用及其重要性的研究,當前進行的實驗是整個研究計劃的一部分。父母對嬰兒說話的方式,嬰兒反應的方式,對孩子的早期發育都同樣地重要。"我們研究嬰兒怎樣認識帶聲調語言的發音,以及何時能分辨出不同的發音。

您的寶寶
一出生就可以來登記!

來參加嬰兒科研
和您的寶寶一起來開心
還得到

✓ 嬰兒精品
✓ 一份特制的禮物
✓ 嬰兒科學家科研獎狀
✓ 交通費 ($20)

預知詳情或登記參加科研,請電: 9772 6448 (Nan/囡)或者填好旁邊的表格,按表格上地址寄回(無須付郵資)。

我對這些研究很感興趣
母親的姓名 ....
父親的姓名 ....
電話 ....
手機 ....
電子郵箱 ....
地址 ....
.... 郵碼 ....
寶寶的姓名 ....
寶寶的出生日期 ....性別 ....
家庭里使用的語言 ....
你從哪裡得悉我們的研究/取得這本小冊子? ....

回郵地址:
MARCS BabyLab,
Bankstown Campus - Building 24
University of Western Sydney
Locked Bag 1797
Reply Paid 70475
PENRITH SOUTH DC NSW 1797

153

Appendix A.2: Recruitment Advertisement

Appendix B: Praat Scripts

Appendix B.1 : Exporting F0 information

#Does batch processing of file duration and F0 statistics for adult female voice

#This script works all .Pitch files within a specified directory. The input must be a file which consists only of a vowel whose duration,

pitch and formants are to be measured.

#For each file processed, the script returns one line of these ordered values to file pitch.csv: filename, duration of the file, mean F0, standard deviation of F0, maximum F0,

minimum F0, the range in F0 (max - min), the time of max F0 normalised as proportion of file length, the time of min F0

normalised as proportion of file length, the timepoint in seconds for max F0, the timepoint in seconds for min F0, the time

in seconds 15 ms after start of file, the time in seconds at midpoint, the time in seconds 15 ms before end of file,

F0 estimate at 15 ms into file, F0 estimate at midpoint, F0 estimate at 15 ms before end of file.

In all values output, F0 and formants are expressed to nearest Hz, and time is in seconds (to nearest ms).

#Reads all files in directory into object list

directory$ = "C:\Documents and Settings\Desktop\Small Sound Files\"

Create Strings as file list... list 'directory$'/*.wav

numberOfFiles = Get number of strings

for ifile to numberOfFiles

   select Strings list

   fileName$ = Get string... ifile

   Read from file... 'directory$'/'fileName$'

#Get pitch analysis with min F0 of 75 Hz and max F0 of 800 Hz, i.e. appropriate for adult female voice

To Pitch... 0.01 125 800

#Get duration of the vowel file and its starting and ending times (in seconds)

vduration = Get duration

#Get pitch statistics for the vowel file (in Hz)

maxF0 = Get maximum... 0 0 Hertz Parabolic

maxF0time = Get time of maximum... 0 0 Hertz Parabolic

minF0 = Get minimum... 0 0 Hertz Parabolic

rangeF0 = maxF0 - minF0

minF0time = Get time of minimum... 0 0 Hertz Parabolic

meanF0 = Get mean... 0 0 Hertz

stdevF0 = Get standard deviation... 0 0 Hertz

#Express times for max and min F0 as ratio of file length

maxF0timeratio = maxF0time / vduration

minF0timeratio = minF0time / vduration

```
#Get time and F0 estimates for 15 ms past start, for midpoint, and for 15 ms before end of vowel
 startpoint = 0.015
 midpoint = vduration / 2
 endpoint = vduration - 0.015
startF0 = Get value at time... startpoint Hertz Linear
midF0 = Get value at time... midpoint Hertz Linear
endF0 = Get value at time... endpoint Hertz Linear
#Write pitch results to file
 fileappend  pitch.csv 'fileName$' , 'vduration:3', 'meanF0:0' , 'stdevF0:0' , 'maxF0:0' , 'minF0:0' ,
'rangeF0:0' , 'maxF0timeratio:3' , 'minF0timeratio:3' , 'maxF0time:3' , 'minF0time:3' , 'startpoint:3' ,
'midpoint:3' , 'endpoint:3' , 'startF0:0' , 'midF0:0' , 'endF0:0' 'newline$'
endfor
#Remove all objects from objects window
select all
Remove
```

Appendix B.2 : Exporting F1, F2 and F3 information

#For each file processed, the script returns these ordered values to file formant.csv: filename, mean F1, mean F2, mean F3

F1 at startpoint, F1 at midpoint, F1 at endpoint, F2 at startpoint, F2 at midpoint, F2 at endpoint, F3 at startpoint, F3 at

midpoint, F3 at endpoint.

#Reads all files in directory into object list

directory$ = "C:\Documents and Settings\Desktop\Small Sound Files"

Create Strings as file list... list 'directory$'/*.wav

numberOfFiles = Get number of strings

for ifile to numberOfFiles

  select Strings list

  fileName$ = Get string... ifile

  Read from file... 'directory$'/'fileName$'

#Read name of selected Sound into variable

audio$ = selected$ ("Sound")

#Get pitch analysis with min F0 of 100 Hz and max F0 of 800 Hz, i.e. appropriate for adult female voice

To Pitch... 0.01 125 800

#Get duration of the vowel file and its starting and ending times (in seconds)

vduration = Get duration

#Get pitch statistics for the vowel file (in Hz)

maxF0 = Get maximum... 0 0 Hertz Parabolic

maxF0time = Get time of maximum... 0 0 Hertz Parabolic

minF0 = Get minimum... 0 0 Hertz Parabolic

rangeF0 = maxF0 - minF0

minF0time = Get time of minimum... 0 0 Hertz Parabolic

meanF0 = Get mean... 0 0 Hertz

stdevF0 = Get standard deviation... 0 0 Hertz

#Express times for max and min F0 as ratio of file length

maxF0timeratio = maxF0time / vduration

minF0timeratio = minF0time / vduration

#Get time and F0 estimates for 25 ms past start, for midpoint, and for 25 ms before end of vowel

 startpoint = 0.025

 midpoint = vduration / 2

 endpoint = vduration - 0.025

startF0 = Get value at time... startpoint Hertz Linear

midF0 = Get value at time... midpoint Hertz Linear

endF0 = Get value at time... endpoint Hertz Linear

#Select the object whose filename is entered into variable called audio

157

select Sound 'audio$'

#Downsamples file to 11000 Hz sampling rate (appropriate for 5 formant analysis of adult female) and does formant analysis.

Resample... 11000 50

To Formant (burg)... 0.01 5 5500 0.025 50

#Gets F1 values in Hertz at 25 ms into file, midpoint of file, and 25 ms before end of file.

startF1 = Get value at time... 1 startpoint Hertz Linear

midF1 = Get value at time... 1 midpoint Hertz Linear

endF1 = Get value at time... 1 endpoint Hertz Linear

#Gets F2 values in Hertz at 25 ms into file, midpoint of file, and 25 ms before end of file.

startF2 = Get value at time... 2 startpoint Hertz Linear

midF2 = Get value at time... 2 midpoint Hertz Linear

endF2 = Get value at time... 2 endpoint Hertz Linear

#Gets F3 values in Hertz at 25 ms into file, midpoint of file, and 25 ms before end of file.

startF3 = Get value at time... 3 startpoint Hertz Linear

midF3 = Get value at time... 3 midpoint Hertz Linear

endF3 = Get value at time... 3 endpoint Hertz Linear

#Gets mean values for F1, F2 and F3 for the file.

meanF1 = Get mean... 1 0 0 Hertz

meanF2 = Get mean... 2 0 0 Hertz

meanF3 = Get mean... 3 0 0 Hertz

#Write formant results to file

fileappend /formant.csv   'fileName$' , 'meanF1:0' , 'meanF2:0' , 'meanF3:0' , 'startF1:0' , 'midF1:0' , 'endF1:0' , 'startF2:0' , 'midF2:0' , 'endF2:0' , 'startF3:0' , 'midF3:0' , 'endF3:0' 'newline$'

endfor

#Remove all objects from objects window

select all

Remove

Appendix C: Administrative Forms

Appendix C.1 : Participant Information Sheet

Infant and Parent Information Statement 科研说明

Title of Study: Infant-, foreigner-, and adult-directed speech in three tone languages: Mandarin, Cantonese and Thai.

This study investigates features of how people talk to infants, and specifically at 3-, 6-, and 9-months-old. In this study your infant will lie comfortably in their own pram/baby capsule so that s/he faces you, in a quiet room. You will be asked to include a list of words in your conversation or story telling to your infant and your speech will be recorded in 2 sessions of around 10 minutes each when your baby is 3-, 6-, and 9-months-old.
该科研主要研究父母对与 3，6，和 9 个月的婴儿的通话方式。此科研需要在婴儿 3，6，和 9 个月时在您和婴儿合适的时间录音。每次需要您对婴儿进行 2 个 15 分钟的谈话。

When your child is 9 months of age, you will be asked to come into the university so that your infant could participate in a speech discrimination task. Your infant will be seated on your lap in a quiet room and listening to some Cantonese sounds. This task is of a brief duration around 5mins per session with a maximum of around 5 sessions.
当您的婴儿 9 个月时，您的婴儿需要到西雪梨大学在 Bankstown 的 MARCS 聪学科研所。 需要您的婴儿座在您的双膝上会听到一系列广东话的声音。每次 5 分钟需要 5 次。

Between sessions as much time as needed will be given for feeding your child, changing their nappy, sleeping etc, so that your child remains happy, attentive and alert. You are free to interrupt or terminate any session or from the study if your child becomes distressed, or if you whish to terminate for any undisclosed reason. If you choose to terminate a session or discontinue the study, this will not affect your current or future relationship with the University of Western Sydney.
如果您或婴儿需要时间休息，喂奶，换尿布，午睡，或其他原因您可在任何时间停止录音或科研。 您可以在任何时间无原因的退出科研， 您和西雪梨大学或 MARCS 科研所的关系不会得到任何影响。

There are no known risks involved in this study. The results will assist us in further understanding the function and purpose of speech directed to infants and its role in early language development. Your details and that of your infant's will be kept confidential and only those researchers listed below and those authorised by the lab director will have access to that information. If your results are published in professional journals and/or conferences your anonymity will be preserved (by coding of participant numbers not using names) so that neither you nor your infant can be identified.
该科研对您的婴儿无任何危害。该科研成果会提高我们对婴儿发育的认识。若该科研成果会在有关科研杂志或科研会议上发表，我们不会发表您和您婴儿身份。

Before signing the consent form (over), please take time to ask Nan Xu any questions that you may have. If you have a question at a later time please call Nan Xu on (02) 9772 6535 or 0403 005 116. This study if supervised by Prof. D. K. Burnham Director if MARCS Auditory Laboratories Tel. 02 9772 6681.
您若有意帮助该科研项目，请在家长同意信上签字。在签字前您若有任何问题请在签字前向徐囡询问。若您以后有问题请打徐囡（02）9772 6535 询问。该科研是由 Burnham 教授兼 MARCS 聪学科研所理事长指导的 （电 02 9772 6681）。

NOTE: This study has been approved by the University of Western Sydney Human Research Ethics Committee or Panel (indicate Committee or Panel). The Approval Number is HERP05/039 If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee/Panel through the Research Ethics Officers (tel: 02 4736 0883 or 4736 0884). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome. 该科研得到西雪梨大学道德委员会的批准， 批准号码 HERP05/039。您若对该科研有任何不满请电 02 4736 0883 或 02 4736 0884。

Appendix C.2: Guardian Consent Form

PARENTAL (OR GUARDIAN) CONSENT FORM – for infant participant

1. I 我 _____ agree to allow my 同意我 _____ months old 个月的，baby 婴儿 _____, to participate in the study as mentioned in the information statement 参加该科研。

2. I acknowledge that I have read the Information Statement, which explains the aims of the experiment and the nature and the possible risks of the investigation, and the statement has been explained to me to my satisfaction. 我读过也明白科研说明。

3. Before signing this Consent Form, I have been given the opportunity of asking any questions relating to any possible physical and mental harm the child might suffer as a result of participation and I have received satisfactory answers.
在签字前我得到询问问题的时间也得到满意的答复。

I understand that I can withdraw the child from the experiment at any time without prejudice to my or the child's relationship to University of Western Sydney.
我明白我可以任何时间把我的婴儿退出科研，我和我的婴 儿与西雪梨大学或 MARCS 科研实的关系不会得到任何 影响。

5. I agree that research data gathered from the results of the study may be published, provided that neither I nor the child can be identified.
我同意该科研成果可以被发表， 但我和婴儿的身份不会被发表。

6. I understand that if I have any questions relating to the child's participation in this research, I may contact Nan Xu on 02 9772 6535 or 0403 005 116 who will be happy to answer them.
我明白我若对于该科研有任何问题我可以给徐囡打电 02 9772 6535 或手提 0403 005 116。

7. I acknowledge receipt of a copy of the Information Statement.
我承认我得到一份科研说明。

NOTE: This study has been approved by the University of Western Sydney Human Research Ethics Committee or Panel (indicate Committee or Panel). The Approval Number is HERP05/039 If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee/Panel through the Research Ethics Officers (tel: 02 4736 0883 or 4736 0884). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome. 该科研得到西雪梨大学道德委员会的批准， 批准号码 HERP05/039。您若对该科研有任何不满请电 02 4736 0883 或 02 4736 0884。

Signature of parent/guardian 家长名 _____

Please PRINT name 家长姓 _____

Date 日期 _____

Appendix C.3: Infant Details Form

Infant Details Form

Your baby's name/婴儿姓名 _____

Baby's gender/婴儿性别:　　　男　　　女

Your name/母亲姓名: _____

Phone Number/电话: _____

Baby's date of birth/婴儿出生日: _____

Baby born to full term?　　　　　　Yes　　or　　No （please circle）

婴儿是足月（38-42 周）出生吗?　　是　　否

If you answered 'No' please specify number of weeks/

如不是足月，请表明婴儿出生的周数 _____

Feeding: Breast fed or Bottle fed/贵子吃 母乳 或 奶粉？_____

Baby's birth weight/婴儿出生体重: _____

Baby's current weight/婴儿如今体重: _____

Language(s) spoken at home/家庭用语言:　　1 _____

　　　　　　　　　　　　　　　　　　　　2 _____

　　　　　　　　　　　　　　　　　　　　3 _____

Has baby suffered from any ear infections?　　Yes　　or　　No　　(please circle)

婴儿有过耳炎吗 （包括中和内耳）?　　　　有　　　　　没有

If you answered 'Yes', how many? _____

有过几次? _____

Have you had your baby's hearing tested?　Yes　　or　　No　　(please circle)

婴儿作过听力检查吗?　　　　　　　　有　　　　没有

Does baby have any known hearing difficulties?　　Yes　　or　　No　　(please circle)

婴儿有听力缺欠吗?　　　　　　　　　　有　　　　没有

Is there a history of hearing difficulties in baby's family?　　Yes　　or　　No (please circle)

婴儿家族有听力缺欠吗?　　　　　　　　　　　　有　　　　没有

Appendix C.4: Mothers' Language Exposure and Infants' Language Input Form

<div style="border:1px solid black; text-align:center;">

QUESTIONNAIRE

</div>

谢谢您有意帮助我们对婴儿语言发展的研究。 我们先需要关于您和婴儿的有关语言方面的信息，请您帮助。您所提供的信息会保密的。您的姓名和身份不会留下纪录。请您认真填写所于下列问题，这对我们了解婴儿的语言发展非常重要。

有关您的传记/BIOGRAPHICAL DETAILS (PARENT)

1.　　Country of Birth/ 您出生的国家: _____

2.　　If you were not born in Australia, how old were you when you arrived in Australia? 如果您不是出生在澳大利亚，您来澳的年龄？ _____

3.　　How long have you lived in Australia?　　_____year (s), _____month(s)
我居住澳洲_____年，_____月。

4.　　What is the official language of your country of origin?
我出生国的官方语言是:_____

5.　　What is your native language (i.e. the language you learned first)?
我的母语是:　　_____

6.　　Which language(s) have you spoken since before age 5?
我五岁前会这些语言(包括方言): _____

7.　　When you talk to yourself in your head, which language do you use?
当您思考问题时您用那种语言?_____

8.　　Please specify all languages that you began to learn after the age of 5 (English included).
请您把您五岁后学的所有语言包括方言写下。

_____

_____

10. Did you live in any other English-speaking countries before you came to Australia?
   Yes   No

除澳洲外，您在其它英语系国家居住过吗？   有   没有

If you answered Yes to question 12, how old were you when you first arrived in an English-speaking country? _____years, _____months

若有，请您写下您最先到达英语系国的年龄：___岁_____月。

11. Did you study/learn English before you arrived in Australia?   Yes   No

您来澳前学过英语吗？   是   否

If you answered Yes to question 13, how old were you when you began to study/learn English? __
_____year(s), _____month(s)

您若是学过英语，请表明您从___岁,_____月开学英语？

For how long did you study English before your arrival in Australia?
_____year(s), _____month(s)

您在来澳前学过_____年，___月英语？

12. Have you left Australia to go overseas since you first arrived here?   Yes   No

您来澳后是否曾离开过澳洲？   是   否

If you answered Yes to question 14, how long did you spend overseas and what language(s) did you speak while overseas?
_____year(s), _____month(s), _____week(s）

您如果离开过澳州，请您表明您曾离澳_____年_____月_____周。

Language(s)/您曾离澳时用了那些语言（包括方言）：

_____

_____

13. What language(s) are spoken at the place where you live?

您在家里用何种语言？

_____

_____

14. What language(s) are spoken at the place where you work (if applicable)?

您在工作用何种语言？

_____

163

15. Please specify the language(s) you use when speaking to family members who don't live in your household？您与不再同您一起住的亲人用何种语言？

_____

_____

16. Please specify the language(s) you use when speaking to friends who don't live in your household.   您与不再同您一起住的朋友用何种语言？

_____

_____

17. Do you ever put English words in a sentence with your native language?
   Yes___ No___
您再使用母语时是否带有英语？   是   否

If yes, how often do you put English words in a sentence with your native language? ("1" indicates that you never put English words and "10" indicates that you always put English words in a sentence with your native language).
若是，请您表明您使用英语的程度。
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
never   sometimes   always
从不   有时   总是

18. Are you more comfortable at conversing in your native language or English?   Native
        English
一般来讲您使用那种语言方便？   母语   英语

19. How would you rate your pronunciation of English? ("1" indicates 'very poor compared to my Cantonese' and "10" indicates 'same as my Cantonese').
        请您表明您的英语语音程度。
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
very poor   average   same as my native language
非常差   一般   和母语相同

20. How would you rate your understanding of spoken English? ("1" indicates very 'poor compared to my native language' and "10" indicates 'same as my native language')
请您表明您使用英语理解问题的程度。
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
very poor   average   same as my native language
非常差   一般   和母语相同

164

21. How would you rate the amount of words you know in English compared to how many you know in your native language?

请您表明您掌握英语词汇程度。

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
very poor   average   same as my native language
非常差   一般   和母语相同

22. How would you rate your ability to put words together like a native speaker of English?

请您表明您对英语组词的能力。

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
very poor   average   same as my native language
非常差   一般   和母语相同

LANGUAGE INPUT TO YOUR BABY
对婴儿的语言沟通

23. Please specify the language(s) you use/have used when speaking to your baby?

您对婴儿使用那些语言（包括方言）？ _____

_____

24. When you speak to your baby, do you put any English words in a sentence?   Yes   No

您对婴儿沟通时使用英文吗？   是   否

If yes, can you please rate your level of English usage

若是，请表明您使用英语的程度。

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
never   sometimes   always
从未   有时   忠是

25. If you sing to your baby, which language(s) do you use?

您若对婴儿唱歌时会用那些语言？

_____

Can you indicate your level of English use when singing to your infant.

您对婴儿唱歌时使用英语的程度。

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐
Never   sometimes   always
从未   有时   忠是

26.   Do all family members speak to your baby in your native language (including grandparents, siblings, and aunts/uncles)?   Yes   No

您的家庭成员都是用您的母语同婴儿沟通吗？   是   否

Please specify what language(s) these family members use, their relationship to your baby (for e.g. older sibling, English) and how many hours per week they are directly engaged (i.e. speaking) with your baby

请表明哪些家庭成员经常同婴儿沟通，使用语言，和每周平均沟通时间？

家人同婴儿关系 ＿＿＿＿＿＿＿＿＿   使用语言＿＿＿＿＿＿＿＿   时间＿＿＿＿＿＿＿＿＿ 。

家人同婴儿关系＿＿＿＿＿＿＿＿＿   使用语言＿＿＿＿＿＿＿＿   时间＿＿＿＿＿＿＿＿＿ 。

家人同婴儿关系＿＿＿＿＿＿＿＿＿   使用语言＿＿＿＿＿＿＿＿   时间＿＿＿＿＿＿＿＿＿ 。

27.   Does your baby attend day-care, either at a day-care centre or at a relative /friends' house?

                        Yes   or   No＿＿

        您的婴儿会经常有他人看管吗？   是   否

If you answered Yes, how many hours per week does your baby spend at day-care? 若是，请表明平均

每周几小时婴儿会在那里？   ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

Which language(s) does your baby hear at day-care?

您的婴儿会在那里听到哪些语言？   ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

28.   Of all the language input your baby hears, what percent of this language input is Cantonese? ("0" indicates that your baby hears Cantonese 0% of the time i.e. never and "100" indicates that your baby always hears Cantonese).

请表明（用百分数）您的婴儿会听到多少广东话？

0% ☐ 10 ☐ 20 ☐) ☐ 40 ☐ 50 ☐ 60 ☐ 70 ☐)0 ☐ 90 ☐ 100% ☐

If you answer was 0%, please indicate what language your infant hears.

若 0%，婴儿会听那些语言？ ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

MASS MEDIA

媒体影响

29.   Do you watch Cantonese channels or English channels on TV?

您看电视时选择哪种语言？广东话＿＿＿＿＿＿＿＿英语＿＿＿＿＿＿其他

＿＿＿＿＿＿＿＿广东话 ＿＿＿＿＿＿＿＿小时

＿＿＿＿＿＿＿＿英语 ＿＿＿＿＿＿＿＿小时

＿＿＿＿＿＿＿＿其他 ＿＿＿＿＿＿＿＿小时

30.   您婴儿会听到广东话电视节目几小时？   _____

您婴儿会听到英语电视节目几小时？   _____

您婴儿会听到其他语言电视节目几小时?   _____

How many hours do you spend listening to the radio per day?

您听广播时选择哪种语言？广东话_____英语_____他

_____广东话 _____小时

_____英语   _____小时

_____其他   _____小时

32.   您婴儿会听到广东话广播节目几小时？   _____

您婴儿会听到英语广播节目几小时？   _____

您婴儿会听到其他语广播视节目几小时？   _____

MUSIC

音乐

33.   Do you listen to music with lyrics sung in Cantonese or English?

您听音乐时选择哪种语言？广东话_____英语_____其他

_____广东话 _____小时

_____英语   _____小时

_____其他   _____小时

34.   您婴儿会听到广东话音乐几小时？   _____

您婴儿会听到英语音乐几小时？   _____

您婴儿会听到其他语音乐几小时？   _____

Thank you for your participation and interest in this study

多谢您的合作

Appendix D: Toy Stimuli used to illicit the Target Words

Toy lion used to illicit the Cantonese tone 55 word [si] meaning *Lion*

Toy rattle snake used to illicit the Cantonese tone 25 word [sei] meaning *dead* and [dük] meaning *poison*
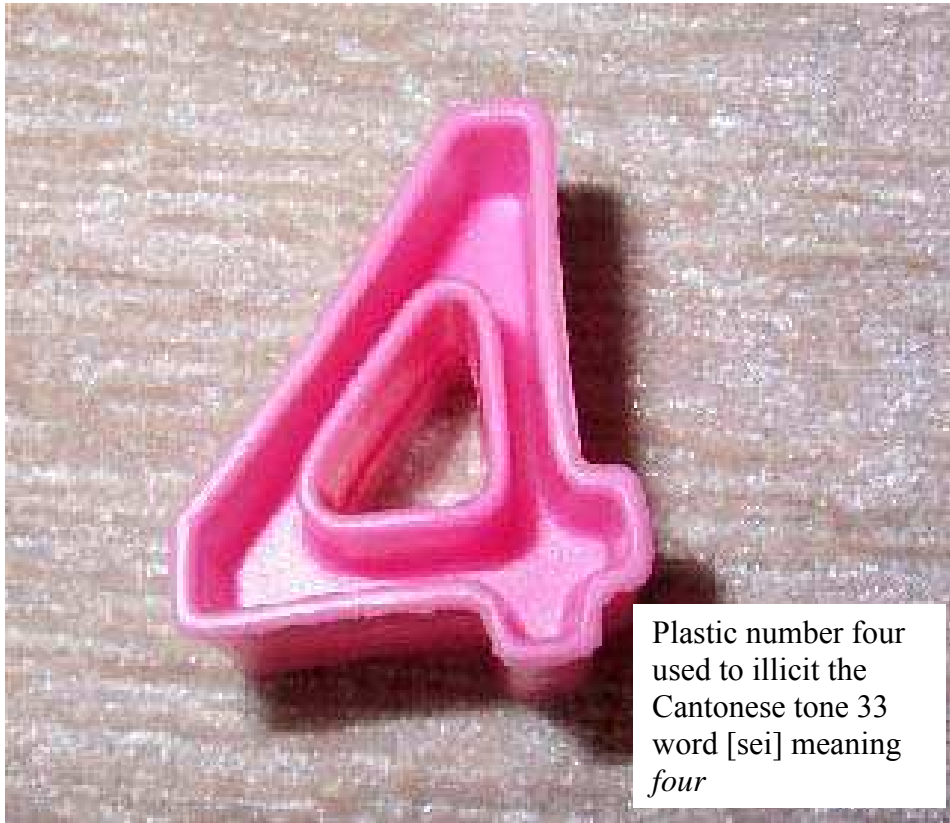
Plastic number four used to illicit the Cantonese tone 33 word [sei] meaning *four*

Coaster of a picture of Sydney used to illicit the Cantonese tone 23 word [sei] meaning *city*

Clock used to illicit the Cantonese tone 21 word [sei] meaning *time*

Plastic food items in a container used to illicit the Cantonese tone 22 word [sei] meaning *food*
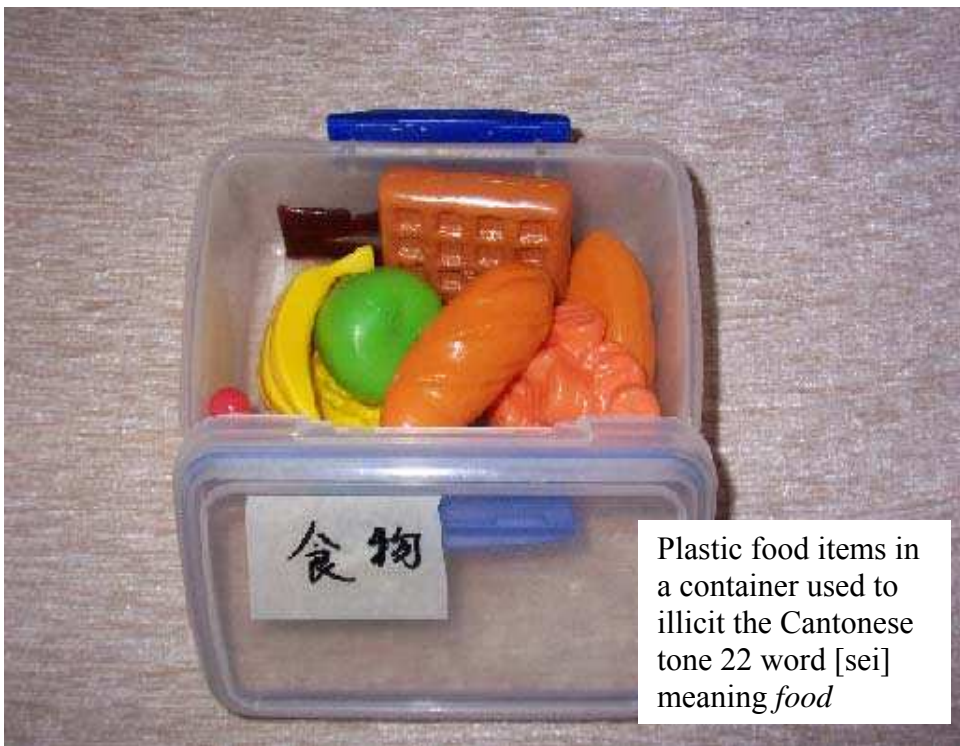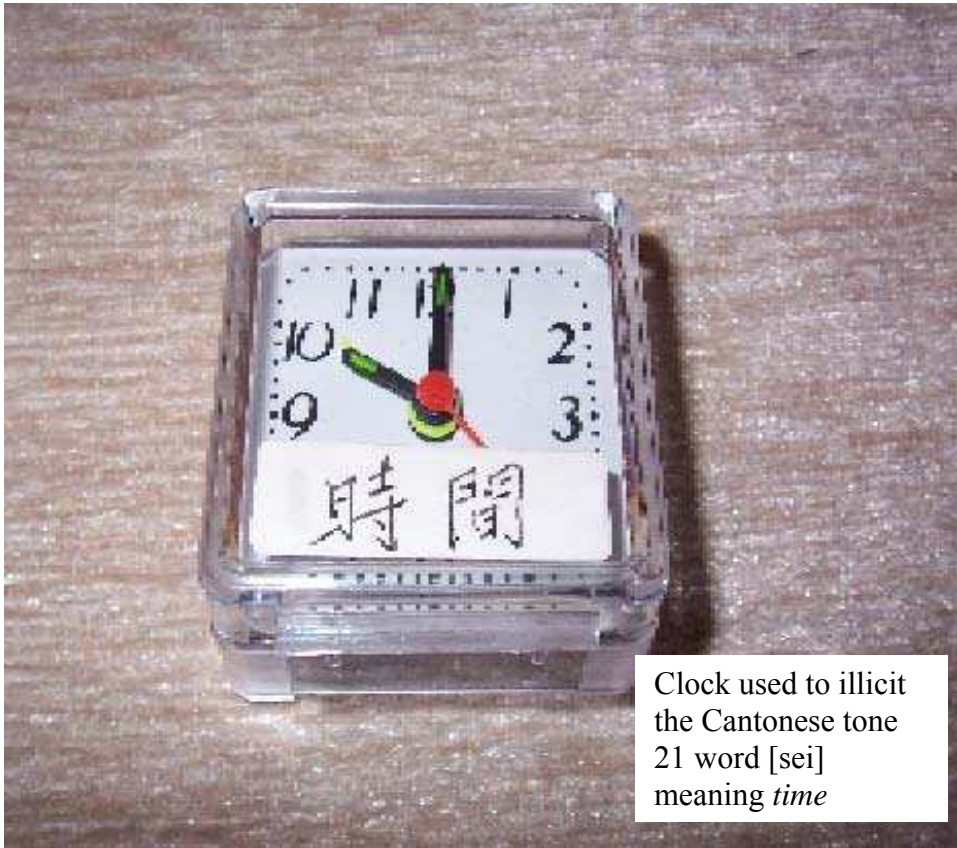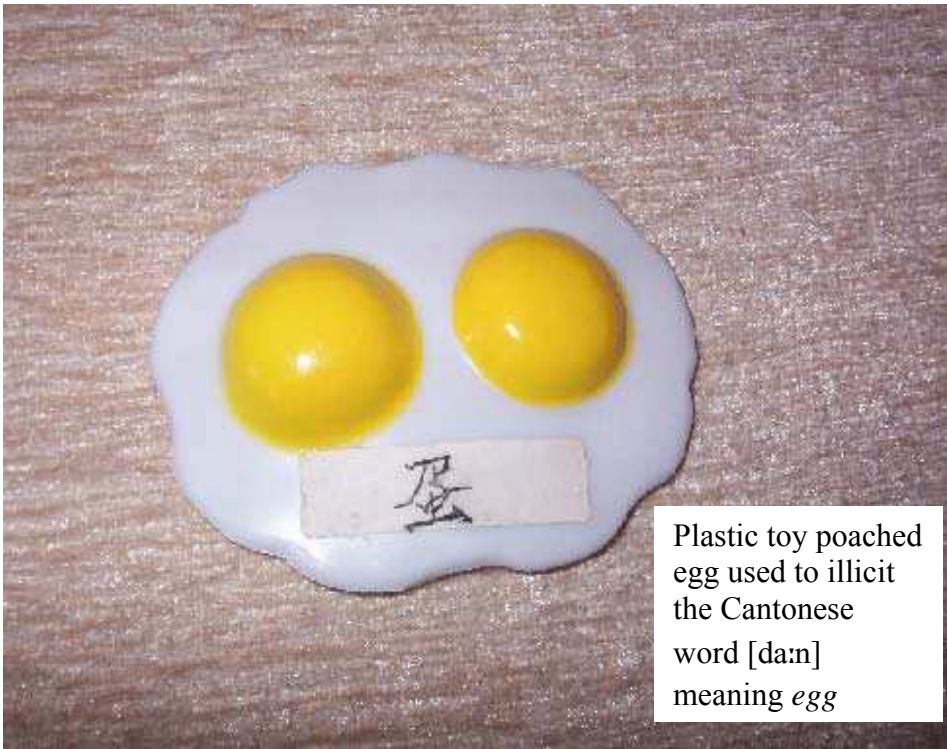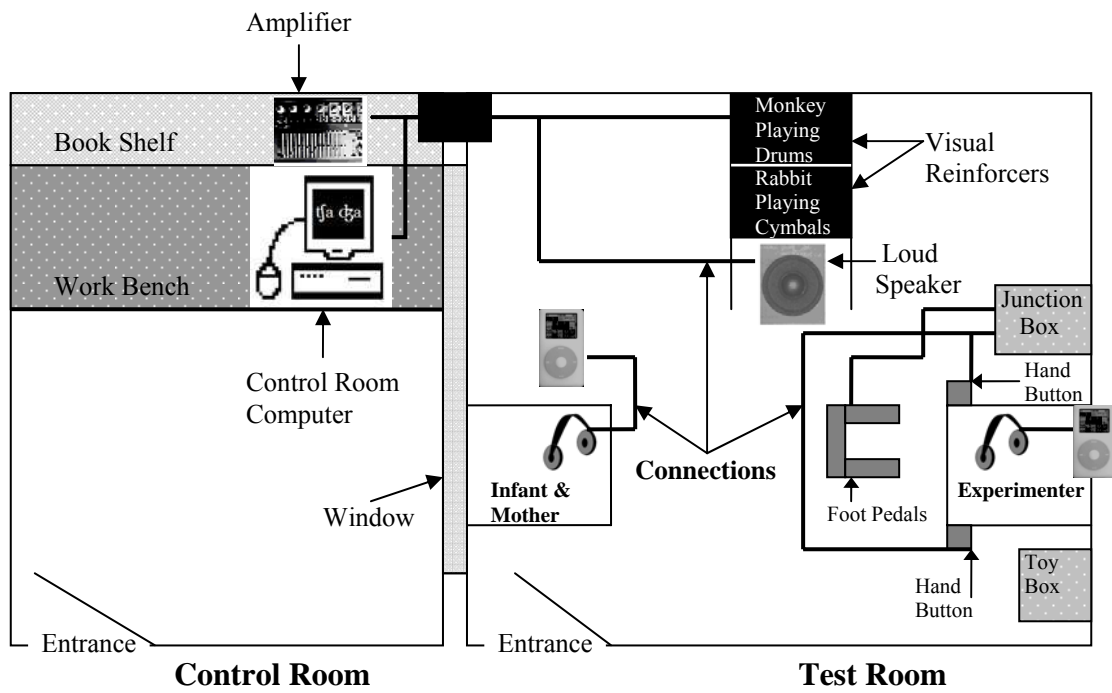
Toy butterfly used to illicit the Cantonese word [dip] meaning *butterfly*

Plastic toy poached egg used to illicit the Cantonese word [daːn] meaning *egg*

Appendix E: Schematic Representation of the Conditioned Head-Turn Setup

Appendix F : Why use Conditioned Head-Turn?

In light of these results together with the (Liu et al., 2003) study which suggests a positive correlation between the degree of mothers' hyperarticulation and their infants' ability to discriminate native consonant contrasts, a study investigating a similar correlation between tone hyperarticulation and infant's native consonant discrimination ability was conducted.

However, it is important to first consider the various paradigms and methods available for investigating infants' speech discrimination as well as a justification for the method chosen for this study. Since the latter half of the last Century, there has been many experimental paradigms developed for infant perception studies. This began with the visual preference paradigm (Fantz, 1963) which was later modified to enable experiments in auditory perception. There are two dominant paradigms for investigating infant auditory perception: the habituation or visual fixation procedure and the conditioned head-turn procedure (CHT).

Appendix F.1 : Visual Habituation/Fixation Procedure

The habituation paradigm in infant auditory discrimination experiments usually involves infants fixating on a static pattern presented on a screen while listening to the same auditory stimulus. Over time, infants will become bored and stop fixating on the visual stimulus, and this is used as an indicator of habituation. Infant usually looks away until they are able to perceive a change in the auditory stimulus, in which time they tend to go back to fixating on the screen again.

The usual method of visual habituation begins by capturing the infant's attention using flashing lights or looming objects. When the infant's gaze is centred on the screen, then a static image is presented – which is often a black and white checker board but can be other images – at the same time as an auditory stimulus. As soon as the infant looks away for a criterial number of seconds, predetermined by the experimenter (usually between 1 and 2 seconds), the flashing lights are reactivated to capture the infant's attention again and the same process continues. In the mean time, the experimenter using a computer program tracks and measures the infant's visual

"fixation" time (time looking at the screen). The experiment begins when the infant's fixation time falls below a predetermined set criterion which is used as a measure of habituation (this is usually fixation times of below 50% of the first two trials in two consecutive trials). During the test phase, the infant is presented with both familiar and novel sounds (sounds not presented during the habituation/training phase). The infant's fixation time is recorded for both types of trials and compared usually with the criterion. Any increases in visual fixation time for novel but not control stimulus in experimental trials over the habituation criterion is used as an indication of discrimination of the auditory stimuli.

An alternative two screen habituation/preference method involves having two screens with the same visual display but playing different auditory stimuli, so that only one sound is associated with one screen. A flash light placed in between the two screens is used to centre the infant's attention between trials and once a trail begins the infant will hear only one auditory stimulus when looking at one screen and another while looking at the other screen. This procedure is more suited for studies interested in measuring infants' preference for different types of speech sounds/registers such as IDS and ADS, or speech with high versus low affect. The double screen procedure is a useful method for preference tasks since the traditional visual habituation procedure doesn't account for sound preferences and only reports a difference score on the duration of visual fixation.

Appendix F.2 : Advantages and Disadvantages

In general, infants enjoy the habituation procedure as the infant's response controls subsequent trials/delivery auditory and visual information. A great advantage of the visual habituation can be used across most age groups even with very young weeks old infants, as long as they have acquired visual acuity and can see the visual display. As with other tasks, with increased mobility it is difficult to make toddlers perform any task.

A criticism levelled at the habituation method is that the criterion for comparing new with habituated stimuli is often artificially low, so the test phase must include both new and habituated stimuli, and related to this issue is the "dangerous" practice of not

identifying and excluding non-habituators from the study (Cohen, 2004). Other points to consider when using the habituation method is that it often runs the course of early preference for familiar then novel stimulus and this is particularly so for young infants and with complex stimuli that involve sound and movement of one or more objects (Cohen, 2004). Implicit in all this is the issue of how much exposure to each stimuli should infants receive and the effect of this repeated exposure (Werker, Polka, & Pegg, 1997).

Appendix F.3 : Conditioned Head Turn Procedure

The traditional Conditioned Head-turn (CHT) procedure is conducted by two experimenters and is primarily used for assessing auditory perception and speech perception in infants (Werker et al., 1997). In CHT experiments, infants are conditioned to turn their head toward a particular direction when they hear a change in the auditory stimuli by been positively reinforced by a visual reinforcer. In a typical set up, the mother, infant, and an experimental assistant (E1) are seated in a sound-attenuated test room, while a second experimental assistant (E2) controls the computer and observes the infant either through a one-way mirror, peephole, or on real-time video. In the test room, both the parent and the experimenter's assistant are listen to masking tapes via headphones so that they are deaf to the auditory stimulus presentation that the infant is hearing to prevent them from influencing the infant's performance. In the control room, E2's task is to observe the infant carefully and initiate a trial by pressing a button whenever the infant is watching E1 – but not totally engaged by the toys being manipulated by E1. Once the trial is initiated, it is then under the computer program's control whether a change trial, in which the speech stimulus changes, or a control trial, in which there is no change in the speech stimulus is presented. E2 also monitors the infant's behaviour and pushes a button if a head-turn occurs in the direction of the visual reinforcer.

In a typical design, the infant sits on his/her mother's lap facing E1 who engages the infants' attention with various toys. A loudspeaker and visual reinforcer box are both located to one side of the infant. Speech stimuli are presented through the

loudspeaker, while the infant is shaped to turn their head towards the visual reinforcer (mechanical toys in a plexiglass box) accompanying a change in the auditory stimulus. If the infant makes a correct head-turn, the visual reinforcer box is illuminated and the toy becomes visible and animated. In addition to the visual reinforcer, E1 also provides positive social reinforcement in the form of smiles and praises the infant for producing a correct head-turn. No reinforcement is given for incorrect head-turns (i.e., no activation of the visual reinforcer – toys). From an operant conditioning perspective, the infant learns that the consequence for turning towards the visual reinforcer when there is a change in auditory stimuli is rewarded by toys being activated and this positive reinforcement ensures future head-turning behaviour towards the visual reinforcer in response to changes in auditory stimulus. This is assuming that the infant is able to discriminate between the test stimuli.

There are typically two stages in the procedure. After familiarising the infants to the auditory stimuli in the training stage, the conditioning stage begins by presenting the infant with a background auditory stimulus and every trial is a change trial where there is a change in auditory stimulus. Head-turns are initially shaped by activation of the visual reinforcer contiguously in time with the change in auditory stimulus. This alone is usually enough to attract the infant's attention to turn towards the visual reinforcer independently, however, occasionally the E1 need to guide the infant's attention towards the reinforcer when the infant is not doing this independently. Then through several trials, the experimenter gradually lengthens the delay between the auditory stimuli change and the activation of the reinforcement box to encourage the infant to make independent anticipatory head-turns (i.e., head-turn after a stimulus change but preceding the activation of the visual reinforcer). Through several trials, most infants can be conditioned to perform the head-turn response when a change in auditory stimuli is detected. To move into the next stage, infants need to make a preset performance criterion which is judged appropriate by the experimenter as an indication that conditioning has taken place. This is usually a set number of consecutive anticipatory head-turns over a set number of trials. Infants that meet this criterion move to the next stage, the actual experimental/testing stage, in which change and control (no

176

change) trials are presented in random order. E2 monitors the infant's behaviour and records head-turns by pressing a button whenever the infant produces a head-turn to the reinforcer regardless of whether the visual reinforcer becomes activated to not. If the button press occurs within a criterion window of 4-6 seconds (approximately 3 change stimulus presentations), the reinforcer is activated and E1 verbally praises the infant ('Good boy/girl'), smiles, and claps. This is recorded as a 'Hit' and only hit conditions are positively reinforced by the visual reinforcer and by social reinforcement. If no head-turn occurs during a change trail, a 'Miss' is recorded. If the infant turns his/her head during a control trial, this is a 'False positive'. The computer also automatically records 'correct rejections' where the infant doesn't make a head-turn during a control trial (no change in auditory stimulus).

The conditioned head-turn paradigm appears to be best used with 6- to 10-months-old infants. Younger infants have not acquired enough physical control to perform head turns while older more mobile children become less attracted to the reinforcer for prolonged periods of time (Werker et al., 1997). The method is also best used with simple stimuli such as consonant vowel syllables and suitable with complex speech phenomenon such as suprasegmental information like intonation (Werker et al., 1997).

Appendix F.4 : Advantages and Disadvantages

The first strength of CHT procedure is that infants seem to enjoy this procedure perhaps because they are in control of the reinforcer – the CHT procedure is fun, interactive, interesting, and the delivery of the reinforcer is dependent in part on their response. Infants who are interested typically learn the association between the change in sounds and the activation of the reinforcer very quickly, within about 15 trials. Also, as (Werker et al., 1997) had pointed out, since random head-turns are not always accompanied by change trials, there is independence between stimulus and response which allows for the distinction between responses that inform perceptual ability versus those that relate to disinterestedness – which is not always easy to tease apart in visual habituation paradigms. Secondly, the CHT procedure is that the stimulus and reinforcer

are independent events, in contrast to them being intertwined as in the habituation procedure. Finally, in the CHT procedure each infant receives multiple test trials, making it relatively easy to evaluate whether an individual infant can detect a change in the stimulus.

Unfortunately, as with other infant procedures, the CHT also suffers from variable attrition rates, which ranges from 5% to 50% or more are possible (Liu et al., 2003), and the method is not particularly suitable for testing infants younger than about 6 months who have not developed muscle tone to produce CHTs or older toddlers who are mobile and tend to get bored with the task more quickly. From a technical perspective, the CHT procedure is also not very useful for testing speech patterns longer than a syllable, such as multi-syllabic words and melodic patterns. Also, because only the 'hit' condition is associated with positive reinforcement, theoretically infants should be more biased toward making head-turns, therefore 'false-alarms' should be expected to out way 'correct-rejections' since only performing head-turns are reinforced. This means that some measure of a proportion of hits plus correct-rejections over false-alarms may be needed to measure the infant's task performance more accurately. Another problem associated with CHT procedure is that it doesn't directly access "what" is actually perceived.

Appendix F.5 : The Method of Choice for This Study

The purpose of the current study is to test Cantonese language environment infants' ability to discriminate native Cantonese consonant contrasts, and to examine any correlations between infants' discrimination ability with the degree of their mother's tone hyperarticulation. In searching for an appropriate procedure to use for this study, three things are of particular interest. Firstly, the study is only interested in infants' ability to discriminate monosyllabic Cantonese sounds and not preferences for any particular auditory information. Secondly, the procedure needs to measures that are graded across infants in reflecting a variance in discrimination abilities, such that infants with better abilities should have scores higher or lower than infant's with poorer abilities. Thirdly, and perhaps of the least importance is that the procedure needs to

yield results that are comparable with previous studies to allow comparisons to be made between the results of the current study and previous studies.

Given the purpose and needs of the study, the conditioned head-turn paradigm was preferred as the method of choice. The CHT procedure allows for a comparison among participants by providing a graded number of CHTs as a measure of discrimination which can be used as an indication of infants' sensitivity to the auditory stimuli – in this case native Cantonese consonants. The CHT procedure was also the method used by in their study investigating a relationship between vowel hyperarticulation and infants' discrimination for native speech sounds. The infants to be tested in this study are between 6- and 9-months-old and the CHT method is well suited to infants of this age range. The CHT paradigm was also the paradigm used in the Liu et al (2003) study which correlated infants' native consonant discrimination ability with their mothers' vowel hyperarticulation measures in Mandarin IDS. As this study is interested in correlating infants' vowel and tone hyperarticulation measures in Cantonese IDS, adopting the CHT procedure will allow for direct comparisons between this study and the Liu et al (2003) study. Finally, the Conditioned Head-Turn procedure will be used in this study for the convenience of having only one experimenter to conduct the study.

Appendix F.6 : References

Burnham, D., & Dodd, B. (1998). Familiarity and novelty preference in infants' auditory-visual speech perception: problems, factors, and a solution. In C. Rovee-Collier, L. Lipsitt & H. Hayne (Eds.), Advances in infancy research volume 12. Stamford: Ablex Publishing Corporation.

Cohen, L. B. (2004). Uses and Misuses of Habituation and Related Preference Paradigms. Infant and Child Development, 13, 349-352.

Fantz, R. L. (1963). Pattern Vision in Newborn Infants. Science, 140(3564), 296-297.

Kuhl, P. K. (1985). Methods in the Study of Infant Speech Perception. In G. Gottlieb & N. A. Krasnegor (Eds.), Measurement of Audition and Vision in the First Year of Postnatal Life (pp. 223-252). New Jersey: Ablex Publishing Corporation.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. Developmental Science, 6(3), F1-F10.

Polka, L., Jusczyk, P. W., & Rvachew, S. (1995). Methods for studying speech perception in infants and children. Speech perception and linguistic experience: Issues in cross-language research, 49–89.

Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedures as a method for assessing infant speech perception. Early Development & Parenting, 6(3-4), 171-178.

Abramson, A. S. (1979a). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie J. A. Henderson*. Bangkok: Indigenous Languages of Thailand Research Project.
Abramson, A. S. (1979b). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research* (pp. 127-134): Academic Press.
Abramson, A. S. (1986). The Thai tone space. *Haskin Laboratories: Status Report on Speech Research SR-85, 85*, 105-114.
Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development, 19*(4), 401-418.

Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *The Journal of the Acoustical Society of America, 105*, 1095.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience. *Child Development, 52*, 1135-1145.

Barrie, M. (2003). Contrast in Cantonese vowels. *Toronto Working Papers in Linguistics, 20*, 1-19.

Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *Acoustical Society of America, 116*(3), 1739–1748.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter & Co.

Bernard, J. (1970). Towards the acoustic specification of Australian English. *Zeitschrift für Phonetik, 2*(3), 113–128.

Bernard, J., & Mannell, R. (1986). *A study of /h_d/ words in Australian English*: Working Papers of the Speech, Hearing and Language Research Centre.

Bird, K., Hadzi-Pavlovic, D., & Isaac, A. (1990). PSY: A program for contrast analysis (Version 5.1.2): © School of Psychology, University of New South Wales.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America, 67*, 648.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 4.5.13) [Computer Program].

Bohannon III, J. N., & Marquis, A. L. (1977). Children's Control of Adult Speech. *Child Development, 48*(3), 1002-1008.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101* (4), 2299-2310.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science, 5*(1), 72-83.

Bredberg, G. (1985). The anatomy of the developing ear. In S. E. Trehub & B. Schneider (Eds.), *Auditory development in infancy*. New York/London: Plenum Press.

Brunelle, M. (2003). *Tone Coarticulation in Northern Vietnamese.* Paper presented at the Proceedings of the 15th International Conference of phonetic Sciences.

Burnham, D., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language, 18*(2), 231-260.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., et al. (1996). *Perception of lexical tone across languages: Evidence for a linguistic mode of processing.* Paper presented at the ICSLP 96 Proceedings, Fourth International Conference on Spoken Language Philadelphia, PA, USA.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 259-280). Amsterdam: John Benjamins.

Chan, W. S., Chuang, C. K., & Wang, W. S.-Y. (1975). Crosslanguage study of categorical perception for lexical tone. *Journal of the Acoustical Society of America, 58*, 119.

Chao, Y. R. (1930). A system of "tone letters". *Le Maître Phonétique, 30*, 24-27.

Chao, Y. R. (1947). *Cantonese primer*. New York: Greenwood Press.

Chen, M. Y. (2000). *Tone Sandhi*. Cambridge, UK: Cambridge University Press.

Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three Facial Expressions Mothers Direct to Their Infants. *Infant and Child Development, 12*, 211–232.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America, 95*(3), 1570-1580.

Clark, J., & Yallop, C. (1995). *An introduction to phonetics and phonology (2nd Ed.)*. Oxford, UK: Blackwell Publishers.

Cohen, L. B. (2004). Uses and Misuses of Habituation and Related Preference Paradigms. *Infant and Child Development, 13*, 349-352.

Colombo, J., & Horowitz, F. D. (1986). Infants' Attentional Responses to Frequency Modulated Sweeps. *Child Development, 57*(2), 287.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development, 20*(4), 477-488.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584-1595.

Cooper, R. P., & Aslin, R. N. (1994). Developmental Differences in Infant Attention to the Spectral Properties of Infant-directed Speech. *Child Development, 65*(6), 1663-1677.

Cox, F. (1996). *An Acoustic Analysis of Vowel Variation in Australian English. Unpublished PhD Thesis.* Macquarie University, Sydney.

Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association, 37*(03), 341-350.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of phonetics, 21*(1-2), 103-108.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

Ellen, G. B., & Anderson, A. H. (1983). The Unintelligibility of Speech to Children. *Journal of Child Language, 10*(2), 265-292.

Engle, M. E. (1979). *Do fathers speak motherese? An analysis of the language environments of young children.* ProQuest Information & Learning, US.

Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language, 25*(2), 219-234.

Englund, K., & Behne, D. (2005). Infant Directed Speech in Natural Interaction-Norwegian Vowel Quantity and Quality. *Journal of Psycholinguistic Research, 34*, 259-280.

Englund, K., & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development, 15*(2), 139-160.

Fant, G. (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), *Manual of Phonetics* (pp. 173-276). Amsterdam: North-Holland Publication. Co.

Fantz, R. L. (1963). Pattern Vision in Newborn Infants. *Science, 140*(3564), 296-297.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181-195.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development, 60*(6), 1497-1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*(3), 657-674.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279-293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477-501.

Field, A. P. (2005). *Discovering Statistics Using SPSS*: Sage Publications Inc.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*, 1029-1044.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*(17), 1-189.

Furrow, D., & Nelson, K. (1986). A further look at the motherese hypothesis: A reply to Gleitman, Newport & Gleitman. *Journal of Child Language, 13*(1), 163-176.

Gandour, J. T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics, 9*, 20-36.

Gandour, J. T. (1983). Tone perception in far east languages. *Journal of Phonetics, 11*, 149-175.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception a multidimensional scaling investigation. *language and Speech, 21*(1), 1-33.

Gandour, J. T., Weinberg, B., Petty, S. H., & Dardarananda, R. (1988). Tone in Thai Laryngeal Speech. *Journal of Speech and Hearing Disorders, 53*, 23-29.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11*(1), 43-79.

Goldsmith, M. (1994). Tone languages. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics*. Oxford: Pergamon Press.

Golinkoff, R. M., & Alioto, A. (1995). Infant-Directed Speech Facilitates Lexical Learning in Adults Hearing Chinese: Implications for Language Acquisition. *Journal of Child Language, 22*(3), 703-726.

Golinkoff, R. M., & Ames, G. J. (1979). A Comparison of Fathers' and Mothers' Speech with Their Young Children. *Child Development, 50*(1), 28-32.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal Speech to Infants in a Tonal Language: Support for Universal Prosodic Features in Motherese. *Source: Developmental Psychology, 24*(1), 14-20.

Han, M. S., & Kim, K. O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics, 2*, 223-232.

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics, 17*(2), 155-184.

Hashimoto, O.-K. Y. (1972). *Studies in Yue Dialects 1: Phonology of Cantonese* (Vol. 1). Cambridge: Cambridge University Press.

Hepper, P. G., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive Infant Psychology, 11*(3), 147.

Hirsh-Pasek, K., & Treiman, R. (1982). Doggerel: Motherese in a new context. *Journal of Child Language, 9*(1), 229-237.

Holzrichter, A. S. (2000). Language acquisition by eye. In C. Chamberlain, J. P. Morford & R. I. Mayberry (Eds.), *Child-directed signing in American Sign Language* (pp. 25-40). Mahwah, NJ: Lawrence Erlbaum Associates.

Hombert, J. M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (Ed.), *Tone: A linguistic survey*. New York: Academic.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations of the development of tones. *Language, 55*(1), 37-58.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.): Duxbury Press Belmont, Calif.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*, 3-42.

Ikeda, Y., & Masataka, N. (1999). A variable that may affect individual differences in the child-directed speech of Japanese women. *Japanese Psychological Research, 41*, 203-208.

Iverson, J. M., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14*(1), 57-75.

Kaplan, P. S., Bachorowski, J.-A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Science, 13*(3), 268.

Kaplan, P. S., Dungan, J. K., & Zinser, M. C. (2004). Infants of Chronically Depressed Mothers Learn in Response to Male, But Not Female, Infant-Directed Speech. *Developmental Psychology, 40*(2), 140-148.

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., & Cooper, R. P. (1995). Dishabituation of visual attention by infant- versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior and Development, 18*, 209-223.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*(1), 55-68.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America, 72*(2), 353-365.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104-117.

Kitamura, C., & Burnham, D. (1998). The infant's response to vocal affect in maternal speech. In C. Rovee-Collier (Ed.), *Advances in Infancy Research* (Vol. 12, pp. 221-236).

Kitamura, C., & Burnham, D. (2003). Pitch and Communicative Intent in Mother's Speech: Adjustments for Age and Sex in the First Year. *Infancy, 4*(1), 85-110.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a

function of infant age and sex in a tonal and non-tonal language. *Infant Behavior & Development, 24,* 372-392.

Kuhl, P. K. (1978). Predispositions for the perception of speech-sound categories: A specifies-specific phonomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities: Early behavioral assessment.* Baltimore: University Park Press.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science, 277*(5326), 684-686.

Ladefoged, P. (1982). *A course in Phonetics (2nd ed.).* New York: Harcourt Brace Joanovich.

Ladefoged, P. (2001a). *A course in Phonetics (4th ed.)* (Fourth Edition. ed.). Boston: Heinle & Heinle, a division of Thomson Learning, Inc.

Ladefoged, P. (2001b). *Vowels and Consonants.* Oxford: Blackwell Publishers Ltd.

Ladefoged, P. (2003). *Phonetic Data Analysis: An introduction to fiedlwork and instrumental techniques*: Blackwell Publishing Ltd.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross linguistic comparison. *Journal of Psycholinguistic research, 25*(5), 527-542.

Lennes, M. (2002). Mietta's Praat scripts.

Li, C. M., & Thompson, S. A. (1978). The acquisition of Tone. In V. Fromkin (Ed.), *Tone: A linguistic survey.* New York: Academic.

Li, P.-Y. (2004). *Perceptual analysis of the six contrastive tones in Cantonese.* Paper presented at the Inter-nations Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1-F10.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43*(4), 912-917.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2004). *The effects of intonation patterns on lexical tone production in Cantonese.* Paper presented at the International symposium on tonal aspects of languages: With an emphasis on Tone Languages.

Masataka, N. (1992). Motherese in a signed language. *Infant Behavior and Development, 15*(4), 453-460.

Masataka, N. (1996). Perception of motherese in a signed language by 6-month-old deaf infants. *Developmental Psychology, 32,* 874-879.

Masataka, N. (1998). Perception of Motherese in Japanese Sign Language by 6-Month-Old Hearing Infants. *Developmental Psychology, 34*(2), 241-246.

Massaro, D. W., Cohen, M. M., & Tseng, C.-Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13,* 267-289.

Mathews, S., & Yip, V. (1994). *Cantonese: a comprehensive grammar.* NY: Routledge.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-264.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143-178.

Mitchell, R. W. (2000). Americans' Talk to Dogs: Similarities and Differences With Talk to Infants. *Research on Language and Social Interaction, 34*(2), 183-210.

Mitchell, R. W., & Edmonson, E. (1999). Functions of repetitive talk to dogs during play: Control, conversation, or planning. *Society and Animals, 7*, 55–81.

Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., & Charnvivit, P. (2002). *Perception of tones and vowel quantity in Thai.* Paper presented at the Proceedings of ICSLP, Denver, USA.

Mok, P. K. P., & Hawkins, S. (2004). *Effects of phonemic vs. allophonic density and stress on vowel-to-vowel coarticulation in Cantonese and Beijing Mandarin.* Paper presented at the 2004 International Symposium on Chinese Spoken Language Processing.

Moore, D. S., Spence, M. J., & Katz, G. S. (1997). Six-Month-Olds' Categorization of Natural Infant-Directed Utterances. *Developmental Psychology, 33*(6), 980-989.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition.* Cambridge: Cambridge University Press.

Nwokah, E. E. (1987). Maidese versus motherese--is the language input of child and adult caregivers similar? *Lang Speech, 30*(3), 213-237.

Ohala, J. J. (1978). Production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5-39). New York: Academic.

Papousek, M., & Bethesda, M. D. (1987). *Melodies in Motherese in Tonal and Nontonal Languages: Mandarin Chinese, Caucasian American, and German.*

Papousek, M., Bornstein, M. H., Nuzzo, C., Papousek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development, 13*(4), 539-545.

Papousek, M., & Hwang, S.-F. C. (1992). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics, 12*(4), 481-504.

Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development, 14*(4), 415-440.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15*(3), 325-345.

Phillips, J. R. (1973). Syntax and Vocabulary of Mothers' Speech to Young Children: Age and Sex Comparisons. *Child Development, 44*(1), 182-185.

Pisoni, D. B. (1973). Perceptual Processing Time for Consonants and Vowels. *The Journal of the Acoustical Society of America, 53*, 369.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustic Society of America, 61*(5), 1352-1361.

Polka, L., & Werker, J. F. (1994). Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435.

Querleu, D., Renard, X., & Versyp, F. (1981). Les perceptionsauditives du foetus humain. *Médecine et Hygiène, 39*(2101-2110).

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crepin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 28*(3), 191-212.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., Vervoort, P., & Crepin, G. (1986). Commentary. Can the fetus listen and learn. *BJOG: An International Journal of Obstetrics and Gynaecology, 93*(4), 411-412.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349.

Read, C., Yun-Fei, Z., Hong-Yin, N. I. E., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31-44.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication, 6*(4), 343-352.

Rose, P. (1993). A linguistic-phonetic acoustic analysis of Shanghai tones. *Australian Journal of Linguistics, 13*, 185-220.

Rose, P. (2000). *Hong Kong Cantonese citation tone acoustics: A linguistic tonetic study.* Paper presented at the Proceedings of 8th Australian International Conference of Speech Science and Technology, Canberra.

Ross, E. D., Edmondson, J. A., & Seiburt, C. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics, 14*, 283-302.

Rvachew, S., Mattock, K., Polka, L., & Menard, L. (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *Journal of the Acoustical Society of America, 120*, 2250-2259.

Sach, J., & Devin, J. (1976). Young children's use of age appropriate speech styles in social interaction and role-playing. *Journal of Child Language, 3*, 81-98.

Scarborough, H., & Wyckoff, J. (1986). Mother, I'd still rather do it myself: Some further non-effects of 'motherese'. *Journal of Child Language, 13*(2), 431-437.

Schwanhäußer, B., & Burnham, D. (2005). *Lexical tone and pitch perception in tone and non-tone language speakers.* Paper presented at the In Proceedings of the 9th European Conference on Speech Communication and Technology, Bonn, Germany.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*(2), 145-156.

Shepard, R. N. (1978). The circumplex and related topological manifolds in the study of perception. *Theory construction and data analysis in the behavioral sciences*, 29-80.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy, 3*(3), 365.

Snow, C. E. (1972). Mothers' Speech to Children Learning Language. *Child Development, 43*(2), 549-565.

Snow, C. E., & Ferguson, C. A. (Eds.). (1977). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and nontone language. *Journal of Chinese Linguistics, 21*(1), 143-163.

Stevens, K. N. (1999). *Acoustic phonetics*. Massachusetts: Massachusetts Institute of Technology.

Sullivan, J. W., & Horowitz, F. D. (1983). The Effects of Intonation on Infant Attention: The Role of the Rising Intonation Contour. *Journal of Child Language, 10*(3), 521-534.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics (4th Ed.)*. MA: USA: Allyn & Bacon: A Pearce Education Company.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579-590.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy, 7*(1), 53-71.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion? *Psychological Science, 11*(3), 188-195.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review, 9*, 335-340.

Trainor, L. J., & Zacharias, C. A. (1998). Infants prefer higher-pitched singing. *Infant Behavior and Development, 21*(4), 799-805.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49*(1), 2-7.

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica, 33*, 368-392.

Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica, 34*, 93-107.

Wang, C., & Seneff, S. (2000). *Improved Tone Recognition by Normalizing for Coarticulation and Intonation Effects.* Paper presented at the Sixth International Conference on Spoken Language Processing, Beijing, China.

Wang, W. S.-Y. (1967). Phonological Features of Tone *International Journal of American Linguistics, 33*(2), 93-105.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences, 280*(1), 61-72.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America, 113*(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*(6), 3649-3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning, 54*(4), 681–712.

Weppelman, T. L., Bostow, A., Schiffer, R., Elbert-Perez, E., & Newman, R. S. (2003). Children's use of the prosodic characteristics of infant-directed speech. *Language & Communication, 23*(1), 63-80.

Werker, J. F., Gilbert, J. H. V., Humphrey, G. K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*, 349-355.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology Revue Canadienne de Psychologie, 43*(2), 230-246.

Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development, 17*(3), 323-333.

Werker, J. F., & Polka, L. (1993). Developmental changes in speech perception. New challenges and new directions. *Journal of Phonetics, 21*(83-101).

Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedures as a method for assessing infant speech perception. *Early Development & Parenting, 6*(3-4), 171-178.

Werker, J. F., & Tees, R. C. (1984a). cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*, 49-63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*(6), 1866-1878.

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2004). *Polly wanna cracker? On talking to infants and parrots. Refereed poster presentation.* Paper presented at the ICIS.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*(2240-2253).

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

Zee, E. (1999). *An acoustical analysis of the diphthongs in Cantonese.* Paper presented at the Conference proceedings from ICPhS, San Francisco, USA.

Zee, E. (2003). *Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers*. Paper presented at the 15th ICPhS.