The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese

Johanna G. Barry[a)]
Department of Otolaryngology and the Department of Linguistics and Applied Linguistics, University of Melbourne, Parkville, 3010, Australia

Peter J. Blamey
Department of Otolaryngology, University of Melbourne, Parkville, 3010, Australia

(Received 7 October 2003; revised 31 May 2004; accepted 14 June 2004)

This paper reports on a methodology for acoustically analyzing tone production in Cantonese. F0 offset versus F0 onset are plotted for a series of tokens for each of the six tones in the language. These are grouped according to tone type into a set of six ellipses. Qualitative visual observations regarding the degree of differentiation of the ellipses within the tonal space are summarized numerically using two indices, referred to here as Index 1 and Index 2. Index 1 is a ratio of the area of the speaker's tonal space and the average of the areas of the ellipses of the three target tones making up the tonal space. Index 2 is a ratio of the average distance between all six tonal ellipses and the average of the sum of the two axes for all six tone ellipses. Using this methodology, tonal differentiation is compared for three groups of speakers; normally hearing adults; normally hearing children aged from 4–6 years; and, prelinguistically deafened cochlear implant users aged from 4–11 years. A potential conundrum regarding how tone production abilities can outstrip tone perception abilities is explained using the data from the acoustic analyses. It is suggested that young children of the age range tested are still learning to normalize for pitch level differences in tone production. Acoustic analysis of the data thus supports results from tone perception studies and suggests that the methodology is suitable for use in studies investigating tone production in both clinical and research contexts. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779272]

PACS numbers: 43.70.Jt, 43.70.Ep, 43.66.Ts [AL] Pages: 1739–1748

I. INTRODUCTION

Studies of the acquisition of lexical tone by young children have used auditory transcription analysis of tonal productions, either in free speech samples (e.g., Tse, 1978; Clumeck, 1980; Tse, 1991) or in words elicited as tokens in citation form using picture- or object-naming tasks (e.g., Li and Thompson, 1977; So and Dodd, 1995). No matter how the speech samples are elicited, these studies agree that the first tones emerge very early in the course of speech acquisition, at about the same time as the first recognizable vowels are observed. Two tones are acquired at once, probably reflecting the fact that tone is a relative not an absolute phenomenon. Finally, children are observed to achieve mastery of the full toneme inventory long before they master consonant phoneme inventories.

Auditory analysis requires the segmentation and appraisal of utterances produced by a young child with direct reference to phonologically significant entities in the mature linguistic system. Given the fact that linguistic tone develops through exploitation of pitch—a phonetic consequence of voiced speech—and that it is a relative not an absolute entity, it is interesting to note that in all materials investigating tone acquisition in young infants, there has been little discussion about the features in an infant's phonetic voicing patterns which changed to result in an utterance becoming recognizably tonal.

Rose (1989) noted that tone is a product of an interaction between F0 and other acoustic parameters in speech. It is possible that part of pitch patterning in a young infant's speech becoming amenable to tone labeling is a reflection of the development of increased control of prosodic features such as syllable duration which is coupled with the acquisition of some readily identifiable vowel phones.

A related issue concerns the assignment of an F0 contour to a tone category. This issue has been ignored in tone acquisition studies possibly because the rapid rate and accuracy of acquisition renders it irrelevant or too difficult to pursue. Yet as Kent and Murray (1982) note such phonological style analyses of a young infant's articulations necessarily assume that it is appropriate to categorize the vocalizations of an infant into segments with a phonemically relevant referent where in fact there is no guarantee that these actually do correspond to real segments in the infant's developing linguistic system. They observe that the emergence of early phonemic contrasts results, in part, from the fact that certain sound patterns occur more frequently in speech than others. Thus a vowel heard in a given utterance may reflect an accidental co-occurrence of the correct position of the articulators with a vowel function at that point in time and may not be an indication that the infant has acquired that phone in his/her developing phonemic inventory (Keating, 1980). On this basis, Kent and Murray argue the need for acoustic in-

---

[a)] Current address: Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3RQ United Kingdom; Electronic mail: johanna.barry@psy.ox.ac.uk

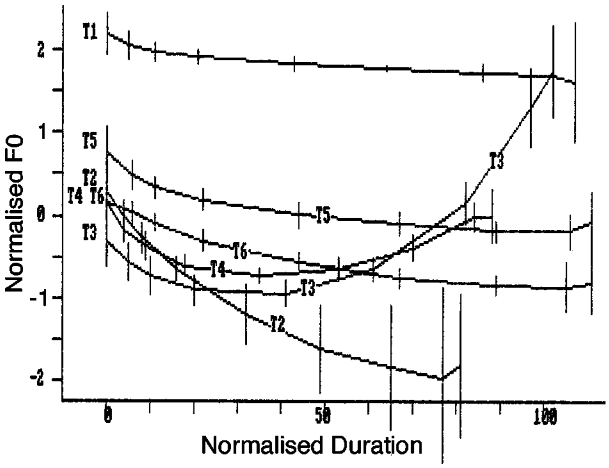

FIG. 1. The six tones in Cantonese normalized for duration and pitch across 10 speakers (five male and five female) (Rose 2000, p. 202). Vertical bars show one standard deviation above and below the mean. T1 is a high level tone (55); T2, a low-falling tone (21); T3 is a high-rise tone (25); T4, a low-rise tone (23); T5 a mid-level tone (33); and, T6 a low-level tone (22) (reproduced with permission from ASSTA and Rose).

vestigations into speech development to verify the perceptual biases of the observer.

There are six tonemes in Cantonese which are labeled; Tone 55, a high-level tone; Tone 25, a high-rise; Tone 33, a mid-level tone; Tone 21, a low-fall tone; Tone 23, a low-rise tone; and Tone 22, a low-level tone. As illustrated in Fig. 1, there are considerable perceptual similarities between some of the tonal contrasts in Cantonese and with the exception of Tone 55, Cantonese tones are primarily distinguished by differences in F0 offset.

Little is known about tone acquisition or indeed tone production in profoundly deaf speakers and still less is known about the effect of using a cochlear implant on the development of tone. Available reports, however, suggest that despite their impaired access to auditory input and the perceptual similarity of the some of the tones in Cantonese, profoundly hearing-impaired Cantonese speakers are able to successfully acquire complete or near complete toneme inventories (Dodd and So, 1994). Dodd and So hypothesize that the children in their study acquired tone successfully because (a) tone has a high functional load in Cantonese; (b) tone production is emphasized during speech habilitation; and (c) tones occur in the low frequency end of the speech spectrum where profoundly hearing-impaired speakers have most residual hearing.

The research reported here describes the development of a methodology for the investigation of tone production by Cantonese speakers using acoustic measurements. The aim of the research was to supplement findings based on more traditional auditory transcription analyses.

From a clinical perspective, an acoustic analysis of the data is attractive because it provides a means for tracking and quantifying changes in development of tone production in response to habilitation or maturational effects. It also avoids issues associated with the use of purely phonological analyses for describing developing linguistic systems. Finally, in the case of Cantonese where some tones are perceptually similar, it provides supplementary support for conclusions based on auditory analyses of the data.

A. Analyzing tone production acoustically

To study tone production and tone development acoustically, a methodology is required which (a) is sufficiently robust to permit the observation and comparison of changes in F0 usage in children with widely varying speech production skills; (b) is readily subjected to a range of statistical analyses; and (c) can be compared with results from auditory analyses of the same sets of data.

In one of the few acoustic studies of tone/intonation production, Kent and Murray (1982) investigated intonation patterns in infants acquiring English. F0 contours were given descriptive labels such as flat, or rising-falling and the frequency with which these patterns were observed in the infant's vocalizations was compared at 3, 6, and 9 months.

Among tonal languages, Gandour *et al.* (1988) analyzed tone productions by one electrolaryngeal and two esophageal speakers of Thai, both acoustically and perceptually, and compared the data with that recorded for five normal speakers of Thai. Like the Kent and Murray study of intonation patterns in infants, the acoustic analyses were highly descriptive in form and based on comparative descriptions of the contours produced by the different speakers.

Wong and Diehl (1999) described the investigation of tone production in speakers affected with Parkinson's disease. They provided little detail about their methodology, but reported that these speakers had a restricted tonal space which made correct labeling of their tones more difficult for normal Cantonese speakers. The concept of tonal space will be referred to again in the context of the present study. It is defined as "...the set of articulatory and auditory dimensions by which the speaker is constrained in production and perception" (Abramson, 1986, p. 105). It is interesting to note that Abramson, like Wong and Diehl, also noted an important role for tonal space in the correct perception of tones.

Rose (1987, 1993) developed a more quantitative approach to the acoustic analysis of tone production. In this approach, values of F0 are measured at fixed points along a tone contour. They are expressed as a multiple dispersion from a mean value of F0 and plotted against similarly normalized values of durational differences (Fig. 1). The aim in this analysis is to abstract away from between-speaker differences, while retaining the linguistically significant features in a language's tonal phonology. A more precise description of the linguistic–tonetic form of the tonemes in a language's toneme inventory can thus be achieved. For this analysis, it is important that normalization procedures are performed on transcriptionally equivalent words only, i.e., on units which constitute the same tonal targets for all speakers. If not, there is a risk that linguistically significant differences between speakers will be normalized out, leading to spurious conclusions.

Rose's approach might be suitable for comparing the tonal productions of post-linguistically deafened cochlear implant users with those of normally hearing speakers. However, it cannot be used with prelinguistically deafened cochlear implant users because not enough is known of their tone perception or tone production abilities to be categorically sure that they (a) would produce the same variety of linguistic tone as their hearing-impaired peers or (b) would produce tonal patterns in response to the same tonal targets as their normally hearing peers.

B. An alternative approach to the acoustic analysis of tone production

Gandour (1978) has shown that five dimensions defining acoustically measurable differences between tones account for a listener's perceptual judgments about tone. These are labeled interpretatively as (a) average pitch, (b) direction, (c) length, (d) extreme endpoint, and (e) slope. A plot of F0 offset versus F0 onset captures all dimensions except length, for each tone type. In such a plot, level tones would be expected to fall midway between the two axes at different points corresponding to different values of average F0. Rising tones would be expected to cluster closer to the *y* axis, while falling tones would cluster closer to the *x* axis.

Ellipses around each tone type can be calculated by determining the distribution of points around a mean for each tone. This is done by first calculating the directions for the major and minor axes for each ellipse using a principal components analysis of the points for the corresponding tone. The data points are projected on to each axis and the standard deviations of the projections calculated. The ellipse radii are then set equal to twice the standard deviation to encompass approximately 95% of the projections on to each axis. Tonal ellipses calculated in this way visually summarize the locations in the plot where each of the six tonemes clusters. These plots are based on categories of perceptual relevance to the listener. They are predicted to be well differentiated where a speaker uses consistent patterns of pitch for each tone. Furthermore, the degree of differentiation should correlate with results based on auditory analyses of the same data. The higher the number of correct tones produced by a speaker, the greater the degree of differentiation of the ellipses in the tone plot.

C. Indexing degrees of differentiation between tones

Apart from describing tone differentiation with respect to visual comparison of the spread and degree of overlap between tonal ellipses, two indices were also developed which provide a numerical measure of the degree of tonal differentiation demonstrated by a speaker.

1. Index 1—Measuring tone differentiation within the tonal space

By acoustic measurement, the tones 55, 25, and 21 are the most differentiated of the six tones in Cantonese. Lines joining the center points of these three tonal ellipses form a triangle. This triangle effectively spans the mean range of F0 used by a speaker and defines the extent of the speaker's tonal space. The area of each ellipse describes the degree of variation of F0 used for each target tone. Tonal differentiation across the tonal space is thus a function of the area of the tonal space (*At*) and the spread of F0 usage for the three tones making up that area (i.e., the average area $Ae_{1,2,4}$, of the ellipses for tones 55, 25, and 21, where 1 = Tone 55, 2 = Tone 25, and 4 = Tone 21),

$$\text{Index}\ \ 1=\frac{At}{Ae_{1,2,4}}. \tag{1}$$

Where $At/Ae_{1,2,4}>2$, the likelihood of overlap between these ellipses is low. For ratios $At/Ae_{1,2,4}\leqslant 2$, tone ellipse overlap is likely between at least two of the ellipses; the lower the ratio, the greater the degree of overlap between the three ellipses.

2. Index 2—Measuring differentiation among tonemes

Like the ellipse area, the lengths of the two axes around which the ellipse is drawn describe the degree of variation in pitch used for each tone. The distances between the six tone ellipse centers determine the degree of difference in the average values of F0 for each tone. Differentiation within the tonal space can thus be expressed as the ratio of the average of the lengths of the two axes for the six tones (Ave Ax_{1+2}) and the average distance of the centers of the six tone ellipses from each other (Ave Dist.).

$$\text{Index}\ \ 2=\frac{\text{Ave Dist.}}{\text{Ave Ax}_{1+2}}. \tag{2}$$

Index 2 has several advantages over Index 1. It relies on measures for all tones rather than three. It is sensitive to differences in pitch height and contour individually. For example, in the case where a child controls pitch height but not pitch contour, tones 55, 25, and 21 would all lie on the diagonal of the plot and the area of the triangle defining the tonal area would be zero. Since Ave Ax_{1+2} is determined from calculated standard deviation units around the mean of the distribution of points in the F0 onset/offset space, Index 2 is analogous to the signal detection measure d' (*d* prime).

D. Summarizing the approach to acoustic analysis of tone

The focus of our approach to the acoustic study of tone production involves observing the degree of differentiation between tone types produced by an individual speaker. The greater the differentiation among the ellipses, the more likely it is that a speaker has a consistent tone target, i.e., has acquired a toneme. Because this approach is at the level of the individual, there is no requirement for a speaker to have the same tonal target as any other speaker, making it suitable for use with young prelinguistically deafened implant users. In this paper, we shall illustrate how this approach can be used to describe differences in tone production in three groups of speakers: (a) normally hearing adults; (b) young normally hearing children aged between 4–6 years who are reported to have mastered tonal production at this age; and, (c) young prelinguistically deafened implant users.

II. METHOD

A. Subjects

This study included five normally hearing adults: (two male, three female); eight normally hearing children (three male, six female); and, 16 prelinguistically profoundly hearing-impaired implant users (seven male, nine female). All participants were native speakers of Cantonese. The age range for each group of speakers was 23–40 years (adults); 3 years 8 months to 6 years (normally hearing children); and 4 years 2 months to 11 years 3 months (implant users).

All implant users received their implant at the Prince of Wales Hospital (PWH) in Shatin, Hong Kong. They were fitted with either a Nucleus CI-22M (six subjects) or CI-24M (10 subjects) electrode manufactured by Cochlear Limited. At the time of testing, seven children were using the SPEAK speech processing strategy in a SPECTRA™ speech processor (McDermott *et al.*, 1992). The remaining nine children were using the ACE speech processing strategy in a SPRINT™ speech processor (Vandali *et al.*, 2000). The pulse rate for the SPEAK users was 250 Hz. The pulse rate used by the ACE users varied according to mapping. Seven children used a pulse rate of 900 Hz, one child used a pulse rate of 720 Hz and one child used a pulse rate of 1200 Hz. The pure tone average (i.e., average hearing loss in dB HL at 500, 1000, and 2000 Hz) for the group as measured prior to implant in the better ear was 101dB HL (s.d. 11 dB HL). The age at implant ranged from 2 years 3 months to 6 years 9 months. The number of months experience using the implant ranged from 15 months to 73 months.

All implant users received habilitation within an oral–aural scheme. Subsequent to receiving the implant, the children underwent a 2-year weekly program of habilitation at PWH. At the time of testing, 10 children attended schools dedicated to the education of the deaf. Three children attended an integrated kindergarten and were withdrawn everyday for extra habilitation. The remaining three children attended normal primary schools and received extra habilitation from the local audiological service.

B. Recording procedures

1. Materials

The words used in the test were collected as part of a battery of tests investigating tone perception and tone production skills in Cantonese-speaking children. Testing was carried out in a soundproof room at PWH. For the speech production task, the children were required to produce tones in citation form. These were elicited by means of a picture-naming task involving 15 presentations of each tone type on a range of syllables. There were 90 items in total. The words in the test were chosen because they had been shown to be frequently used by Cantonese-speaking children aged three and older (Lee, 2003). They were presented to the children as black and white line drawings arranged two to a page. Where children could not spontaneously name a picture, the principal investigator would try to elicit the word by asking a question, e.g., ‘‘Is this a boy? No it’s a ... .’’ Some items in the test involved tones presented on closed syllables ending in [−p, −t, or −k]. These tones are referred to as clipped or stopped. They are characterized by a shorter duration and were excluded from further acoustic analysis because of differences in form.

2. Recording and pitch extraction

Data from the children were first recorded on to audiotape using either a Portable Cassette Recorder Sony TM5000EV with an internal microphone or an AIWA Millennium Collection JXM2000 Walkman recorder and microphone. When using the latter, the microphone was fitted to the child’s collar during recording. Speech samples were then digitized using ‘‘Speech Analyzer’’ (1999) a freeware software for the Microsoft Windows™ operating system developed as part of the acoustic speech analysis project at the Summer Institute of Linguistics. The speech for the normally hearing adults was recorded directly into ‘‘Speech Analyzer’’ using an electret condenser 600 ohms microphone. A 16-bit monochannel, 22.05 kHz sampling rate was used for digitizing all speech samples. Utterances were then phonetically labeled by auditory analysis to segment the initial consonant from the following rhyme. Phonetic labels and pitch values (sampled at 1 ms intervals) were exported using Speech Analyzer’s ‘‘Timetable’’ format and downloaded into a program called ‘‘Pitch.’’ This program extracted values of F0 at onset and offset in a format which was for suitable further analysis.

3. Determining the onset and offset of F0

‘‘Pitch’’ identified the approximate boundaries for each syllable rhyme based on changes in amplitude. These boundaries were checked manually and adjusted to align with points of onset and offset where post-consonantal fluctuations or fluctuations due to vocal fold closure were negligible. In general, F0 onset was readily identifiable and corresponded to the point where the vowel amplitude reached a post-consonantal maximum. Identification of the pitch offset was more complex, particularly for the tones produced by the implant users. Typically, it corresponded with the point where the vowel amplitude had declined to 50% of the maximum. This criterion could not be applied to syllables ending on a nasal. Word-final nasals tend to have a low amplitude relative to the preceding vowel. The F0 offset marker was manually adjusted to a point near the end of the syllable which was representative of the observed trend in the preceding contour.

Test–retest reliability of this approach for marking tone boundaries was performed on data from four speakers (one adult, one normally hearing child, and two implant users). The time separation between the two analyses was more than 2 years. Intra-Class Correlation values (ICC) using a one-way random effects model were calculated for the six mean tone onsets and offsets and the 12 axis lengths. ICCs calculated for F0 onset and offset were above 0.90 for all tones, indicating high test–retest reliability. Similarly ICCs calculated for all tone axis lengths except for Tone 25 were above 0.9. The ICCs for the Tone 25 axes were 0.80 and 0.78. Though lower than the other values calculated, they still indicate acceptable test–retest reliability.

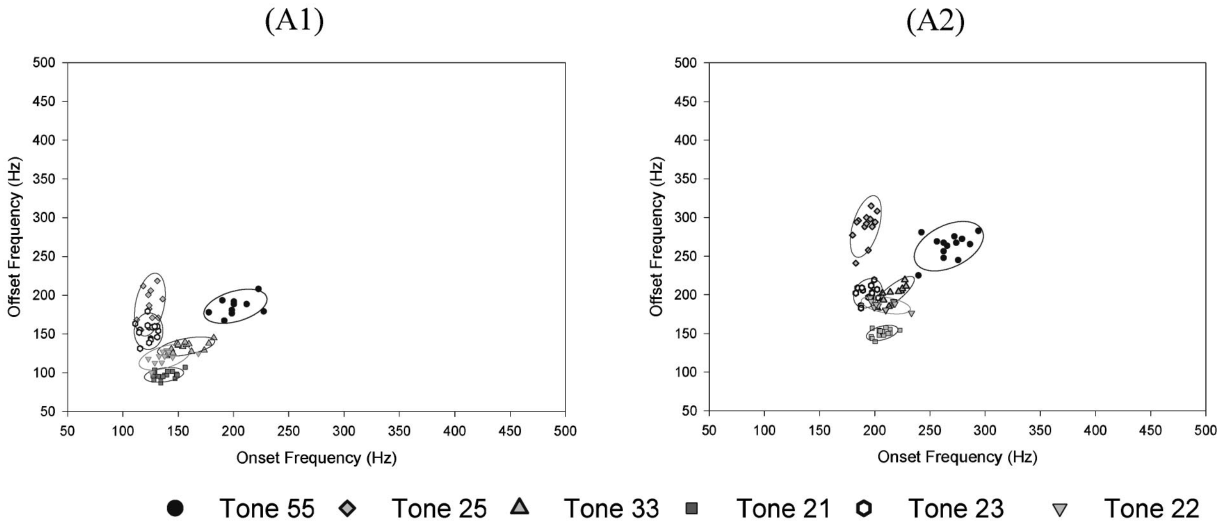

FIG. 2. Tone ellipse plots for two adult speakers. A1 is a male normally hearing adult. A2 is a female normally hearing adult. The plots with the most differentiated ellipses for the group are presented in each case. These plots demonstrate the triangular shape of the tonal space in Cantonese as well as the tightly defined tonal ellipses typical of adult speakers. Some crowding of the three tones (Tone 33, 23, and 22) is also apparent in the center of the tonal space.

C. Data analysis

Three forms of analysis were undertaken to compare tone production in the three groups of speakers. First, as described in the Introduction (Sec. I B), F0 offset versus F0 onset values for all speech tokens were plotted to encompass approximately 95% of the tones tokens making up each tone type. Second, the following parameters were calculated: (a) the lengths of the axes of the tonal ellipses; (b) the areas of the tonal ellipses; and (c) the distances between the center points of each of the tonal ellipses. These parameters were used to calculate Index 1 and Index 2 [Eqs. (1) and (2) above] which in turn function to numerically summarize the degree of tonal differentiation observed in an individual's tonal plot. Third, an auditory transcription analysis of the speech production data for each of the three groups of speakers was undertaken. These data were summarized as tone confusion matrices. Percent correct tones (PCT) was also calculated for each child.

D. Auditory transcription analysis

As illustrated in Fig. 1, there is considerable crowding in the Cantonese tonal system since five of the six tones in Cantonese have similar onsets. Further, tones 25 and 23, and tones 33 and 22 have similar contours and are primarily distinguished by their different offsets. The acoustic similarities among these tones mean that the Cantonese tonal system represents a considerable challenge for reliable transcription. Moreover, the first author who performed the auditory analyses of the data is not a native-speaker of Cantonese. To ensure consistent auditory transcription of the data, protocols for tone labeling were established with reference to linguistic-tonetic description of Cantonese tones developed by Rose (2000). For example, to be labeled a high rise Tone 25, a tone contour had to be uttered with a sharply rising contour and the F0 offset had to be closer to Tone 55 than Tone 33. Similarly, to be labeled as a low-fall Tone 21, a tone had to have a clearly falling contour shape ending with an audibly lower offset than a Tone 22. It is more difficult to label tones in citation form than in sentence context (Wong and Diehl, 2003) and to aid the transcriber in normalizing to a speaker's typical pitch range, all tones produced by a single speaker were transcribed in a block.

For implant users who had good control of F0 across the syllable, but did not use consistently differentiated pitch levels across a series of tone productions, it was often very difficult to identify a pitch range for the speaker or to label specific tones within it. Nonetheless an auditory analysis of F0 was attempted for such speakers despite some doubt about whether it was appropriate to do so. Implant users with poor control of prosodic features such as syllable duration often used pitch patterns that varied unpredictably across the syllable. No attempt was made to perform an auditory transcription of the F0 contours produced by these speakers.

Inter-rater reliability measures were calculated for the first author and a native speaker trained in phonetic transcription (Shriberg and Lof, 1991). Tone transcriptions were compared for one normally hearing child and two implant users. The children varied in intelligibility from highly intelligible to fairly unintelligible. An agreement of 81.8% was observed for the normally hearing child, 80% for the implant user with a high level of intelligibility and 67.6% agreement for the implant user with a low level of intelligibility. These are within inter-transcriber measures reported for transcription of segmental phonemic contrasts in English (Shriberg and Lof, 1991) but less than was reported for transcription of Cantonese tonal contrasts by So and Dodd (1995). They observed 100% agreement between two native-speakers of Cantonese transcribing tones produced by normally hearing Cantonese-speaking children. This may be an overestimation of inter-transcriber reliability, given that the word lists were small and familiar to both transcribers.

III. RESULTS

Tone productions for each speaker were plotted. The plots with the most clearly differentiated tonal ellipses are shown for each group of normally hearing speakers (Fig. 2,

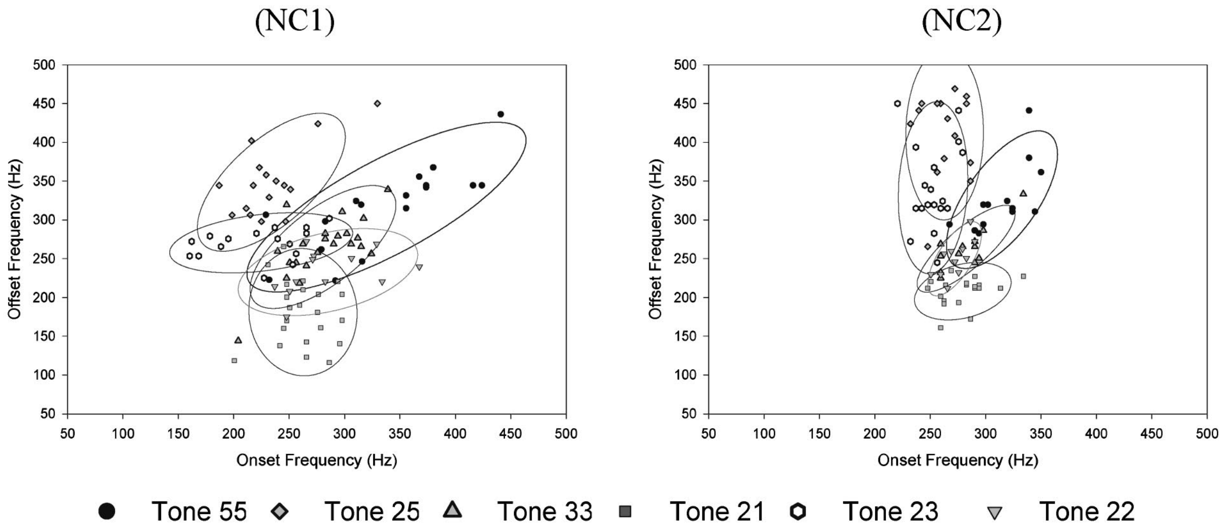

FIG. 3. Tone ellipse plots for two normally hearing children. There was considerable variation apparent in the tone plots from this group and the plots shown here are from children who demonstrated the greatest tonal differentiation. They illustrate the characteristic features of plots from this group, namely: large tonal ellipses located in a large tonal space and reduced tonal differentiation relative to the tonal plots from the adult speakers.

adults; Fig. 3, children).

Tone plots for two male implant users are shown in Fig. 4. The plot for CI1 was typical of this group of speakers. CI2 by contrast demonstrated relatively good tonal differentiation. Only three other implant users in the present study exhibited similar levels of differentiation.

Shapiro–Wilk tests for normality were performed on Index 1 and Index 2; tone ellipse area; *At*, the tonal space; the average of the sum of the two axis lengths for each tone; and, the 15 intertone distances. Only the ellipse area for Tone 33 and four of the 15 intertones distances demonstrated near normal distribution. Statistical analysis of the data thus required the use of the nonparametric Kruskal–Wallis test. Box plots (Figs. 5 and 6) summarize medians and variation around the medians across the three groups for Index 1 and Index 2 (Fig. 5); and, for the largest and smallest tone ellipse areas and the tonal space *At* (Fig. 6).

A. Statistical analysis of the data

1. Statistical comparison of results

Group of speaker and tone area for each tone type were entered into a Kruskal–Wallis nonparametric analysis. Significant differences in median tone area for the three groups of speakers were found for all tones. These differences ranged in significance from [χ^2 (2)$=$14.028; $p<0.001$, Tone 25] to [χ^2 (2) 8.354; $p<0.05$, Tone 33]. As illustrated in Fig. 6 adult speakers had the smallest median ellipse areas, normally hearing children had the largest median ellipse areas.

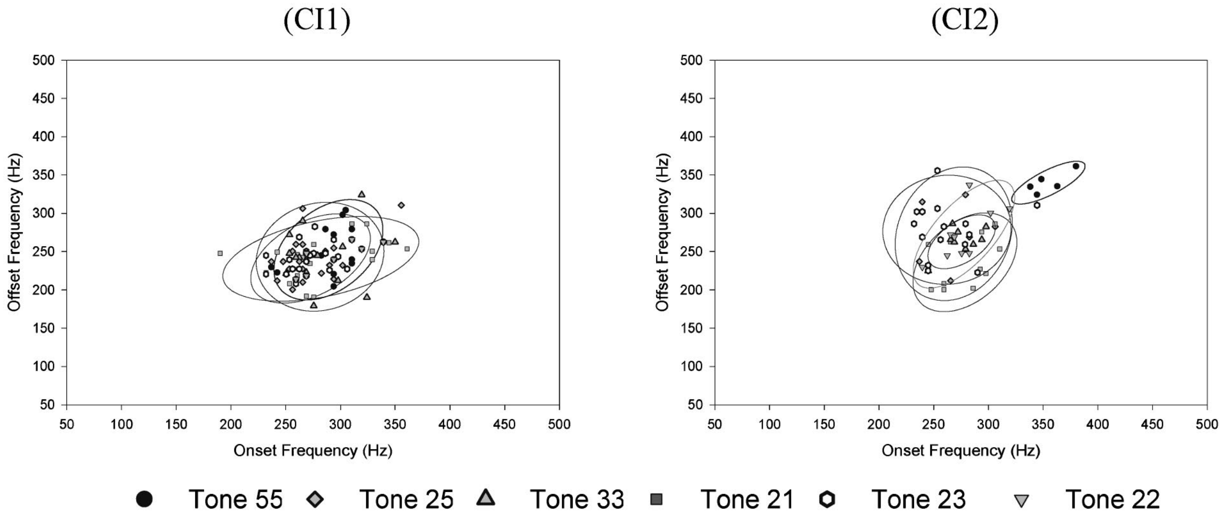

FIG. 4. Tone ellipse plots for two implant users. The plot shown for CI1 is typical of the majority of implant users in the study. The neat centrally located tonal ellipses reflect similar values of F0 onset and offset for most tone tokens regardless of type. The tonal ellipses all cluster in one location in the plot indicating little tonal differentiation. CI2 was one of four implant users who demonstrated reasonable tonal differentiation. The tonal ellipses are more dispersed than they are for CI1. The most clearly differentiated tone is tone 55 which appears as a distinct ellipse separated from the other five tonal ellipses. This suggests a clear opposition between the features high and nonhigh.

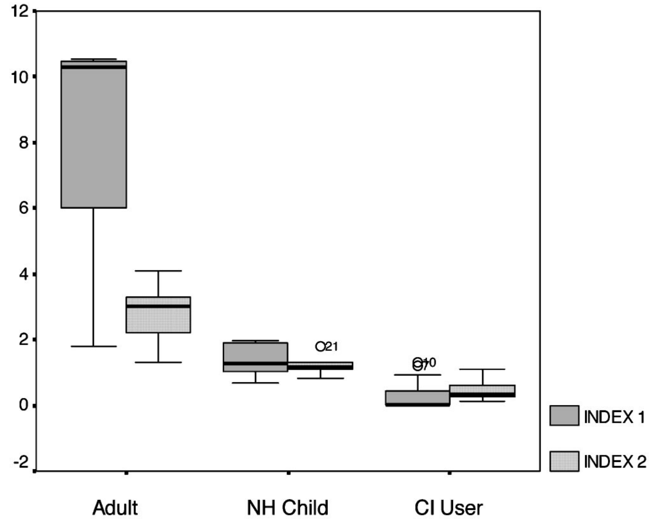

FIG. 5. Box plots for Index 1 and Index 2. The box indicates the interquartile range of values obtained with the median (indicated by the solid horizontal line). The range of measurement is shown by whiskers extending from the box except for points more than 1.5 (indicated by "○") or 3 box lengths (indicated by "*") from the upper or lower edge of the box.

No comparison of the lengths of the ellipse axes was performed because these values were used to calculate the ellipse areas and a similar significant relationship would be expected.

2. *Differences in tone ellipse separation for the three groups of speakers*

Subject group and intertone distances were entered into a Kruskal–Wallis nonparametric analysis. A significant difference in 14 of the 15 median tone distances for the three groups of speaker was observed. These differences ranged from [χ^2 (2)$=$18.592; $p<0.0001$ for the median intertone distance between Tones 25 and 21] to [χ^2 (2) 7.865; $p<0.05$ for the distance between Tones 23 and 22]. No significant difference in the median tone distance between Tone 33 and Tone 22 was observed.

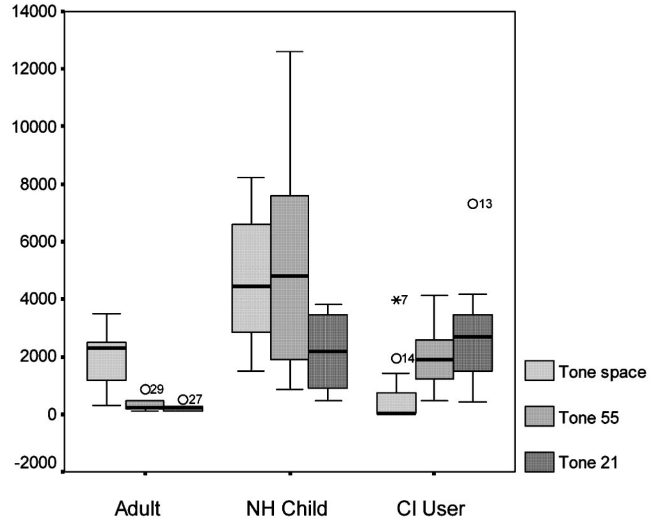

FIG. 6. Box plots summarizing the variation in tonal space and the range of variation in tone ellipse areas for the three groups of speakers. To indicate variation in tone ellipse area, box plots are shown for Tone 55 which was typically one of the larger tonal ellipses together with box plots for Tone 21 which was one of the smaller ellipses among the three groups of speaker.

TABLE I. Tone confusion matrix from an auditory analysis of tones produced by normally hearing adults. Confusions are shown as a proportion of total target productions for each tone.

	Actual					
Target	55	25	33	21	23	22
55	0.97		0.02	0.01		
25		0.88		0.05	0.02	
33			0.94	0.02	0.01	0.03
21				0.90		0.10
23		0.05	0.01		0.94	
22			0.07	0.04	0.01	0.88

The implant users tended to have the smallest intertonal distances with medians ranging from 10.1 Hz (Tone 55–Tone 23) to 32.0 Hz (Tone 55–Tone 23). The range of median intertone distances for the normally hearing children from 147.2 Hz (Tone 25–Tone 21) to 16.9 Hz (Tone 33–Tone 22) as compared with 83.5 Hz and 16.63 Hz, respectively, for the adult speakers.

Tonal space, *At*, and speaker group were entered into a nonparametric Kruskal–Wallis analysis. A significant difference between the median tonal spaces for the three groups of speakers was observed [χ^2 (2) 16.891, $p<0.001$]. The median tonal spaces were 40.23 Hz^2 (implant users); 4452.81 Hz^2 (normally hearing children); and 1965.22 Hz^2 (normally hearing adults).

3. *Indexing differentiation*

Values of Index 1 and speaker group were entered into a Kruskal–Wallis nonparametric analysis. A significant difference in median values for Index 1 was observed [(χ^2 (2) 18.746 $p<0.001$]. The median values for Index 1 were 0.015 (cochlear implant users); 1.27 (normally hearing children); and 10.28 (normally hearing adults).

Index 2 and speaker group were entered into a Kruskal–Wallis analysis. A significant difference in Index 2 among the three groups of speakers was observed [χ^2 (2) 20.866 $p<0.001$], medians: 0.32 (implant users); 1.17 (normally hearing children); and 3.01 (adults). The order of magnitude for both indices was CI users <NH children <NH adults.

B. Auditory transcription analysis of tone production

The average PCT (percent correct tones) observed for the normally hearing children by auditory analysis was 78%. The observed range of PCT produced was 60%–93%. A Pearson's product moment correlation analysis on PCT and the area of the tonal space *At* showed a significant positive correlation between *At* and PCT ($r=0.843, p<0.01$).

A confusion matrix was prepared for the tones produced by all the normally hearing adults (Table I) and children (Table II). Relative to the children, fewer confusions are apparent in the adult data though the trends are similar with most confusions being apparent between tones 33 and 22, and tones 25 and 23. In all, where the children were required to produce target tones 33 and 22, 20% were confusable,

TABLE II. Tone confusion matrix from an auditory analysis of tones produced by normally hearing children. Confusions are shown as a proportion of total target productions for each tone.

	Actual					
Target	55	25	33	21	23	22
55	0.89		0.11			
25	0.02	0.70	0.01	0.07	0.19	0.01
33	0.04		0.78	0.01		0.17
21			0.02	0.80	0.03	0.15
23	0.01	0.14	0.01	0.01	0.69	0.14
22			0.22	0.04		0.74

16% of tones 2 and 5 were confusable, and 10% of tones 4 and 6 were confusable. This compares with 5%, 6%, and 6%, respectively, in the adult data.

The average PCT produced by the implant users was 38% correct with a range of 21–59 % correct. A confusion matrix for the data is provided in Table III. Contour tones had the lowest percent correct scores and typically they were perceived as either Tone 33 or Tone 22.

IV. DISCUSSION

This study compared tone productions by children using a cochlear implant with two groups of normally hearing speaker. The aim was to demonstrate the application of a methodology for the acoustic analysis of tone production. In this methodology, F0 offset was plotted against F0 onset for a series of tone tokens and the degree of differentiation between ellipses encompassing each tone target was observed qualitatively and then summarized numerically. Since F0 onset and offset were the only points plotted, the emphasis in the methodology was on observing pitch level differences between tone types. Some contour effects were also reflected by overall position of individual ellipses in the tone plot.

This approach is quite logical for the analysis of Cantonese, since pitch level has been suggested to be perceptually more salient to the Cantonese-speaker than pitch contour (Gandour, 1983). Given the relative perceptual importance of pitch level to speakers of Cantonese and the fact that pitch level differences between some tones are quite small (Fig. 1) one would predict that the range of onset and offset values for each of the six citation form tones would necessarily be narrow to maintain pitch level differences, i.e., all tokens for a particular tone target would congregate in tightly defined ellipses covering restricted areas of the tonal plot.

TABLE III. Summary of tone confusions from an auditory analysis of tones produced by children using a cochlear implant. Confusions are shown as a proportion of correct tones per total number of target tones.

	Actual					
Target	55	25	33	21	23	22
55	0.39	0.03	0.40	0.02	0.02	0.14
25	0.10	0.19	0.36	0.07	0.13	0.15
33	0.06	0.01	0.64	0.05	0.04	0.20
21	0.04	0.03	0.29	0.23	0.06	0.36
23	0.02	0.09	0.33	0.04	0.22	0.28
22	0.04	0.0	0.29	0.06	0.03	0.58

This prediction was confirmed in part by the tone production data for the adult speakers. Among this group of speakers, greatest overlap was observed between tones 33, 23, and 22. Some overlap was also observed between tones 25 and 23.

The plots obtained for the normally hearing children and the implant users were quite different to those obtained for the group of normally hearing adults. First, the areas of the tone ellipses for both groups of children were larger than they were for the older speakers. Second, the median size of the tonal space *At* for each of the three groups of speakers were significantly different from each other.

The tonal space for the implant users was smaller than it was for either group of normally hearing speaker. Combined with the relatively diffuse tonal ellipses, this resulted in little or no clear differentiation among the tones produced by these children. By contrast median value of *At* for the normally hearing children was over twice as large as median *At* observed for the normally hearing adults (4452.81 Hz^2 versus 1965.22 Hz^2). This suggests that among normally hearing speakers, there is a direct relationship between the spread of pitch used for each tone type and the size of the tonal space in which the tonal ellipses are located. The young children in this study had diffuse tonal ellipses but maintained a reasonable degree of differentiation by virtue of their broad tonal spaces.

Given the similarities in F0 among some tones in Cantonese, it is interesting to note how the normally hearing children had significantly larger tonal ellipse areas than the implant users. This suggests a greater spread of pitch usage for each tone type. The normally hearing children are reported to master tone production early and because of having access to all available auditory information about tone, one might have predicted that their tonal ellipses would have been smaller than those of the implant users.

Testing conditions for the two groups of children were kept constant, however the normally hearing children were younger than many of the implant users and were less used to such tasks. These factors may partially explain some of the variation observed between the two groups of children. It is further possible that the younger normally hearing children are still developing control of vocal fold use during speech further explaining some of the differences in the spread of their tonal ellipses relative to the other groups of speakers. Nonetheless, the degree of difference in the size of the tonal spaces and tonal ellipses for the three groups of speakers suggests that there are real differences in tone production abilities among the three groups. These results further suggest that it is not enough to evaluate differentiation among individual tones through examination of the spread of F0 usage. Measures of tone ellipse separation within the tonal space and degree of ellipse overlap are also required.

Index 1 and Index 2 were developed to meet this requirement and to summarize the interplay between the ellipse area and tonal space in determining the degree of tone differentiation. The three groups of speakers were clearly identifiable based on their observed values of Index 1 and Index 2. The medians of both indices were significantly greater for the adults than for the normally hearing children.

In the case of the implant users, the median for Index 1 was close to zero and for Index 2 was barely above noise. Both values indicate little to no differentiation in the implant users' production of tone.

A. Auditory transcription analysis of the data

Normally hearing children averaged 78% correct tone production, though there was considerable variation. Among the implant users, there was an average PCT of 38% (range 21%–59%). This is barely above chance level. Two points can be made with respect to these data.

First, Wong and Diehl (1999) found that the size of the tonal space of the speaker seems to affect the relative ease with which the different tones produced by a Cantonese speaker can be perceived. This observation also holds true for the participants in this study. In the case of the young normally hearing children a significant positive correlation was observed between the size of the speaker's tonal space and the percentage correct tones produced by the speaker ($p < 0.01$).

Second, these observations are significantly different to results from previous studies investigating tone development in children acquiring Cantonese. Typically, it is reported that Cantonese-speaking children acquire a complete toneme inventory by two years of age and make very few tonal errors thereafter (So and Dodd, 1995). Similarly children with impaired hearing have been reported to be able to acquire most if not all tones in Cantonese (Dodd and So, 1994). Could the differences in the results presented here reflect the fact that the auditory analyses were performed by a non-native speaker of Cantonese? This does not seem to be an adequate explanation, given that the auditory analyses are supported by acoustic analysis of the same data.

B. A conundrum: Tone production is mastered before tone perception

The literature suggests that tone production is mastered early by both hearing and deaf children learning Cantonese. By contrast, adultlike tone perception skills do not develop until around the age of 10 years (Ching, 1984). Ching concluded that the younger children in her study were generally able to identify gross pitch contour differences but they still had to develop the capacity to normalize for pitch level differences across tones.

In the case of profoundly hearing-impaired children, Ching (1988) reported that they perform at chance levels on tone perception tests except when listening to larynx tones presented as sinusoids. ''High'' was the best perceived tonal feature and the children showed a basic opposition in their tone identification between high and nonhigh. Further, Dodd and So (1994) also noted that tone production skills were better than tone perception skills in the hearing-impaired children in their study.

Despite some discussion about the exact relationship between speech perception and the development of spoken language, it is generally agreed that perception must precede production. How is it that the acquisition of good tone production abilities in normally hearing children seems to outstrip the acquisition of accurate tone perception abilities? Or, in the case of children with impaired hearing, how is it that despite a reduced ability to accurately perceive tonal contrasts, these children reportedly acquire and accurately produce most tonal contrasts in Cantonese?

The acoustic analysis of the tone production data indicated significant differences in tone production abilities between profoundly hearing-impaired prelinguistically deafened speakers, normally hearing children and normally hearing adults. The primary tonal confusions by auditory analysis among the normally hearing children were between tones 33 and 22, tones 23 and 25, and tones 21 and 22. This was also reflected in individual plots of F0 offset versus F0 onset for each of the children such that there was considerable overlap for tones 33 and 22 and tones 23 and 25 and also some overlap in the ellipses produced for tones 21 and 22 (refer to Fig. 3 for two examples). Thus, while it is true that tone confusion results are dependent on the skills of the transcriber to correctly identify and label the tones produced, acoustic analysis of the same data supports the results from auditory analyses.

Among the implant users, only four children demonstrated any meaningful tonal differentiation (CI2 in Fig. 4 is an example of one of these children). However, auditory analysis showed that though most of the implant users produced a range of F0 contours that could be labeled as corresponding to a particular toneme in Cantonese, in fact, they did not occur with sufficient frequency in the phonologically correct context to be judged as having been acquired. Auditory analyses averaged across the group of implant users showed that the likelihood of correct production of a particular tone type on the correct syllable was barely above chance at 38%. Once again these results match reported tone perception results for implant users. Ciocca *et al.* (2002) using a simple, forced-choice tone identification test involving minimal pair alternatives found that most participants performed below chance on most of the 15 tone contrasts in Cantonese. Further, like the hearing-impaired children in Ching's study (1988), they noted that implant users were more successful at identifying tones with larger separations in average pitch levels. This effect is evident in the plot for CI2 (Fig. 4) where the Tone 55 ellipse is quite distinct from the other tonal ellipses in the plot.

Overall the results from the acoustic analysis support results from auditory analyses. Furthermore, results from both methodologies reflect observations from studies investigating tone perception abilities in normally hearing children and children using a cochlear implant. Moreover the large tonal spaces noted for the young normally hearing children in this study may provide interesting evidence from the analysis of speech production materials to support Ching's hypothesis (1984) that Cantonese-speaking children of this age are still acquiring skills for normalization of pitch level differences among tones.

Taken together, these results suggest a close link between a child's tone perception abilities and the subsequent development of tone in spoken language. From these data, the possibility that mastery of tone production outstrips the development of adultlike tone perception abilities is not tenable and the potential conundrum disappears.

V. CONCLUDING REMARKS

A key aim for this study was to develop a means for analyzing tone production acoustically to support or enhance auditory analyses of speech production materials. F0 offset versus F0 onset was plotted to examine general tendencies in F0 usage for different tone types. Two indices were also developed for numerically summarizing the degree of differentiation between tone ellipses. Using this approach, it was shown that normally hearing adult speakers have small ellipses located in a relatively small tonal space. Young normally hearing children have large tonal ellipses together with a large tonal space. These children are reported to have mastered tone production by the age of two yet the acoustic data suggest that their production skills are still maturing. An interesting aspect of these results is that they support conclusions from a variety of studies investigating Cantonese tone perception.

There are considerable differences between the auditory analyses reported here and results from previous studies investigating tone production by young speakers of Cantonese. These may reflect the fact that the analyses were performed by a non-native speaker of Cantonese. However, acoustic analyses of the data provide effective back-up for the auditory analyses performed and raise some interesting questions about the nature of tone perception. For one thing, an important role for linguistic experience when labeling tones is suggested by the fact that young children who have such broadly spread overlapping tonal ellipses are judged to have mastered tone production by native speakers. This raises a further question regarding the degree to which tone perception is categorical.

In summary, this approach to the acoustic analysis of tone supports what is known about tone perception abilities in young speakers and enhances understanding about tone production based on auditory analyses. It is proposed that the procedure has application both in cross-sectional studies to observe differences between groups of speakers and in longitudinal studies to observe developmental changes in tone production. It may also have application for aiding in tone habilitation. A prototype habilitation tool employing this methodology has since been developed which can be downloaded. (URL http://www.psych.ox.ac.uk/oscci/johanna%20barry/v2/index.htm)

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of the children, their families and the staff of the Prince of Wales Hospital Shatin New Territories Hong Kong. In particular, we thank Dr. Kathy Lee who provided the list of words used in data collection and Grace Chan who phonetically transcribed samples for obtaining inter-rater transcription measures. Dr. Johanna Barry was funded by an Australian Postgraduate Award and a Victoria Fellowship. Further support was also provided from an Australian National Health and Medical Research Council (NHMRC) project Grant No. 970257 and the Bionic Ear Institute, Melbourne. Dr. Peter Blamey was supported by an Australian NHMRC Senior Research Fellowship (No. 208998). The research was conducted with the approval of the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital (No. 99/376H) and the Clinical Research Ethics Committee at the Chinese University of Hong Kong (No. CRE 9271).

Abramson, A. S. (**1986**). ‘‘The Thai tonal space,’’ Status Report of Speech Research (No. 85), Haskins Laboratories, pp. 105–114.

Ching, T. (**1984**). ‘‘Lexical tone pattern learning in Cantonese children,’’ Lang. Learn. Commun. **3**, 243–414.

Ching, T. (**1988**). ‘‘Lexical tone perception by Cantonese deaf children,’’ in *Cognitive Aspects of the Chinese Language Vol. 1*, edited by I. M. Liu, M. J. Chen, and H. C. Chen (Asian Research Service, Hong Kong), pp. 93–102.

Ciocca, V., Francis, A., Aisha, R., and Wong, L. (**2002**). ‘‘The perception of Cantonese lexical tones by early-deafened cochlear implantees,’’ J. Acoust. Soc. Am. **111**, 2250–2256.

Clumeck, H. (**1980**). ‘‘The acquisition of tone,’’ in *Child Phonology: Production Vol. 1*, edited by G. H. Yeni-Konshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), pp. 257–275.

Dodd, B. J., and So, L. K. H. (**1994**). ‘‘The phonological abilities of Cantonese-speaking children with hearing loss,’’ J. Speech Hear. Res. **37**, 671–679.

Gandour, J. (**1983**). ‘‘Tone perception in Far Eastern languages,’’ J. Phonetics **11**, 149–175.

Gandour, J., Weinberg, B., Petty, S. H., and Dardarananda, R. (**1988**). ‘‘Tone in Thai alaryngeal speech,’’ J. Speech Hear Disord. **53**, 23–29.

Gandour, J. T. (**1978**). ‘‘The perception of tone,’’ in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 41–76.

Keating, P. (**1980**). ‘‘Patterns of fundamental register and vocal registers,’’ in *Infant Communication: Cry and Early Speech*, edited by T. Murry and J. Murry (College-Hill, Houston, Texas), pp. 209–233.

Kent, R., and Murray, A. D. (**1982**). ‘‘Acoustic features of infant vocalic utterances at 3, 6, and 9 months,’’ J. Acoust. Soc. Am. **72**, 353–365.

Lee, K. (**2003**). ‘‘The construction and validation of a speech perception test for Cantonese-speaking children,’’ Ph.D. thesis, Chinese University of Hong Kong.

Li, C. N., and Thompson, S. A. (**1977**). ‘‘The acquisition of tone in Mandarin-speaking children,’’ J. Child Lang **4**, 185–199.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). ‘‘A new portable sound processor for the University of Melbourne/Nucleus Limited multi-electrode cochlear implant,’’ J. Acoust. Soc. Am. **91**, 3367–3371.

Rose, P. (**1987**). ‘‘Considerations in the normalization of the fundamental frequency of linguistic tone,’’ Speech Commun. **6**, 343–351.

Rose, P. (**1989**). ‘‘On the non-equivalence of fundamental frequency and linguistic tone,’’ in *Prosodic Analysis and Asian Linguistics to Honour R. K. Sprigg. Pacific Linguistics C-104*, edited by D. Bradley (Dept. of Linguistics, Research School of Pacific and Asian Studies, Australian National University, Canberra), pp. 55–82.

Rose, P. (**1993**). ‘‘A linguistic-phonetic acoustic analysis of Shanghai tones,’’ Aust. J. Ling. **13**, 185–220.

Rose, P. (**2000**). ‘‘Hong Kong Cantonese citation tone acoustics: A linguistic tonetic study,’’ Proceedings of the 8th Australian International Conference on Speech Science and Technology, Canberra, pp. 198–203.

Shriberg, L. D., and Lof, G. L. (**1991**). ‘‘Reliability studies in broad and narrow phonetic transcription,’’ Clin. Ling. Phonetics **5**, 225–279.

So, L. K. H., and Dodd, B. J. (**1995**). ‘‘The acquisition of phonology by Cantonese-speaking children,’’ J. Child Lang **22**, 473–495.

Speech tools support (**1999**). *Speech Analyzer*, Ver. 1.5, Test version 10.6 (Summer Institute of Linguistics Dallas, Texas).

Tse, A. C.-Y. (**1991**). ‘‘The acquisition process of Cantonese phonology—a case study,’’ M. Philosophy, University of Hong Kong, Hong Kong.

Tse, J. K.-P. (**1978**). ‘‘Tone acquisition in Cantonese: A longitudinal case study,’’ J. Child Lang **5**, 191–204.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). ‘‘Speech perception as a function of electrical stimulation rate: Using the Nucleus-24 cochlear implant system,’’ Ear Hear. **21**, 608–624.

Wong, P., and Diehl, R. L. (**2003**). ‘‘Perceptual normalization for inter-and intra-talker variation in Cantonese level tones,’’ J. Speech Hear. Res. **46**, 413–421.

Wong, P., and Diehl, R. L. (**1999**). ‘‘How Parkinson’s disease affects tone language and [what it] tells us about tone comprehension,’’ http://www.acoustics.org/press/137th/wong.html