The Role of Amplitude Envelope in Cantonese Lexical Tone Perception: Implications for Cochlear Implants

Yining V. Zhou, Brett A. Martin

Speech-Language-Hearing Sciences, Graduate Center, City University of New York, USA

yzhou@gc.cuny.edu, bmartin@gc.cuny.edu

Abstract

Amplitude envelope co-varies with F0 in Mandarin lexical tones. It can cue lexical tone perception in a tone language such as Mandarin which uses different pitch contours for phonemic contrasts. The current study investigated whether amplitude envelope could also aid the perception of lexical tones in a language such as Cantonese which uses both pitch contour and relative pitch height for phonemic contrasts. Stimuli containing only the amplitude envelope of Cantonese lexical tones were used, and native speakers of Cantonese identified the lexical tones contained in the stimuli with above-chance accuracy. Thus, amplitude envelope could cue pitch contour and pitch height for lexical tone perception. Amplitude envelope was found to co-vary with F0 in Cantonese lexical tones, and the degree of their covariability was a fair predictor of Cantonese lexical tone perception using amplitude envelope alone. Clinical implications of these findings for cochlear implants were discussed.

Index terms: lexical tone perception, amplitude envelope, F0, Cantonese, cochlear implant

1. Introduction

Amplitude envelope (AE) is defined as the slow fluctuation in the overall amplitude of a sound at the rate of 2 to 50 Hz [1]. The role of AE in lexical tone perception has been investigated in Mandarin [2][3][4]. A lexical tone (abbreviated as ***tone*** hereafter) is a pitch variation pattern associated with a word and contributes to the core meaning of the word. In Mandarin, for example, the syllable /ji/ with a falling tone represents the word "idea," whereas the same syllable with a level tone signifies "medicine." There are four lexical tones in Mandarin, each with a distinct pitch contour: a flat level contour for Tone 1, a rising contour for Tone 2, a falling-rising contour for Tone 3, and a falling contour for Tone 4 [2][3][4]. Native listeners of Mandarin could use the AE cue alone to identify the four Mandarin tones with approximately 60% accuracy in a four-alternative, forced-choice (4AFC) task (chance = 25%) [2][3][4]. Furthermore, F0 co-varied with AE in Mandarin tones [3]. The degree of this covariability was a good predictor of Mandarin tone identification in native speakers using the AE cue alone. That is, tones with a higher degree of covariability were identified with greater accuracy, and vice versa [3].

Previously published studies on the role of AE in tone perception focused on Mandarin. Each Mandarin tone has a distinct pitch contour, and thus only pitch contours are used for phonemic tonal contrasts. Some tone languages use both pitch contour and relative pitch height for phonemic tonal contrasts [5][6][7]. In Cantonese, for example, there are three basic pitch contours that are phonemically contrastive: falling, rising, and level. Relative pitch height is also used for phonemic contrasts: high falling vs. low falling, high rising vs. low rising, and upper middle level vs. lower middle level [5][6][7]. Thus, there are six phonemically contrastive tones in Cantonese: Tone 1 (High Falling), Tone 2 (High Rising), Tone 3 (Upper Middle Level), Tone 4 (Low Falling), Tone 5 (Low Rising), and Tone 6 (Lower Middle Level).

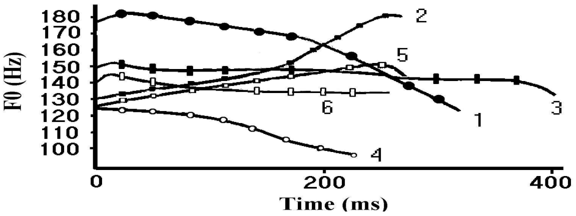

Figure 1: F0 of Cantonese tones (adapted from [6]).

The purpose of the current study is to determine whether AE alone can cue both pitch contour and pitch height in tone languages such as Cantonese which uses both pitch contour and pitch height for phonemic contrasts. Four questions will be answered: (1) Does F0 co-vary with AE in Cantonese tones? (2) Can native Cantonese-listeners use the AE cue alone to identify Cantonese tones above chance level? (3) Can the degree of covariability between the F0 and the AE of Cantonese tones be used to predict Cantonese tone identification in native Cantonese listeners using the AE cue alone? (4) What are the relative contributions of AE contour and height to Cantonese tone perception using AE?

To answer the first question, a tone production experiment was conducted. Cantonese words differing only in tones were produced by two native speakers, and an acoustical analysis was then performed to measure the co-variability of F0 and AE for each of the six Cantonese tones. To answer the second question, a tone perception experiment was conducted, and stimuli containing only the AE cue for Cantonese tone perception were used. The results from the tone production experiment were compared with the outcomes of the tone perception experiment to answer the third question. Finally, a multidimensional scaling analysis of the perceptual patterns observed in the perception experiment was conducted to answer the fourth question.

2. Methods

2.1 Participants

Twenty-five native Cantonese-speaking adults (eleven males and fourteen females) between 28 and 50 years of age participated in this study. All participants had pure-tone air conduction thresholds of 25 dBHL or better bilaterally at octave frequencies between 250 Hz and 8000 Hz. They also had type A tympanograms and present ipsilateral acoustic reflexes at 90dBHL at 1000 Hz.

Two participants (one male and one female) produced the stimuli for this study. Three other participants (one male and two females) served as naïve judges in assessing the perceptibility of the stimuli. The remaining twenty adults participated in the perceptual experiment.

2.2 Stimuli

Stimulus recording was done with a Technica Audio AT892CT4 head-mounted microphone, a pre-amplifier, a Dell Dimension E521 computer and the software Sound Forge. Six written Chinese characters, each representing the syllable /ji/ with one of the six Cantonese tones, were presented individually on a computer screen. The talkers read each character aloud, as if they were reading a vision exam chart. The characters were presented ten times, in a random order.

The rationale for using tones produced with the carrier syllable /ji/ in isolation was the following. First, it could control for several variables affecting Cantonese tone perception, such as co-articulation, intonation and stress [8]. Second, the syllable /ji/ represents the typical syllabic structure of a Cantonese word (i.e., the CV syllable). Third, the syllable /ji/ can constitute a meaningful word in Cantonese with each Cantonese tone. Lastly, the syllable /ji/ has been used in numerous studies on Cantonese tone perception studies [8][9]. Therefore, the use of this carrier syllable would facilitate comparison across studies.

Editing was done using the Praat software and according to [3]. The ten tone tokens produced by each talker for each tone type were averaged in amplitude and F0 as described below. First, each of the ten tokens was divided into twenty equal time frames, and the root-mean-square (RMS) amplitude and mean F0 within each time frame was computed. Second, for each tone type produced by each talker, the RMS amplitude within each time frame was averaged across the ten tokens, yielding a ten-token average. Similarly, the mean F0 within each time frame was also averaged across the ten tokens. Third, the token with the least deviation from the ten-token average was selected as the prototype of that tone of that talker. With six tones in Cantonese and two talkers in the current experiment, a total of twelve tone prototypes were selected.

The prototypes were then equalized in duration to remove the duration cue. The procedure consisted of two steps. First, the average duration of all sixty tone tokens produced by the female talker was used as the target duration. Second, the duration of each tone prototype produced by the female talker was increased or decreased to the target duration using the linear interpolation method described in [3]. This procedure was also used to equalize the duration of the tone prototypes produced by the male talker.

The perceptibility of the duration-equalized tone prototypes was rated by three naïve native listeners of Cantonese. The prototypes were presented to the three judges in a sound-attenuated booth via TDH-50 headphones, with the peak intensity level of each prototype calibrated to 70 dBSPL. The stimuli were blocked by talker because native listeners of Cantonese needed to hear several tokens produced by a talker to familiarize themselves with the talker's pitch range [10]. With one male talker and one female talker in the current study, two blocks of prototypes were presented to the judges in a random order. Within each block, the six prototypes were presented in a random order five times, with an interstimulus-onset-interval of five seconds. The judges were instructed to circle the corresponding Chinese character on the answer sheet upon hearing each stimulus.

The performance of the three judges on this 6AFC tone identification task ranged from 83% to 90% (mean = 86%; SD = 0.49; chance = 16.67%). These results were in line with findings from recent studies on Cantonese tone perception [9]. Thus, all twelve prototypes were accepted as natural tone stimuli in the perception experiment.

Using the Praat software and following the procedure in [3], AE was extracted from each of the tone prototypes using half-wave rectification and a 50-Hz lowpass filter (EIIR, 96 dB/octave). The extracted AE was a sequence of amplitude samples from the original waveform of each tone prototype. Half of the amplitude samples in the sequence were randomly selected, and the sign of each selected amplitude sample was reversed. This process transformed each tone prototype into a wide-band noise, known as the signal-correlated noise (SCN) [2][3][4]. Twelve SCNs were thus generated from the twelve tone prototypes. These SCNs, which only retained the AE of the original tones, were labeled ***processed tones*** in this study and used as stimuli in the perception experiment. The tone prototypes were labeled ***natural tones***.

Since the role of AE in Cantonese tone perception is the focus of this study, the peak intensity level of each processed tone was examined, and verified to be identical to that of the corresponding natural tone. The male talker's high falling tone had the highest peak intensity level (~70 dBSPL), followed by the female high falling tone (~69 dBSPL).

The processed tones and the natural tones were then separated into two sets. Within each set, the stimuli were further blocked by talker. Thus, there were a total of four blocks of stimuli for the perception experiment.

2.3 Experimental procedure

In the practice session, each participant listened to four blocks of stimuli presented in a random order, via a pair of TDH-50 headphones, at the original peak intensity levels of the stimuli. That is, the computer sound volume control was calibrated in such a way that a stimulus with an original peak intensity level of 70 dBSPL, for example, retained its peak intensity level of 70 dBSPL at the level of the headphones. Two tokens of each stimulus were included in each block, and the stimulus presentation procedure was the same as in the perceptibility rating session. Verbal feedback was provided. The practice session lasted between ten to fifteen minutes per listener. The actual experiment was then conducted in the same way as in the practice session, with the exception that ten tokens of each stimulus were presented in each block and no feedback was provided. The actual experiment lasted approximately forty minutes per listener, including a break between blocks.

3. Results and Discussion

3.1 Covariability of F0 and AE in Cantonese tones

Using the procedure described in Section 2.2, the average F0 contour and AE of each of the six Cantonese tones produced by each talker was computed. The female talker's average F0 contour and AE for each tone were highly correlated with those of the male talker according to the Pearson linear correlation analysis (r = 85%, p <0.001). Therefore, the two talkers' F0 contour and AE for each tone type were combined for further analysis, as displayed in Figure 2.

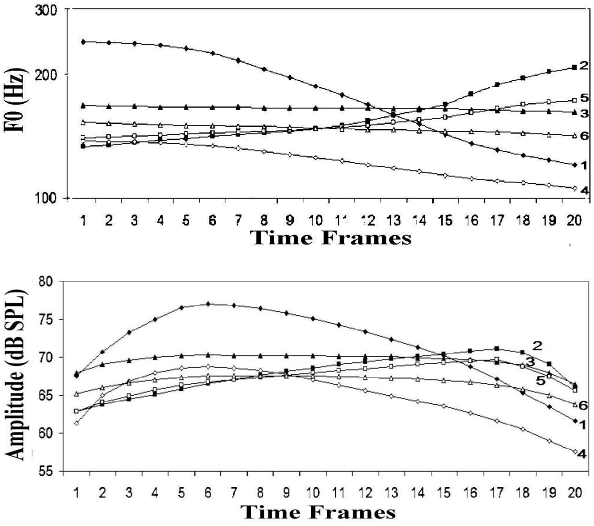

Figure 2. F0 (top) and AE contours of Cantonese tones.

The statistical correlation between the F0 contour and the AE of each of the six Cantonese tones was estimated using the Pearson linear correlation analysis. Pearson's correlation coefficient for each tone ranged from 0.45 to 0.71 (mean = 0.56; SD = 0.13; p = 0.045), which suggested that F0 co-varied with AE in Cantonese tones. Thus, the answer to the first research question of the current study is positive.

There were significant differences across tone types. In particular, the falling tones exhibited a higher correlation coefficient (0.71) than the rising tones (0.53) which in turn demonstrated a higher coefficient than the level tones (0.45). These correlation patterns were similar to those observed in [3][4]. Thus, F0 not only co-varied with AE in Cantonese tones, but also showed similar co-variability patterns as in Mandarin.

3.2 Cantonese tone identification using AE

As illustrated in Figure 3, the twenty participants' overall identification of each of the six processed tones was above the chance level of 16.7% (mean = 40%; range: 21% to 52%; SD= 0.098; $t_{(19)}$ = 10.6; $\chi^2_{(5)}$ = 1214; p < 0.0001). Thus, AE alone provided a sufficient cue for native listeners to identify Cantonese tones with above-chance accuracy, which answered the second research question of the current study positively. However, the participants identified the natural tones significantly better than the processed tones (mean difference = 38%; $F_{(1, 38)}$ = 280; p < 0.0001). Therefore, the AE cue is clearly not the only cue for Cantonese tone identification in individuals with normal hearing.

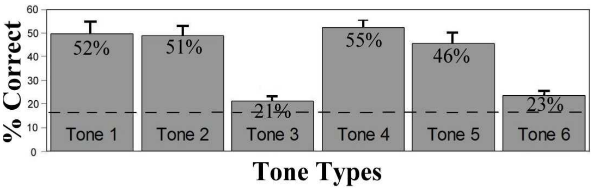

Figure 3: Identification of processed tones (chance level and standard error represented by dotted line and error bars).

Each listener's percent correct measure was computed for each talker and each tone type, and was analyzed using a two-way repeated-measures ANOVA. No significant effect of talker was observed (mean difference = 1.3%; $F_{(1, 19)}$ = 0.11; p = 0.35). That is, the tones produced by the male talker did not produce significantly different results than those produced by the female talker. The interaction between talker and tone type did not reach statistical significance ($F_{(5, 95)}$ = 1.4; p=0.27).

The effect of tone type was statistically significant ($F_{(5, 95)}$ = 29.59; p<0.0001). Specifically, the falling tones (Tones 1 & 4) were identified with greater accuracy than the rising tones (Tones 2 & 5) which were in turn identified with higher accuracy than the level tones (Tones 3 & 6), as illustrated in Figure 3. These findings were consistent with the results from similar studies in Mandarin [2][3][4].

3.3 The role of F0-AE covariability in Cantonese tone identification using AE

As illustrated in Figure 4, the falling tones (Tone 1 and Tone 4), with the highest F0-AE correlation (r = 0.66 and 0.75), were identified with the highest accuracy (52% and 55% correct). The rising tones (Tone 2 and tone 5), with the second highest F0-AE correlation (r = 0.54 and 0.60), yielded the second best identification performance (51% and 46% correct). The level tones (Tone 3 and Tone 6), with the lowest F0-AE correlation (r = 0.47 and 0.46), produced the lowest identification results (21% and 23% correct). Thus, the results of Cantonese tone identification using AE alone could be somewhat predicted on the basis of the F0-AE covariability, which answered the third research question of the current study positively.

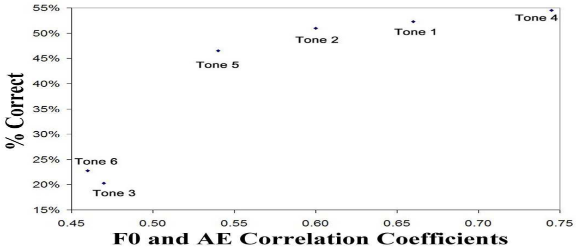

Figure 4: Identification of processed tones as a function of F0-AE covariability.

However, the predictive power of the F0-AE covariability is neither linear nor perfect. For instance, Tone 4 (Low Falling) had an F0-AE correlation coefficient that was significantly higher than that of Tone 2 (High Rising) (r=0.75 vs. 0.60; p = 0.0001), but Tone 4 was not identified significantly better than Tone 2 (55% vs. 51% correct; p = 0.79). In contrast, Tone 5 (Low Rising) was identified significantly better than Tone 3 (Upper Middle Level) (46% vs. 21%; p = 0.0001), but the F0-AE correlation coefficient for Tone 5 was not significantly different from that of Tone 3 (Upper Middle Level) (0.60 vs. 0.47; p = 0.45). Thus, the F0-AE covariability could not be used to fully predict the identification of Cantonese tones using AE alone. This finding was in line with similar studies in Mandarin [3][4].

3.4 Relative contribution of AE contour and height to Cantonese tone identification using AE

The perceptual patterns in the current Cantonese tone identification experiment using AE are displayed in Table 1. The listeners identified the contour of a Cantonese tone using the AE cue alone with 54.2% accuracy (chance = 33.4%; SD = 0.07; $t_{(19)}$ = 13.8; p < 0.001). They differentiated high tones from low tones with 59.1% success using the AE cue (chance = 50%; SD = 0.08; $t_{(19)}$ = 5.2; p = 0.002). The SPSS multidimensional scaling analysis of these perceptual patterns suggested two important dimensions in Cantonese tone identification using AE: the contour of the AE (i.e., falling, level and rising) and the relative height of the AE. AE contour alone accounted for 76% of the variances. Combined together, AE contour and height could explain 93% of the variances. Thus, AE contour seemed to contribute more than AE height to Cantonese tone identification using AE. In other words, AE contour appeared to be a more salient cue than AE height for Cantonese tone identification using AE.

Target Tones	**Perceived as**					
	T1 High Falling	**T2** High Rising	**T3** Upper Middle Level	**T4** Low Falling	**T5** Low Rising	**T6** Lower Middle Level
T1	**49.5%**	1.3%	7.0%	39.0%	0.3%	3.0%
T2	3.3%	**48.8%**	5.5%	5.0%	35.8%	1.8%
T3	4.5%	23.0%	**21.0%**	11.5%	26.3%	13.8%
T4	16.0%	0.8%	10.8%	**52.0%**	4.8%	15.8%
T5	2.6%	37.0%	7.8%	3.0%	**45.3%**	4.5%
T6	1.8%	16.0%	14.8%	11.0%	33.5%	**23.0%**

Table 1. Patterns of Cantonese tone identification using AE

4. Conclusion

All four research questions of the current study have been answered positively. First, the current study indicated that F0 co-varied with AE in Cantonese lexical tones. This is a unique contribution of the current study to the research on the role of AE in tone perception. Thus, the covariability between F0 and AE is not unique in Mandarin tones, but may be universal across tone languages. However, this needs to be verified empirically in future studies involving different types of tone languages.

Second, the current study suggested that AE alone can aid tone identification in a language such as Cantonese which uses both pitch contour and pitch height for phonemic tonal contrasts. This is another unique contribution of the current study to the research on the role of AE in tone perception. Since tone languages in the world use only tone height, tone contour or a combination of both for phonemic tonal contrasts [5], it may be possible for AE to cue tone perception in all tone languages. Once again, this needs to be verified empirically in future studies.

Third, the outcome of the current Cantonese tone identification experiment using AE could be somewhat predicted on the basis of the degree of the F0-AE covariability. That is, a tone with a higher degree of F0-AE covariability tended to be identified with greater accuracy than a tone with a lower degree of F0-AE covariability. However, the relationship between the degree of the F0-AE covariability and the current Cantonese tone identification was not linear. The predictive power of F0-AE covariability seemed to be constrained by other factors yet to be discovered.

Lastly, the current study was the first to tease apart the relative contribution of AE contour and AE height to tone perception using AE. Specifically, AE contour could provide a perceptual cue for tone contour, whereas the relative AE height could serve as a perceptual cue for tone height. Relatively speaking, AE contour seemed to be a more salient cue than AE height for Cantonese tone perception using AE. This is the third unique contribution of the current study to the research of tone perception using AE.

An implication of the current findings pertains to the improvement of speech processors for cochlear implants. Cochlear implant users who speak a tone language have difficulties perceiving tones and therefore overall speech in their native language due to the limitations of currently available commercial speech processing strategies in encoding F0 [2][4][8][11]. Experimental speech processing strategies have demonstrated promising clinical results in improving Mandarin tone perception in cochlear implant users by enhancing the encoding of the AE cue [11]. These experimental strategies could potentially be adapted to enhance the encoding of the AE cue in Cantonese tones and thus improve tone perception in Cantonese-speaking implantees, given the current finding that AE could also cue Cantonese tone perception. In broader terms, if future studies reveal that AE could cue tone perception in any tone language, the above-mentioned experimental speech processing strategies could potentially be adapted to enhance the encoding of the AE cue in any tone language, and thus would improve tone perception in cochlear implantees who speak any tone language.

5. Acknowledgements

The authors thank Winifred Strange, Glenis Long and three anonymous reviewers for their helpful suggestions and comments.

6. References

[1] Rosen, S., “Temporal information in speech: acoustic, auditory and linguistic aspects”, Philosophical Transactions: Biological Sciences, 336, 367-373, 1992.

[2] Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D., “Importance of tone envelope cues in Chinese speech recognition”, Journal of Acoustical Society of America, 104(1), 505-510, 1998.

[3] Fu, Q. J., and Zeng, F. G., “Identification of temporal envelope cues in Chinese tone recognition”, Asia Pacific J. Speech, Language and Hearing, 5, 45-47, 2000.

[4] Kuo, Y., Rosen, S., and Faulkner, A., “Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants”, Journal of the Acoustical Society of America, 123(5), 2815-2824, 2008.

[5] Yip, M., Tone. Cambridge: Cambridge University Press, 2002.

[6] Fok, Y.Y., “A Perceptual Study of Tones in Cantonese”, Hong Kong: University of Hong Kong Press, 1974.

[7] Gandour, J. T., “Perceptual dimensions of tone: evidence from Cantonese”, Journal of Chinese Linguistics, 9 20-36, 1981.

[8] Yuen, K. C., Tong, M. C. F., van Hasselt, C., Yuan, M., Lee, T, and Soli, S. D., “Cantonese lexical tone recognition from frequency-specific temporal envelope and periodicity components in the same versus different noise band carriers”, Cochlear Implants International, 10(S1), 148-158, 2009.

[9] Khouw, E., and Ciocca, V., “Perceptual correlates of Cantonese tones”, Journal of Phonetics, 35(1), 104-117, 2007.

[10] Wong, P. and Diehl, R., “Perceptual normalization for inter- and intra-talker variation in Cantonese level tones”, Journal of Speech, Language, and Hearing Research, 46, 413-421, 2003.

[11] Luo, X., and Fu, Q. J., “Enhancing Chinese tone recognition by manipulating amplitude enveloped: implications for cochlear implants”, Journal of Acoustical Society of America, 116(6), 3659-3667, 2004.