Interspeech 2018
2-6 September 2018, Hyderabad

Weighting Pitch Contour and Loudness Contour in Mandarin Tone Perception in Cochlear Implant Listeners

Qinglin Meng[1,2], *Nengheng Zheng*[3], *Ambika Prasad Mishra*[2], *Jacinta Dan Luo*[2], *Jan W. H. Schnupp*[2]

[1] Acoustics Lab, School of Physics and Optoelectronics, South China University of Technology, China
[2] Hearing Research Group, Biomedical Science Department, City University of Hong Kong, Hong Kong SAR of China
[3] College of Information Engineering, Shenzhen University, China
mengqinglin@scut.edu.cn, nhzheng@szu.edu.cn, wschnupp@cityu.edu.hk

Abstract

Previous investigations found that loudness-contours within individual Mandarin monosyllables can drive categorical perception of Mandarin tone for cochlear implant (CI) users, while in normal hearing (NH) subjects the pitch contour is phonologically acknowledged to be the dominant cue. Here we further examine the weighting strategy of pitch induced and loudness induced contour identification on Mandarin tone perception by CI users. Twenty-seven versions of the disyllabic utterance /Lao3 Shi/ with orthogonally manipulated loudness-contour and pitch-contour of the voiced portion of the second monosyllable /Shi/ served as the stimuli to both CI and NH subjects. In Mandarin, if /Shi/ is pronounced with high-flat-pitched Tone 1 the word means "teacher", with rising Tone 2 it means "well-behaved", or with falling Tone 4 it means "always". CI users generally had poorer word-recognition scores and their inter-subject variance was large. While NH subjects recognized tone reliably based on pitch-contour, half of the CI users relied on pitch-contour, the other half on loudness-contour, implying systematic differences in pitch coding in their CI processing. This paradigm of orthogonal manipulation of pitch and loudness contours could be developed into improved audiometric tests of Mandarin tone perception and pitch coding with CIs.

Index Terms: pitch perception, loudness contour, cochlear implant, Mandarin tone

1. Introduction

Mandarin tone is a linguistic term that refers to a phonological category that distinguishes words based on the pitch contour of voiced duration [1]. A classic example is /ma/, which, with the first, high, flat tone /mā/ means "妈; mother", with the 2nd, rising tone /má/ means "麻; numb", with the 3rd, falling-then-rising tone /mǎ/ means "马; horse", and with the fourth, falling tone /mà/ means "骂; curse" [2]. While pitch is recognized as the dominant cue, there are also systematic differences in the amplitude contour and duration of Mandarin tones.

Pitch was defined as "that auditory attribute of sound according to which sounds can be ordered on a scale from low to high" by the American National Standards Institute [3]. The major cue for pitch is temporal periodicity [4], which is usually quantified by the fundamental frequency (F0). The periodicity is reflected not only by the important but not necessarily present (think about the missing fundamental phenomenon) fundamental component but also by harmonics which may distribute widely in the frequency domain, either in a place mode or in a temporal mode in the cochlea [4, 5]. Normal hearing (NH) listeners can integrate acoustic cues even from noise-corrupted sounds to detect pitch variation within (e.g., for Mandarin tone) or between (e.g., for music melody) sounds with little effort.

However, pitch is not the only dimension along which the auditory brain can perceive variations from low to high. Loudness and timbre variation over time can also be perceived as contours [6, 7]. Nevertheless, in NH listeners, loudness contours and timbre contours have much less influence than pitch variation on Mandarin tone or musical melody perception, as demonstrated in studies using constant-pitch [7], degraded-pitch [8], whisper [9], or inharmonicity [10].

CIs have been successful in helping severe-to-profound hearing impaired patients regain speech perception ability, but results remain highly variable. Also, because of the coarse spectral and temporal resolution of the artificial electric stimulation, pitch discrimination in CI users is relatively poor, and it is reasonable to assume that CI users may place greater weight on loudness cues during contour perception tasks than NH listeners. This has been supported by several previous CI studies, both for musical melodies [6, 11] and Mandarin tone contours [8, 12] using either explicit loudness contour manipulation [6, 8] or implicit manipulation of loudness-related acoustic cues [11, 12]. Here we build on these studies and examine the weighting of loudness contour and pitch contour cues by Mandarin-speaking CI users on in lexical tone perception.

We recently proposed a loudness contour manipulation algorithm [8], "Loudness-Tone", which allows loudness variations to be changed according to the instantaneous F_0 contour, either in a co-varying or in a conflicting manner. We found that co-varying conditions could enhance tone recognition, while conflicting conditions would increase the rate of tone confusions and sometimes led subjects to judge the Mandarin tone according to the modified loudness contour instead of the unchanged pitch contour.

In this study, we developed a disyllabic word test. The second syllable was manipulated, varying in pitch contour and loudness contour independently. Listeners were then asked to make lexical category judgments on the manipulated syllables.

10.21437/Interspeech.2018-1245

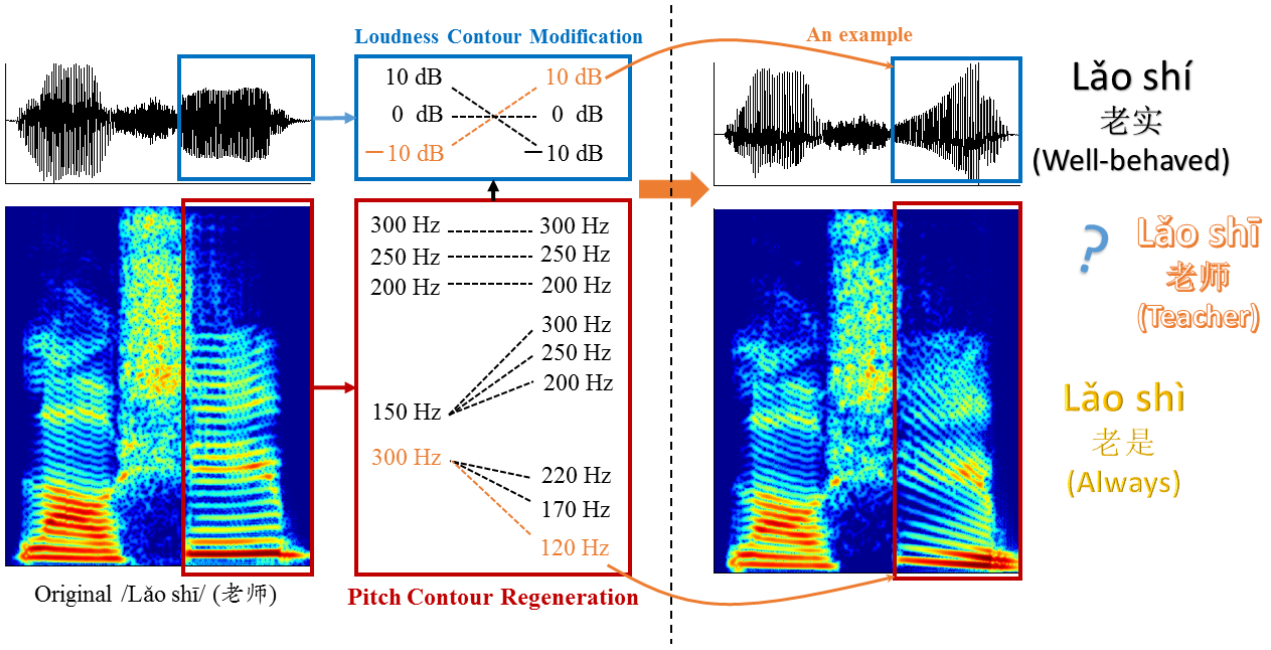

Figure 1: *Disyllabic stimuli synthesis conditions and a synthesis example.*

This approach was inspired by another recent study, which examined word duration effects on tone recognition in CIs [13].

2. Methods

2.1. Materials

Twenty-seven disyllabic words were synthesized from a recording of the word /Lǎo shī/ (老师) recorded by a female speaker, using the STRAIGHT toolbox (17/09/2005). The STRAIGHT vocoder decomposes speech into excitation information and smoothed time-frequency spectral information, and can resynthesize speech from these parameters which is heard as natural as the original speech [14]. Prior to resynthesis, source excitation parameters (F_0 and aperiodic components) or formant filter characteristics can be modified.

Figure 1 shows the synthesis conditions and a synthesis example for the current study. The voiced portion of the second syllable (i.e. /Shi/ in the original /Lǎo shī/) was resynthesized to have three different Mandarin tones each with three F_0 contour options, i.e., Tone 1 (the high-flat tone) with 300 to 300 Hz, 250 to 250 Hz, and 200 to 200 Hz; Tone 2 (the rising tone) with 150 to 300 Hz, 150 to 250 Hz, and 150 to 200 Hz; and Tone 4 (the falling tone) with 300 to 220 Hz, 300 to 170 Hz, and 300 to 120 Hz. Then this portion of the synthesized signal would be scaled by three different intensity gradients (i.e., from −10 to 10 dB, from 0 to 0 dB, and from 10 to −10 dB) to create a different loudness contours. This yielded 27 stimulus conditions (9 pitch × 3 loudness contours). All information within the original /Lǎo Shī/ signal other than the loudness and pitch contour of the second voiced segment were kept unchanged. One synthesis example with a rising loudness contour and a falling pitch contour in the voiced portion of /Shi/ is shown on the right of Figure 1.

Our subjects were asked to categorize each of these 27 stimuli according to whether they heard them as /Lǎo Shī/ (老师, teacher, 1st tone), /Lǎo Shí/ (老实, well-behaved, 2nd tone), or /Lǎo Shì/ (老是, always, 4th tone).

2.2. Subjects

We tested these 27 stimuli on four NH subjects (including the first author) and four bilaterally deaf, unilaterally implanted CI subjects (see Table 1). Subjects (except the first author) were paid for their participation, and all provided informed consent in accordance with the local institutional review board. Two subjects used devices from Cochlear, the other two devices from Nurotron. Subjects were tested with their clinical default device settings.

Table 1: *CI user demographic information*

Subject	Gender	Age(yr)	CI Experience (yr)	Etiology
C2	M	24	15	Drug-induced
C22	M	37	1.25	Sudden deafness
C23	F	29	3	Sudden deafness
C25	F	38	6	Sudden deafness

2.3. Procedure

Each CI subject took part in one training and one formal testing session. For the NH subjects there was only one formal testing session. In a session, each of the 27 stimuli were presented in a random trial order. A three-alternative-forced-choice paradigm was used: each stimulus was presented one or two times, and thereafter the subject was asked to indicate whether they heard the stimulus as the word /Lǎo Shī/ (teacher), Lǎo Shí/ (well-

behaved), or /Lǎo Shì/ (always) by selecting written Mandarin words on a computer screen with the mouse. All sounds were presented at a comfortable level (approximately 70 dB SPL) through an audio interface (Focusrite Scarlett 2i4) and a loudspeaker (Genelec 8010A) in a soundproof room. No feedback about the correctness of the response was given.

2.4. Statistical analysis

For the purpose of reporting the results, we shall refer here as a "**pitch-tone**" the conventional term of lexical tone represented by pitch variation, and as a "**loudness-tone**" the systematic flat, rising or falling loudness contour introduced into the target syllable [8].

We compared observed word choices against the null hypothesis that, if a subject were to choose randomly, they would identify the "correct" word on a third of trials on average, and the number of "correct" choices observed would follow a binomial distribution with n=27 and p=1/3. The probability of observing 15 or more "correct" choices out of 27 by chance is then as small as 0.0144. We want to establish whether "correct" response rates are significant either according to a pitch-tone or a loudness tone criterion, so we perform two "multiple comparisons", but the probability of 15 or more correct choices is still lower than a Bonferroni corrected α of 0.025. Consequently subjects can be said to be significantly guided by pitch-tone or loudness-tone respectively if their choices followed that cue in at least 15 of the 27 trials (55.56%).

3. Results

All of the NH subjects identified all the stimuli according to pitch-tone, without any effect from the modified loudness contour. This tells us that NH subjects place absolutely dominating weight on the pitch contour during this word identification task. They reported that all stimuli sounded natural to them, which validates the effectiveness of the STRAIGHT vocoder combined with the designed F_0 contours and intensity gain contours.

The results for the four CI subjects' tone identification results are listed in Table 2.

Each of the 27 stimuli has one pitch-tone and one loudness-tone, as described in section 2.1.

The CI subjects' response for each stimulus is shown in the right four columns. Two subjects (C2 and C25) got high scores (74% and 78% respectively; much higher than the 56% significance threshold) according to the pitch-tone, which means they two identified the Mandarin tones mainly according to the pitch contour. The other two subjects (C22 and C23), in contrast, got significantly high scores (63% and 81% respectively) according to the loudness-tone, which indicates that they identified the Mandarin tone mainly according to the loudness contour. Thus, Mandarin speaking CI users appear to differ in how they weight "pitch-contour" vs "loudness contour" cues when identifying lexical tone.

As mentioned above, the four NH subjects always identified the Mandarin tones according to pitch, irrespective of loudness contour manipulations. For all of the CI listeners, when pitch contour and loudness contour were co-varying (see green shading in Table 2), the choice followed the pitch cue 78% of the time, but when they were conflicting (see blue shading in table 2), the choice followed the pitch cue only 46% of the time. Results for individual subjects are listed at the bottom of Table 2. It seems loudness manipulation has very little effect on subjects C2 and C25, but marked effects on subjects C22 and C23. Under the six most-conflicting conditions (with dark blue shading), i.e., rising-pitch-tone with falling-loudness-tone and falling-pitch-tone with rising-loudness-tone, C22 and C23 (the loudness-tone dominating subjects) identified all tones of the six stimuli according to loudness-contour (see results with orange shading).

Reliance on pitch or loudness cues respectively covaried with which manufacturer had supplied the CIs, but at this point we have far too few subjects to say whether that is coincidental.

Table 2: *Tone identification results*

	Pitch-Tone (F_0 range in Hz)	Loudness-Tone	C2	C22	C23	C25
1	1(200-200)	2	1	2	2	2
2	1(200-200)	1	1	2	1	2
3	1(200-200)	4	1	4	4	2
4	1(250-250)	2	1	2	2	2
5	1(250-250)	1	1	2	1	1
6	1(250-250)	4	1	1	4	1
7	1(300-300)	2	1	1	2	1
8	1(300-300)	1	1	1	1	1
9	1(300-300)	4	1	4	1	1
10	2(150-200)	2	2	1	2	2
11	2(150-200)	1	2	1	2	2
12	2(150-200)	4	4	4	4	2
13	2(150-250)	2	1	2	2	2
14	2(150-250)	1	2	2	2	2
15	2(150-250)	4	4	4	4	2
16	2(150-300)	2	1	2	2	2
17	2(150-300)	1	4	4	1	2
18	2(150-300)	4	2	4	4	2
19	4(300-220)	2	2	2	2	1
20	4(300-220)	1	1	1	2	4
21	4(300-220)	4	4	1	4	1
22	4(300-170)	2	4	2	2	4
23	4(300-170)	1	4	2	1	4
24	4(330-170)	4	4	4	4	4
25	4(300-120)	2	4	2	2	4
26	4(300-120)	1	4	1	2	4
27	4(300-120)	4	4	4	4	4
Loudness-Tone matching percentage (%)			41	**63***	**81***	33
Pitch-Tone matching percentage (%)			**74***	30	44	**78***
-When co-varying conditions (Green, %)			**78***	**56***	**100***	**78***
-When conflicting conditions (Blue, %)			**72***	17	17	**78***

* The asterisk marks significantly high scores according to a binomial distribution.
The purposes of these shadings are explained in the text.

4. Discussion

Although the dominating effect of pitch contour perception and its related acoustic cues on Mandarin tone category are well recognized [1, 2], some secondary cues, e.g. intensity or amplitude contours [8, 15], or timbre contours created by

dynamic changes in formant or frequency spectrum distributions [9, 10], as well as duration cues [13] are known to co-vary with pitch contours cues for Mandarin lexical tone, and can be helpful especially when pitch cues are corrupted acoustically in normal-hearing, or poorly encoded in CI users.

In a previous study [8], we proposed an algorithm to modify the loudness contour of Mandarin monosyllables and found that loudness contours could be used in CI Mandarin tone recognition and its enhancement. Here we developed a disyllabic word test based on 27 stimuli derived from the Mandarin utterance /Lǎo Shi/, each of which has a unique voiced portion of /Shi/ with one conventional "pitch-tone" and one so-called "loudness-tone". The Mandarin tones of 1 (flat), 2 (rising), and 4 (falling) for /Shi/ were selected to give the stimuli clearly distinct meanings which Mandarin speakers will be highly familiar with, and which are representative of real-world lexical distinctions they need to make thousands of times a day. Embedding the loudness and pitch contour manipulations into a natural word recognition task makes this test easier to administer to typical Mandarin speaking CI users who have little practice in, and little need for, identifying the tone numbering conventions for syllables presented in isolation.

Independently manipulating pitch and loudness contours makes it possible to investigate how CI users weight the pitch contour and loudness contour cues during Mandarin tone recognition. We found that the NH control group entirely relied on pitch contour information. In contrast, the performance of CI was much less consistent, and there was marked inter-subject variability. Two patients placed greater weight on loudness contour and the other two placed greater weight on pitch contour. The perceptual weighting strategy difference may be attributable to differences in their respective pitch discrimination abilities. Greater reliance on loudness contour cues could be (but does not necessarily have to be) a result of poorer pitch encoding. In the future, we plan to administer pitch discrimination tasks alongside of this contour weighting experiment in more CI patients to test whether increased loudness cue weighting indeed correlates with poor pitch discrimination.

This paradigm of disyllabic words combined with orthogonal manipulation of pitch and loudness contour can be generalized in future audiological measurement and rehabilitation training. Specifically, there are two potential aspects. Firstly, stimulus sets like this one can be used not just to measure Mandarin tone recognition abilities, but also to train patients to make greater use of a particular set of cues, depending on what is achievable given their specific circumstances. Secondly, tests and stimuli like the ones introduced here may turn out to provide a quick and easily administered indirect measure of pitch discrimination ability if the suspected correlation between high pitch weighting in this task and some additional pitch discrimination tasks is confirmed. With these aims in mind we will be developing additional materials like these /Lǎo Shi/ disyllables, and are planning further experiments to validate and promote their use in audiometric tests and rehabilitation.

5. Conclusions

1. Loudness contour can influence Mandarin tone recognition for CI users but not influence NH subjects. This is consistent as previous research.

2. There is a large inter-subject variance in the perceptual weighting strategy of pitch-contour or loudness-contour respectively for Mandarin tone recognition among CI users. Some CI subjects appear to rely mostly on pitch contour, and others mostly on loudness contour. In contrast, NH subjects consistently rely entirely on pitch contour.

3. This novel paradigm of disyllabic words combined with orthogonal manipulation of pitch and loudness contour may prove useful in improved audiometric tests and rehabilitation materials for Mandarin tone recognition ability and pitch coding ability with CIs.

6. Acknowledgements

We thank all the subjects participated in the experiments. We appreciate the reviewers' comments very much. This work is jointly supported by NSF of China (Grant No. 11704129 and 61771320), the Fundamental Research Funds for the Central Universities (SCUT), State Key Laboratory of Subtropical Building Science (SCUT, Grant No. 2018ZB23), and Shenzhen Science and Innovation Funds (JCYJ 20170302145906843). N. H. Zheng and J. W. H. Schnupp are corresponding authors.

7. References

[1] M. Yip, *Tone*: Cambridge University Press, 2002.
[2] F. Zeng, "Tonal Language Processing," *Acoustics Today,* vol. 4, pp. 26-27, 2012.
[3] A. ANSI, "American National Standard Acoustical Terminology," *ANSI Sl,* pp. 1-1994, 1994.
[4] J. Schnupp, I. Nelken and A. King, *Auditory neuroscience: Making sense of sound*: MIT press, 2011.
[5] A. J. Oxenham, "How We Hear: The Perception and Neural Coding of Sound," *Annu Rev Psychol,* vol. 69, pp. 27-50, 2018.
[6] X. Luo, M. E. Masterson and C. C. Wu, "Contour identification with pitch and loudness cues using cochlear implants," *J Acoust Soc Am,* vol. 135, pp. EL8-14, 2014.
[7] J. H. McDermott, A. J. Lehr and A. J. Oxenham, "Is relative pitch specific to pitch?" *Psychological Science,* vol. 19, pp. 1263-1271, 2008.
[8] Q. Meng, N. Zheng and X. Li, "Loudness Contour Can Influence Mandarin Tone Recognition: Vocoder Simulation and Cochlear Implants," *IEEE Trans Neural Syst Rehabil Eng,* vol. 25, pp. 641-649, 2017.
[9] Z. A. Liang, "The auditory discrimination basis of tone recognition in Standard Chinese," *Acta Physiol. Sinica.,* pp. 85-92, 1963.
[10] M. J. McPherson and J. H. McDermott, "Diversity in pitch perception revealed by task dependence," *Nature Human Behaviour,* vol. 2, p. 52, 2018.
[11] M. Cousineau, L. Demany, B. Meyer, and D. Pressnitzer, "What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant," *Hear Res,* vol. 269, pp. 34-41, 2010.
[12] L. Ping, N. Wang, G. Tang, T. Lu, L. Yin, W. Tu, and Q. Fu, "Implementation and preliminary evaluation of 'C-tone': A novel algorithm to improve lexical tone recognition in Mandarin-speaking cochlear implant users," *Cochlear implants international,* vol. 18, pp. 240-249, 2017.
[13] S. Peng, H. Lu, N. Lu, Y. Lin, M. L. Deroche, and M. Chatterjee, "Processing of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients," *Journal of Speech, Language, and Hearing Research,* vol. 60, pp. 1223-1235, 2017.
[14] H. Kawahara, H. Banno, T. Irino, and P. Zolfaghari, "Algorithm amalgam: morphing waveform based methods, sinusoidal models and STRAIGHT," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, 2004, p. I-13.
[15] D. H. Whalen and Y. Xu, "Information for Mandarin tones in the amplitude contour and in brief segments," *Phonetica,* vol. 49, pp. 25-47, 1992.