INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Production of Sustained Vowels and Categorical Perception of Tones in Mandarin among Cochlear-Implanted Children

Wentao Gu[1], *Jiao Yin*[1] *and James Mahshie*[2]
[1]Nanjing Normal University, China
[2]George Washington University, USA
wtgu@njnu.edu.cn, 1533346190@qq.com, jmahshie@gwu.edu

Abstract

This study investigated both production and perception of Mandarin speech, comparing two groups of 4-to-5-year-old children, a normal-hearing (NH) group and a cochlear-implanted (CI) hearing-impaired group; the perception ability of the CI group was tested under two conditions, with and without hearing aids. In the production study, the participants were asked to produce sustained vowels /a/, /i/ and /u/, on which a set of acoustic parameters were then measured. In comparison to the NH group, the CI group showed a higher F_0, a higher H1–H2, and a smaller acoustic space for vowels, demonstrating both phonatory and articulatory impairments. In the perception study, the identification tests of two tone-pairs in Mandarin (T1-T2 and T1-T4) were conducted, using two sets of synthetic speech stimuli varying only along F_0 continua. All groups/conditions showed categorical effects in perception. The CI group in the unimodal condition showed little difference from normal, while in the bimodal condition the categorical effect became weaker in identifying the T1-T4 continuum, with the category boundary more biased to T4. This suggests that bimodal CI children may need more fine grain adjustments of hearing aids to take full advantage of the bimodal technology.

Index Terms: cochlear-implanted children, bimodal, Mandarin, vowel production, categorical perception of tones

1. Introduction

Cochlear implants (CIs) are surgically implanted electronic devices that help people with severe sensorineural hearing loss regain partial hearing abilities. With the aid of CIs, speech perception has been greatly improved in those who suffer significant hearing loss, especially in children.

Current CIs were originally designed and optimized for hearing-impaired subjects speaking non-tone languages in which the fundamental frequency (F_0) of speech is mainly used for sentential intonation conveying syntactic structures and emotional states, with little role in distinguishing word meaning. The coding strategy of CIs has focused on conveying temporal envelope while the fine structure of sounds has not been coded explicitly due to technological constraints. As a result, most current CIs do not provide F_0 information directly [1, 2].

In tone languages, the F_0 of speech conveys not only sentential intonation but also lexical tones that differentiate word meanings. For example, Mandarin is a well-known tone language in which each syllable can have four lexical tones – T1 (high level), T2 (rising), T3 (dipping, or falling-rising), and T4 (falling), differing from each other mainly in F_0. Due to the lack of direct coding of F_0 in current CIs, implanted children have less access to the cues of lexical tones, and thus show lower perceptual accuracy of tones than their normal-hearing (NH) peers [1], leading to potential difficulties in speech communication.

It has been reported that 3-year-old normal-hearing (NH) Mandarin-speaking children have largely acquired the ability to perceive all four tones, with higher perceptual accuracy on T1, T2, T4, and lower accuracy on T3 [3]. Research examining tone perception by children with CIs suggests that T4 is identified more accurately than other tones [2, 4, 5], mainly because T4 has the shortest duration and the greatest amplitude change, thus providing additional perceptual cues when pitch is absent.

It is only very recently that research examined children's categorical perception of Mandarin tones along a continuum varying only in F_0. Chen et al. [6] showed that NH children's perception of the T1-T2 continuum became categorical by 4-year old. Our recent study [7] found that the 4-5 year-old children with CIs using bimodal technology showed similar categorical effects as the NH peers in tone identification tests for the T1-T2 and T1-T4 continua, differing in the position of category boundaries and the degree of categorical effects. However, only two repetitions of stimuli were tested in [7]. To get statistically more reliable results, the present study further conducted the experiments using ten repetitions of the stimuli.

In recent years, the use of bimodal technology (i.e., CIs in one ear and hearing aids in the other) has shown to be more effective for perception of tones than CIs alone. It was reported that children with unilateral CIs who wore hearing aids (HAs) on the other ear showed significantly better identification of Mandarin tones than with CIs alone [8]. Therefore, this study aimed at CI children using bimodal technology, and compared their perception in two conditions (with/without HAs).

As for speech production, while speech outcomes for the population of children with CIs are improved when compared to those of children using hearing aids, differences from normal still exist. There also remain aspects of speech production that have not been well studied in children with CIs. In particular, there is significant evidence suggesting that the speech and voice characteristics of children with CIs differ from NH children [9-13]. However, there are few reported studies that systematically examine voice characteristics of these children or the acoustic underpinnings of these characteristics.

2. Method

2.1. Participants

Two groups of children (i.e., NH and CI) were recruited. Each group consisted of 10 native Mandarin-speaking children (7M, 3F) between the ages of 4;1 and 5;6 ('years; months') – this age setting was adopted since previous research has shown that NH children's tone perception becomes categorical by 4-year old [6]. All children in the NH group were from a local kindergarten,

http://dx.doi.org/10.21437/Interspeech.2017-1698

without any reported history of speech, hearing or cognitive disorders.

All children in the CI group were from a local hearing rehabilitation center. They had profound hearing losses (≥ 91 dB HL) in both ears prelingually, and had no reported history of other cognitive or physical disabilities. All of them were bimodal technology users, with hearing aids on the left ear and CIs on the right. Moreover, they were all implanted before the age of 3 years, and the duration of their CI use was at least 18 months; the children with this CI setting were chosen to ensure the effectiveness of CI according to previous studies [2, 5, 14]. As shown in Table 1, the CI group was basically homogenous in the variables described above. The rightmost column shows the duration of HA use before/after implantation.

Table 1: *Background information of the CI group.*

ID	Gender	Age	Age of activation	Duration of CI use	Duration of HA use
1	M	4;1	1;5	2;8	0;0 / 2;9
2	F	4;9	2;3	2;6	0;8 / 0;11
3	M	4;8	2;9	1;11	0;4 / 1;1
4	M	4;2	1;2	3;0	0;6 / 1;1
5	F	5;6	3;0	2;6	2;2 / 1;9
6	M	4;8	2;4	2;4	1;9 / 2;0
7	F	5;1	2;1	2;11	1;2 / 1;8
8	M	4;6	1;10	2;8	0;5 / 1;9
9	M	5;4	2;6	2;10	0;11 / 2;6
10	M	4;11	1;1	3;10	0;0 / 0;7

2.2. Production experiment

The production experiment was conducted in a quiet room. Each participant was asked to produce three sustained Mandarin vowels /a/, /i/ and /u/ as long as possible (> 4s) in a comfortable manner, with at least five tokens for each. All utterances were recorded using a portable digital recorder (Zoom H4n). Before the formal experiment, each participant was trained for the experimental procedure by their teachers, who used different gestures to elicit the three vowels. After the experiment, a small gift was given to each participant.

For each participant, the two speech samples which the first and second authors judged most stable were selected out of the five tokens for each vowel. The central steady-state 3 second interval of each vowel was analyzed and the following acoustic parameters were obtained using Praat [15] and VoiceSauce [16]: mean F_0 related to prosody, mean F_1 and F_2 related to vowel articulation, and jitter, shimmer, HNR, and H1–H2 related to voice quality. To minimize the influence of the formants on H1 and H2, the value of H1–H2 was only measured for the low vowel /a/ which had the highest F_1 (thus farthest from H1 and H2). Based on F_1 and F_2 of the three corner vowels, triangular Vowel Space Area (VSA) was also calculated [17].

2.3. Perceptual experiments

The speech materials produced by a native female speaker consisted of three isolated Mandarin syllables: 妈 "mā" (mother), 麻 "má" (sesame), and 骂 "mà" (scolding), sharing the same sonorant consonant and vowel, but carrying different tones (i.e., T1, T2, and T4, respectively). We also prepared three computer-generated pictures representing 'mother', 'sesame', and 'scolding', respectively, which were used to obtain responses from the children.

F_0 values of these syllables were extracted at 10ms intervals using autocorrelation analysis in Praat. By extracting the F_0 values at ten equally-spaced points in the syllables, we obtained time-normalized F_0 contours of the three syllables. As shown in Fig. 1, taking the time-normalized F_0 contours of the T1 and T2 syllables, respectively, as the 1st and 11th steps, an 11-step rising F_0 continuum was constructed by successive linear interpolation between the two extremes, with a 7.1 Hz stepwise F_0 difference at the onset point and a 1.8 Hz stepwise F_0 difference at the offset point. As shown in Fig. 2, taking the time-normalized F_0 contours of the T1 and T4 syllables, respectively, as the 1st and 11th steps, an 11-step falling F_0 continuum was constructed by successive linear interpolation, with a 2.4 Hz stepwise F_0 difference at the onset point and an 8.0 Hz stepwise F_0 difference at the offset point.

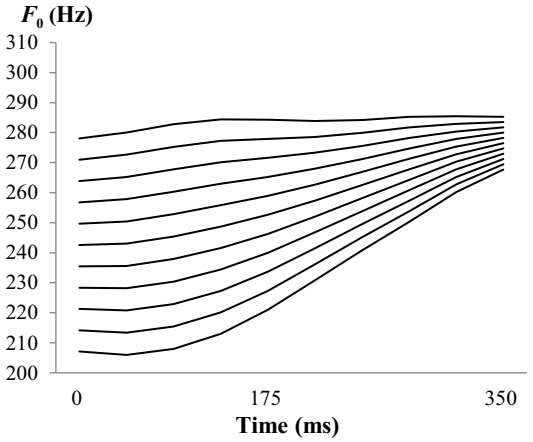


Figure 1: *The rising F_0 continuum between T1 and T2.*

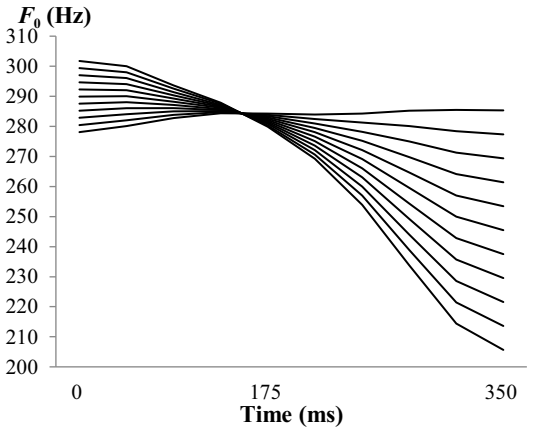


Figure 2: *The falling F_0 continuum between T1 and T4.*

In an experiment for categorical perception, only one acoustic parameter (here, F_0) is typically varied along a continuum, while all other acoustic parameters are kept intact. Therefore, the T1 syllable "mā" was adopted as the base syllable after time-normalization to 350ms, a length that sounded acceptable for all three tones. This token was then resynthesized with each F_0 contour from the two F_0 continua, using the PSOLA synthesis technique implemented in Praat. This resulted in two continua of speech stimuli, i.e., T1-T2 continuum and T1-T4 continuum. This method ensured the

naturalness of the synthetic speech stimuli, which was also confirmed by listening.

Due to the lack of discrimination ability for 4-5 year-old children [7], only identification tests were conducted in the present study, one for T1-T2 continuum and the other for T1-T4 continuum. In each identification test, five practice stimuli randomly selected from the continuum were presented first to adapt the participants to the task and focus their attention, followed by 10 repetitions of the 11 stimuli (thus totally 110 stimuli) presented in a random order. The responses for the five practice stimuli were not included in the experimental results. To counterbalance the order of the two identification tests, the participants in each group were divided into two sub-groups: one received the T1-T2 continuum first and then the T1-T4 continuum; the other received the tests in the reverse order.

Before the formal experiment, each participant was familiarized with the experimental procedure and the three response pictures. All stimuli were presented using E-Prime 2.0 Professional software. After a stimulus was presented through loudspeakers, the participant was asked to select from two pictures the one that matched what they heard. After selection was completed, the next stimulus was presented. Brief breaks were allowed during the experiment, and a small gift was given to each participant after the experiment.

For each participant in the CI group, the experiments were done twice in two separate days (with 4-5 days in between), once in the bimodal condition as usual (+HA), and the other in the unimodal condition (–HA, i.e., with HAs taken off). The order of the bimodal and unimodal conditions was also counterbalanced among all participants in the CI group.

3. Results

3.1. Results of production experiment

Table 2 shows some acoustic parameters of the sustained vowels, with the significance levels for the differences between the NH and CI groups indicated. The parameters that did not show significant differences are not listed in the table.

As shown, the CI group tends to have higher F_0 than the NH group (significant for /a/ and /i/). For the formants F_1 and F_2, high values tend to be lowered, and low values tend to be raised in the CI group compared to NH peers. Especially, the highest values of F_1 and F_2, corresponding to /a/ and /i/, respectively, decrease significantly in the CI group, indicating that the tongue of the CI group is not as low (for /a/) and as anterior (for /i/) as that of the NH group, i.e., there is an undershooting of tongue, resulting in a significantly reduced Vowel Space Area (VSA), as shown in Fig. 3, where the triangles composed of solid and dashed lines correspond to the NH and CI groups, respectively.

For H1–H2, which is closely related to the looseness of glottis and hence to the degree of breathiness, the NH and CI groups show negative and positive values, respectively, with a marginally significant difference, suggesting that the CI group produces more breathy voice than the NH group.

Table 2: *Mean acoustic parameters of sustained vowels.*

Feature	NH			CI		
	/i/	/a/	/u/	/i/	/a/	/u/
F_0 (Hz)	313*	289*	326	365*	333*	358
F_1 (Hz)	469	1355*	488*	478	1230*	574*
F_2 (Hz)	3429**	2028†	1021†	3142**	1876†	1197†
H1–H2 (dB)	—	-0.43†	—	—	1.82†	—
VSA (Hz^2)	1,060,134***			668,273***		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.08$.

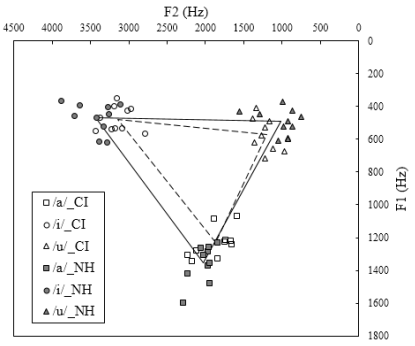
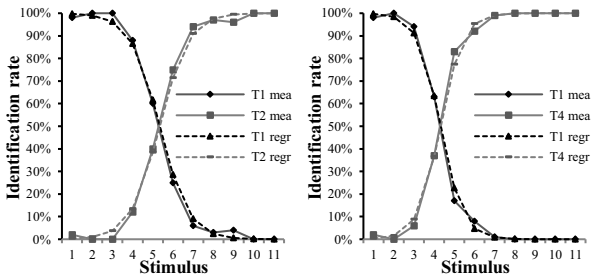


Figure 3: *Acoustic vowel space for NH and CI groups.*

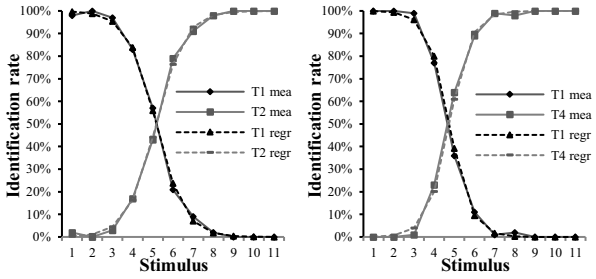
In sum, the CI group show differences from the NH group during sustained vowel production involving articulation (F_1, F_2), prosody (F_0), and voice quality (H1–H2).

3.2. Results of perceptual experiments

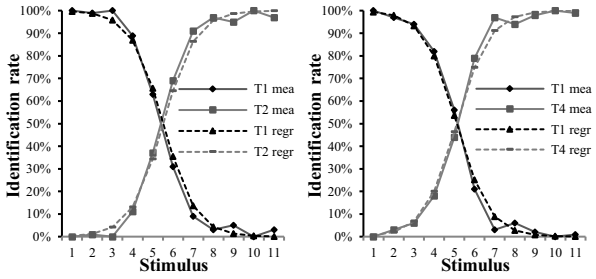
The identification test involved selecting one out of two candidate tones. Because the sum of the identification rates for the two candidate tones equals 1, only the identification rate for one tone (say, T1) needs to be calculated. The rate of a stimulus being identified as T1 is defined.



(a) NH group



(b) CI group without HA



(c) CI group with HA

Figure 4: *Curves of tone identification rates. ('mea' and 'regr' indicate measured and regression.)*

Table 3: *Parameters of identification regression curves.*

Group	T1-T2			T1-T4		
	b_1	W_{cb}	X_{cb}	b_1	W_{cb}	X_{cb}
NH	1.68	1.31	5.30	1.78	1.23*	4.31*
CI–HA	1.40	1.56	5.17	1.82†	1.21†	4.76
CI+HA	1.24	1.77	5.52	1.24†	1.78*†	5.12*

* $p < 0.05$ between NH and CI+HA.

† $p < 0.05$ between CI–HA and CI+HA.

To permit a more precise analysis, we approximated the identification rates using a double-variable logistic regression [18]: $\ln[P/(1-P)] = b_0 + b_1 X$, where X is the step number of the continuum, b_1 is the slope of the identification curve, and thus $|b_1|$ reflects the steepness of the category boundary – a steeper boundary indicates a stronger categorical effect. When setting $P = 0.5$, we get the position of category boundary $X_{cb} = -b_0/b_1$. Also, Hallé et al. [19] defined the width of category boundary W_{cb} as the distance between the steps corresponding to $P = 0.75$ and $P = 0.25$. A smaller W_{cb} indicates a steeper category boundary. Thus, the slope $|b_1|$ and the width W_{cb} are two parameters representing the degree of categorical effects.

Figure 4 shows the identification rates for the T1-T2 (left panels) and T1-T4 (right panels) continua in the NH and CI groups. For the CI group, the results in both conditions (–HA and +HA) are presented. The solid lines indicate the measured rates, while the dashed lines indicate the curves after logistic regression ($p < 0.01$). All curves show a typical *S*-shape, suggesting the effect of categorical perception.

Table 3 lists the parameters b_1, W_{cb} and X_{cb} of all regression curves. Paired *t*-tests were conducted between –HA and +HA conditions for the CI group, whereas independent *t*-tests were conducted between the NH group and the CI group in each condition. The parameters that show significant differences ($p < 0.05$) are indicated in Table 3.

For the T1-T2 continuum, there is no significant difference in any parameter between any two groups/conditions, though from the mean values there is a slight tendency for the CI+HA condition to result in a rightward shift of the boundary that reflects a weaker categorical effect. For the T1-T4 continuum, there is no significant difference in any parameter between NH and CI–HA, whereas CI+HA shows a less steep slope $|b_1|$, a wider boundary width W_{cb}, and a right-shifted (i.e., more biased to T4) boundary position X_{cb} in comparison with NH or CI–HA. Taken together, these suggest that the CI group with HAs has a weaker categorical effect in identifying T1-T4 than the other group/condition, and is less sensitive to falling pitch than the NH group.

4. Discussion and conclusions

The production experiment using sustained vowels revealed that the children with CIs showed differences from the normal-hearing ones in almost all aspects of speech production, including articulation (undershooting of tongue), prosody (higher F_0), and voice quality (more breathy voice). This suggests that impaired hearing, even with cochlear implants and hearing aids, tends to result in impairments in speech production. Our finding of higher F_0 in the CI group coincides with previous reports [11-13].

With 10 repetitions of stimuli being tested, the perceptual experiments in the present study are believed to give more reliable results than in [7]. The experiments found that both NH and CI (with/without HAs) children at the age of 4–5 years exhibited categorical effects in the perception of T1-T2 and T1-T4 continua. The CI group in the unimodal condition showed little difference from the NH group, suggesting that their ability to categorically perceive tones has benefitted from implantation. In the bimodal condition, the CI group showed similar results as the NH group in identifying the T1-T2 continuum, but the categorical effect degraded in identifying the T1-T4 continuum. A possible reason is that T4 (falling) is psychologically less distinctive than is T2 (rising) from T1 (level), because T1 is subject to falling as a result of F_0 declination universally coded in continuous speech.

Since all children with CIs in the present study were bimodal technology users who had access to F_0 through their aided ears with residual hearing, it was expected that their categorical perception in the bimodal condition be stronger than in the unimodal condition. However, this was not supported by the experimental results that the categorical effect of perception in the bimodal condition were even weaker than in the unimodal condition which was comparable to that of the normal hearing group. This seems to contradict our knowledge of the effectiveness of bimodal technology, but is quite possibly due to the fact that most users of bimodal technology in China have not had their hearing aids fitted with the aid of professional audiologists, and thus hearing aids may not coordinate well with CIs. Considering that CI+HA is the daily setting of the bimodal technology users, their impairments in speech production as we observed may also be a result of this less than optimal use of the bimodal technology. To take full advantage of the bimodal technology, hearing aids should be fitted appropriately in conjunction with the child's implant mapping.

5. Acknowledgements

This work is supported by the Major Program of the National Social Science Fund of China 13&ZD189.

6. References

[1] L. Xu and N. Zhou, "Tonal languages and cochlear implants," in F.-G. Zeng, A. N. Popper, and R. R. Fay (Eds.), *Auditory Prostheses: New Horizons*. New York: Springer, 2011, pp. 341–364.

[2] P. Wong, R. G. Schwartz, J. J. Jenkins, "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children," *Journal of Speech, Language, and Hearing Research*, 48: 1065–1079, 2005.

[3] S.-C. Peng, J. B. Tomblin, H. Cheung, Y.-S. Lin, and L.-S. Wang, "Perception and production of Mandarin tones in prelingually deaf children with cochlear implants," *Ear and Hearing*, 25(3): 251-264, 2004.

[4] A. Li, N. Wang, J. Li, J. Zhang, and Z. Liu, "Mandarin lexical tones identification among children with cochlear implants or hearing aids," *International Journal of Pediatric Otorhinolaryngology*, 78: 1945–1952, 2014.

[5] Y. Lu, *Study on the Acquisition of Mandarin Tones in the Cochlear-Implanted Children*. MA thesis, Tianjin: Tianjin Normal University, 2014.

[6] F. Chen, N. Yan, L. Wang, T. Yang, J. Wu, H. Zhao, and G. Peng, "The development of categorical perception of lexical tones in Mandarin-speaking preschoolers," *Proc. INTERSPEECH,* Dresden, Germany, 2015.

[7] W. Gu, J. Yin, and J. Mahshie, "Categorical perception in two pairs of Mandarin tones among bimodal cochlear implanted children," *Proc. ISCSLP*, Tianjin, China, 2016.

[8] Y. P. Chang, R. Y. Chang, C. Y. Lin, and X., Luo, "Mandarin tone and vowel recognition in cochlear implant users," *Ear and Hearing*, 37(3): 271–281, 2016.

[9] M. Evans and D. D. Deliyski, “Acoustic voice analysis of prelingually deaf adults before and after cochlear implantation,” *Journal of Voice*, 21(6): 669–682, 2007.
[10] J. M. Lenden and P. J. Flipsen, “Prosody and voice characteristics of children with cochlear implants,” *J Commun Disord*, 40(1): 66-81, 2007.
[11] G. J. Valero, J. M. Rovira, and L. G. Sanvicens, “The influence of the auditory prosthesis type on deaf children's voice quality,” *International Journal of Pediatric Otorhinolaryngology*, 74(8): 843-8, 2010.
[12] N. Baudonck, E. D'Haeseleer, and I. Dhooge, and K. V. Lierde, “Objective vocal quality in children using cochlear implants: a multiparameter approach,” *Journal of Voice*, 25(6): 683-591, 2011.
[13] I. Barbu, *Listener Ratings and Acoustic Characteristics of Intonation Contours Produced by Children with Cochlear Implants and Children with Normal Hearing*. B.A. Thesis, Los Angeles: University of California, 2012.
[14] K. Y. S. Lee, C. A. van Hasselt, S. N. Chiu, and D. M. C. Cheung, “Cantonese tone perception ability of cochlear implant children in comparison with normal hearing children,” *International Journal of Pediatric Otorhinolaryngology*, 63: 137–147, 2002.
[15] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer* [Computer program], Version 6.0.14, http://www.praat.org/.
[16] Y.-L. Shue, *The Voice Source in Speech Production: Data, Analysis and Models*. PhD dissertation, Los Angeles: UCLA, 2010.
[17] G. S. Turner, K. Tjaden, and G. Weismer, “The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis,” *Journal of Speech, Language, and Hearing Research*, 38: 1001-1013, 1995.
[18] Y. S. Xu, J. T. Gandour, and A. L. Francis, “Effects of language experience and stimulus complexity on the categorical perception of pitch direction,” *Journal of the Acoustical Society of America*, 120(2): 1063–1074, 2006.
[19] P. A. Hallé, Y.-C. Chang, and C. T. Best, “Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners,” *Journal of Phonetics*, 32: 395–421, 2004.