

The Effects of Intrinsic Acoustic Cues on Categorical Perception in Children with Cochlear Implants

Kejuan Cheng¹ & Xiaoxiang Chen¹

¹School of Foreign Languages, Hunan University, Changsha, China

Correspondence: Xiaoxiang Chen, School of Foreign Languages, Hunan University, South Lushan Road, Yuelu District, Changsha 410082, China. E-mail: laxiaoxiangchen@sina.com

Received: June 2, 2020 Accepted: July 1, 2020 Online Published: July 8, 2020

doi:10.5539/ijel.v10n5p110 URL: <https://doi.org/10.5539/ijel.v10n5p110>

Abstract

Many previous studies researched the influence of external cues on speech perception, yet little is known pertaining to the role of intrinsic cues in categorical perception of Mandarin vowels and tones by children with cochlear implants (CI). This study investigated the effects of intrinsic acoustic cues on categorical perception in children with CIs, compared to normal-hearing (NH) children. Categorical perception experiment paradigm was applied to evaluate their identification and discrimination abilities in perceiving /i/-/u/ with static intrinsic formants and Tone 1 (T1)-Tone 2 (T2) with dynamic intrinsic fundamental frequency (F0) contours. Results for the NH group showed that vowel continuum of /i/-/u/ was less categorically perceived than T1-T2 continuum with significantly wider boundary width and less alignment between the discrimination peak and the boundary position. However, a different categorical perception pattern was depicted for the CI group. Specifically, the CI group exhibited less categoricalness in both /i/-/u/ and T1-T2. It suggested that the effects of intrinsic acoustic cues on categorical perception was proved for the normal-hearing children, while not for the hearing-impaired children with cochlear implants. In conclusion, acoustically dynamic cues can facilitate categorical perception of speech in NH children, whereas this effect will be inhibited by difficulties in processing spectral F0 information as in the CI users.

Keywords: categorical perception, intrinsic cues, Mandarin tones, Mandarin vowels

1. Introduction

Categorical perception (CP) in speech sounds refers to the phenomenon that the acoustic stimuli which vary along a physical continuum of equal intervals are perceived as discrete categories, and the differences between categories are more discriminable than within categories (Harnad, 2003; Reetz & Jongman, 2009; Zhang et al., 2016). To date, categorical perception has extended from stop consonants to vowels, and to tones, and has extended from speech to non-speech. Categorical perception research started its way at Haskins Laboratories. Dating back to Liberman, Harris, Hoffman and Griffith (1957), we found people perceived sounds that varied along a continuum abruptly not gradually. Accordingly, Liberman proposed that perception of stop consonants was “categorical”. The essential indicators of CP were summarized as follows: 1) the identification boundary position and boundary width of one category shifts to another category sharply; 2) the peak of the discrimination curve and its alignment with the categorical boundary; 3) the between-category accuracy is higher than that of the within-category (Liberman et al., 1957; Zhang et al., 2016). Quite a few conclusions have been made: perception of consonants is universally believed as categorical (Liberman et al., 1957; Miller & Eimas, 1977); for contour tones, it is concluded they are categorically perceived (e.g., Abramson, 1979; Francis et al., 2003; Xu et al., 2006; Hallé et al., 2004; Peng et al., 2010 etc.), while the perception of level tones is continuous (Abramson, 1979; Francis, Ciocca, & Ng, 2003); vowels chiefly are considered as being less categorical or even continuous (Fry et al., 1962).

Although extensive research has been carried out on categorical perception of consonants and tones, far too little attention has been paid to vowels. In contrast with the “continuous” view, Hallé (2004) believed perception of vowels was quasi-categorical, in line with the results in Zhang et al. (2016) and in Zhang and Shi (2014). In comparison, as a reference, the perception of vowels and consonants was reported to be categorical for the cochlear implanted children (Bouton, Serniclaes, Bertoncin, & Colé, 2012; Gu et al., 2016). As regards Mandarin tones, the study of Chen et al. (2017) on normal children uncovered that the perception of T1-T2 continuum

became categorical by the age of 4 yr old. Instead, children with severe hearing impairment have considerable difficulty in learning the tone system. In contrast to regular hearing-aids, cochlear implants bypassed the damaged hair cells and directly stimulated the auditory nerve by converting the mechanical sound energy into an electrical stimulus (Lee, van Hasselt, & Tong, 2010). Gu, Yin and Mahshie (2016) investigated the manner of categorical perception of T1-T2 and T1-T4 by the CI children, obtaining that both groups showed categorical effects. Nevertheless, some scholars argued that the gain in tone perception was not satisfactory (Wong & Wong, 2004; Aisha, 2000).

Despite the studies on categorical perception were abundant, previous studies mostly focused on the extrinsic factors of categorical perception, such as language experience, speech complexity, signal duration, and aging effect (e.g., Chen, Peng, Yan, & Wang, 2017; Francis et al., 2003; Hallé et al., 2004; Peng et al., 2010; Pisoni, 1975; Repp, Healy, & Crowder, 1979; Wang et al., 2017; Xu, Gandour, & Francis, 2006; Zhang et al., 2016). To compare categorical perception of vowels with tones is thus a good attempt to test the effects of intrinsic acoustic cues on categorical perception. Formants and fundamental frequency (F0) both are spectral/frequency information. F0 is a highly attributable cue to tones, whereas F1 and F2 are important for vowels. They are in contrast in their states.

In consideration of formant, it is inherently steady-state property as in monophthongs. Contrastively for diphthongs, the direction of lower formants will change continuously or even sharply, thus being a dynamic cue (Chen, Zhang, Wang, & Peng, 2019). Nearey (1989) has hypothesized that vowels' inherent static cue (i.e., formants of monophthongs) was crucial for the identification and discrimination, and dynamic changes within the inherent cue, like formants of diphthongs, would affect the manner of vowel perceptions. Differences between categorical perception patterns of monophthongs and vowels carried in a CVC construction supportively illustrate the foregoing statement. Being carried in CVC syllables, the original steady-state formant would generate a transition part due to the surrounding consonants. Therefore, in the study of Studdert-Kennedy (1976), he found stronger categoricalness in the CVC syllable than in the V carrier syllable. Chen et al. (2019) addressed the intrinsic factors in the categorical perception of vowels, implying that the lack of categoricalness of monophthongs was factually due to the steady-state of formant. In the same way, F0 contours are treated as a dynamic-state cue (Chen et al., 2019; Wang, 1967). With these regards, it is understandable that the perception of tones with different F0 contours are categorical (e.g., Wang, 1976; Xu et al., 2006; Peng et al., 2010; Chen et al., 2017), while the perception of level tones is continuous (e.g., Abramson, 1979; Francis et al., 2003). The former is dynamically changed in its intrinsic F0 contour, while the latter remains steady.

Regarding children with cochlear implants, studies stated that current implant systems didn't provide fine spectral or temporal information (Mckay, 2005). For hearing-impaired children, even with the help of CIs, the gain in tone perception remained unsatisfactory (Cheung, Wong, Lam, Lee, & van Hasset, 2002; Ciocca, Aisha, Francis, & Wong, 2002; Lee, van Hasselt, Chiu, & Cheung, 2002; Aisha, 2000; Wong & Wong, 2004). Lee et al. (2010) explained the pitch information essential for tonal languages seemed not to be explicitly represented in the electrical stimulation via current cochlear implant systems. For the CI users, spectral information is degraded (Chatterjee & Peng, 2008), along with difficulty in processing spectral information (Friesen et al., 2001; Lee et al., 2010; Petersen et al., 2015). Luo, Fu, Wu and Hsu (2009) brought additional evidences, in which the CI users performed better on vowel recognition than tone recognition, but were still able to score above 60% on average on tone recognition in quiet. In brief, conditions in the CI group are more complex.

Amid the existing literature, the studies of CP of tones and vowels if not most examined extrinsic factors, and adults were their priority. Yet, studies concerning how the intrinsic cues (static and dynamic) affect the categorical perception are nearly blank. Worse still, the effects in children, especially in the hearing-impaired children, are also less discovered. Originally, using the traditional paradigm of CP, the present study aims to address how the CI group, compared to the NH group, perceive /i-/u/ and T1-T2, and thus to examine how the intrinsic acoustic cues affect categorical perception.

2. Mandarin Vowels and Tones

2.1 Mandarin Vowels

Acoustically, vowels are specified by intensity, formant and duration. Formant is a concentration of acoustic energy, reflecting the way air from the lungs vibrates in the vocal tract. It appears as a peak in the frequency spectrum, and is closely related to vowel quality. The first three formants are the primary basis for vowels, i.e., F1, F2 and F3. Here, F1 and F2 will be used as two cues to select the most natural vowel token.

In Lin's (2007) opinion, a 5-vowel system was more acceptably to define and describe the Mandarin vowel system, namely, /i/, /y/, /u/, /ə/ and /a/. Alternatively, in Mandarin Chinese, Lin and Wang (2013) believed there

are mainly 7 monophthongs, namely [i], [u], [y], [ɤ], [a], and two apical vowels: front /ɿ/ and back /ʮ/. Table 1 gives averaged values of the first three formants from the recordings of 16 Beijing dialect speaking women. /i/ and /u/ are two of the most frequent vowel phonemes in the world (Zee & Lee, 2007). Mandarin /i/ is a front unrounded vowel with a higher value of F2 while a lower value of F1. On the contrary, /u/ is a back rounded vowel with both lower F1 and F2 values. They are primarily distinct from each other in F2.

Table 1. Formant values of the 7 Mandarin vowels from Lin and Wang (2013)

Formant (Hz)	[i]	[u]	[y]	[ɤ]	[a]	[ɿ]	[ʮ]
F1	320	420	320	750	1280	420	370
F2	2800	650	2580	1220	1350	1630	2180
F3	3780	3120	3700	3030	2830	3130	3210

Note. F1 = the first formant; F2 = the second formant; F3 = the third formant.

2.2 Mandarin Lexical Tones

Tone is a term used in phonology and phonetics referring to the distinctive pitch level of a syllable. Difference in tone is caused by pitch variations, which are produced by changes in the tension of the glottal folds that cause variations in fundamental frequency (F0) during voiced intervals of speech. Almost 60%–70% of the world's languages are tone languages (Yip, 2002), and over half of the world's people speak a tone language (Fromkin, 1978).

In Mandarin tone system, there are four lexical tones, carried by monosyllables. The four tones in terms of their contours are, respectively, level tone, rising tone, dipping tone and falling tone. According to Chao (1948), the relative pitch values of these four tones can be represented through a 1–5 scale, specifically indicating the relative starting and ending pitch of each tone. 1 refers to the lowest pitch, and 5 refers to the highest pitch. The four tones are corresponding to 55, 35, 214 and 51 respectively. To illustrate the relationship between the tone pitch and its meaning, an example of /ba/ is presented in the following table.

Table 2. Descriptive features of Mandarin tones

Tone type	Pinyin	Meaning	F0 contour	Pitch Value
Tone 1	/bā/	八(eight)	high level	55
Tone 2	/bá/	拔(to pull out)	mid rising	35
Tone 3	/bǎ/	把(a handful of)	low dipping	214
Tone 4	/bà/	爸(father)	high falling	51

In the present study, the perceptual performance for two continua, i.e., /i/-/u/ with static cue and T1-T2 with dynamic cue, are compared across two children groups: normal-hearing children listeners and cochlear implanted children listeners. In view of previous findings, the following hypotheses are made: 1) The NH group will perceive T1-T2 in a more categorical manner than vowels due to their different state of intrinsic acoustic cues (i.e., static versus dynamic); 2) Potential discrepancies are expected within the CI group, where they will perform both vowels and tones continua in a less categorical fashion due to their hearing disability.

3. Methodology

Before the beginning of the formal tests, a pilot study was done to detect problems during the whole experimental design, and then to make timely revision. Afterwards, also prior to the formal experiment procedure, an experimenter was responsible for giving a subject prescreening to avoid any individual who is unable to fulfill the basic cognitive development criteria. To minimize unexpected errors, the experimenter together with the teacher gave an extra lesson on /i/, /u/, T1 and T2, and at the same time an illustration session was ready to further their understanding of the procedure.

3.1 Participants

Two groups of children (i.e., NH and CI) were recruited for this study. The NH group consisted of 10 native Mandarin-speaking children (5M, 5F) with age range from 4;0 to 7;10 recruited from either kindergarten or primary school in Changsha. They had no reported history of speech, hearing or cognitive disorders, nor any music learning experience according to school and parental reports. In the CI group, 10 native Mandarin-speaking children (8M, 2F) between the ages of 4;9 and 8;8 were recruited from a local hearing

rehabilitation center in Changsha. Their duration of use was from 1;8 to 3;7 with a mean duration of 2;2. They were reported to be prelingually and congenitally impaired with profound hearing losses (> 90 dB HL) in either the right or the left ear, and had no reported history of other cognitive or physical disabilities. All of them were unimodal technology users, with only cochlear implant in the impaired ear. Moreover, they were all implanted before the age of 7 yr old, and the duration of their CI usage was at least 8 months. For more information see Table 3.

A background questionnaire was collected to gather information about their use of dialects and musical training experience. The questionnaire was written in Chinese, and was prepared for their parents or teachers to fill out before the children started their test. Parents' Notice was designed to make sure that they know what the experiment was about and how it would be proceeded, as well as any other details they deserved to know. If holding no disagreement, they would sign their names on it. We received informed consent. Every child was paid for their participation.

Table 3. Background information of the CI group

ID	Gender	Age	Age of activation	Duration of use	Implant device	Averaged hearing loss
1	M	4;9	1;2	3;7	MED-ELSonata	Profoundly
2	F	5;8	3;1	2;7	AB (Harmony)	Profoundly
3	M	5;1	1;7	3;6	MED-ELSonata	Profoundly
4	M	7;6	6;3	1;3	AB	Profoundly
5	M	5;10	4;9	1;1	AB (Meirenyu)	Profoundly
6	M	8;8	7;0	1;8	AB	Profoundly
7	M	6;2	4;7	1;8	AB (Meirenyu) Q90	Profoundly
8	M	5;7	3;6	2;1	AB (Meirenyu)	Profoundly
9	M	6;10	4;7	2;6	AB	Profoundly
10	F	5;6	3;9	1;9	CochlearCP802	Profoundly

Note. CI = cochlear implant.

3.2 Materials

Stimuli along the vowel and tone continuum in the present study was synthesized from the recorded samples of tone produced by a female native Mandarin speaker who is from northern mainland China.

For the /i/-/u/ continuum, /i/ was firstly produced in high level, and a set of 9 vowel stimuli were arranged in equal F1 (i.e., 12 Hz) and F2 (i.e., 296 Hz) acoustic intervals from /i/ to /u/. The frequency of F1 was from 317 Hz to 414 Hz, and from 2901 Hz to 536 Hz in F2. Owing to that the formant was a level property, the ending frequency was kept the same to the starting frequency. /i/-/u/ continuum was constructed based on the natural samples of /i/. The stimulus duration was interpolated to be 350 ms, with amplitude fixed at 70 dB. The third, fourth and fifth formants were fixed at 3957 Hz, 4766 Hz, and 4914 Hz respectively, which were derived from recorded samples. The major steps of synthesizing the stimuli with a formant synthesizer in Praat were listed as follows: 1) Sound normalization. In order to minimize the potentially confounding effects of duration, the duration of stimuli was normalized to 350 ms; 2) Synthesizing speech continuum based on a Praat script. Using /i/ as the basis for manipulation, a 9-step continuum was created by setting the designated values, regarding /i/ as Number 1 stimulus (the onset of the continuum) and /u/ as Number 9 stimulus (the end of the continuum). Figure 1 shows the schematic diagram of these stimuli.

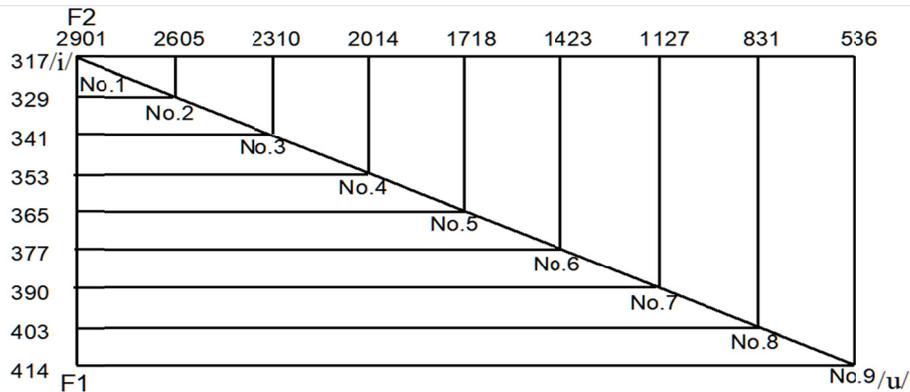


Figure 1. Schematic diagram of the vowel continuum

Stimuli along the tone continuum were synthesized from the recorded samples of Mandarin syllable /bā/ (T1, means “eight”), and /bá/ (T2, means “to pull out”) produced by the same speaker. Out of the five tokens recorded, only the best one was selected as the source and target sound. The /bā/ was set as the target sound, and /bá/ as the source sound. A set of 9 tone stimuli were arranged in equal F0 acoustic steps (i.e., 9 Hz) from /bá/ to /bā/. The starting frequency of F0 was from 210 Hz to 282 Hz, and the ending frequency was always 282 Hz. T1-T2 continuum was constructed based on the natural speech templates of two tones. After duration normalization, it was re-synthesized by applying the pitch-synchronous overlap and add (PSOLA) method (Moulines & Laroche, 1995) implemented in Praat (Boersma & Weenink, 2019). The normalization process of duration (i.e., 350ms) and intensity (70 dB) was the same as in the /i/-/u/ continuum. The synthesizing procedure was: 1) Fixing the pitch contour of Tone 1 to the level frequency of 282 Hz and the starting frequency of Tone 2 to 210 Hz; 2) Setting the number of pitch points to three, with the first one at the starting position, the second one at the 50 ms position, and the third one at the ending position; 3) Synthesizing the continuum using a Praat script based on the formula $210 \text{ Hz} + 9 \text{ Hz} * (\text{Stimulus Number} - 1)$. The schematic diagram is presented in Figure 2. All these manipulations were done on Praat (Boersma & Weenink, 2019).

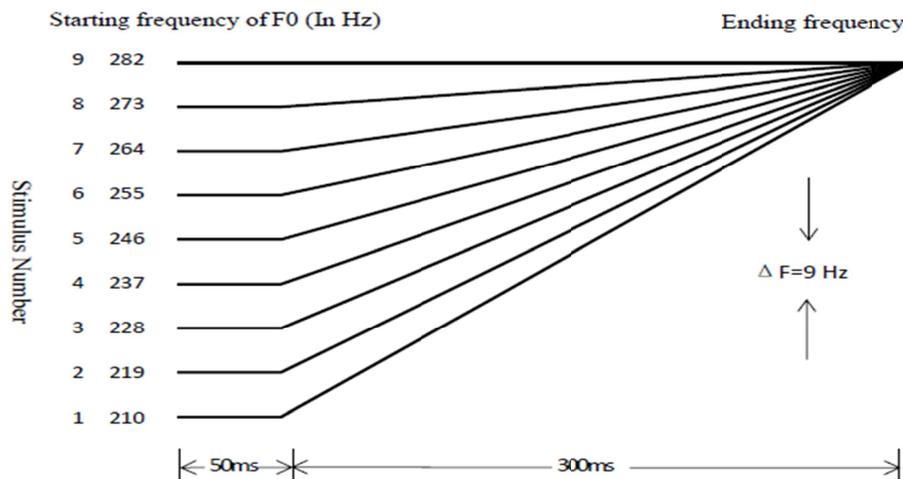


Figure 2. The schematic diagram of T1-T2 continuum

3.3 Procedure

All the participants were individually tested in a soundproof room by an experienced experimenter. They were instructed to complete two tasks including identification task and discrimination task. Each task had two separate blocks: one for the /i/-/u/ continuum and one for the T1-T2 continuum. They were required to respond as correctly and quickly as possible. In both tasks, stimuli were presented through a laptop running the software of E-prime (Schneider, Eschman, & Zuccolotto, 2002). Prior to the recording, all participants were given a brief

introduction of the experiment again to ensure that each participant knew how to make response in this experiment.

3.3.1 Identification Task

Before the experimental blocks, 9 tokens (Number 1 to 9) in each continuum were used as practicing trails. It aimed to familiarize participants with the experimental procedure and the labels used in the tasks, thus obtaining a stabilized performance in the formal experiments. The results from the practice session were not included in the following statistical analyses.

In the formal task, a 2AFC (two-alternative forced choice) test was used, during which the participants were instructed to identify each stimulus with labels on the screen (see Figure 3). The experimenter clicked the corresponding keys for them: “1” for “/i/” in the vowel continuum, “T1” in the tone continuum, and “2” for “/u/” and “T2” respectively. Vowel /i-/u/ and T1-T2 were separated into two different blocks, which were randomly presented to the participants. In the testing blocks, each stimulus was repeated 6 times, yielding a total of 54 experimental trials for each continuum. All the stimuli were randomly and automatically presented to the participants, without limitation on reaction time. Feedback was provided for the practicing trails, while no feedback was given during the testing blocks. It took each participant approximately 60 minutes to finish.



Figure 3. Picture for /i/ and /u/ (one the left) and T1 and T2 (on the right) on the screen

3.3.2 Discrimination Task

Similarly, prior to the testing procedure, a practicing session consisted of 23 stimuli (1-1, ..., 9-9, 1-3, ..., 7-9 and 3-1, ..., 9-7) was set.

The AX test design was used in the discrimination task. Stimuli were presented in pairs with a fixed inter-stimulus interval (ISI = 600 ms), and the participants were required to tell whether the two stimuli in each pair are the “same” or “different” by pointing to the pictures (see Figure 3). The experimenter helped to press the corresponding keys (“1” for “same” and “2” for “different”). Vowel /i-/u/ and T1-T2 were separated into two different blocks. Since AX test paradigm was applied, there were 7 units paired by two steps, and each unit had four types, i.e., AA, BB, AB, BA. Thus, a total of 23 stimuli pairs in each continuum were presented to the participants in a random order. More specifically, there were 14 pairs consisted of two different stimuli (different pairs) either in forward order (1-3, 2-4, ..., 6-8, 7-9) or in reverse order (3-1, 4-2, ..., 8-6, 9-7), and 9 pairs consisted of stimuli paired with themselves (same pairs) (following Chen et al., 2016). Each pair was repeated 4 times, yielding a total of 92 experimental trials in each block for each participant. The order of the two blocks was also randomized for each subject. Considering the particularity of the targeted participants, reaction time was not limited. Feedback was given in the practice session but not in the testing blocks. In order to arrange some rest for participants to avoid fatigue, they had a 20s’ rest after each time they finished a block. The entire session lasted about 90 minutes.



Figure 4. Picture for “same” (smiling face) and “different” (sad face) on the screen

3.4 Data Analysis

To investigate the difference between the categorical perception of /i/-/u/ and T1-T2, 3 essential parameters of CP function were obtained by E-prime, Excel and SPSS: boundary position, boundary width, and overall discrimination accuracy, as well as between- and within-category discrimination accuracies (Peng et al., 2010; Zhang et al., 2016; Chen et al., 2017).

The identification score was defined as the percentage of which the participants identified a certain stimulus as /i/ or /u/, or T1 or T2. In the study, the categorical boundary positions were obtained from the crossover points of identification curves of /i/-/u/ and T1-T2. The width of the categorical boundary was the linear distance between 25% and 75% of identification scores, which indicated the sharpness of this curve (Hallé et al., 2004). Probit analysis of the identification function was applied to assess and count each individual’s boundary position and boundary width (Finney, 1971). The narrower the boundary width was, the faster the change rate from one phoneme category to another was.

The discrimination score (P value) was the accuracy rate of responses, which was calculated according to formula (1) (Xu et al., 2006). Discrimination pairs in the present study were divided into 7 comparison units that consisted of four types each (i.e., AB, BA, AA, and BB). The 7 comparison units were further divided into between-category comparisons (P_{bc}) and within-category comparisons (P_{wc}) based on the positions of categorical boundaries (Chen et al., 2017). For example, if for a participant’s identification boundary position was 4.5, then the between-category pairs would be any unit that covered this point, namely 3-5 and 4-6 units, while the remaining 5 units belonged to the within-category pairs. The averaged accuracy rate of these pairs was the between- and within-category accuracy rate respectively. Here was the formula:

$$P = P('S'/S) \times P(S) + P('D'/D) \times P(D) \quad (1)$$

$P('S'/S)$ represented the percentage of participants’ “same” responses to all the same (S) trails (AA or BB) and $P('D'/D)$ represented the percentage of participants’ “different” responses to all the different (D) trails (AB or BA). $P(S)$ and $P(D)$ were the percentages of “same” (AA or BB) and “different” (AB or BA) pairs in each unit, respectively. In this experiment, both $P(S)$ and $P(D)$ were equal to 50%.

4. Results

4.1 Identification and Discrimination Curves

Identification and discrimination curves for /i/-/u/ and T1-T2 continua are shown in Figure 5. In terms of sharpness of the identification curves, the NH group exhibits steep slopes in tones continuum, while a less steep slope is shown in the vowels continuum. On the contrary, gentler slopes are seen in both vowels and tones continua for the CI group. As also presented in Figure 6, peaks are clearly shown in T1-T2 discrimination curves for both groups, but no clear peaks are observed in /i/-/u/ discrimination curves. Besides, the peak of the discrimination curve of T1-T2 corresponds well to the crossover position of the identification curves for the NH group. That is, for the NH group, the maximum accuracy appears around at the stimulus pair 5-7, and the boundary position is approximately at No.6 stimulus. As for the CI group, though with a peak in the discrimination curve, it doesn’t align with the boundary position in a fair way. In conclusion, clear alignment is found in the T1-T2 identification only for the NH group, while no salient peaks are found in vowels discrimination curves.

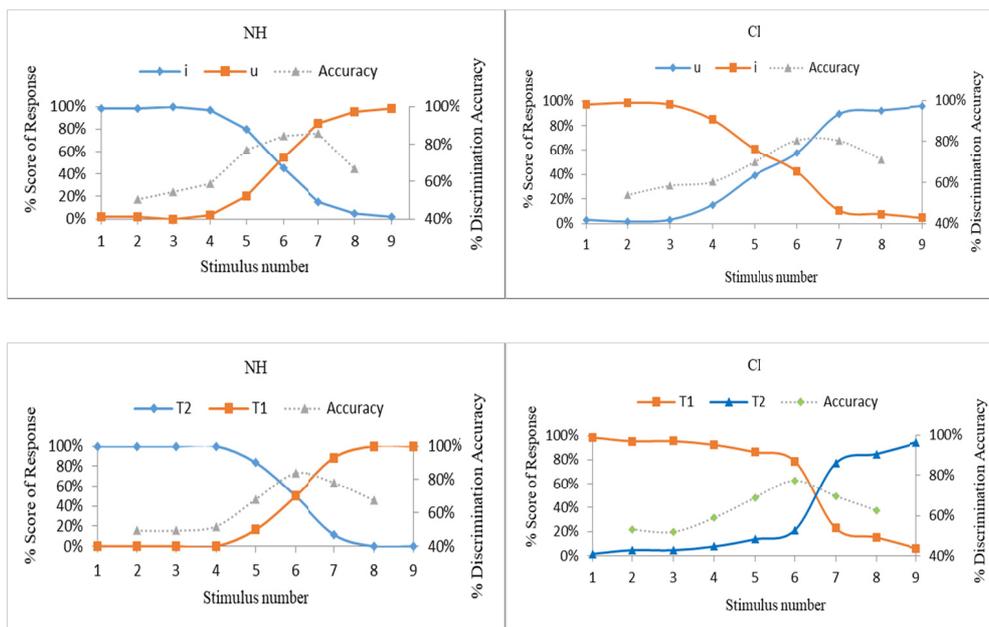


Figure 5. Identification and discrimination curves of both continuum for each group

Note. The left y-axis represents the percentage scores of /i/ or /u/, or T1 or T2 response, while the right y-axis represents the overall discrimination accuracy.

4.2 Positions and Widths of Categorical Boundary

The mean boundary positions and widths of the two continua are shown in Table 4, and are further depicted in Figure 6. Two two-way mixed design ANOVAs were conducted to examine the differences.

A two-way mixed design ANOVA was conducted, with the continua type (/i-/u/ and T1-T2 continua) as the within-subject factors and group as a between-subject factor. Greenhouse-geisser correction method was used when appropriate to correct violations of sphericity following Zhang et al. (2016). Results of the two-way mixed design ANOVA revealed no significant main effect of group ($F(1,18) = 2.177, p = 0.157$), nor significant main effect of continuum type ($F(1,18) = 0.958, p = 0.341$), but unveiled significant interactions of continuum type by group ($F(1,18) = 4.9, p = 0.04$). Post hoc analyses reveal that for the NH group, the boundary width in vowels was significantly wider than that in tones ($p < 0.05$), while for the CI group, boundary width in vowels was not significantly differed from that in tones ($p > 0.05$). And boundary width of vowels was not significantly different between groups ($p > 0.05$), while significant difference was shown in width of tones between the two groups ($p < 0.05$).

Another two-way mixed design ANOVA was assumed to examine the difference in boundary position. Results suggested no significant main effect of type ($F(1,18) = 3.301, p = 0.086$), nor no significant interactions ($F(1,18) = 1.397, p = 0.123$), and also showed nonsignificant main effect of group ($F(1,18) = 0.01, p = 0.923$).

Results indicated that the two groups differed not significantly either in the boundary positions, or in the boundary width of /i-/u/, but they differed significantly in the boundary width of T1-T2. And for the NH group, the vowel width was significantly wider than that of tones, whereas for the CI group, no such significant effect was observed. Therefore, these results induce that the CI group performs similarly in /i-/u/ and T1-T2 identification tasks, whereas the NH group exhibits a significantly sharper slope in the T1-T2 identification curve than in the /i-/u/ identification curve.

Table 4. Derived categorical boundary positions and widths of /i-/u/ and T1-T2 for each group

Group	Position		Width	
	i-u	T1-T2	i-u	T1-T2
NH	5.9084	5.9546	1.3845	0.7959
CI	5.5052	6.2989	1.5224	1.75

4.3 Discrimination Accuracy

The overall discrimination accuracies of /i/-/u/ and T1-T2 discrimination are depicted in Figure 7, and discrimination accuracies of between-category and within-category for the two groups are demonstrated in Figure 8. The mean overall discrimination accuracies are 68% (/i/-/u/) and 64% (T1-T2) for the NH group, and 68% (/i/-/u/) and 63% (T1-T2) for the CI group.

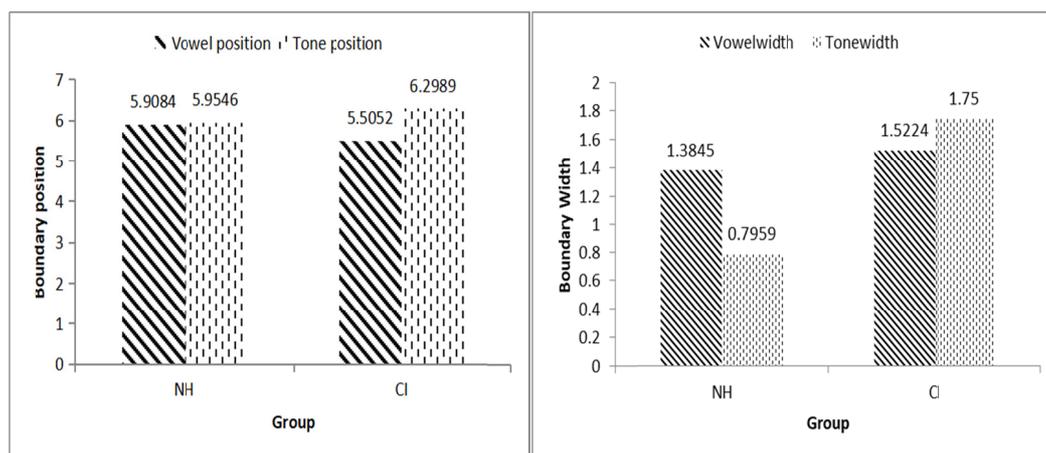


Figure 6. Derived boundary position and width in /i/-/u/ and T1-T2 continua for the NH and CI groups

A two-way mixed design ANOVA was conducted to examine the differences in their overall discrimination accuracies, with the continua type (/i/-/u/ and T1-T2 continua) as the within-subject factor and group as the between-subject factor. In the same way, Greenhouse-geisser correction method was used when appropriate to correct violations of sphericity. Results of the two-way mixed design ANOVA revealed significant main effect of continuum type ($F(1,18) = 16.978, p = 0.001$), but showed no significant main effect of group ($F(1,18) = 0.048, p = 0.829$), nor significant interactions of continuum type by group ($F(1,18) = 0.047, p = 0.831$). These suggested that for both groups, they performed significantly more accurately in /i/-/u/ than in T1-T2.

Since better discrimination performance across category boundaries than within the same category is an important parameter of categorical perception, these 7 stimulus pairs were further divided into between-category pairs and within-category pairs. The peakedness was calculated by the difference between P_{bc} and P_{wc} . And as shown in Figure 8, the mean between-category accuracies in vowels and tones for the NH group are 83% and 76% respectively, and for the CI group are 79% and 70% respectively. The mean within-category accuracies in vowels and tones for the NH group are 62% and 59%, and for the CI group are 64% and 61%.

Further Paired Samples T-tests were done to detect their difference in the between- and within-category accuracies. Results for the NH group demonstrated significant difference between P_{bc} and P_{wc} in /i/-/u/ and in T1-T2 [($t(9) = 5.306, p < 0.001$) and ($t(9) = 3.329, p = 0.009$), respectively]. For the CI group, statistical analysis presented significant difference between P_{bc} and P_{wc} in /i/-/u/ but only marginally significant difference was found in T1-T2 [($t(9) = 4.601, p = 0.001$) and ($t(9) = 2.167, p = 0.058$), respectively]. Conclusively, for both groups, the between-category accuracy is strongly and significantly higher than that of the within-category accuracy no matter in the /i/-/u/ or in the T1-T2 discrimination, except for the CI group in perceiving T1-T2.

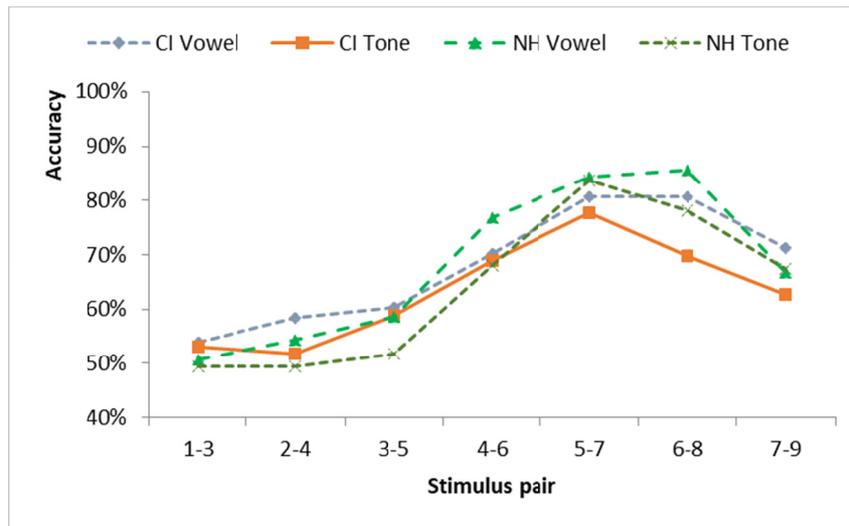


Figure 7. The overall discrimination accuracies in /i/-/u/ and T1-T2 continua for the NH and CI groups

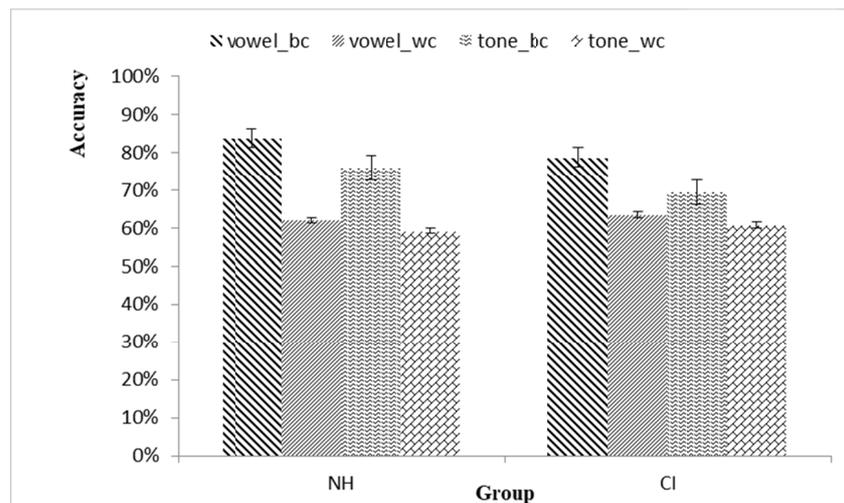


Figure 8. Within-category and between-category discrimination accuracies in /i/-/u/ and T1-T2 continua for the NH and CI groups

5. Discussion

This study investigated the effects of intrinsic acoustic cues on categorical perception in children with cochlear implants, compared with normal-hearing children. Results for the NH group unveiled that vowels /i/-/u/ continuum exhibited wider boundary width, ambiguous peak, and unclear correspondence between the identification boundary and the discrimination peak. Therefore, though showing some categorical effects (i.e., significant peakedness), vowels /i/-/u/ was still less categorically perceived than T1-T2 continuum, which was in line with the hypothesis previously stated. Contrastively, within the CI group, there was no significant difference between their boundary widths of /i/-/u/ and T1-T2 continua. Worse still, the vowel continuum showed neither clear peaks nor indisputable correspondences, while in the tone perception tasks, a peak in the discrimination curve was shown but it didn't correspond to the boundary position. Difference in the peakedness was only significant for the /i/-/u/ discrimination rather than for the T1-T2 discrimination. Results for the CI group illustrated that the CI children behaved in a less categorical manner both in perceiving /i/-/u/ and T1-T2, which also paralleled the hypothesis.

5.1 Static VS Dynamic Cues

Formants of vowels are static cues compared with F0 of contour tones, although they both belong to the spectral

information (Chen et al., 2019; Chen, Zhu, & Wayland, 2017; Zhang et al., 2017). Formants are steady-state properties, with few inflections with time going (Chen et al., 2019; Nearey, 1989). However, the F0 feature of contour tones is a dynamic-state property (Aramson, 1978; Chen et al., 2017; Wang, 1967). In Chen et al. (2017), they viewed that steady vs. dynamic acoustic may be responsible for the contrasting effects of stimulus duration on vowel and pitch direction perception.

The results of this experimental study depict that when the internal acoustic cue, i.e. formants, between monophthongs is in a relatively steady state, the categoricalness of perception decreases accordingly; on the contrary, when the internal cue, i.e., F0 contours, between tones dynamically change, the categoricalness of perception will be significantly improved. The current findings also echo the previous study (Chen et al., 2019). In Chen et al. (2019), the perception of monophthongs with static formants was significantly more categorical than that of monophthong-diphthong continuum with dynamic formants changes. Moreover, studies of level tones and contour tones provide supplementary evidences for this effect (e.g., Abramson, 1979; Francis et al., 2003; Xu et al., 2006; Hallé et al., 2004; Peng et al., 2010 etc.). These studies altogether discover that perception of tones with different contours is categorical (Peng et al., 2010; Xu et al., 2006; Hallé et al., 2004), while perception of level tones was continuous (Abramson, 1979; Francis et al., 2003).

It's worth noting that significant differences in positions amid tasks of /i-/u/ and T1-T2 identification were not present in this study, whereas both groups manifested significant differences between between-category and within-category accuracies in /i-/u/ and in T1-T2 discrimination. For the NH group, significant difference in widths between the two continua was displayed. These results might imply that the effects of intrinsic acoustic cues on categorical perception is in operation, and it plays its role mostly in the boundary width. Boundary width is closely related to the rate of detecting the changes between each sound pair. Therefore, it means that this effect influences the normal-hearing children's rate of perceiving the acoustic changes between sound stimuli. Chen et al. (2019) compared the indicators of categorical perception, and concluded that boundary width was the one of the decisive indicators to evaluate categoricalness in perception. Accordingly, the significantly narrower width in T1-T2 means more categorical, which effectively signals the presence of the effects of intrinsic acoustic cues on categorical perception.

5.2 Spectral Information Processing

In addition, in the current study, the conditions in the CI group was, to some extent, different from those in the NH group. That is, categoricalness in the perception of either vowels or tones was not significantly different and not satisfactory enough. To be specific, they showed significantly wider widths in both vowels and tones identification than the control group in T1-T2, and no significant differences in their boundary widths and positions were seen between the identification of /i-/u/ and T1-T2; the identification boundary and the discrimination peak didn't align well in the two continua. And worse still, strongly significant difference in the peakedness was observed in discriminating /i-/u/, but not found in T1-T2; overall discrimination accuracy in /i-/u/ was significantly higher than that in T1-T2. The analyses of boundary widths then suggested that the CI group perceived the changes of /i-/u/ and T1-T2 at a significantly lower rate compared to the NH group, but they performed with similar boundary widths in /i-/u/ compared to T1-T2.

Previously, studies observed that for hearing-impaired children, improvements were seen in perception of consonants and vowels (Lee et al., 2005). /i-/u/ by the CI group in this study was perceived in a similar manner with the NH group, which mirrored the results in Munson and Nelson (2005). The current findings of vowels perception by CI children testified the improvement of cochlear implants in processing formants information.

Nevertheless, the coding strategies of CIs have focused on conveying speech envelope information, while the fine structure of sounds (e.g., F0) has not been coded due to technological constraints (Gu et al., 2017). Results in this study are compatible with previous studies that the gain in tone perception tends to be unsatisfactory for those hearing-impaired children irrespective of the aid given by CIs (Wong & Wong, 2004; Cheung et al., 2002; Ciocca et al., 2002; Lee et al., 2002; Aisha, 2000). Significant reduction in performance in T1-T2 by the CI group in this study is also similar to the results of Luo et al. (2009). In their study, results indicated that CI users performed better on vowel recognition than tone recognition. More importantly, spectral information is degraded in CIs (Chatterjee & Peng, 2008), with CI users having a limited number of spectral channels available when compared to NH listeners (Friesen et al., 2001). In their study, CI recipients reportedly had difficulty with the recognition of some prosodic cues, especially those features closely related to fundamental frequency (F0). As cited in Lee et al. (2010), the pitch information was not explicitly represented in the electrical stimulation via CIs. In addition, evidence from MMN response verified that CI users were not sensitive to pitch deviants (Petersen et al., 2015). Therefore, worse performance in T1-T2 could be ascribed to their difficulty in processing fundamental

frequency information. It is likely that their difficulty in processing F0 information affects the way in which the intrinsic dynamic acoustic cue functions in the categorical perception of T1-T2. In a word, this effect is not manifested well in the CI group due to their difficulty in processing spectral information, especially in processing F0.

5.3 Acoustic Difference Detection

Intriguingly, /i/-/u/ exhibited higher discrimination accuracy for both groups. For one reason, this might be due to the relative larger step size difference in formants within each vowel sound pair, namely around 296 Hz in F2 and 12 Hz in F1, but only 9 Hz in F0 in T1-T2 continuum. As illuminated in Liu (2013), the just-noticeable difference (JND) of lexical pitch perception was 4-8 Hz. Chen et al. (2018) employed tonal comparisons with varying acoustic intervals as 3-step and 4-step stimuli in categorical perception of lexical tones, because 2-step pairs might be too small to perceive for amusics who were impaired in musical pitch perception. Petersen et al. (2015) discovered that weaker brain responses and poorer behavioral performance were true for CI users' discrimination of small changes in pitch. And, the lower-level acoustics underlies higher-level phonological categories (Chen et al., 2018). With these regards, in the current study, the results of the better discriminability on the vowel continuum might be attributable to the larger step size. T1-T2, due to the smaller changes in pitch, exhibits poorer performance in discrimination. Besides, Zhang et al. (2016) reported that vowel perception might be strongly influenced by pitch properties of lexical tones. The perception of /i/-/u/ continuum in the current study, therefore, might be posed as a result of the impacts from the original high-level tone as recorded prior to the identification and discrimination tasks. The flattened pitch remains steady in high level tone, which is congruous with the static formant in vowels. The overlaps of that kind of property could subserve the perception of the vowel stimuli synthesized under the condition of the high-level tone, hence the higher scores of discrimination could be graded for the vowel continuum instead of the tonal continuum. In addition, cochlear implants can detect higher frequency information moderately better, while worse in the detection of lower frequency information. In this study, F2 of /i/ and /u/ is much higher than the F0 of T1 and T2, along with the fact that the current processing schemes in CI do not provide optimum information about F0 compared to formants (Mckay, 2005), which in some way explains the discrepancy between them.

Importantly, though the discrimination accuracy was higher in /i/-/u/ than that in T1-T2, it was still lower than the accuracy for adults in the study of Chen et al. (2019). In addition, for the CI group, although they performed similarly in discrimination accuracy and boundary width in /i/-/u/ and T1-T2, they differed in the correspondence between discrimination peaks and boundary positions. Altogether, the categoricalness of /i/-/u/ and T1-T2 by the CI group is not as strong as the perception of T1-T2 by the NH group.

6. Conclusion

This study investigated the effects of intrinsic acoustic cues on categorical perception in children with cochlear implants, compared to normal-hearing children. Results showed that vowels /i/-/u/ continuum was less categorically perceived than T1-T2 continuum for the NH group. However, the CI group perceived T1-T2 and /i/-/u/ in a similarly less categorical way. In conclusion, the effects of intrinsic acoustic cues on categorical perception is proved to be true for normal-hearing children, while not for the hearing-impaired children due to the interference of their difficulty in processing spectral information. In a nutshell, dynamic acoustic cues can facilitate categorical perception of speech, which however might be hindered by difficulties in processing F0 information.

The present study is a good attempt to investigate the effects of intrinsic acoustic properties on categorical perception. It helps to decide whether categorical perception is affected differently by different states of inherent cues (i.e., static or dynamic) for the CI users compared to normal children. This will generate important theoretical contributions to further understanding of the relationship between acoustic property and perceptual mechanism. Drawing upon two stands of research into perception of speech, it can generate some pedagogical implications for children with special educational needs. By comparing their performances, we can find out which acoustic property causes greater difficulty for hearing-impaired children with cochlear implants and to what extent the hearing-impaired group differs from normal children. Clinically, it can exert suggestions for the research and development of cochlear implants. In the future work, attention should be paid to the comparison between the categorical perception of monophthong and diphthong, where different states of formant are exhibited. It is also possible to take other influential factors into consideration, such as the duration of use of CIs and the complexity of speech signals (i.e., speech versus non-speech). To compare categorical perception of spectral information (e.g., vowels or tones) with temporal information (e.g., consonants) by children with CIs is another concern to be addressed in the future. Furthermore, attempts should turn into the Event-related Potentials

(ERPs) experiments to make up for the shortage of behavioral experiments.

Funding

This project was supported by Hunan Provincial Social Science Fund (18YBA071) and by Humanities and Social Science Foundation of Chinese Ministry of Education (20YJC740041).

Acknowledgments

We are very grateful to the following people for their professional comments and very valuable suggestions concerning this paper. They are Prof. Fei Chen; Dr. Yuxiao Yang; Dr. Junzhou Ma; Jiaqiang Zhu and Huan Mei.

References

- Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language & Speech*, 21(4), 319–325. <https://doi.org/10.1177/002383097802100406>
- Abramson, A. S. (1979). *The noncategorical perception of tone categories in Thai* (pp. 91–100). *Frontiers of speech Communication Research*. London: Academic Press.
- Aisha, R. (2000). *Tone perception of Cantonese-speaking children with cochlear implant*. Bachelor of Science (Speech and Hearing Sciences) thesis. University of Hong Kong, Hong Kong.
- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Version 6.0). Retrieved September 25, 2019, from <https://www.fon.hum.uva.nl/praat/>
- Bouton, S., Serniclaes, W., Bertoncin, J., & Colé, P. (2012). Perception of speech features by French-speaking children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 55(1), 139–153. [https://doi.org/10.1044/1092-4388\(2011/10-0330\)](https://doi.org/10.1044/1092-4388(2011/10-0330))
- Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/harvard.9780674732889>
- Chatterjee, M., & Peng, S. C. (2008). Processing f0 with cochlear implants: modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235(1–2), 0–156. <https://doi.org/10.1016/j.heares.2007.11.004>
- Chen, F., & Peng, G. (2018). Lower-level acoustics underlie higher-level phonological categories in lexical tone perception. *The Journal of the Acoustical Society of America*, 144(3), EL158–EL164. <https://doi.org/10.1121/1.5052205>
- Chen F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, 1413–1434. <https://doi.org/10.1017/S0305000916000581>
- Chen, F., Zhang, H., Wang, S. Y., & Peng, G. (2019). Intrinsic cues and vowel categorical perception. *Linguistic Sciences*, 18(4), 410–425 (in Chinese). <https://doi.org/10.7509/j.linsci.201808.032304>
- Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in crosslinguistic categorical perception of pitch directions. *PLoS ONE*, 12(7), e0180656. <https://doi.org/10.1371/journal.pone.0180656>
- Cheung, D. M. C., Wong, T. K. C., Lam, S. L. M., Lee, K. Y. S., & van Hasselt, C. A. (2002). Comparison of speech coding strategies in Cantonese children with cochlear implants. *Cochlear Implants-An Update*, 147–151. <https://doi.org/10.1002/cii.193>
- Ciocca, V., Aisha, R., Francis, A., & Wong L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *Journal of the Acoustical Society of America*, 111, 2250–2256. <https://doi.org/10.1121/1.1471897>
- Finney, D. J. (1971). *Probit Analysis* (3rd ed.). Cambridge: Cambridge University Press.
- Francis, A. L., Ciocca, V., & Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, 65(7), 1029–1044. <https://doi.org/10.3758/BF03194832>
- Friesen, L. M., Shannon, R. V., Baskent, D. et al. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, 110, 1150–1163. <https://doi.org/10.1121/1.1381538>
- Fromkin, V. (1978). *Tone: a linguistic survey* (1st ed.). New York: Academic Press. <https://doi.org/10.1016/B978-0-12-267350-4.50005-4>

- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171–189. <https://doi.org/10.1177/002383096200500401>
- Gu, W. T., Yin, J., & Mahshie, J. (2017). *Categorical Perception in Two Pairs of Mandarin Tones among Bimodal Cochlear Implanted Children*. International Symposium on Chinese Spoken Language Processing. IEEE. <https://doi.org/10.1109/ISCSLP.2016.7918376>
- Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese Tones by Mandarin Chinese v.s. French Listeners. *Journal of Phonetics*, 32, 395–421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)
- Harnad, S. (2003). Categorical perception. In *Encyclopedia of Cognitive Science* (01/12/03). Nature Publishing Group: Macmillan.
- Lee, K. Y. S., & van Hasselt, C. A. (2005). Spoken word recognition in children with cochlear implants: a five-year study on speakers of a tonal language. *Ear Hear*, 26, 30S–37S. <https://doi.org/10.1097/00003446-200508001-00005>
- Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N. & Cheung, D. M. C. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *International Journal of Pediatric Otorhinolaryngol*, 63(2), 137–147. [https://doi.org/10.1016/S0165-5876\(02\)00005-8](https://doi.org/10.1016/S0165-5876(02)00005-8)
- Lee, K. Y. S., van Hasselt, C. A., & Tong, M. C. F. (2010). Lexical Tone Perception Ability of Profoundly Hearing-Impaired Children: Performance of Cochlear Implant and Hearing Aid Users. *Otology & Neurotology*, 31(7), 1079–1087. <https://doi.org/10.1097/MAO.0b013e3181eb3242>
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368. <https://doi.org/10.1037/h0044417>
- Lin, T., & Wang, Y. J. (2016). *Phonetics Course* (6th ed.). Beijing: Peking University Press (in Chinese).
- Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge, UK: Cambridge University Press.
- Liu, C. (2013). Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *The Journal of the Acoustical Society of America*, 134(4), 3011–3020. <https://doi.org/10.1121/1.4820887>
- Luo, X., Fu, Q. J., Wu, H. P., & Hsu, C. J. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hearing Research*, 256, 75–84. <https://doi.org/10.1016/j.heares.2009.07.001>
- Mckay, C. M. (2005). Spectral Processing in Cochlear Implants. *International Review of Neurobiology*, 70, 473–509. [https://doi.org/10.1016/S0074-7742\(05\)70014-3](https://doi.org/10.1016/S0074-7742(05)70014-3)
- Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *The Journal of the Acoustical Society of America*, 61(3), 835–845. <https://doi.org/10.1121/1.381373>
- Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication*, 16(2), 175–205. [https://doi.org/10.1016/0167-6393\(94\)00054-E](https://doi.org/10.1016/0167-6393(94)00054-E)
- Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners with cochlear implants. *Journal of the Acoustical Society of America*, 118, 2607–2617. <https://doi.org/10.1121/1.2005887>
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088–2113. <https://doi.org/10.1121/1.397861>
- Petersen, B., Weed, E., Sandmann, P., Brattico, E., Hansen, M., Sørensen, S. D., & Vuust, P. (2015). Brain responses to musical feature changes in adolescent cochlear implant users. *Frontiers in Human Neuroscience*, 9(FEB), 1–14. <https://doi.org/10.3389/fnhum.2015.00007>
- Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, 3(1), 7–18. <https://doi.org/10.3758/BF03198202>
- Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, Production, Acoustics, and Perception*. Chichester: Wiley-Blackwell.
- Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 5(1), 129–145. <https://doi.org/10.1037/0096-1523.5.1.129>

- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide* (Psychology Software Tools Inc., Pittsburgh).
- Studdert-Kennedy, M. (1980). Speech perception. *Language & Speech*, 23(1). <https://doi.org/10.1177/002383098002300106>
- Wang, S. Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33(2), 93–105. <https://doi.org/10.1086/464946>
- Wang, S. Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 28, 61–72. <https://doi.org/10.1111/j.1749-6632.1976.tb25472.x>
- Wang, Y. X., Yang, X. H., Zhang, H. et al. (2017). Aging Effect on Categorical Perception of Mandarin Tones 2 and 3 and Thresholds of Pitch Contour Discrimination. *American Journal of Audiology*, 26(1), 1–9. https://doi.org/10.1044/2016_AJA-16-0020
- Wong, A., & Wong, L. (2004). Tone perception of Cantonese-speaking prelingually hearing-impaired children with cochlear implants. *Otolaryngol Head & Neck Surgery*, 130(6), 751–758. <https://doi.org/10.1016/j.otohns.2003.09.037>
- Xu, Y., Gandour, J., & Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of Acoustic Society of America*, 20, 1063–1074. <https://doi.org/10.1121/1.2213572>
- Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164559>
- Zee, E., & Lee, W.-S. (2007). *Vowel Typology in Chinese* (pp. 1429–1432). In Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI).
- Zhang, H., Chen, F., Yan, N., Wang, L., Chen, Y., & Shi, F. (2016). *The effects of tone categories on the perception of Mandarin vowels*. 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP). <https://doi.org/10.1109/iscslp.2016.7918413>
- Zhang, H., Chen, F., Yan, N., Wang, L., Shi, F., & Ng, M. L. (2016). The Influence of Language Experience on the Categorical Perception of Vowels: Evidence from Mandarin and Korean. *Interspeech*, 873–877. <https://doi.org/10.21437/Interspeech.2016-887>
- Zhang, H., & Shi, F. (2014). The perceptual study of /a/ and /u/ in Beijing Mandarin. *Experimental Linguistics*, 3(1), 61–67 (in Chinese).

Copyrights

Copyright for this article is retained by the authors, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).