

# The Influence of Language Experience on the Categorical Perception of Vowels: Evidence from Mandarin and Korean

Hao Zhang <sup>1, 2</sup>, Fei Chen <sup>1</sup>, Nan Yan <sup>1</sup>, Lan Wang <sup>1</sup>, Feng Shi <sup>2</sup>, Manwa L. Ng <sup>3</sup>

 CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
College of Chinese Language and Culture, Nankai University, Tianjin, China
Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong

zhanghao@siat.ac.cn, nan.yan@siat.ac.cn

### **Abstract**

Previous research on categorical perception of speech sounds has demonstrated a strong influence of language experience on the categorical perception of consonants and lexical tones. In order to explore the influence of language experience on vowel perception, the present study investigated the perceptual performance for Mandarin and Korean listeners along a vowel continuum, which spanned three vowel categories /a/, /3/, and /u/. The results showed that both language groups exhibited categorical features in vowel perception, with a sharper categorical boundary of /3/-/u/ than that of /a/-/3/. Moreover, the differences found between the two groups revealed that the Korean listeners' perception tended to be more categorical along the /a/-/a/-/u/ vowel continuum than that of the Mandarin listeners. Furthermore, the Mandarin listeners tended to label stimuli more often as /a/ and less often as /u/ than the Korean counterparts. These perceptual differences between the Mandarin and Korean groups might be attributed to the different acoustic distribution in the F1×F2 vowel space of the two different native languages.

**Index Terms**: vowel perception, language experience, Mandarin, Korean

#### 1. Introduction

Categorical perception (CP) in speech sounds refers to the phenomenon that the acoustic stimuli which vary along physical continuum of equal intervals are perceived as discrete categories, and the differences between categories are more discriminable than within categories [1, 2]. The well-established psychophysical experimental paradigm for speech perception was firstly established by Liberman and his colleagues [3], including identification and discrimination tasks to label phonemes and tell them apart. They also generalized the characteristics of CP as follows: the categorical boundary was sharply defined by the identification function; a marked accuracy peak in discrimination function was close to the position of categorical boundary [3, 4].

The phenomenon of CP in speech perception has been well documented with respect to the perception of stop consonants and lexical tones [3-10]. However, whether perception of vowels is categorical is still under debate. In earlier studies, the mode of vowel perception was dubbed "continuous perception" in stark contrast to CP [11, 12]. These studies showed that the identification function of vowels exhibited a gradual slope and the discrimination functions did

not show any marked peaks close to the positions of categorical boundary. On the other hand, categorical-like perception of vowels was reported by some studies with peaks of discrimination accuracy around the vowel boundaries [13-15]. Moreover, our previous study also demonstrated CP features along a vowel continuum from the Mandarin vowels /a/ to /u/, in which the discrimination peak was well aligned with the location of categorical boundary (see Fig. 1) [16]. Consequently, it is reasonable to assume that perception of vowels is much less categorical than the classical CP in stop consonants, although not entirely continuous.

Language experience has exerted a crucial influence on CP of speech sounds [6-10]. For instance, typical CP of a consonant continuum /r-l/ was shown for American English, but not for Japanese listeners as there is no such phonemic contrast in Japanese [6]. Cross-language research on the perception of lexical tones showed evidence of strong CP in speech stimuli for tonal language listeners but not for nontonal language ones [9, 10]. In addition, the influence of tone inventory on pitch perception was demonstrated by considering different categorical boundaries between Mandarin and Cantonese listeners [8]. However, few studies examined CP of vowels associated with cross-language exploration. Stevens et al. [12] found different identification and discrimination functions between Swedish and American English listeners along the vowel continuum /i/-/y/-/u/, all of which are phonemic contrasts in Swedish but not in English. For the identification functions, the crossover points from /i/ to /y/ and from /y/ to /u/ were approximately one step higher along the continuum for the American than for the Swedish listeners. For the discrimination functions, the discrimination accuracy of all stimulus comparisons was consistently lower for the American listeners than the Swedish.

The vowels /a/ and /u/ are detected in almost all languages and dialects around the world, so the perceptual performance of these general vowels may shed some light on the perception of vowels at a more global level. Moreover, to extend our previous study [16], we included another vowel category /3/ for a 13-step continuum from /a/ to /u/ in the present study, as an attempt to confirm whether perception of the vowel continuum also exhibited characteristics of CP with an additional /3/ choice. The acoustic span between Mandarin vowels /a/ and /u/ is considerably large, and there is some acoustic distribution of the phoneme /3/ between the two corner vowels [17]. Furthermore, Korean listeners were selected as a comparison group to conduct a cross-language study of vowel perception, because of the similarities and

differences of vowel phonemic inventory in these two languages along the vowel continuum in the study. It has been demonstrated that Korean has the same vowel phonemes of /a/ and /u/ as Mandarin, and the acoustic distributions of /a/ and /u/ overlap in the F1×F2 vowel space [18, 19]. However, the vowel phoneme /3/ differs slightly in the two languages, as the main allophone is [x] in Mandarin, whereas it is [ə] in Korean. As shown in Fig. 2, there is only a small portion of the acoustic distribution of Mandarin [x] is located within the vowel continuum from /a/ to /u/, whereas the core distribution of Korean [ə] is located exactly within the continuum (see Fig. 2). In short, the differences of phonetic details of Mandarin and Korean offered a unique chance to reveal the influence of language experience on vowel perception.

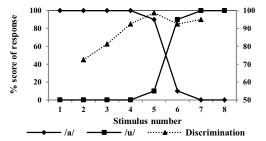


Figure 1: Identification and discrimination curves for vowel perception between /a/ and /u/. The left y-axis and the right y-axis indicate the percentage of identification response and the discrimination accuracy rate respectively. (From Zhang [16])

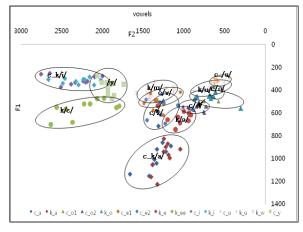


Figure 2: The acoustic space distributions for main vowels in Chinese and Korean, where "C" stands for Chinese vowel and "K" stands for Korean vowel. The x-axis and y-axis indicate the second and the first formant dimensions respectively. (From Lee [18])

## 2. Methods

## 2.1. Participants

Two groups of participants were recruited in this study: The Mandarin and Korean listener groups. For the Mandarin group, there were sixteen volunteers (8 females, M = 21.04 years, SD = 1.08) who were born and raised in Beijing, with Mandarin as native language for both the participants and their parents. For the Korean group, another sixteen volunteers (8 females, M = 23.63 years, SD = 2.27) were recruited with Korean as their

native language. All participants reported no history of speech or hearing difficulties. Besides, all participants were right-handed and were paid for their participation.

#### 2.2. Materials

Stimuli along the vowel continuum of the present study were synthesized from the recorded sample of /a/ produced in highlevel tone by a male native Mandarin speaker. A set of 13 vowel stimuli were arranged in equal F1 and F2 acoustic steps from /a/ to /u/. Fig. 3 shows the schematic diagram of these stimuli. Stimulus duration was set to 400 ms, with amplitude fixed at 75 dB. The third, fourth and fifth formants were fixed at 2685 Hz, 3380 Hz, and 4840 Hz respectively, which were derived from recorded samples of the male speaker. All the vowel stimuli were re-synthesized with a cascade formant synthesizer in Praat [20].

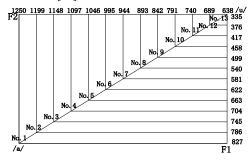


Figure 3: Schematic diagram of the vowel stimuli.

#### 2.3. Procedure

All the participants were instructed to complete two tasks including identification and discrimination tasks. They were required to respond as correctly and quickly as they could. In both tasks, stimuli were presented through a laptop running the software of E-prime.

#### 2.3.1. Identification task

A 3AFC (three-alternative forced choice) test was used in the identification task. During the task, the participants were instructed to identify each stimulus presented through a headphone by selecting one of the three categories: /a/, /3/, and /u/. Two identification procedures (ID 1, ID 2) were designed according to the differences in their selective requirements (i.e. in ID1, the '1', '2', and '3' in the keyboard were required to press for /a/, /3/ and /u/ respectively; while in ID2, the keys '3', '2', and '1' were pressed for /a/, /3/ and /u/ respectively). All the participants were randomly assigned to one of the two procedures. Before the experimental blocks, 12 trials were used as practice to familiarize the participants with the tasks. Each vowel stimulus was repeated 4 times, yielding a total of 52 experimental trials for each participant. All the stimuli in this task were randomly presented to the participants.

#### 2.3.2. Discrimination task

The AX test design was used in the discrimination task. Stimuli were presented in pairs with a fixed inter-stimulus interval (ISI = 500 ms), and the participants were required to decide whether the two stimuli in each pair were the "same" or "different". Also, two procedures (DIS 1, DIS 2) were designed to counterbalance across the participants (likewise, the keys 'F' and 'J' were pressed for different pair conditions

in different procedures). A total of 35 vowel stimuli pairs were presented to the participants in a random order. Among them, 22 pairs consisted of two different stimuli which were separated by two steps (different pairs) either in forward order (1-3, 2-4, ..., 10-12, 11-13) or in reverse order (3-1, 4-2, ..., 12-10, 13-11), and 13 pairs consisted of stimuli paired with themselves (same pairs). Before the experimental blocks, four pairs were used as practice trials to familiarize the participants with the task.

#### 2.4. Data analysis

To investigate the effect of language experience on vowel perception, identification and discrimination data exhibited by each participant were calculated including positions and widths of categorical boundary, maximum scores of vowel identification, and discrimination accuracy.

### 2.4.1. Identification scores

The identification score was defined as the percentage of responses which the participants identified a certain stimulus as /a/, /3/, or /u/. In the study, two categorical boundary positions were obtained: the crossover point of identification curves between /a/ and /3/, and between /3/ and /u/ (see Fig. 5). The widths of the two categorical boundaries were the linear distance between 25% and 75% of identification scores between /a/ and /3/, and between /3/ and /u/ respectively, which indicated the sharpness of the two boundaries [8-10].

#### 2.4.2. Discrimination scores

The discrimination score (P value) was defined as the accuracy rate of responses which was calculated according to the formula described specifically in Xu et al. [10]. Discrimination pairs in the present study were divided into 11 comparison cohorts that consisted of four types of pair-wise comparisons (i.e. AB, BA, AA, and BB) [8]. The 11 comparison cohorts were further divided into between-category comparisons ( $P_{bc}$ ) and within-category comparisons ( $P_{wc}$ ) based on the positions of categorical boundaries (cf. Jiang et al. [21]).

## 3. Results

#### 3.1. Identification and discrimination curves

Identification and discrimination curves associated with the vowels /a/, /a/, and /u/ for different language groups are shown in Fig. 5. As depicted in Fig. 5, prominent discrimination peaks were close to the corresponding identification boundaries for both language groups. For the Korean group, the maximum accuracy in the discrimination tasks was found around pair 7-9 which was close to the boundary position of /a/-/u/, while around the boundary position of /a/-/a/ there was also a peak in discrimination of pair 3-5. For the Mandarin group, although the discrimination peaks were not as prominent as the Korean group, the accuracy reached its maximum around pairs 6-8 and 8-10 which aligned well with boundary position of /a/-/a/ and /a/-/u/ respectively.

The discrimination accuracies of between-category and within-category comparisons for the two language groups are shown in Fig 4. Results of pairwise t-tests revealed that the discrimination accuracy of between-category comparisons was significantly higher than that of within-category both for the Mandarin (t = 4.33, p = 0.001, df = 15) and for the Korean listeners (t = 8.3, p < 0.001, df = 15).

#### 3.2. Positions and widths of categorical boundary

The boundary positions and widths of /a/-/3/ and /3/-/u/ for the two language groups are shown in Table 1. For the boundary positions, a two-way mixed design ANOVA was conducted, with the vowel type (/a/-/3/ and /3/-/u/) as a within-subject factor and group as a between-subject factor. Greenhouse-Geisser correction method was used when appropriate to correct violations of sphericity. The results showed a significant main effect of group (F(1, 30) = 31.89, p < 0.001), and the interactions of vowel type by group (F(1, 30) = 4.69,p = 0.039). Post-hoc analyses revealed that both boundary positions of /a/-/3/ and /3/-/u/ occurred consistently at a smaller stimulus number for the Korean than the Mandarin listeners (all ps < 0.05). As depicted in Fig. 3, the stimuli with smaller numbers were more inclined to /a/, so the smaller number of boundary positions indicated that the Korean listeners labeled stimuli as /a/ less often and as /u/ more often than the Mandarin counterparts.

For the boundary widths, a 2 (vowel type)  $\times$  2 (language group) mixed design ANOVA was conducted, with the vowel type as a within-subject factor and group as a between-subject factor. The result showed a significant main effect of the vowel type (F (1, 30) = 9.68, p = 0.004), but not group (F (1, 30) = 0.31, p = 0.58) and the interaction of vowel type by group (F (1, 30) = 1.45, p = 0.24). Post-hoc analyses revealed that the boundary width of /3/-/u/ was significantly narrower than that of /a/-/3/ for both listener groups (all ps < 0.05).

#### 3.3. Maximum identification scores

The maximum percentages of identification response of /a/ and /u/ were 100% for both groups, whereas that of /3/ was 78.1% and 85.9% for the Mandarin and Korean group respectively. Mann-Whitney nonparametric tests on the maximum identification score of /a/, /3/, and /u/ between the two groups showed that the maximum score of vowel /3/ for the Korean listeners was significantly higher than the Mandarin (p = 0.035), whereas the maximum score of /a/ or /u/ between the two groups was similar, both approximating the ceiling level (all ps > 0.05). Besides, rank-sum tests revealed that both groups showed a lower maximum score of /3/ than that of /a/ and /u/ (all ps < 0.05).

Table 1: Derived categorical boundary positions and widths of /a/-/3/ and /3/-/u/ for each language group.

Language	Position		Width	
group	/a/-/3/	/3/-/u/	/a/-/3/	/3/-/u/
Mandarin	6.59	9.85	1.89	1.36
Korean	4.72	8.68	1.74	1.21

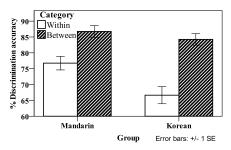


Figure 4: The discrimination accuracy rates of betweencategory and within-category comparisons for the two language groups.

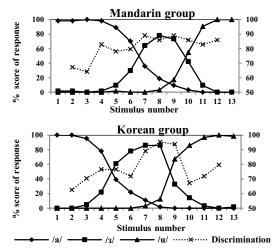


Figure 5: Identification curves and discrimination curves for each language group.

#### 4. Discussion

Perception along the vowel continuum from /a/ to /u/ exhibited some characteristics of categorical perception, especially for the Korean listeners. Moreover, the influence of language experience on perception of vowels can be evidenced by the identification and discrimination functions for the two language groups.

As regards CP, the major outcome was related with better discrimination performance for stimuli pairs with straddling category boundaries than for equivalently separated pairs within the same categories [8]. In the present study, comparatively prominent discrimination peaks were close to the corresponding identification boundaries for both language groups (see Fig. 5) and the discrimination accuracy of between-category comparisons was significantly higher than within-category (see Fig. 4). Therefore, categorical features were shown in perception of the vowel continuum of /a/, /3/, and /u/. Besides, for both language groups, there was a significant difference between the boundary width of /3/-/u/ and /a/-/3/, with a much narrower width of /3/-/u/ than /a/-/3/, which indicated that the categorical boundary of /3/-/u/ was much sharper than that of /a/-/3/ [8]. It was hypothesized that the degree of articulatory discontinuity between speech sounds might be manifested by the sharpness of categorical boundary separating them. More specifically, a higher degree of articulatory discontinuity between phonemes would result in a sharper categorical boundary [11]. The articulatory change in transition from /a/ to /3/ is less prominent than from /3/ to /u/, since the former includes only change of tongue position while the later includes both tongue position and lip aperture.

Furthermore, the perception for the Korean listeners tended to be more categorical along the vowel continuum of /a/, /3/, and /u/ than the Mandarin listeners. Discrimination peaks and the corresponding identification boundaries were better aligned and a larger difference of discrimination accuracies between within-category and between-category comparisons was exhibited for the Korean than the Mandarin group. As mentioned previously, the vowel continuum in this study comprised of the core acoustic distribution of Korean [9], whereas only a small portion of the acoustic distribution of Mandarin [8] was located within this continuum. That is to say, the vowel continuum involves the prototype category of

Korean [ $\mathfrak{a}$ ], whereas only peripheral category of Mandarin [ $\mathfrak{x}$ ]. The signal of speech production and the performance of speech perception are closely coupled and inherently linked [22]. Lacking the perceptual prototype structure of Mandarin [x] might contribute to a more reliant upon acoustic differences while discriminating within-category and betweencategory comparisons for the Mandarin listeners in contrast to the Korean. Therefore, the non-perceptual status of [x] in physical production could map a less categorical performance in psychological perception for the Mandarin listeners. However, the future research should be investigated to explore the perceptual performance of the vowel continuum in the present study for the Korean learners whose native language is Mandarin, and to detect if these Mandarin listeners' perception tends to be more categorical with the increase of linguistic background knowledge of Korean.

Another significant difference between the two language groups revealed that the boundary positions of /a/-/3/ and /3/-/u/ showed more than one step higher for the Mandarin than the Korean group. According to the motor theory proposed by Liberman and his colleagues [23, 24], sensory perception of speech sounds was mediated by motor production. When listeners heard synthetic stimuli from a continuum, they perceived these stimuli by matching them with what would have to be done to articulate them [1]. Since the most familiar sounds for the listeners to pronounce are those from their native languages, vowel systems of different native languages could influence their vowel perception. On the basis of the perceptual assimilation model (PAM) [25, 26], listeners tended to assimilate an unfamiliar phonetic segment to a similar category in their native language. Consequently, Mandarin listeners might attribute vowel category /3/ in the vowel continuum as the allophone [x], whereas Korean listeners might attribute that as the allophone [5]. Moreover, the core distribution of vowel space for Korean [ə] is much closer to /a/ than Mandarin [x] (see Fig. 2). The differences in acoustic distribution of different allophones of /3/ in the two languages could explain the two boundary positions' shifting toward the /a/ direction for the Korean listeners in comparison with the Mandarin, this also provide a support to the perspective that the fine-grained phonetic details of native language could strongly affect the perception of vowels [27, 28].

## 5. Conclusions

Both language groups showed some features of categorical perception along the vowel continuum in this study. Moreover, perceptual performance for the Mandarin listeners was less categorical than for the Koreans, since the discrimination curves for the Korean listeners exhibited relatively prominent peaks around both two categorical boundaries while those for the Mandarin listeners were much smoother. Besides, the Korean listeners tended to show significantly smaller stimulus numbers in both boundary positions than the Mandarin counterparts. The different acoustic distribution of Mandarin and Korean vowels could account for the different performance of categorical perception between the two groups.

## 6. Acknowledgement

This study was jointly supported by grants from the National Natural Science Foundation of China (NSFC 61135003, 91420301 and 61401452), and the Major Project of National Social Science Foundation of China (13&ZD134).

#### 7. References

- [1] S. Harnad, "Categorical perception," *Encyclopedia of Cognitive science*, vol. 67, no. 4, 2003.
- [2] H. Reetz, and A. Jongman, Phonetics: Transcription, Production, Acoustics, and Perception. Chichester, UK: Wiley-Blackwell, 2009.
- [3] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B.C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries," *Journal of Experimental Psychology*, vol. 54, pp. 358-368, 1957.
- [4] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the speech code," Psychological Review, vol. 74, no. 6, pp. 431-61, 1967.
- [5] E. B. Slawinski, and N. L. Lau, "Categorical perception of [ba] and [pa]: integration of acoustical cues," *Journal of the Acoustical Society of America*, vol. 100, no. 4, pp. 2693, 1996.
- [6] K. S. Mackain, C. T. Best, and W. Strange, "Categorical perception of English /r/ and /i/ in Japanese bilinguals," *Applied Psycholinguistics*, vol. 2, no. 4, pp. 369-390, 1981.
- [7] W. S.-Y. Wang, "Language change," Annals of the New York Academy of Sciences, vol. 208, pp. 61-72, 1976.
- [8] G. Peng, H-Y. Zheng, T. Gong, R-X. Yang, J-P. Kong and W. S-Y. Wang, "The influence of language experience on categorical perception of pitch contours," *Journal of Phonetics*, vol. 38, pp. 616-624, 2010.
- [9] P. A. Halle', Y-C. Chang, and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," *Journal of Phonetics*, vol. 32, pp. 395-421, 2004.
- [10] Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *Journal of Acoustical Society of America*, vol. 120, no. 2, pp. 1063-1074, 2006.
- [11] D. B. Fry, A. S. Abramson, P. D. Eimas, and A. M. Liberman, "The identification and discrimination of synthetic vowels," *Language and Speech*, vol. 5, no. 4, pp. 171-189, 1962.
- [12] K. N. Stevens, A. M. Liberman, M. Studdert-Kennedy, and S. E. Öhman, "Crosslanguage study of vowel perception," *Language and Speech*, vol. 12, no. 6, pp. 1-23, 1969.
- [13] H. Fujisaki, and T. Kawashima, "Some experiments on speech perception and a model for the perceptual mechanism," *Annual Report of the Engineering Research Institute*, vol. 29, no. 2, pp. 207-214, 1970.
- [14] D. B. Pisoni, and J. H. Lazarus, "Categorical and noncategorical modes of speech perception along the voicing continuum," *Journal of Acoustical Society of America*, vol. 55, no. 2, pp. 328-333, 1974.
- [15] D. B. Pisoni, "Auditory short-term memory and vowel perception," *Memory & Cognition*, vol. 3, no. 1, pp. 7-18, 1975.
- [16] H. Zhang, and F. Shi, "The perceptual study of /a/ and /u/ in Beijing Mandarin," *Experimental Linguistics*, vol. 3, no. 1, pp. 61-67, 2014. (in Chinese).
- [17] F. Shi, "The vowel pattern of Beijing Mandarin," *Nankai Linguistics*, vol. 1, pp. 30-36, 2002. (in Chinese).
- [18] S. H. Lee, "The prototypical features of Chinese vowel systems for Korean Chinese leaners," in *Proceedings of the 22<sup>nd</sup> North American Conference on Chinese Linguistics (NACCL-22) & the 18<sup>th</sup> International Conference on Chinese Linguistics (IACL-18), May 20–22, Massachusetts, Cambridge, USA, Proceedings, 2010, pp. 169-186. (in Chinese).*
- [19] M. Lee, Korean Chinese learner's Chinese mono vowels. Seoul, Korea: Seoul National University, 2006.
- [20] P. Boersma and D. Weenink, Praat: Doing phonetics by computer (Version 6.0), 2016, http://www. praat. org/, last accessed March 10, 2016.
- [21] C. Jiang, J. P. Hamm, V. K. Lim, I. J. Kirk and Y. Yang, "Impaired categorical perception of lexical tones in mandarinspeaking congenital amusics," *Memory & Cognition*, vol. 40, no. 7, pp. 1109-1121, 2012.

- [22] E. D. Casserly, and D. B. Pisoni, "Speech perception and production," Wiley Interdisciplinary Reviews Cognitive Science, vol. 1, no. 5, pp. 629-647, 2010.
- [23] M. Studdert-Kennedy, A. M. Liberman, K. S. Harris, and F. S. Cooper, "Theoretical notes. Motor theory of speech perception: A reply to Lane's critical review," *Psychological Review*, vol. 77, no. 3, pp. 234-249, 1970.
- [24] A. M. Liberman, and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, no. 1, pp. 1-36, 1985.
- [25] C. T. Best, G. W. McRoberts, and E. Goodell, "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 775-794, 2001.
- [26] C. T. Best, and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, pp. 13-34, 2007.
- [27] C. T. Best, and W. Strange, "Effects of phonological and phonetic factors on cross-language perception of approximants," *Journal of Phonetics*, vol. 20, pp. 305-330, 1992.
- [28] C. T. Best, P. Halle, O. S. Bohn, and A. Faber, "Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages," in *ICPhS 2003 - 15<sup>th</sup> International Congress of Phonetic Sciences, August 3–9, Barcelona, Spain, Proceedings*, 2003, pp. 2889-2892.