A Perceptual Illusion in Native Language: Positional voicing variations in Korean consonants

SEUNG-EUN CHANG (장성은)
UNIVERSITY OF CALIFORNIA, BERKELEY
Introduction

- **Perceptual difference** of native and non-native speakers in Korean
  - Example: three-way stop contrasts

Chang (2011)
Introduction

- Dupoux et al. (1999): Japanese listeners judged that the vowel was present even for an item with no vowel at all between the consonants (e.g. ebzo) as well as for the item with full vowel (e.g., ebuazo).

- Jacobs (2014): The final consonant of *koop* ‘to buy’ in Dutch is produced differently /p/ or /b/, depending on the subsequent sounds but Dutch listeners report hearing the same sound [p] in both cases, and the variations are neutralized by native Dutch listeners.
Introduction

Research question

- Perceptions on the positional voicing variations in Korean consonants
  - Modern Korean has nineteen consonant phonemes. All stop and affricate consonants among them are voiceless in word-initial position. However, lax stops and affricates are voiced in intervocalic position (e.g., kake /kəkə/ ‘store’, pâta /pədə/ ‘sea’, munche /munĉe/ ‘problem’) (Sohn, 2001).
  - Korean speakers actually produce “voiced” sounds
Experiment: A discrimination test for positional voicing variations in Korean consonants (i.e., voiceless obstruents vs. voiced obstruents) was performed on Korean listeners and non-Korean listeners (i.e., English and Chinese listeners).

- **English**: A two-way contrast: voiced and voiceless (VOT & F0)
- **Chinese**: A two-way contrast: voiceless unaspirated and voiceless aspirated (VOT)

- Korean: phonetic but not phonemic
- **English**: phonemic (big/pig)
- **Chinese**: neither phonemic nor phonetic

- voicing contrast in stop and obstruent
Methodology: Participants

- Native speakers of
  - 10 Korean
  - 10 Chinese
  - 10 English

- 21-30 years old
- 8 females + 2 males
- English & Chinese speakers (Students of Korean language class at UC, Berkeley)
Methodology: Stimuli

- $C_1$ = voiceless obstruents [k, t, p, ch],
- $C_2$ = VOT manipulated obstruents of $C_1$
- $V_1$ and $V_2$ = [u] (Klatt, 1975; Weismer 1979: VOT before high vowels is significantly longer than VOT before low vowels).

<table>
<thead>
<tr>
<th>Same Vowel context</th>
<th>/kuku/</th>
<th>/tutu/</th>
<th>/pupu/ ‘married couple’</th>
<th>/chuchu/ ‘stock holder’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different Vowel context</td>
<td>/koki/ ‘meat’</td>
<td>/tetu/ ‘soybean’</td>
<td>/pope/ ‘treasure’</td>
<td>/chechu/ ‘talent’</td>
</tr>
</tbody>
</table>

- $V_1$ and $V_2$ are different
- real words whose two vowel heights were same or as close to each other as possible (e.g., mid-mid or mid-high, excluding high-low) to avoid the excessive distortion of the original sound of the words.

: Uttered by a female Korean speaker were used as the original stimulus items
Methodology: Speech stimuli

The stimuli were digitized at 16kHz/16 bits. Using Praat, ten /C₁V₁C₂V₂/ synthetic stimuli differing in VOT of C₂ by 20ms steps (ranging from -90ms to +90ms) were created from each of eight original sound objects (Boersma & Weenink, 2011). The VOT length of C₁ was reserved as it is in the original sound (i.e., 50ms ~ 95ms) in order to be heard as naturally as possible.

Fig.1: A sample of /gu/-/ku/ voicing continuum of C₂ with waveforms (upper part) and spectrograms (lower part).
Methodology: Procedures

- Participants heard 2 repetitions of each stimulus, for a total of 160 tokens (8 stimuli * 10 continuum * 2 repetition).
- The stimuli were presented on a laptop using a Sennheiser headphone.
- A two-second interval between each stimuli.
- The participants were provided with an answer sheet and asked to circle “same” if the two consonants of $C_1V_1C_2V_2$ sequence are the same or circle “different” if two consonants were different.
- We emphasized that the experiment was not aimed at measuring their linguistic competence and that the number of “same” and “different” answers did not need to be balanced.
Results

- Question: The effects of VOT length on the perception of Korean consonants voicing by language group, VOT variations, and vowel context

- A univariate ANOVA - **dependent variable**: the percentage of responses collected for the stimuli with different VOT length
  - **Independent variables**: (i) language group (Korean, English, Chinese), (ii) VOT variation (-90ms, -70ms, -50ms, -30ms, -10ms, +10ms, +30ms, +50ms, +70ms, +90ms), (iii) vowel context (same-vowel vs. different-vowel)
Results

- The significant main effect was found for all factors:
  - **language group**: [F (2, 420) = 425.2, p < 0.001], with more “same” responses for Korean listeners than English listeners, and English listeners chose more “same” responses than Chinese listeners
  - **VOT**: [F (2,420) = 18.1, p < 0.001], with most “same” responses for +90ms and least for -30ms
  - **vowel context**: [F (1, 420) = 156.9, p < 0.001], with more “same” responses for same-vowel context than different-vowel context

- Significant two-way effect was also found for language and vowel context [F (2,420) = 107.7, p < 0.001], between language and VOT [F (2,420) = 2.1, p = 0.004], and between VOT and vowel context [F (2,420) = 13.3, p < 0.001].
Results

All participants responded $C_1$ and $C_2$ are “same”

All participants responded $C_1$ and $C_2$ are “different”

Fig. 2: Mean percentage of responses of Korean, Chinese, and English listeners on the Korean obstruents by vowel context type (Error bars: 95% CI)
Results

All participants responded $C_1$ and $C_2$ are “same”

All participants responded $C_1$ and $C_2$ are “different”

Fig.3: Mean percentage of responses of Korean, Chinese, and English listeners on the VOT continuum of Korean obstruents.
Results

- Successful perception of Chinese and English speakers: VOT length of $C_1$ in $C_1V_1C_2V_2$ ranged from 50ms to 95ms in the experiment.

- Consistent with the previous research that voicing perception of Mandarin Chinese was almost identical to that of English listeners in the perception task of English voicing contrast: Both English and Chinese listeners judged the stimulus as voiceless when the VOT is longer than 40ms and all other stimulus as voiced (Bond & Fokes, 1991).
Results

- Further investigation is needed
  - Different stop contrast system in three languages
    - **Chinese** - aspiration only (VOT)  
    - **English** - voicing (VOT, $f_0$)  
    - **Korean** - at least three acoustic cues, i.e., VOT, $f_0$, and $H_1-H_2$ (the relative amplitude of the first versus the second harmonic of the voice source)
  
- VOT for the production of Mandarin Chinese stops /p,t,k/ are significantly longer than their equivalents in English (Chao & Chen, 2008).

- How is the perception of Chinese and English listeners on the “consonant voicing” similar?
Results

- Further investigation is needed
  - Voicing contrast is not phonemic in either Korean or Chinese
    However, Chinese listeners are found to be better at distinguishing voicing than Korean listeners are. Why?

- Possible conjecture
  - If VOT is the only acoustic cue to contrast stops in the native language, the listeners might be better equipped with the perception test varying the VOT length than the listeners of a language whose stop contrast requires other acoustic cues than VOT.
Conclusion

Non-native listeners are more sensitive to the **positional voicing contrast** of Korean obstruents than native listeners are, and that non-phonemic sound contrasts of Korean positional voicing variation might induce the misperception of listeners in their native language.

The language teachers’ awareness of their **potential insensitivity to their native language perception** can help provide a deeper understanding of foreign learners’ different perceptions of the target language and guide them more effectively.