The ability to use contextual cues to achieve phonological constancy emerges by 14 months
Ye Feng1,2,4, René Kager3, Regine Lai1, Patrick C. M. Wong1,2
1. Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China.
2. Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China.
3. Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands.
4. Department of Linguistics, Beijing Language and Culture University, Beijing, China.

Introduction

- The ability to map similar sounding sounds to different meanings alone is far from enough for successful speech processing. To overcome variability in the speech signal, young learners must also adapt to the surface variability and achieve phonological constancy.
- Previous studies have shown that infants at 14 months are able to utilize variations in stimulus-internal cues to form phonological categories and to learn words (Apfelbaum and McMurray, 2011; Höhle et al., 2020; Rost and McMurray, 2009, 2010).
- The present study takes into consideration the fact that talker variability can easily lead to acoustic overlap between categories, in which case reliance on stimulus-external or contextual cues is obligatory for successful talker adaptation.

Experiment 1

Method

- **Stimuli**
  - Cantonese Tone 1 (high level tone) vs Tone 3 (mid level tone) carried by /pi/
  - Non-words: /si/, /ji/, /xi/.
  - 3 tokens x 2 tones x 6 speakers (female native speaker of Cantonese) = 36 tokens

  Figure 1. Pitch contours of the 36 tokens in Exp.1, showing the overlap in tokens.

- **Procedures**
  - Habituation-based visual fixation procedure (VFP)
  - Two word-object pairings, the Switch Task

  Table 2. A demonstration of visual and auditory stimuli used throughout the procedures.

Participant

- Cantonese monolingual families
- No prior history of perceptual or neurological disorders

Results

- When contextual cues, infants from all age groups (14 to 24 months) showed improved robustness for adult-like talker adaptation.
- Infants relied on the phonetic information, rather than referential information, when processing contextual cues to track different speakers’ phonetic spaces (tone spaces in this case) and extract the relative pitch height of the target tone produced by each speaker.

Table 3. Average acoustic measurements and standard deviations (in Hz) of the vowels of the 36 tokens in Exp.3.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Range (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker-matched</td>
<td>24 (12 girls)</td>
<td>393–446</td>
</tr>
<tr>
<td>Speaker-mismatched</td>
<td>24 (12 girls)</td>
<td>513–569</td>
</tr>
</tbody>
</table>

Figure 3. Examples of the final stimuli with sound waves (in black), pitch tracks (in blue), test grid- and glosses (a) and (b) are example stimuli of the speaker-matched condition, where the carrier phrases and the target non-words are from the same speaker (Speaker B in these examples). (c) and (d) are example stimuli of the speaker-unmatched condition, where carrier phrases from Speaker A are mismatched with targets from Speaker B (as in this lower mean F0). In this case, the target T1 sounds more like T3 (c) and the target T3 sounds lower than normal (d). (e) and (f) are example stimuli of the speaker-mismatched condition, where carrier phrases from Speaker B are mismatched with targets from Speaker A who has a higher mean F0. Therefore, the target T1 sounds higher than normal (f) or the target T3 sounds closer to a T1 (f).

Figure 4. Mean fixation times to the visual stimulus for same and novel trials in the test phase in Experiment 3 divided by age group within speaker-matched condition on the right and speaker-mismatched condition on the left.

Experiment 2

Method

- **Stimuli**
  - Target non-words same as those in Exp.1
  - Six carrier phrases (adapted from Singh et al., 2016) used as carrier phrases (contextual cues

  Figure 2. Mean fixation times to the visual stimulus for same- and novel-trials in the test phase in Experiment 1 divided by age group (error bars: SD).

Results

- When given contextual cues, infants from all age groups (14 to 24 months) showed improved robustness for adult-like talker adaptation.
- Infants relied on the phonetic information, rather than referential information, when processing contextual cues to track different speakers’ phonetic spaces (tone spaces in this case) and extract the relative pitch height of the target tone produced by each speaker.

Table 3. Average acoustic measurements and standard deviations (in Hz) of the vowels of the 36 tokens in Exp.3.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Range (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker-matched</td>
<td>24 (12 girls)</td>
<td>393–446</td>
</tr>
<tr>
<td>Speaker-mismatched</td>
<td>24 (12 girls)</td>
<td>513–569</td>
</tr>
</tbody>
</table>

Figure 3. Examples of the final stimuli with sound waves (in black), pitch tracks (in blue), test grid- and glosses (a) and (b) are example stimuli of the speaker-matched condition, where the carrier phrases and the target non-words are from the same speaker (Speaker B in these examples). (c) and (d) are example stimuli of the speaker-unmatched condition, where carrier phrases from Speaker A are mismatched with targets from Speaker B (as in this lower mean F0). In this case, the target T1 sounds more like T3 (c) and the target T3 sounds lower than normal (d). (e) and (f) are example stimuli of the speaker-mismatched condition, where carrier phrases from Speaker B are mismatched with targets from Speaker A who has a higher mean F0. Therefore, the target T1 sounds higher than normal (f) or the target T3 sounds closer to a T1 (f).

Figure 4. Mean fixation times to the visual stimulus for same and novel trials in the test phase in Experiment 2 divided by age group within speaker-matched condition on the left and speaker-mismatched condition on the right (error bars: SD).

Experiment 3

Method

- **Stimuli**
  - Single-speaker versions of experiments with the same stimuli and procedures as Exp. 1 and 2.
  - Tested on the 14 months group

Results

- Results of Experiment 3 replicated previous findings on phonological distinctiveness, showing that even when the stimuli were produced by a single speaker, 14-month-old infants cannot reliably integrate Cantonese TI-T3 contrast into different word meanings unless contextual cues were provided, similar to previous results with segmental contrasts.

Figure 5. Mean fixation times to the visual stimulus for same and novel trials in the test phase in Experiment 3 (a) and Experiment 3b (right) (error bars: SD).

References