Infants’ perception of intonation and its use in early word learning

Joseph Butler, Marina Vigário and Sónia Frota
Universidade de Lisboa
Laboratório de Fonética & Lisbon BabyLab (FLUL/CLUL)
Overview

- Early intonation perception
  - Introduction
  - Aims and research questions
  - Methodology
  - Results
  - Discussion

- Early word learning
  - Introduction
  - Methodology
  - Results
  - Discussion

- Overall summary
Introduction

- Little is known about the developmental course of infants’ perception of linguistic intonation
- Previous studies on pitch contrasts focused on acquisition of lexical pitch
  - Lexical pitch accent, as in Japanese
    - Japanese-learning infants discriminate native pitch accent contrast from as early as 4 months, and maintained during the 1st year (Sato et al., 2009)
  - Lexical tone, as in Mandarin, Cantonese
    - Learners of tone languages show stable discrimination of lexical tone in the 1st year, as early as 4 months (Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013)
Introduction

- Previous studies on lexical pitch focused on pitch direction/height only (e.g. Tones 25 vs. 33, Cantonese – Yeung et al, 2013). However, tones are also distinguished by temporal properties, like the temporal location of the pitch turning point (Jongman et al, 2006).

- Recent research on infant perception suggests pitch and duration differences are processed differently.
  - Infants at 7 months prefer pairs of syllables with variable pitch when pitch height in the first syllable, but no preference when duration differs between syllables (Bion et al., 2011).
  - Human infants and nonhumans group sound sequences into trochaic patterns based on pitch, but do not utilize duration as a cue to group sequences in iambic structures (de la Mora et al., 2012).
Introduction

- Recent research on infant perception suggests pitch and duration differences are processed differently
  - Unlike for lexical pitch, word stress discrimination abilities involve sensitivity gains during the first year of life, dependent on exposure to the native language (Skoruppa et al. 2009, 2013). Stress contrasts typically use to a greater extent phonetic cues other than pitch (namely, duration – Beckman, 1986)

- By and large
  - Pitch discrimination dependent on perceptual abilities
  - Duration dependent on language experience
Introduction

- Intonation is a prosodic dimension that varies across languages and impacts upon meaning.
- Unlike lexical pitch, which is meaningful at the word level, intonation conveys phrasal meanings, like sentence type and pragmatic distinctions (Ladd, 2008; Frota, 2002, in press)
- Intonation languages (e.g., English, Portuguese) may use pitch height, pitch direction and pitch timing (temporal location of pitch turning points) to convey phrasal meanings (statement, question, calling contour, information status, etc).
Aims and Research Questions

- We investigated European Portuguese-learning infants’ perception of two native pitch contrasts:
  - The statement/yes-no question distinction, marked by a pitch direction contrast (rising/falling)
  - The broad/narrow focus distinction, signalled by a pitch timing contrast (early/late fall within the syllable)

- We asked (i) whether the developmental perceptual trajectory of the two contrast types was similar and (ii) how it related to previous reports on pitch perception by infants learning lexical pitch systems
The two pitch contrasts in EP

Statement

Yes-no question

Both contrasts have been shown to be perceived by adult native speakers (Falé & Faria, 2005; Frota, 2012)

Illustrated with one-word utterances/intonational phrases (single prosodic word intonational phrases are the most frequent type of IP in Portuguese)

Methodology

- **Experiment 1** (Statement/Yes-No Question distinction)
  - 40 infants
    - 20 younger (8 female, $M = 5$ months 29 days, range 5 months 3 days – 6 months 23 days)
    - 20 older (10 females, $M = 8$ months 12 days, range 7 months 11 days-9 months 29 days)

- **Experiment 2** (Broad/Narrow Focus distinction)
  - 28 infants
    - 12 younger (6 female, $M = 6$ months 30 days, range 5 months 30 days – 8 months 3 days)
    - 16 older (8 females, $M = 12$ months 8 days, range 10 months 16 days-14 months 6 days)
Stimuli

- **Stimuli:** Contrast Statement/Yes-no question (Exp1), Contrast Broad/Narrow Focus (Exp2) – Child directed speech

**Declarative**

<table>
<thead>
<tr>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>3.5</th>
<th>4</th>
<th>4.5</th>
<th>5</th>
<th>5.5</th>
<th>6</th>
<th>6.5</th>
<th>7.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>H+L%</td>
<td>H+L%</td>
<td>H+L%</td>
<td>H+L%</td>
<td>H+L%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mora</td>
<td>mola</td>
<td>rima</td>
<td>lama</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>morb6</td>
<td>mola</td>
<td>Rima</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Interrogative**

**Direction**

**Alignment of the pitch fall**

Phonetic variation relevant to meaning > conveys important distinctions for processing the input language, for communication and social interaction.
# Acoustic analysis

## Exp 1: Statement/Yes-No Question

<table>
<thead>
<tr>
<th></th>
<th>Statements</th>
<th>Questions</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 range 1&lt;sup&gt;st&lt;/sup&gt; syll (Hz)</td>
<td>67</td>
<td>66</td>
<td>0.12, p = .9</td>
</tr>
<tr>
<td>F0 range 2&lt;sup&gt;nd&lt;/sup&gt; syll (Hz)</td>
<td>-25</td>
<td>192</td>
<td>23.46, p&lt;.001</td>
</tr>
<tr>
<td>Final F0 (Hz)</td>
<td>163</td>
<td>380</td>
<td>23.61, p&lt;.001</td>
</tr>
<tr>
<td>Duration (ms)</td>
<td>529</td>
<td>765</td>
<td>11.91, p&lt;.001</td>
</tr>
</tbody>
</table>

## Exp 2: Broad/Narrow Focus

<table>
<thead>
<tr>
<th></th>
<th>Focus</th>
<th>Neutral</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 peak (Hz)</td>
<td>249.79</td>
<td>230.8</td>
<td>7.4, p &lt; .001</td>
</tr>
<tr>
<td>F0 low (Hz)</td>
<td>160.26</td>
<td>161.53</td>
<td>1.05, p = .31</td>
</tr>
<tr>
<td>Duration (ms)</td>
<td>0.59</td>
<td>0.63</td>
<td>3.29, p &lt; .01</td>
</tr>
<tr>
<td>Timing of the fall (ms)</td>
<td>0.27</td>
<td>0.15</td>
<td>11.15, p &lt; .001</td>
</tr>
</tbody>
</table>
**Procedure**

- Modified version of the visual habituation paradigm (Stager & Werker, 1997)
- Habituated with lists of utterances
  - Half the infants habituated with declarative (Exp1)/broad focus (Exp2)
  - Other half habituated with question (Exp1)/narrow focus (Exp2)
- Habituation continued until pre-defined criteria reached
  - Avg looking time to last 4 habituation trials <60% the avg looking time to the first 4 habituation trials
- Test phase presented lists of different utterances in both declarative/question intonation (Exp1)/broad/narrow focus intonation (Exp2)
Procedure

- Each trial lasted 16 seconds
- Presentation of test stimuli counterbalanced (same/switch trial first)
- Look software was used (Meints & Woodford, 2008)
- Looking times recorded and compared
- If sensitive to the intonational contrast, longer looking times to the switch trials should be evident
Lisbon Babylab Visual Habituation Paradigm
Results

- **Exp 1: Statement/Yes-No Question distinction**

  - Significant difference between same and switch test trials ($F(1,36) = 54.18$, $p < .001$, $\eta^2 = .6$)

  - No effect of age group ($F(1,36) = 2.13$, $p = .15$, $\eta^2 = .06$)

  - No interaction between trial type and age group ($F(1,36) = 3.29$, $p = .08$, $\eta^2 = .08$).

  - Paired T-tests: significant difference between same and switch trials for younger ($t(19) = 6.1$, $p < .001$, $d = 1.474$) and older ($t(19) = 4.42$, $p < .001$, $d = 0.816$) groups.

(Frota et al, submitted)
Results

Exp 2: Broad/Narrow Focus distinction

Preliminary analysis:
- Significant interaction between trial type and age group (F(1,24) = 5.81, p < .05, $\eta^2 = .2$)
- No effect of trial type (F(1,24) < 1).
- No effect of age group (F(1,24) < 1).
- Paired T-tests:
  younger group no significant difference between same and switch trials ($t(11) = 1.35, p = .21$),
  older significant difference ($t(15) = 2.29, p < .05$).
Summary

- Infants learning European Portuguese demonstrate a discrimination ability for the statement/question prosodic contrast as early as 5 months, and maintain this ability throughout the first year.

- However, for the broad/narrow focus contrast, infants only demonstrate discrimination by 12 months.
Discussion

- Our findings suggest that the perceptual trajectory of intonation categories depends on the primary cues involved (pitch direction, pitch timing).

- Supports earlier results that show a protracted development of the perception of timing relative to pitch height/direction (Bion et al., 2011)
Discussion

- Our findings suggest that perception of intonation categories based on a pitch direction contrast may be as precocious as lexical tone/lexical pitch accent perception.

- Pitch-dominant contrasts are those that show a precocious development of discrimination abilities versus stress and other duration/timing-related contrasts that seem more dependent on language experience (Skoruppa et al., 2013; Frota et al., submitted)
Interpretation of phonetic variation in word learning

- Acquisition of phonology requires learning to interpret phonetic variation

- Across languages, prosodic properties may vary in their acoustic correlates and the phonological domains they signal (segments distinguish words):
  - Stress signals word contrasts in Spanish, English or EP but not in French (phrasal level); Pitch signals word contrasts in Chinese or Japanese but not in English or EP (phrasal level)
  - Pitch is an important cue to stress in Spanish (Hualde 2005), but not in European Portuguese

- The task for the young learner: which variation is meaningless?; which variation conveys meaning, at what prosodic level (word level/phrasal level)?
Interpretation of phonetic variation in word learning

- European Portuguese (EP) is an intonation language with lexical stress
- **Stress** is a *word level* property
- **Pitch** is a property of *phrase level* phonology
- We tested whether EP learners were sensitive to stress and pitch contrasts in a word learning task, using an eyegaze-based procedure (and compared this with a vowel contrast).
  - Will young learners notice stress differences and/or intonation differences in ‘new words’?
  - When do young learners interpret phonetic variation at the appropriate levels according to the native language?
Interpretation of phonetic variation in word learning

Eyegaze-based procedure (similar to Quam & Swingley 2010) where visual fixation to the labelled picture is the response variable.
Interpretation of phonetic variation in word learning

- **Auditory stimuli: Phonetic properties**
  - **Stress contrast:** penult / final [\textit{milu}] / [\textit{milu}]
    - Stress location cued by relative duration (stressed > 58-132 ms) and by the alignment of the pitch fall (through the stressed syllable)
  - **Pitch contrast:** declarative / interrogative (H+L* L\% / H+L* LH\%)
    - Intonation contrast cued by the low versus rising boundary and by the longer duration of the final syllable in interrogatives (117-165 ms)
  - **Vowel contrast:** change in height and backness [\textit{milu}] / [\textit{malu}]
Interpretation of phonetic variation in word learning

- 146 children between 1;0 and 4;9 were tested, all from monolingual EP homes. 72 children included in the analysis (learned the trained word: >50% fixation to the labelled picture; matching numbers across conditions)
Interpretation of phonetic variation in word learning

Younger: 1-year olds and 2-year olds

Older: 3-year olds and 4-year olds

Prosodic changes: stress (SC) intonation (IC)

Vowel change
Interpretation of phonetic variation in word learning

Our results show that

- Pitch contour variation is regarded as relevant in new words by 1-year olds and 2-year olds, at odds with native language phonology.
- 3-year olds already regard pitch variation as NOT relevant: Stress variation is regarded as meaningful, both by the younger and older age groups (similar to segmental variation)
- The effect of segmental variation is stronger than stress variation in the older group (mutual exclusivity)
- Only at 3;0 do young learners interpret phonetic variation at the appropriate levels according to the native language
- Pitch is a very salient property in more demanding tasks earlier in development
Overall summary

- The development of the perception of different intonational categories varies dependent on the primary cues involved.

- Pitch contour variation is initially regarded as relevant in early word learning, and is only interpreted in line with native phonology later in development.
Thanks to Cátia Severino and Susana Correia

And especially to all the families and nurseries that have taken part in these studies.

Obrigado
Thank you

DEPE: PTDC/CLE-LIN/108722/2008
References

- INTERSPEECH, 1705-1708.