Prosodic encoding of information structure in Mandarin and English: Combining insights from phonetics and psycholinguistics

Iris Chuoying Ouyang
University of Southern California
Aug 14, 2015, National University of Singapore
My research interests, work I’ve done across disciplines and languages

- Phonetics, psycholinguistics, sociolinguistics, using corpora or experiments

- Collaborated on sentence processing (e.g. reference resolution, causality) using eye-tracking techniques

- Also interested in language acquisition, bilingualism

- Dialectology, language contact, historical sound change

- Fieldwork on Chinese (Hakka, Hui, Wu) and Austronesian (Atayal, Bunun, Saisiyat) languages
Information structure in communication/interaction

• There are many aspects of language, e.g. social/cultural, physiological, *communicative*

• We say/write things to be heard/read (by others or ourselves)

• *Information structure is a discourse process*
  – What’s new
  – What’s old/given or assumed
  – What’s contrastive or corrective
Correction to prior discourse
Disagreement with interlocutors

• Prevails in communication
Information structure can be conveyed in many ways

- **Words**
  - ‘yes/no’
  - Utterance-final particles in Chinese, e.g. ‘a 啊’, ‘o 喔’
- **Syntax**, e.g. cleft, word order
- **Prosody** (tone of voice) – intonation and rhythm

A: *Who got the mangos?* B: *Mariana got the mangos.* [New]
A: *Peter got the mangos.* B: *No, MARIANA got the mangos.* [Corrective]
Prosody – *glue* of speech

- Occurs at every level of language
  - Word/phrase structure
    - WHITE house vs. white HOUSE
    - she2 tou0 舌头 ‘tongue’ vs. she2 tou2 蛇头 ‘snake head’
  - Sentence type, e.g. question vs. statement
  - Discourse function, e.g. information structure

- Not limited to spoken language
  - Poetry
  - Implicit prosody
Science knows little about prosody

• Earliest for children
• Extremely difficult for L2 learners
• Challenge to speech technology

• A complex creature
  – pitch/f0, loudness/intensity, length/duration, quality of segments (i.e. vowels and consonants)
Even more unanswered questions with tone languages

Prosody

Post-lexical

Lexical (tone)

English

Chinese
Pitch is essential to tones

Mariana made marmalade
MARIANA made marmalade

ma 妈  ‘mother’
ma 麻  ‘hemp’
ma 马  ‘horse’
ma 骂  ‘scold’

4 tones in Mandarin

English ToBI (tones and break indices) labeling examples
Loudness and length are also used for tones

Loudness is correlated with pitch (Whalen & Xu 1992)

9 tones in Cantonese

si 诗  si 史  si 试  si 时  si 市  si 是  sik 色  sek 石  sik 食

poem  history  try  time  market  yes  color  rock  eat
• **Exp 1:** Prosody affects syllable perception & word recognition in Mandarin (Ouyang & Iskarous, 2012; Ouyang 2013)

• **Exp 2:** But prosody also signals corrective vs. new in Mandarin – how? (Ouyang & Kaiser 2011, 2015)

• **Exp 3:** But we don’t always prosodically emphasize correction – when do we do that? (Ouyang & Kaiser 2014, in press, in prep)
Exp 1: Just how important are prosodic cues in Mandarin comprehension?
Exp 1: Does the prosodic properties of an utterance affect syllabification?

- Why syllable?
  - Fundamental unit; has not been investigated in this regard
  - Special status of syllable in Chinese
    - syllable-timed (Lin & Wang 2007), one syllable = one morpheme
      (DeFrancis 1984), one syllable = one tone (Xu 1998; Gao 2008)

- Which prosodic cues to test?
  - Pitch/f0 contour, because its association with tone category
  - Speech rate, following prior work on English

- How do we manipulate tonal content without compromising segmental content?
Mandarin tonal f0 contours

Tone1+Tone3 and Tone3+Tone1 form f0 patterns that are, shape-wise, similar to Tone4 and Tone2.

Xu (1997)
Exp 1

• **F0 contour in the target word**: 6 steps, on a continuum from a single contour tone to a combination of two level tones (i.e. R-LH or F-HL)

• **Speech rate of the carrier sentence**: slow, medium, fast, and none (i.e. target words in isolation)

Research question: Do local F0 contour and global speech rate influence the perceptual syllable count of a word in Mandarin?
Target words

- Segmental properties – **bisyllabic, original, kept constant,** vowel \([u]\) at both sides of the syllable boundary
- F0 contours – in different degrees of ‘single-tone-ness’

<table>
<thead>
<tr>
<th></th>
<th>High + Low</th>
<th>Low + High</th>
</tr>
</thead>
<tbody>
<tr>
<td>u.u</td>
<td>tʂʰu.u 初舞 (tʂʰu 触)</td>
<td>tsu.u 祖屋 (tsu 足)</td>
</tr>
<tr>
<td>Vu.u</td>
<td>tʂou.u 周五 (tʂou 皱)</td>
<td>pau.u 宝坞 (pau 薄)</td>
</tr>
<tr>
<td>u.uV</td>
<td>ku.ua 孤瓦 (kua 卦)</td>
<td>hu.uo 虎窝 (huo 活)</td>
</tr>
</tbody>
</table>

F0 contours generated using the qTA model (Prom-on, Xu & Thipakorn 2009)
Carrier sentences

• Naturally produced with a placeholder word; recorded separately for LH and HL conditions

• Average syllable length of carrier sentences:
  o Fast rate – 131 ms/syllable
  o Medium rate – 259.5 ms/syllable
  o Slow rate – 441.5 ms/syllable

• Duration of target words is held constant (at 426-556 ms/syllable)
Participant’s task: Forced-choice identification

Before the main experiment, participants had learned the names of the pictures during a familiarization phase.
• Stimuli were originally recorded by a male speaker of Beijing Mandarin

• Participants: 27 native speakers of Mandarin who were born in China and left China no earlier than the age of 22

• 6 target words * 6 levels of f0 contours * 4 levels of speech rate = 144 items

• Each item occurred 4 times
What influences syllable count:
Local f0 contour?
Global speech rate?
Neither (segmental cues only)?

tʂou [H]. u [L] 周五
tʂou [F] 皱
If segmental information dominates syllabification...

If local f0 contours affect syllabification...

If global speech rate affects syllabification...

If f0 contours and speech rate both affect syllabification...

y-axis: % monosyllabic responses
x-axis: f0 contour
lines: speech rate
Results: both local f0 contours and global speech rate matter

• Despite the presence of intact, bisyllabic segmental cues, monosyllabic-word responses increased when the f0 contours were more single-tone-like and when the carrier sentences were slower

• Mixed-effects statistical models were conducted
Prosodic cues interfere with segmental cues in syllable perception

- Note again: The segmental content of a target word was taken from a naturally-produced, bisyllabic token.
- Monosyllabic response = The two [u] vowels across the syllable boundary were considered as one.
Take home messages from Exp 1

• Segments don’t determine syllables
• Prosodic cues (be it local or global) interact with segmental cues, affecting syllabification
  – In Mandarin, this may substantially impact word recognition
• Cue integration as a fundamental property of the language comprehension system
Exp 2: Ok, if pitch patterns are so crucial to Mandarin word recognition, how do they also convey discourse information?
Exp 2 research questions

• **Q#1**: Prosodic differences between information types?
  – corrective vs. new vs. given

• **Q#2**: What cues mark differences between information types?
  – Lengthening?
  – F0 range (max-min) expansion?
  – Intensity range (max-min) expansion?
Prior work on Mandarin

- Little work on **corrective vs. new vs. given** (Chen & Braun 2006, Greif 2010)

  - **Mixed results** on prosodic encoding of information structure in general. Certain information types were reflected in:
    - only duration (Chen & Gussenhoven 2008, Grief 2010), only f0 (Wang & Xu 2006), or both (Jin 1996, Chen 2006, Chen & Braun 2006)

- Little work on **intensity** (Jin 1996, Chen et al 2009)
  - no work on intensity ranges

- Although most researchers agree that **pitch ranges**, instead of **contours**, convey information structure
Production task

• Participants provided spoken instruction to move an object to a location, based on pictures and arrows shown on the screen, e.g. Move the frog next to the jellyfish 把青蛙放到水母旁边

• Computer “responded” to participants’ instructions by moving objects on the screen

• Participants were encouraged to believe they were interacting with another person in another room
Production task
Target nouns were bisyllabic with three kinds of tonal contours
- High-High (香烟, 乌鸦, 青蛙)
- High-Low (蚯蚓, 鹦鹉, 斑马)
- Low-High (鬼屋, 雨衣, 海鸥)

9 target nouns, 12 filler nouns
18 target trials, 36 filler trials
Word frequency was controlled (15.23/million words according to Cai & Brysbaert 2010)
Participants: 5 women and 5 men from Beijing
Correction of **Given** Target

**Move juice next to FROG**

(Correct moving)

**Move FROG next to jellyfish**

[Given]

(Wrong object moves to jellyfish)

**Move FROG next to jellyfish**

[Corrective]

(Correct moving)
**Correction of New Target**

<p>| Move juice next to <strong>barrel</strong> | (Correct moving) |
| Move FROG next to <strong>jellyfish</strong> [New] | (Wrong object moves to jellyfish) |
| Move FROG next to <strong>jellyfish</strong> [Corrective] | (Correct moving) |</p>
<table>
<thead>
<tr>
<th>Correction of <strong>New</strong> Target</th>
<th>Correction of <strong>Given</strong> Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Move juice next to \textit{barrel}}</td>
<td>\textit{Move juice next to \textit{FROG}}</td>
</tr>
<tr>
<td>\textit{(Correct moving)}</td>
<td></td>
</tr>
<tr>
<td>\textit{Move \textit{FROG} next to \textit{jellyfish}} \textit{[New]}</td>
<td>\textit{Move \textit{FROG} next to \textit{jellyfish}} \textit{[Given]}</td>
</tr>
<tr>
<td>\textit{(Wrong object moves to jellyfish)}</td>
<td></td>
</tr>
<tr>
<td>\textit{Move \textit{FROG} next to \textit{jellyfish}} \textit{[Corrective]}</td>
<td>\textit{Move \textit{FROG} next to \textit{jellyfish}} \textit{[Corrective]}</td>
</tr>
<tr>
<td>\textit{(Correct moving)}</td>
<td></td>
</tr>
</tbody>
</table>
Results: Effect of correction

Presence/absence of correction was reflected in all three acoustic dimensions: Corrective words had longer durations, larger F0 and intensity ranges than Non-Corrective words (p’s < .05)
Results: **Effect of newness**

New/given distinction was reflected in duration and F0, but not in intensity: New info had **longer duration and larger F0 ranges** than Given info ($p$’s < .05); intensity ranges did not differ significantly.
Corrective vs. New vs. Given

<table>
<thead>
<tr>
<th></th>
<th>Corrective vs. Non-Corrective</th>
<th>(Non-Corrective) New vs. (Non-corrective) Given</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>F0 range</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Intensity range</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
How are f0 and intensity range expansion achieved?

- **Max f0 raising & Min f0 lowering** for correctiveness
- **Max f0 raising** for newness
  - Similar behavior in coarticulation of lexical tones for aerodynamic reasons? (Xu 1993, 1997)
- **Min intensity lowering** for correctiveness
  - Not seen in tonal phenomena. Socially motivated?
Take home messages from Exp 2

- **Multi-functional acoustic dimensions**
  - Even in a tone language, all three acoustic dimensions (F0, duration, intensity) also encode information structure

- **Specialization of the parameters of encoding**
  - *ranges of f0 and intensity* mark the information structure of discourse
  - the *shapes/contours* carry word-level information about *tones* (Whalen & Xu 1992; Xu 1997)
Exp 3: Do we always prosodically emphasize correction?
Exp 3 research question

Is corrective prosody indeed affected by the speaker’s consideration for the addressee?

– Exp 2 results about min intensity lowering seems to suggest a social factor

– In corrective information structure, an important element directly associated with the addressee is the mistake/misstatement

– How do we manipulate about the addressee in a way that the results can show the speaker cares?
Elisa likes her salad a certain way. She loves vegetables but hates fruit.

She went grocery shopping yesterday evening. I heard that she got some **lettuce** at the farmer’s market.

**No**, she got some **spinach** at the farmer’s market.

**Speaker A**

**Speaker B**

Plausible statement (lettuce)
Corrective response
**Speaker A**

What did Elisa get?

Spinach.

---

**Speaker B**

What did people think that Elisa get?

Lettuce.
She went grocery shopping yesterday evening. I heard that she got some **lettuce** at the farmer’s market.

Elisa likes her salad a certain way. She loves vegetables but hates fruit.

Yes, she got some lettuce at the farmer’s market.

**Plausible statement (lettuce)**

**Non-corrective response**
Collaborative read-aloud task
Exp 3A: Shared Knowledge Version

**Speaker A**

She went grocery shopping yesterday evening. I heard that she got some **apples** at the farmer’s market.

**Speaker B**

*Elisa likes her salad a certain way. She loves vegetables but hates fruit.*

No, she got some **cherries** at the farmer’s market.

**Implausible statement (apples)**

**Corrective response**
She went grocery shopping yesterday evening. I heard that she got some apples at the farmer’s market.

Elisa likes her salad a certain way. She loves vegetables but hates fruit.

Yes, she got some apples at the farmer’s market.

Implausible statement (apples)
Non-corrective response
192 target items, using 6 scenarios and 24 objects
26 participants in total; each person completed 13-47 target items
Norming study was used to establish contextual probability.
Contextual probability of Speaker B’s response was controlled.

<table>
<thead>
<tr>
<th>A states</th>
<th>Implausible A Corrective B</th>
<th>Implausible A Non-Corr B</th>
<th>Plausible A Corrective B</th>
<th>Plausible A Non-Corr B</th>
</tr>
</thead>
<tbody>
<tr>
<td>I heard ‘apples’</td>
<td>I heard ‘apples’</td>
<td>I heard ‘lettuce’</td>
<td>I heard ‘lettuce’</td>
<td>I heard ‘lettuce’</td>
</tr>
<tr>
<td>No, ‘cherries’</td>
<td>Yes, ‘apples’</td>
<td>No, ‘spinach’</td>
<td>Yes, ‘lettuce’</td>
<td></td>
</tr>
</tbody>
</table>

• In another version (Exp 3B: Privileged Knowledge), only Speaker B sees the crucial context ‘Elisa loves vegetables but hates fruit.’ Speaker A doesn’t know the plausibility.
• F0 ranges of target words in the responses were measured
Exp 1
Tone (Lexical)

Exp 2
Information Structure (Discourse)

Prosodic realization of an utterance

Exp 3A&3B
Contextual plausibility (semantic)
Addressee’s knowledge (Interpersonal)
Corrective prosody depends on what the other person knows & says

(Elisa loves vegetables but hates fruit)

Exp 3A
Shared Knowledge

Exp 3B
Privileged Knowledge

I heard... apples lettuce

No, she...
Yes, she...

<table>
<thead>
<tr>
<th>Plausible</th>
<th>Implausible</th>
</tr>
</thead>
<tbody>
<tr>
<td>n.s.</td>
<td>*</td>
</tr>
<tr>
<td>corr</td>
<td>corr</td>
</tr>
<tr>
<td>non-corr</td>
<td>non-corr</td>
</tr>
</tbody>
</table>
Prosodically prominent correction occurs when Speaker is surprised by Addressee

In Exp 3A (**addressee knows what is likely**):

- The **implausible** condition created high surprisal in Speaker B’s response, because Speaker A shouldn’t have made an implausible statement
  - (“You should’ve known better!”)

In Exp 3B (**addressee doesn’t know what is likely**):

- The **plausible** condition created high surprisal in Speaker B’s response, because Speaker A made a plausible statement without knowing the context
  - (“How on earth did you know that?”)
Speaker takes Addressee into account, and in a sophisticated way

- It’s not the case that:
  - Correction always receive prosodic emphasis
  - Contextually implausible misstatements always surprise the responder

- A misstatement elicits prosodic prominence in the correction only when the misstatement is inconsistent with what the interlocutor is expected to know (or not to know)
Take home messages from Exp 3

• New insights gained by combining information structure, contextual plausibility and addressee’s knowledge state.
  – To understand the prosodic encoding of information structure, we need to also consider other factors.

• Interaction-based explanation:
  – The prosodic prominence associated with corrective information reflects the gap between what the speaker had expected the addressee to know and what the addressee appeared to know
Road map revisit, conclusions

• **Exp 1:** Prosody (f0 contour and speech rate) affect syllable perception & word recognition in Mandarin

• **Exp 2:** Nevertheless, prosody (f0 ranges, intensity ranges, and duration) also signal information structure (corrective vs. new vs. given) in Mandarin

• **Exp 3:** From a communicative perspective, the level of prominence reflects the extent to which speakers are surprised by their interlocutors (in English)
Implications (1)
Complexity of human language system

Acoustic dimensions

- Pitch
- Loudness
- Duration

Linguistic functions

- Tone of word
- New information
- Corrective information

(Based on prior research and mine)
Implications (1)
Complexity of human language system

Acoustic dimensions

- Pitch contours
- Pitch ranges
- Loudness contours
- Loudness ranges
- Duration

Linguistic functions

- Tone of word
- New information
- Corrective information

(English)

(Based on prior research and mine)
Implications (1)
Complexity of human language system

Acoustic dimensions
- Pitch contours
- Pitch ranges
- Loudness contours
- Loudness ranges
- Duration

Linguistic functions
- Tone of word
- New information
- Corrective information

Mandarin

(Based on prior research and mine)
Implications (2)
Importance of context in linguistic research

- Advancing our understanding of linguistic elements by studying them in discourse, dialogues, and interpersonal communication.
Future directions

• **Perception** of prosodic cues to information structure
  – Would *pitch* be less important than *duration & loudness* in the comprehension of *information structure* in Chinese? What about English?

• Would **non-linguistic** unexpectedness elicit prosodic prominence too (i.e. a cross-modular process)?
  – Visual, e.g. seeing something surprising while talking

• Testing the interaction-based account in other cultures

• What do bilinguals do? What makes second-language prosody sound non-native?
Thank you
谢谢
terima kasih
நன்றி

chuoyino@usc.edu
www.irisouyang.com

Comments and questions are welcome 😊
Thanks also go to:

• Emily Fedele and Josephine Lim for help with experiment administration
• Sasha Spala for work on data analysis
• Dr. Khalil Iskarous and Dr. Elsi Kaiser for comments and advice
• This project is funded by an NSF Doctoral Dissertation Research Improvement Grant