Infant-directed speech as a window into the dynamic nature of phonology

Reiko Mazuka,[1,4] Andrew Martin,[1]
Yosuke Igarashi,[2] & Akira Utsugi[3]

1.RIKEN Brain Science Institute, 2. Hiroshima University,
3. Nagoya University, 4. Duke University

Theme of LabPhone14

Laboratory Phonology beyond the laboratory:

Quantitative analyses of speech produced outside the phonetics laboratory

- Fieldwork-based studies
- Corpus-based approaches
- Acquisition of L1 phonology/prosody

Background

- Real speech occurs in dynamic contexts.
- Speakers adjust their speech dynamically depending on the contexts.
- Phonology needs to account for dynamic aspects of human speech as well.
- Systematic analysis of specialized speech register can offer a window into the dynamic aspects of speech.

Outline

1. RIKEN Japanese mother-infant conversation corpus (R-JMICC)
2. Exaggeration of intonation (Igarashi et al., JASA, 2013)
3. Realization of phonological rule (Martin, et al., Cognition, 2014)

Input for Learning Japanese

- RIKEN-Japanese Mother-Infant Conversation Corpus

  (Mazuka, et al 2006; Igarashi & Mazuka, 2006)

- Participants
  - 22 mothers
    - with their 18-24 month-old infants (12 females, 10 males)
    - From Tokyo area

- Size

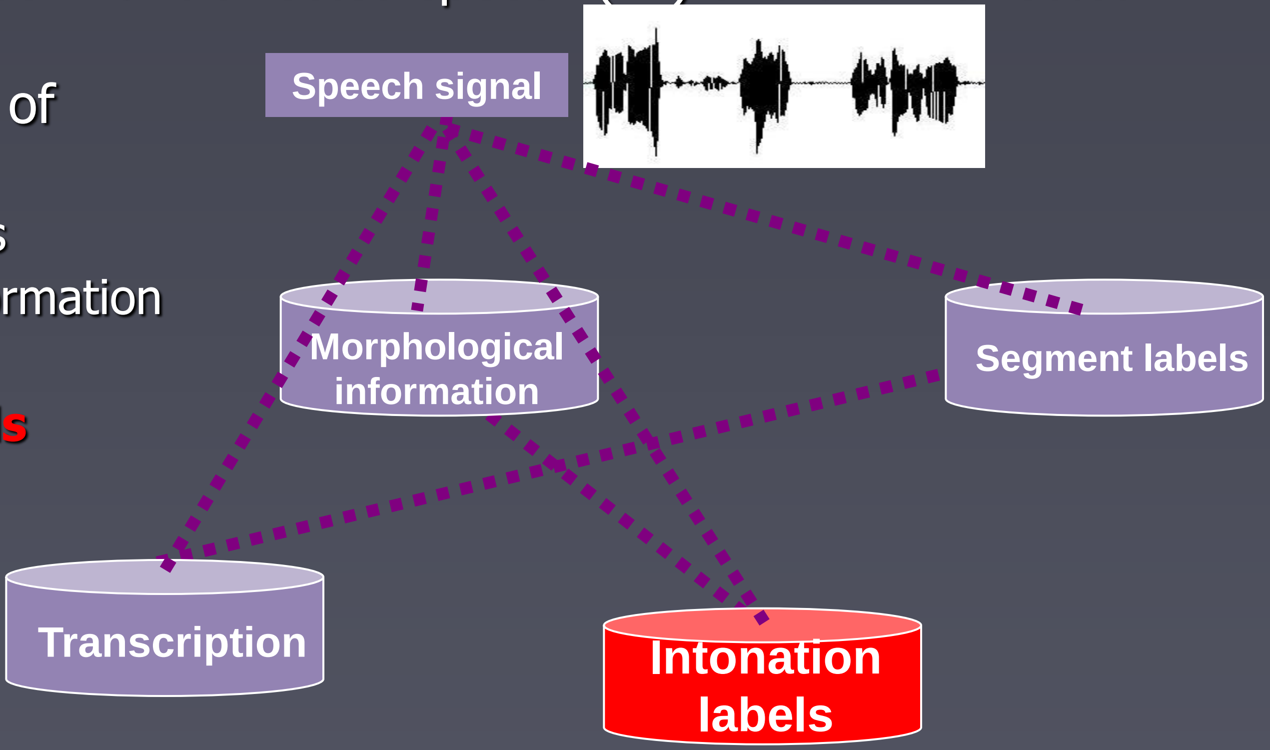
ID	11 hours	50,000 words
AD	3 hours	30,000 words
Overall	14 hours	80,000 words

Tasks (recording environments)

Adult-directed speech	Conversation	Talking with a female experimenter **10 min**	
Infant-directed speech	Book	Playing with picture books **15 min**	
	Toy	Playing with toys **15 min**	

RIKEN Mother-Infant Conversation Corpus

- RIKEN Japanese Mother-Infant Conversation Corpus is a speech database of Japanese infant-directed speech (ID) and adult-directed speech (AD).
- The corpus consists of
  - Speech signals
  - Transcription texts
  - Morphological information
  - Segmental labels
  - **Intonation labels**



Speech signal

Morphological information

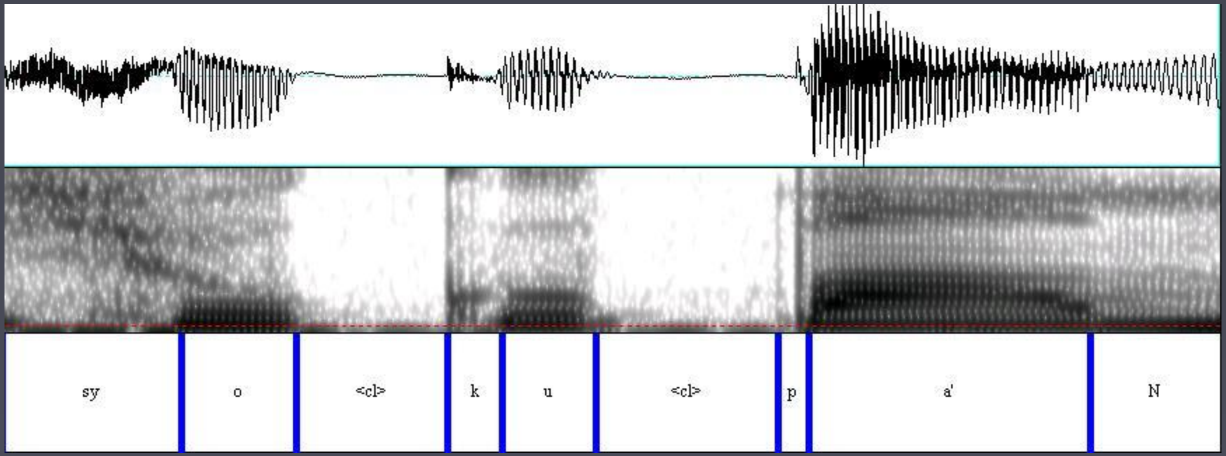
Segment labels

Transcription

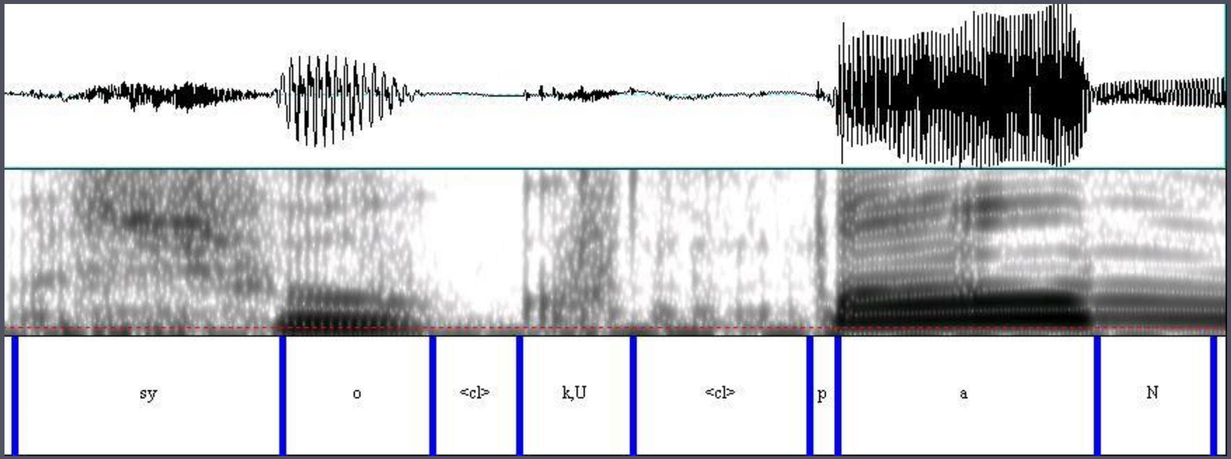
Intonation labels

- The intonation labeling is based on the **X-JToBI** scheme (Maekawa et al. 2002, cf. Venditti 2006)
  - It owes its theoretical foundation to the phonological model of Japanese intonation (Pierrehumbert and Beckman 1988).

Segmental labels

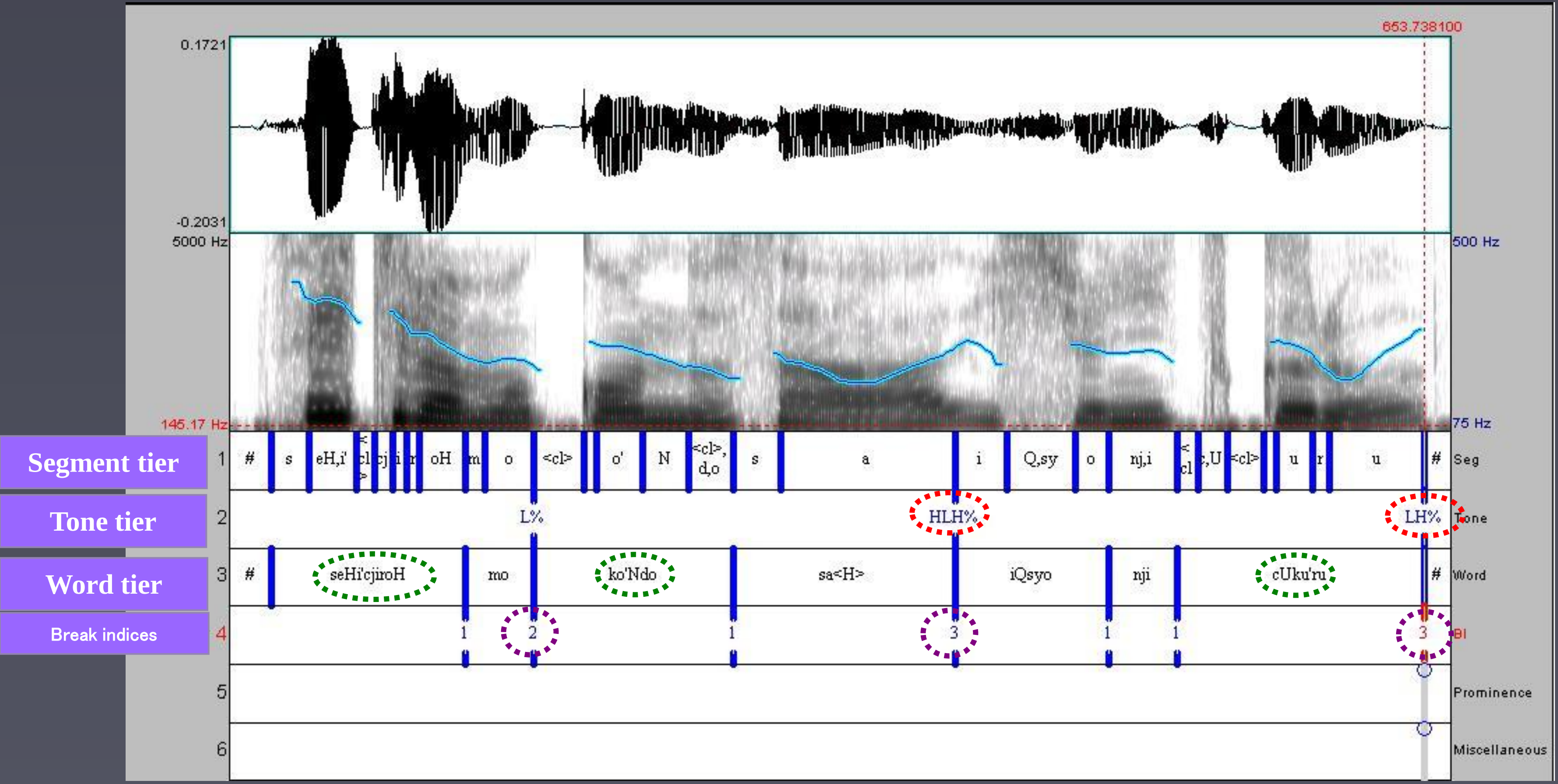


‘bread’ without devoicing
sy o <cl> k **u** <cl> p a N



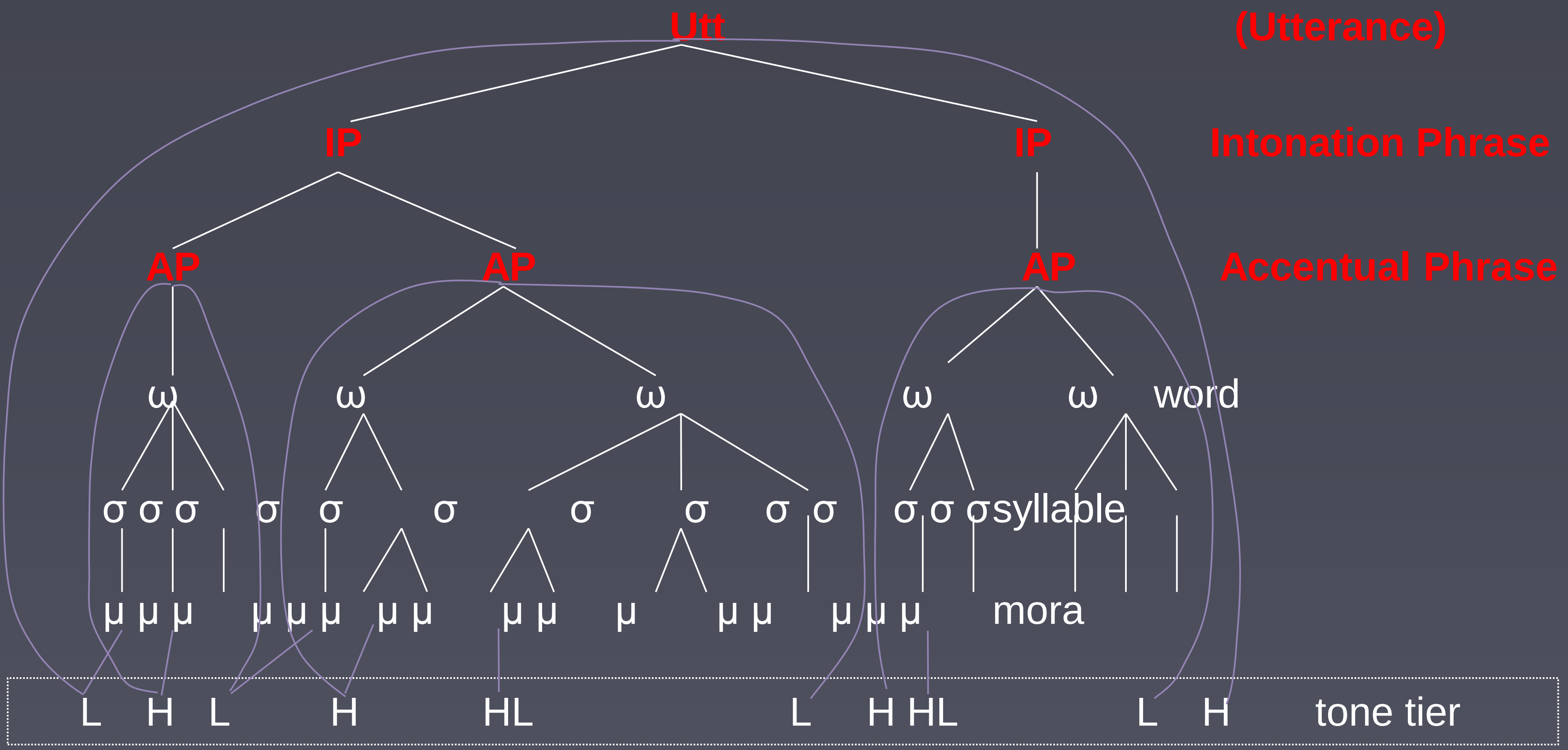
‘bread’ with devoicing
sy o <cl> k **U** <cl> p a N

X-JToBI Coding



「成一郎も今度さあ　一緒に作る？」

Next time, are you going to make it together, Seiichiro?



Utt **(Utterance)**

IP **IP** **Intonation Phrase**

AP **AP** **AP** **Accentual Phrase**

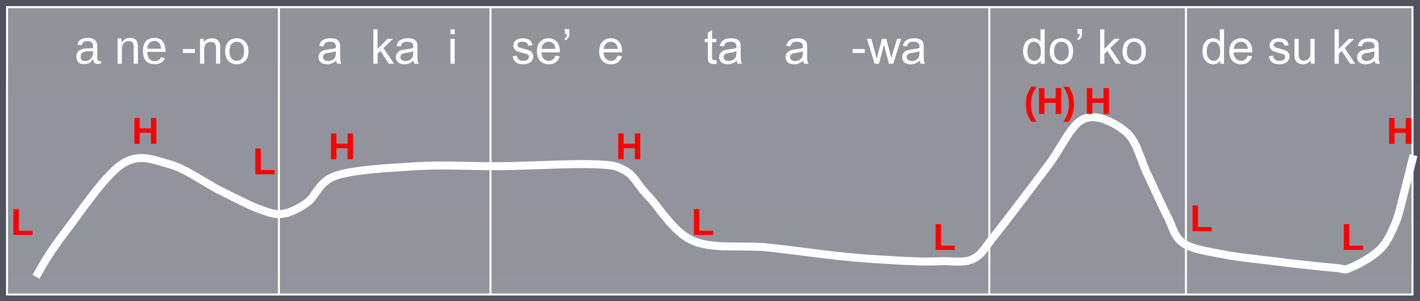
ω ω ω ω ω word

σ σ σ σ σ σ σ σ σ σ σ σ σ syllable

μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ μ mora

L H L H HL L H HL L H tone tier

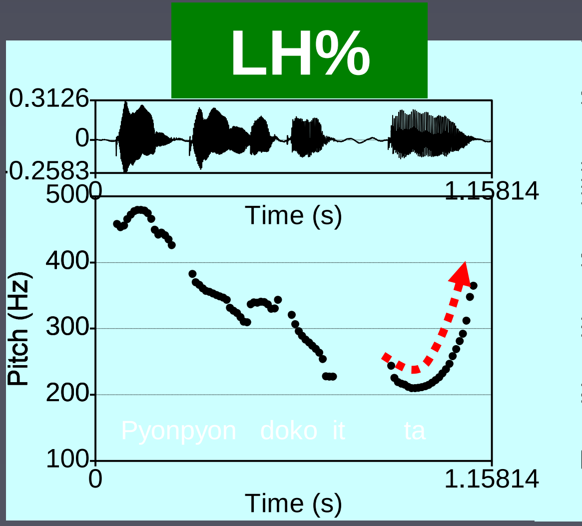
a ne -no a ka i se’ e ta a -wa do’ ko de su ka phoneme tier



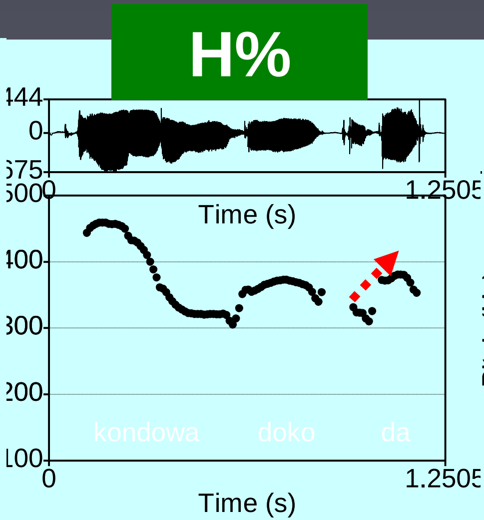
‘Where is big sister’s red sweater?’

What is Boundary Pitch Movement (BPM)?

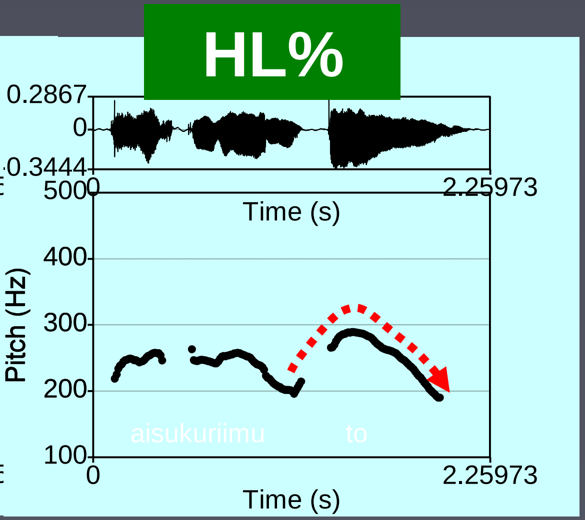
- BPMs are tonal categories that occur **in the final mora of a prosodic phrase** (typically utterance).
- BPMs contribute to the **pragmatic interpretation** of the utterance, such as questioning and continuation (see Venditti et al. forthcoming).
- Main types of BPMs:



"Where's the frog gone?"



"Where is it now?"



"An ice cream and…"

Outline

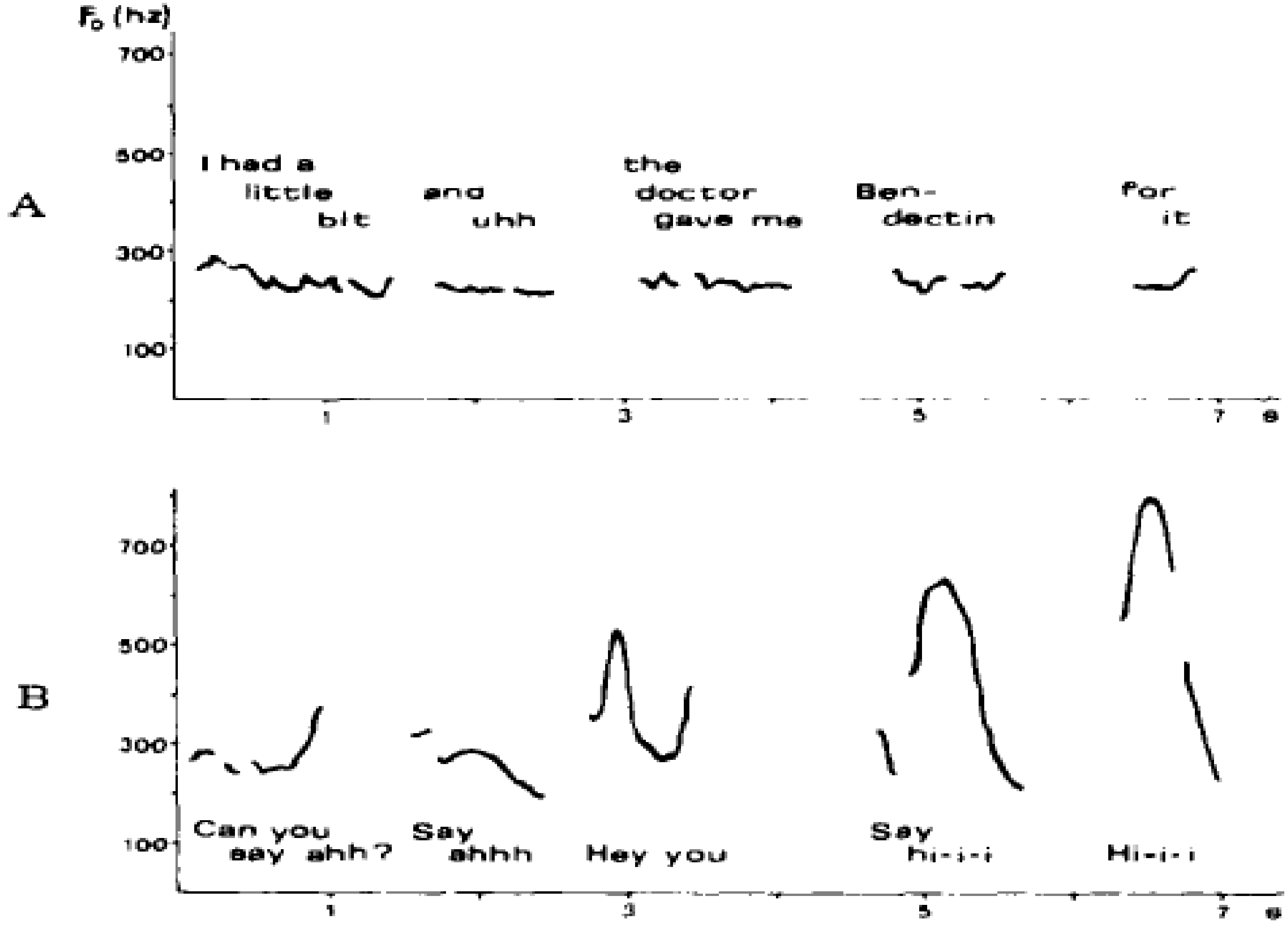
1. RIKEN Japanese mother-infant conversation corpus (R-JMICC)
2. Exaggeration of intonation (Igarashi et al., JASA, 2013)
3. Realization of phonological rule (Martin, et al., Cognition, 2014)

Exaggerated Intonation in IDS?

Igarashi, Nishikawa, Tanaka, & Mazuka, JASA, 2013

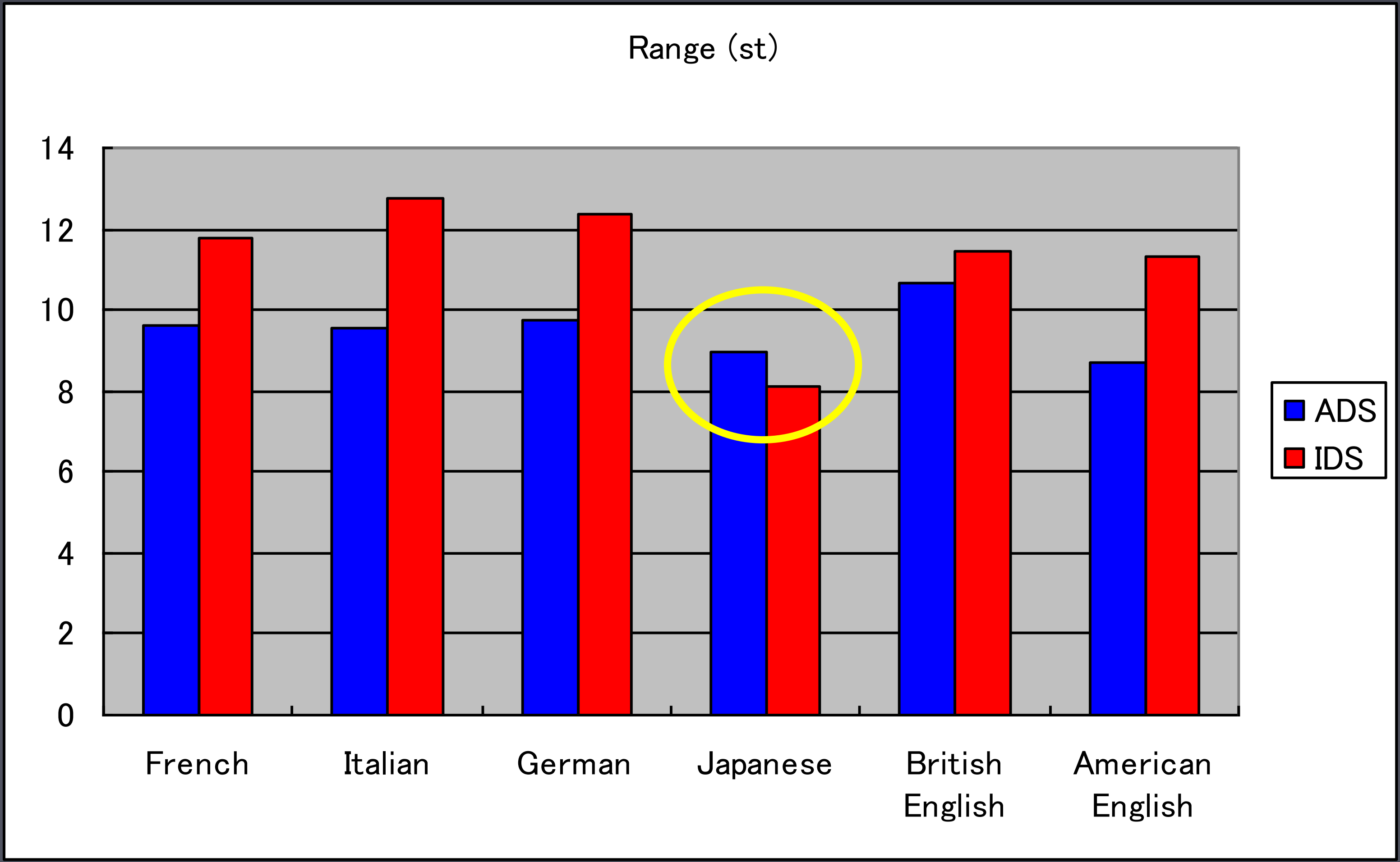
- '*Exaggerated*' *intonation* is one of the most often sited characteristics of IDS prosody (e.g., Fernald et al. 1989).

Pitch Exaggeration in IDS



Fernald (1985)

Fernald et al (1989) found Japanese IDS showed no pitch expansion.

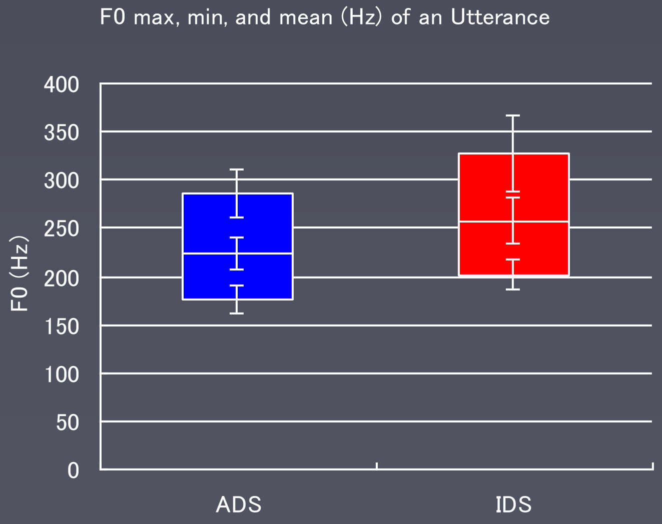


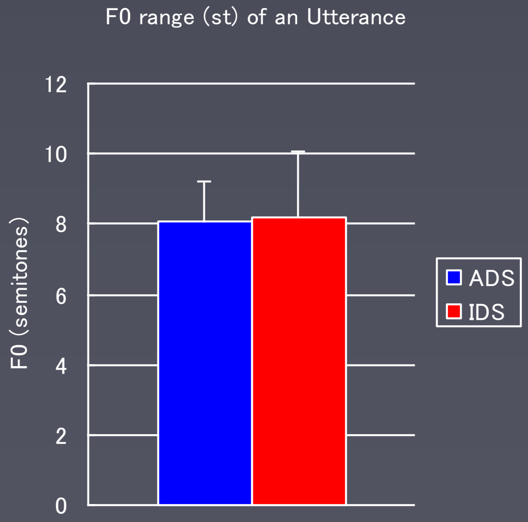
Exaggerated Intonation in IDS?

- *'Exaggerated' intonation* is one of the most often sited characteristics of IDS (e.g., Fernald et al. 1989).
- Based on *physical measurements* of overall fundamental frequency (F0) contours; e.g., 'expanded pitch range', 'higher pitch level.'
- No reference to the *linguistic structure* of a language's intonation.
- No exaggeration found in Japanese IDS.

Analysis of R-JMICC Utterance (Overall)

- When we examine the whole utterance
  - Max, min, mean: AD < ID
  - Range (semi tone): AD = ID





Numerial replication of Fernald et al (1989)

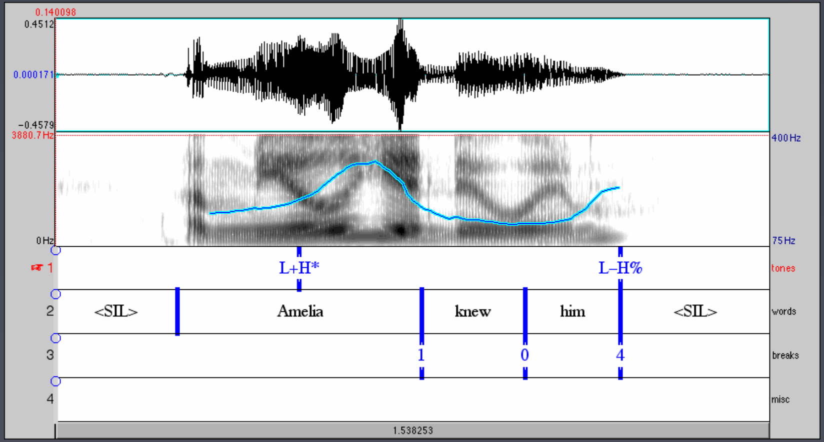
Why?

- Intonation of a language is NOT merely a physical parameter.
- Intonation has a linguistically organized internal structure.
  (e.g. Pierrehumbert 1980; Ladd 1996)
- Examining IDS intonation with reference to the prosodic system of Japanese may reveal language specific nature of Japanese IDS intonation.

Prosodic system of English

- A stress, which is lexically specified, involves larger intensity and longer duration.
- Two major components of intonation
  - Pitch accent: appears on a stressed syllable
  - Boundary tone : appears at a phrase edge

A sample taken from: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm



Ex.) Amelia knew him.

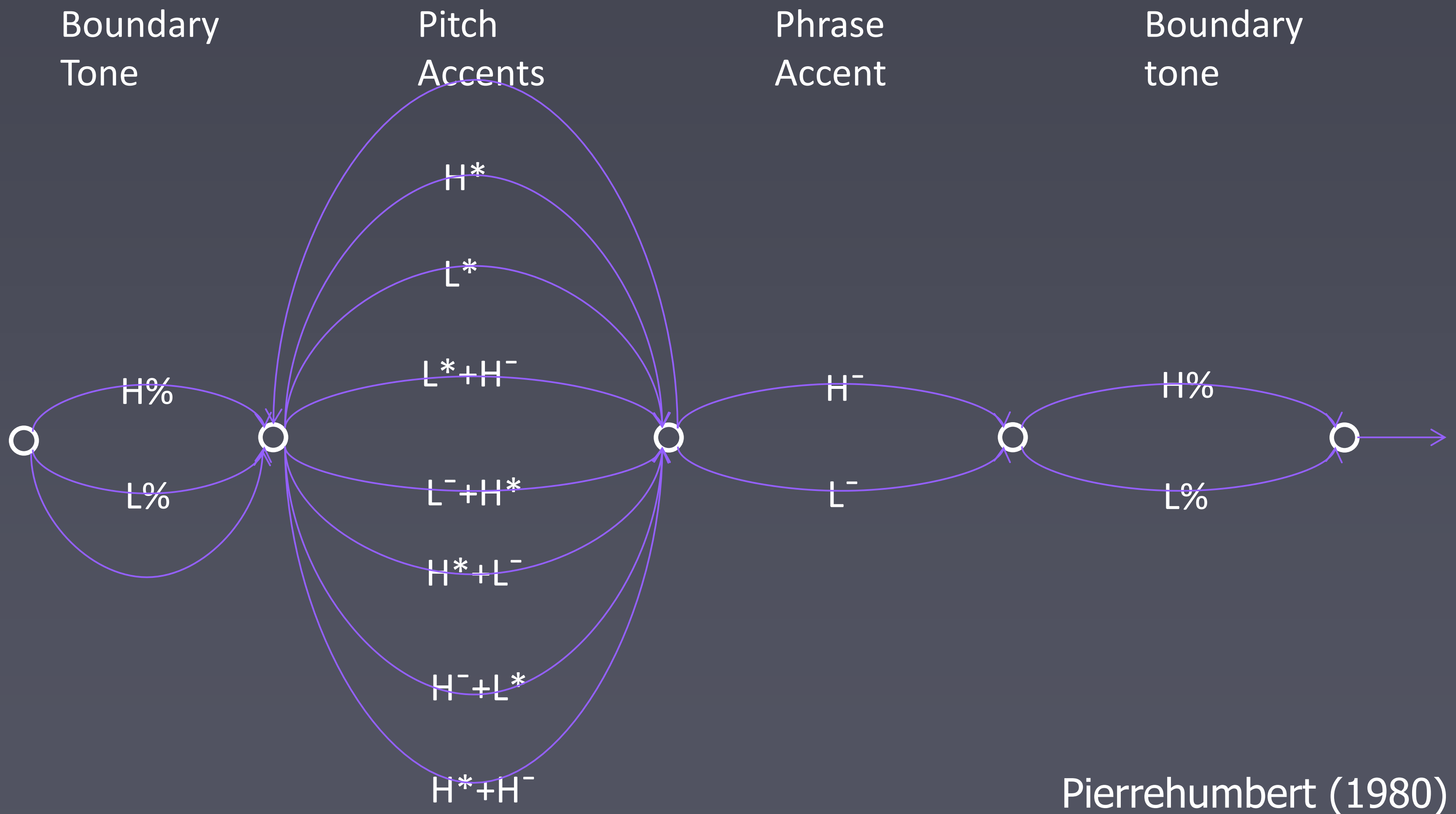
Stress ● ●

Intonation L+H* L-H%

Pitch Accent

Boundary Tone

Finite State Grammar to generate all tunes in English



Pierrehumbert (1980)

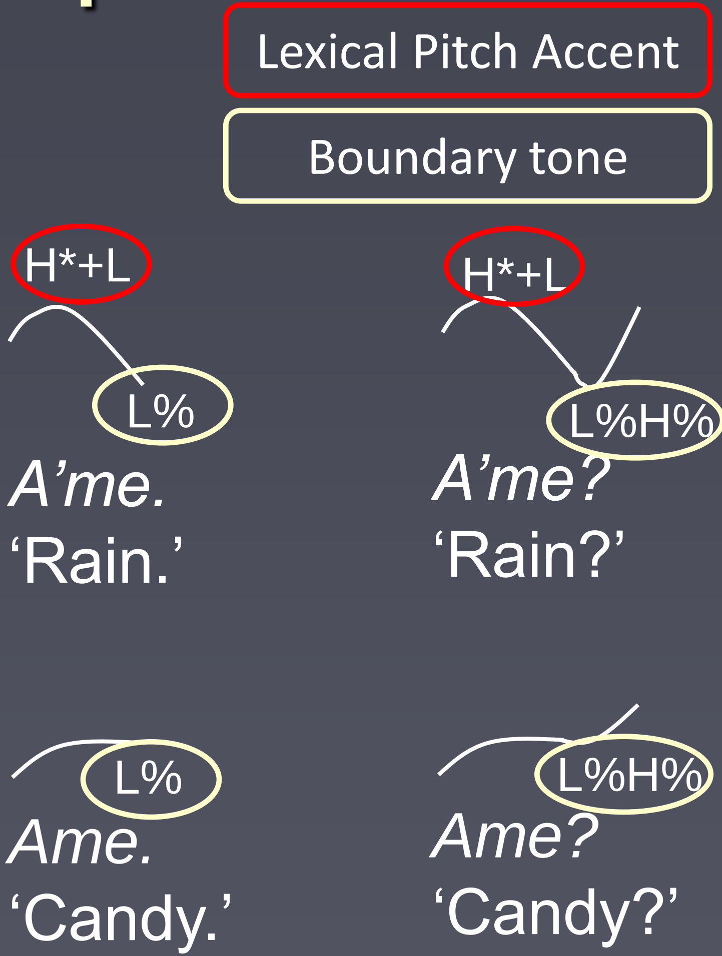
But

- In a language with the edge-prominent prosody, like Japanese & Korean, pitch movement tend to occur phrase finally (e.g., Jun, 2005).

- Boundary pitch movement (BPM)
- LH% -- rising intonation. Question
- HL% -- falling intonation. Turn taking

Prosodic system of Japanese

- A word has a lexically specified pitch shape.
  - Accented word: has the H*+L lexical pitch accent
  - Unaccented word: has no lexical pitch accent
- Intonation is realized through boundary tones



Lexical Pitch Accent

Boundary tone

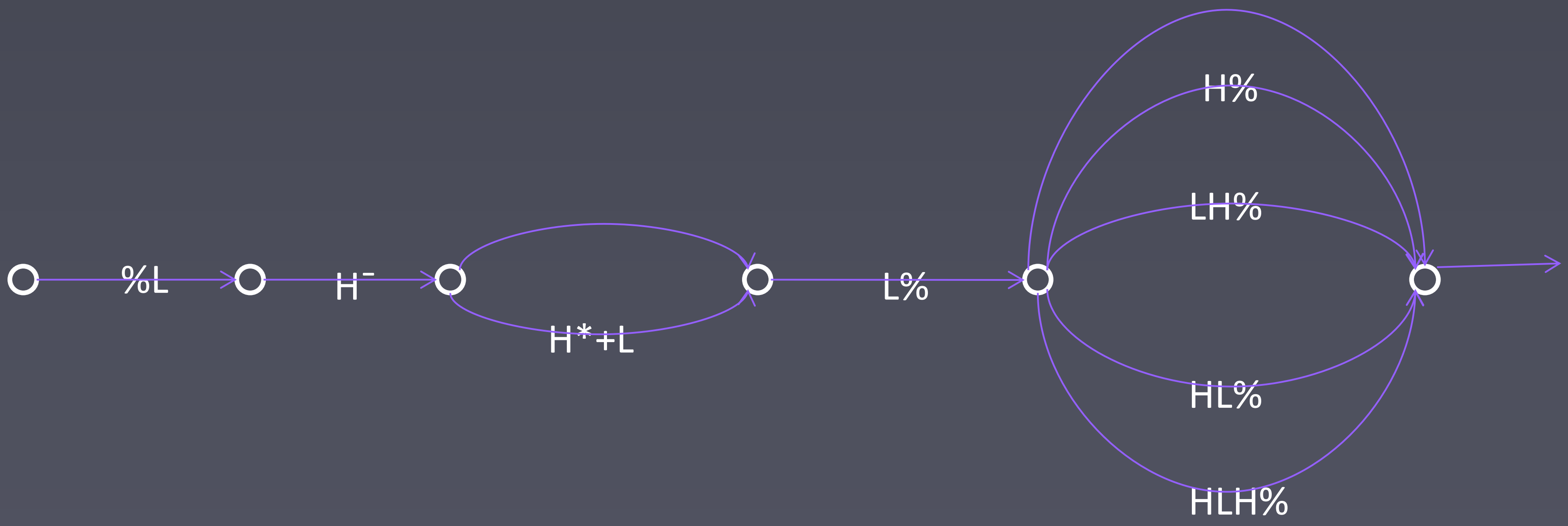
A’me. ‘Rain.’

A’me? ‘Rain?’

Ame. ‘Candy.’

Ame? ‘Candy?’

Finite State Grammar to generate all tunes in Japanese



Boundary Tone

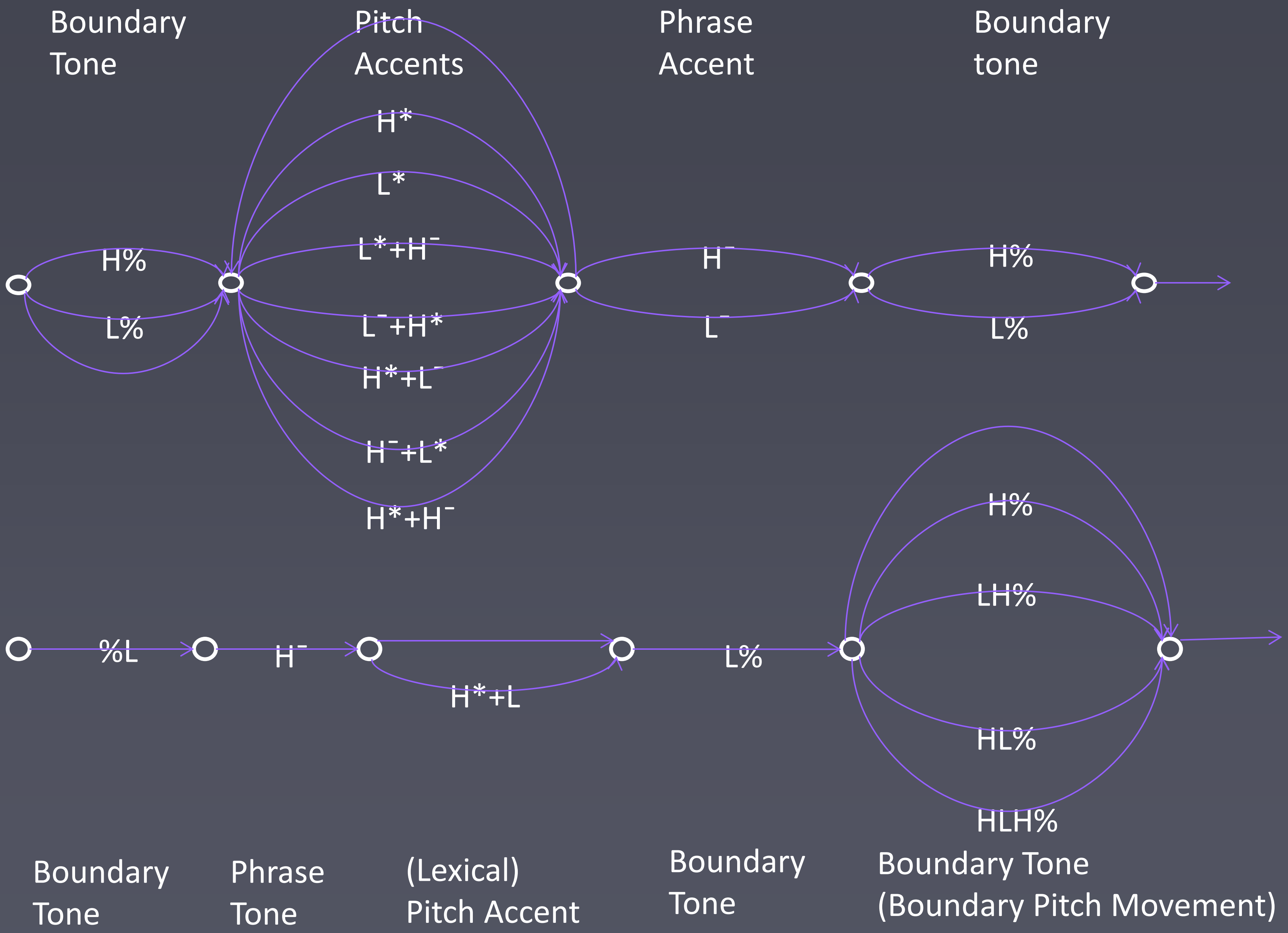
Phrase Tone

(Lexical) Pitch Accent

Boundary Tone

Boundary Tone (Boundary Pitch Movement)

Pierrehumbert & Beckaman (1988), Maekawa et al. (2002)



Boundary
Tone

Pitch
Accents

Phrase
Accent

Boundary
tone

H%

L%

H*

L*

L*+H^-

L^-+H*

H*+L^-

H^-+L*

H*+H^-

H^-

L^-

H%

L%

%L

H^-

H*+L

L%

H%

LH%

HL%

HLH%

Boundary
Tone

Phrase
Tone

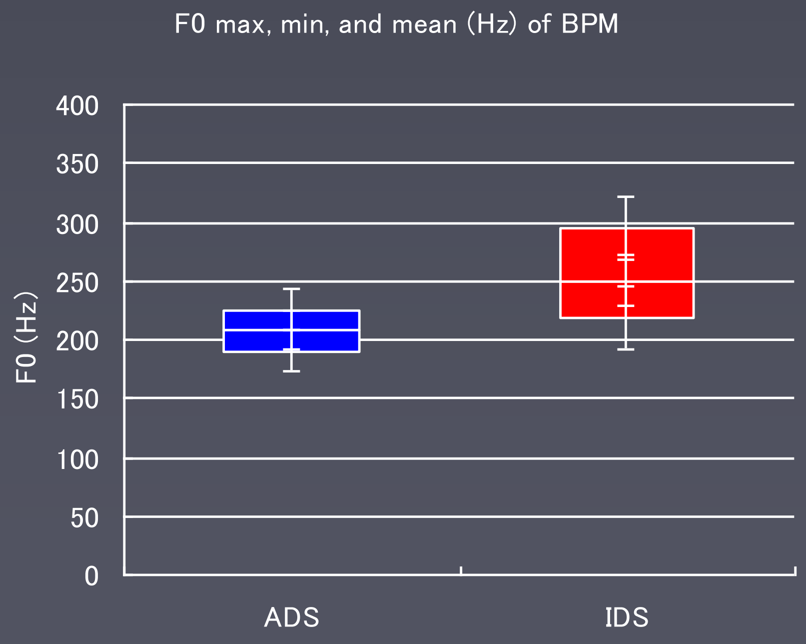
(Lexical)
Pitch Accent

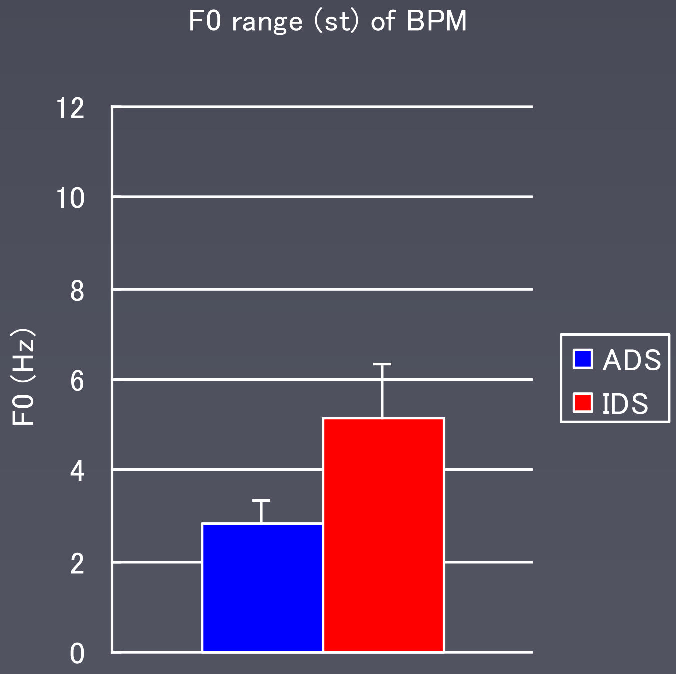
Boundary
Tone

Boundary Tone
(Boundary Pitch Movement)

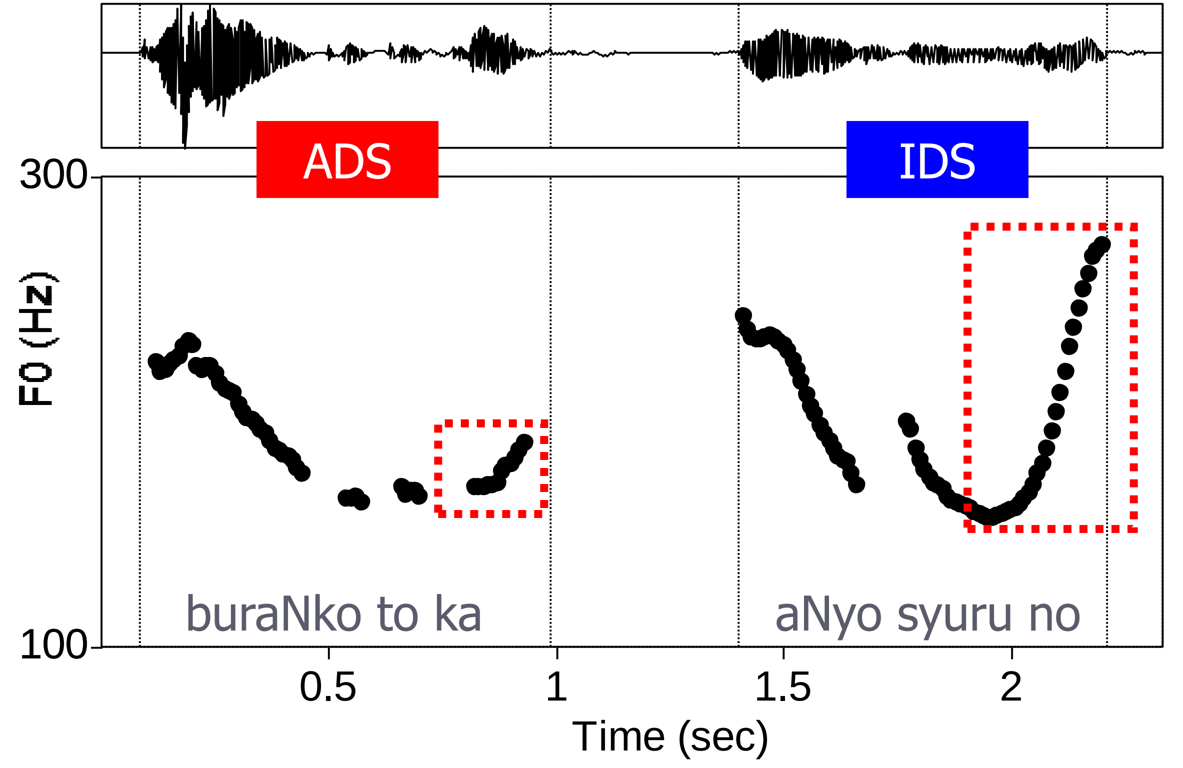
Boundary Pitch Movement

- Syllables with BPM
    - Max, min, mean, range (st): AD < ID

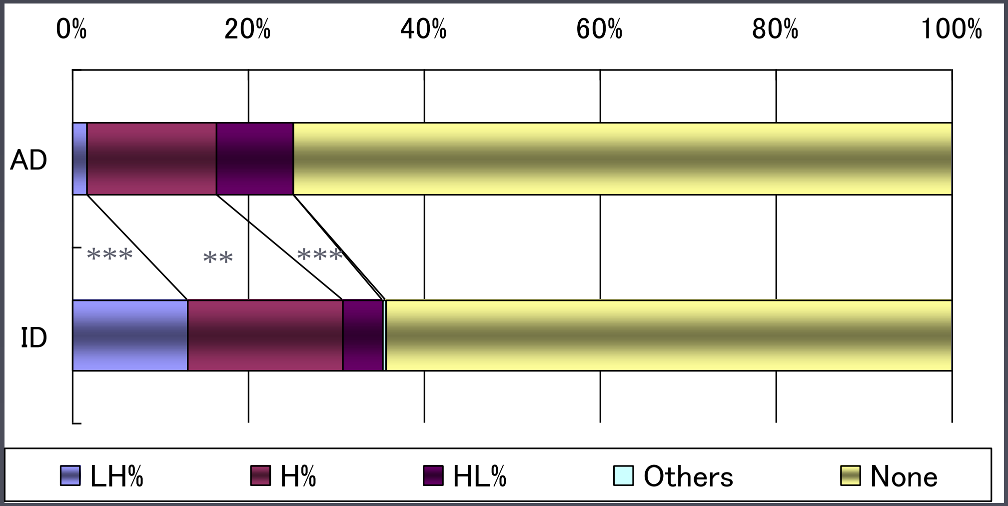

F0 max, min, and mean (Hz) of BPM


F0 range (st) of BPM

Expansion occurs at BPM



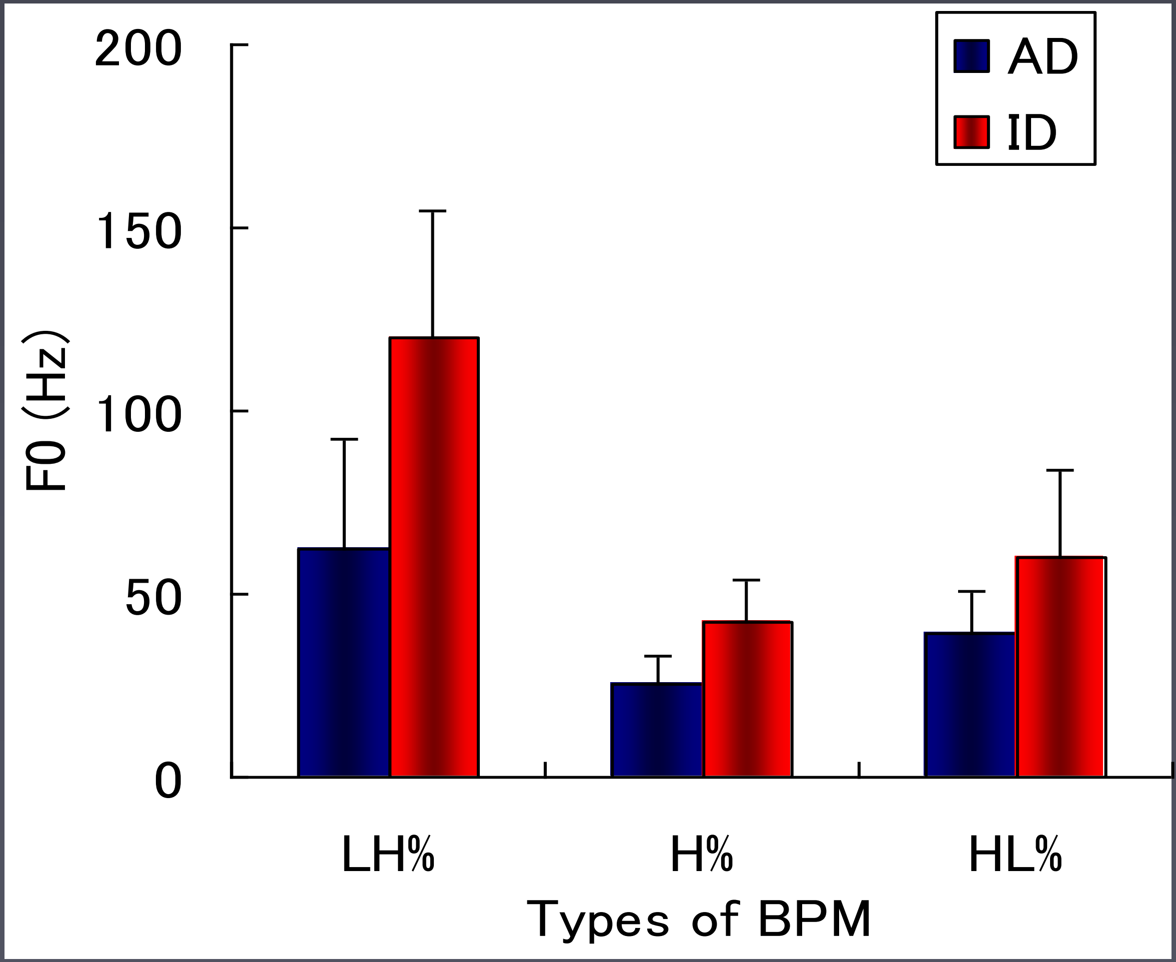
Boundary pitch movements (BPMs)



Frequency of BPMs

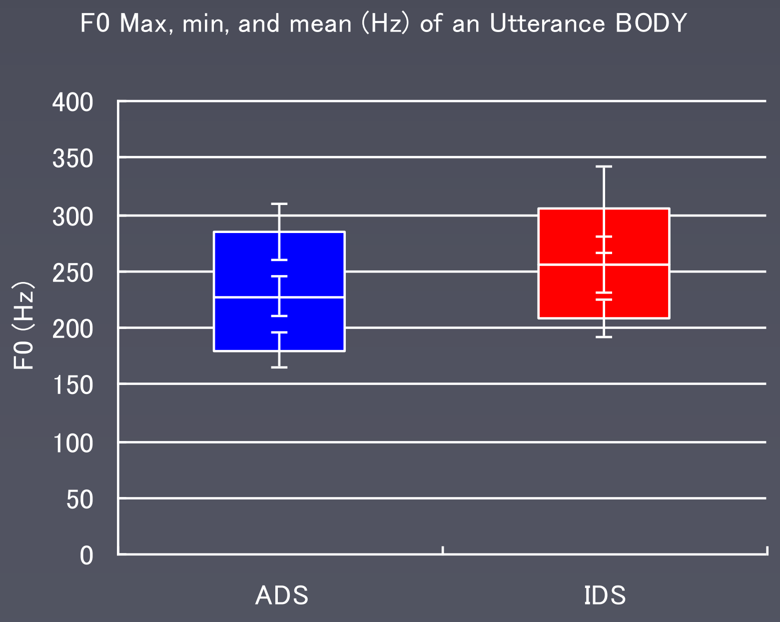
- **More frequent LH% (rise ↗) in ID.**
  - Presumably, more questioning in ID.
- More frequent HL% (rise-fall ↷) in AD.
  - Presumably, more continuation (turn-keeping) in AD

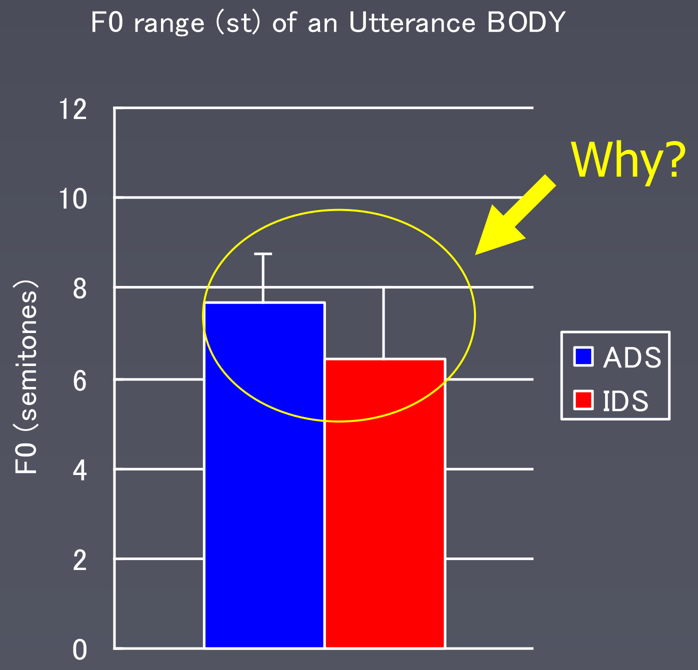
Pitch range is expanded in every BPM type



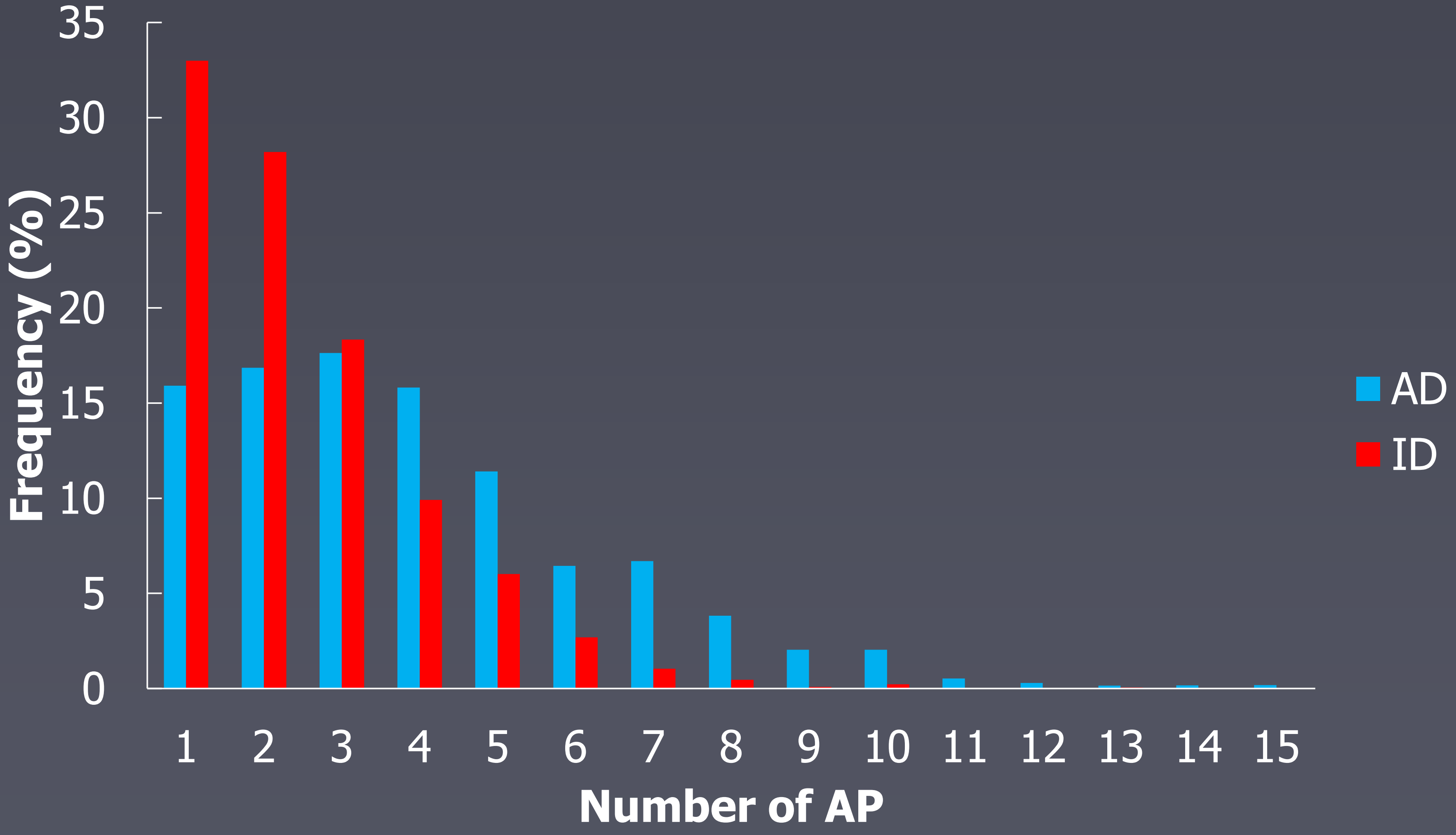
When syllables with BPM are removed

- Body of the utterances
  - “Max, min, mean: AD < ID
  - Range (st): AD > ID

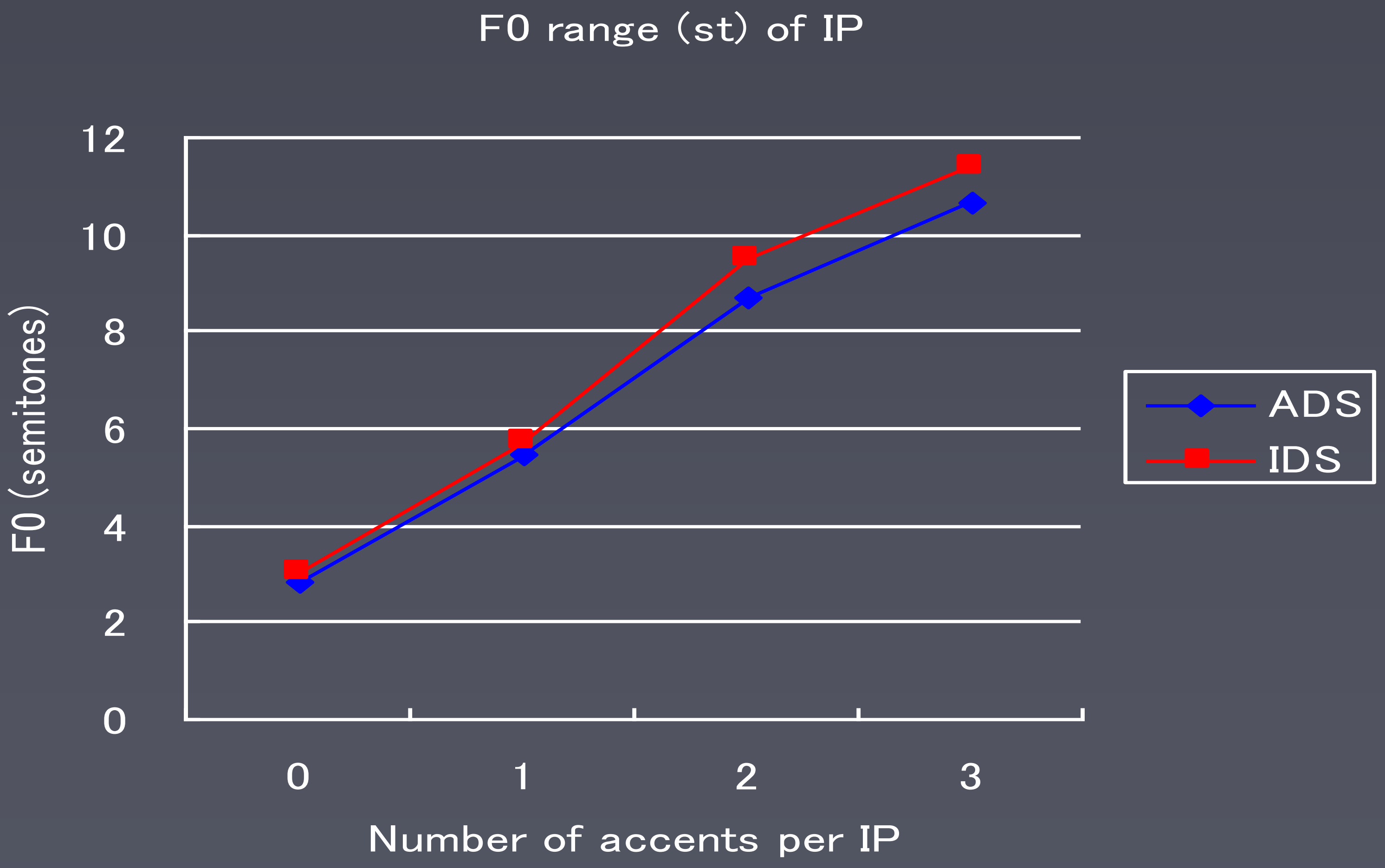




Frequency of the number of AP in an utterance



When IP is long & contain more accented words, pitch expanses



Pitch exaggeration in Japanese IDS

- No pitch exaggeration in overall utterances.
  - (Replication of Fernald et al, 1989)
- But;
- Pitch expansion found in syllables with BPM
  - This expansion *DOES* sound *exaggerated!*
- Larger pitch range in ADS, except for BPM
- ADS utterances are generally longer, contain more accented words, and have larger ranges.
  - This expansion does *NOT sound exaggerated*!

Intonation of Japanese IDS

- Exaggeration of intonation is an example of dynamic nature of prosody.
- Speakers’ desire to exaggerate intonation maybe universal…
- But how to do so is constrained by the prosodic structure of the language.
- Examination of IDS register allowed us to examine this dynamic property.

Outline

1. RIKEN Japanese mother-infant conversation corpus (R-JMICC)
2. Exaggeration of intonation (Igarashi et al., JASA, 2013)
3. Realization of phonological rule (Martin, et al., Cognition, 2014)

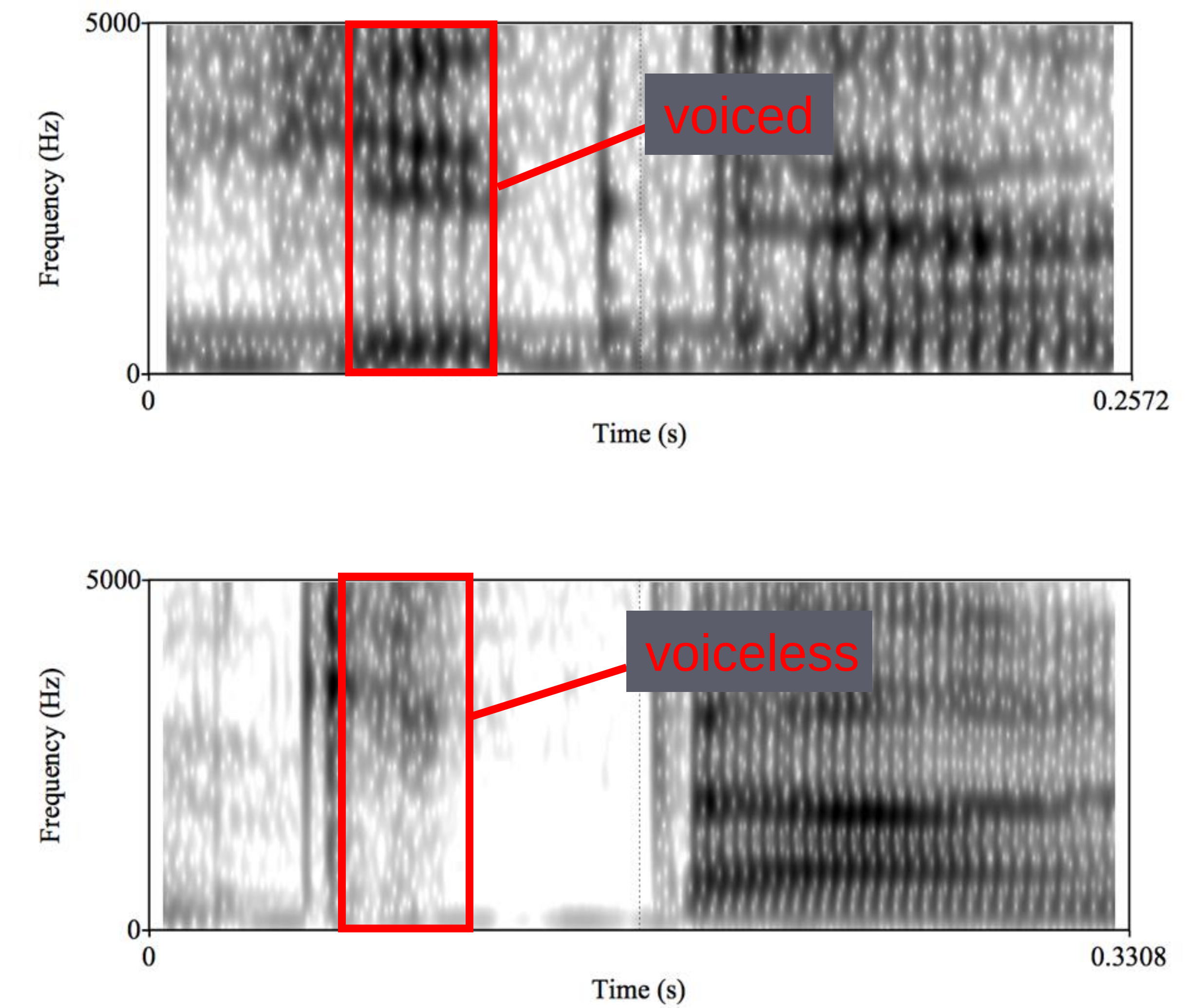
Phonological Rule: Vowel devoicing

Martin, Utsugi, & Mazuka, Cognition, 2014

- Rule:
  - Japanese high vowels /i/ & /u/([ɯ]) are devoiced between voiceless consonants, or following voiceless consonants word finally.
  - "kiki" (emergency) vs "kaki" (oyster)

High vowel devoicing

- Two tokens of *kita* ‘came’



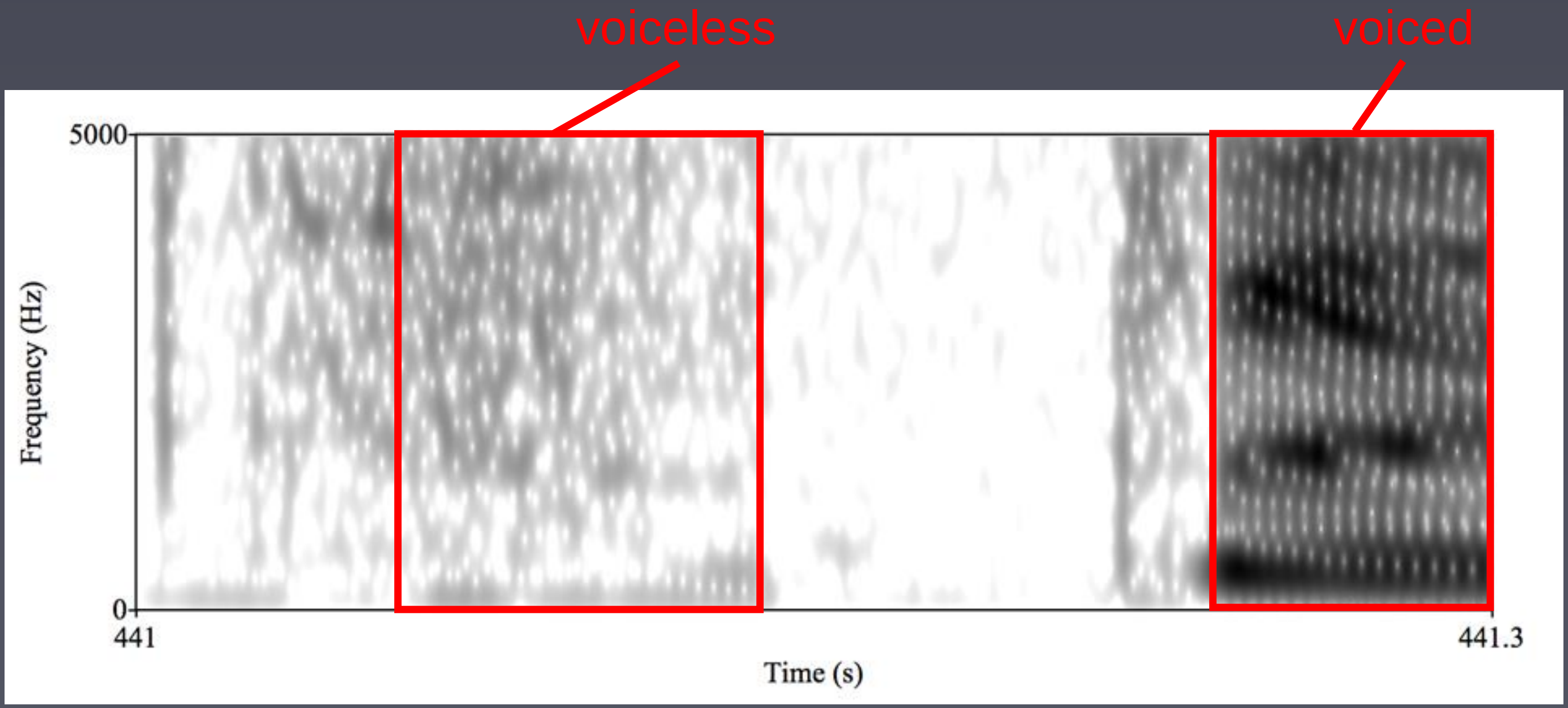
Phonological rule: Vowel devoicing

Martin, Utsugi, & Mazuka, Cognition, 2014

- Rule:
  - In Japanese, high vowels /i/ & /u/([ɯ]) tend to be devoiced between voiceless consonants, or following voiceless consonants word finally.
  - "kiki" (emergency) vs "kaki" (oyster)
- But
  - Non-high vowels can be devoiced also

Non-high vowel devoicing

- Devoiced /o/ in *soto* ‘outside’



Devoicing does not occur 100%

1. Devoicing of high vowels occur at much higher rate than non-high vowels
2. Devoicing rate changes by phonological contexts, speech rate, etc.
3. Speech register is a factor
    - Teachers devoice less when talking to hearing impaired children (Imaizumi, Hayashi, & Deguchi, 1995)

- *What determines devoicing rate?*

Data

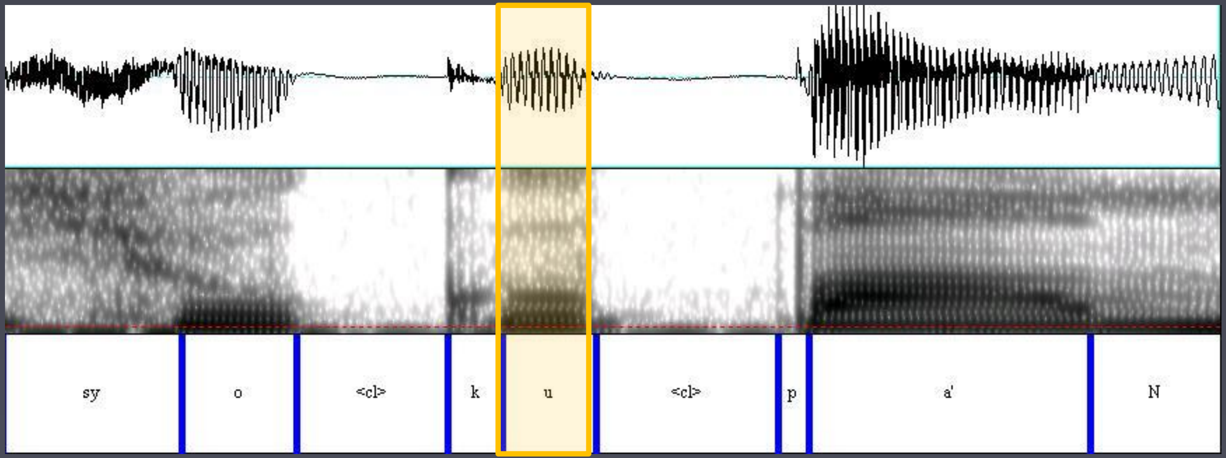
RIKEN Japanese Mother-Infant Conversation Corpus		
IDS	Playing with infant (average age 20 mo)	30 min each
ADS	Speaking with experimenter	10 min each
RS	Reading list of sentences (20 of 22 moms)	10 min each

Data

RIKEN Japanese Mother-Infant Conversation Corpus		
IDS	Playing with infant (average age 20 mo)	30 min each
ADS	Speaking with experimenter	10 min each
RS	Reading list of sentences (20 of 22 moms)	10 min each

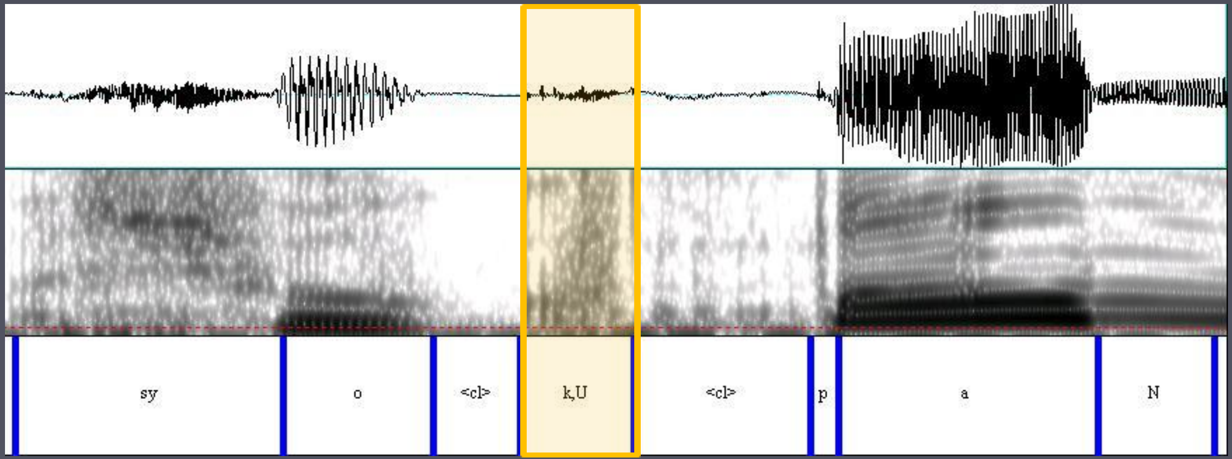
- Every vowel occurring between two voiceless consonants was labeled voiced or voiceless

Segmental labels



‘bread’ without devoicing

sy o <cl> k **u** <cl> p a N



‘bread’ with devoicing

sy o <cl> k **U** <cl> p a N

Data

RIKEN Japanese Mother-Infant Conversation Corpus		
IDS	Playing with infant (average age 20 mo)	30 min each
ADS	Speaking with experimenter	10 min each
RS	Reading list of sentences (20 of 22 moms)	10 min each

- Every vowel occurring between two voiceless consonants was labeled voiced or voiceless

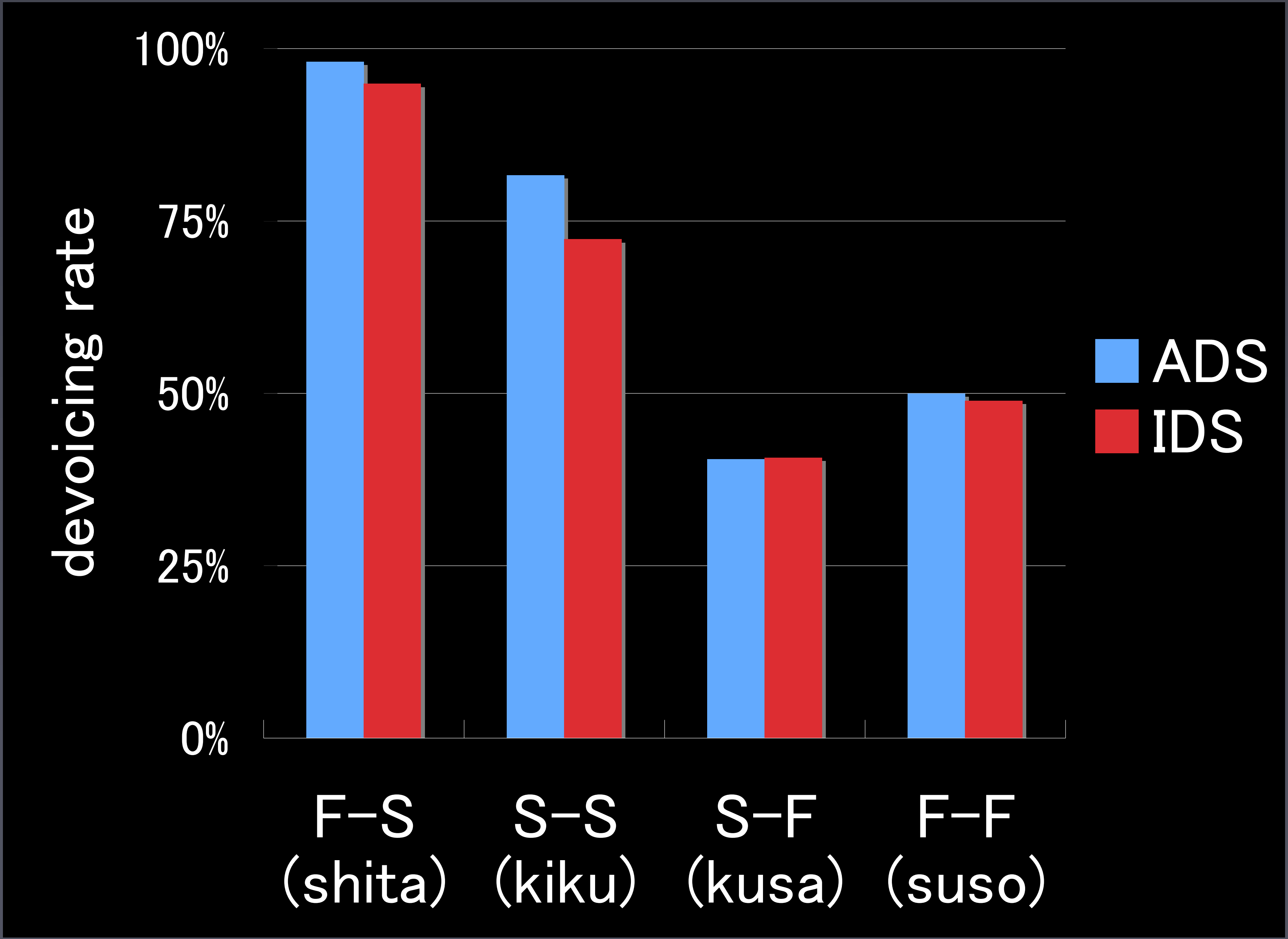
Devoicing rate

$$\frac{\text{number of devoiced vowels}}{\text{number of vowels between voiceless consonants}}$$

Linear Mixed Effects Models

- Dependent variable: voicing
- Fixed factors:
  - Speech style (RS, IDS, or ADS)
  - Vowel
  - Speech rate (moras per second)
  - Breathiness (mean H1-H2 of utterance)
  - Preceding context (affricate, fricative, stop)
  - Following context (affricate, fricative, stop)
  - Preceding * following
- Random factors:
  - Speaker
  - Word

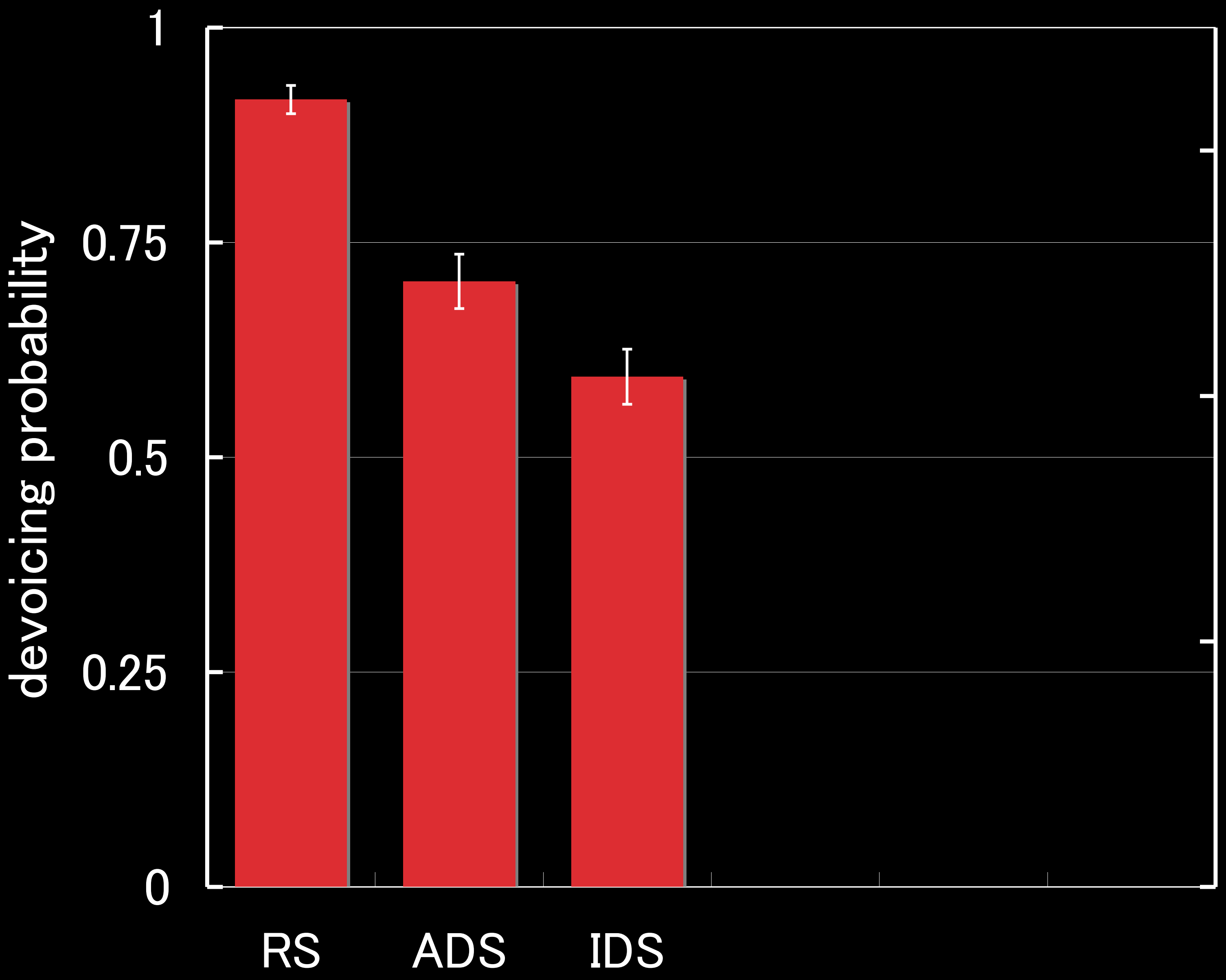
Segmental Contexts Matter



Linear Model Results

- Compared two models for each vowel height: one with SPEECH STYLE predictor, one without
- Including SPEECH STYLE significantly improves the fit of the model
  - High vowels: $\chi^2(1) = 37.4$, $p < 0.001$
  - Non-high vowels: $\chi^2(1) = 32.2$, $p < 0.001$
- Devoicing rate differences across speech styles are not just the result of subsidiary factors

Comparing All Speech Styles



Possible Explanations 1

1. Need to learn a phonological rule
    - Mothers want to provide more consistent input to infants

- Prediction:

  IDS > ADS/RS at least for high vowels

- Results

  ADS > IDS. Not supported by data.

Possible Explanations 2

2. Speakers want to increase intelligibility
   - Devoiced vowels are less intelligible (Gordon, 1988)
   - Adults find it difficult to identify devoiced vowels presented in nonce words (Beckman & Shoji, 1984; Cutler, et al., 2009)

- ► Prediction: ADS > IDS/RS

  High vowels = Non-high vowels

- ► Results:

  High vowels and non-high vowels showed an opposite pattern.

Why do high and non-high vowels behave differently?

- Two types of mechanism
  - Phonological: Speakers intend to devoice vowel, do so at abstract symbolic level (Tsuchida 1997, 2001; Teshigawara, 2002; Varden, 1998, 2010)
  - Phonetic: Speakers intend to voice vowel, but devoice inadvertently through gestural overlap (Imaizumi, Hayashi, & Deguchi, 1995; Jun & Beckman, 1993)
- Both mechanisms may operate in different dialects of Japanese (Fujimoto, et al, 1998; Fujimoto, 2004)

Why in opposite pattern?

- High vowels
  - Adults find it easier to recognize actual words with devoiced vowels (Ogasawara and Warner, 2009)
  - For adults, they are used to hearing words with high vowels devoiced. Devoicing known words -> better intelligibility.
  - For infants, most words are unknown. Devoicing -> worse intelligibility
  - The same goal (increasing intelligibility) results in the opposite results for adults & infants.

Why in opposite pattern?

- Non-high vowels
  - Even adults are not used to hearing words with non-high vowels devoiced.
  - Articulatory factors such as speech rate and breathiness of voice may account for at least part of the difference.

Discussion

- Devoicing rate dynamically changes in three speech registers.
- An opposite pattern of devoicing rate changes were observed in high- and non-high vowels
- Different factors that impact the two types of devoicing were revealed by examining distinct speech registers

Conclusion

- Phonological phenomena have dynamic aspects.
- They are difficult to capture in laboratory recordings.
- Comparisons of speech registers whose function is relatively easily defined can provide an window into dynamical aspects of real speech.

Thank You