

# Prosody: Thinking Outside the Box

## *Lecture 4* *The Sociophonetics of Prosody*

Dafydd Gibbon

Bielefeld University

*Fudan University Summer School: Contemporary Phonetics and Phonology*  
*Shanghai, 7–13 July 2018*

# *Sociophonetics*

## **Selected Approaches to Sociophonetics**

1) Labov's Sociophonetics

2) OSCAR: Phonetic Opinion Survey:

1) The Prosody of Impoliteness

2) Description of Mandarin Tones

3) Prosodic Analysis of Discourse

- Case 1: Prosodic Framing
- Case 2: AM vs. FM Spectra
- Case 3: Accent Constraints
- Case 4: Long FM contours
- Case 5: Emotive FM contours



# ***Sociophonetics: several approaches***

## **1. Methodologies:**

- Interpretative methods:
  - ethnomethodology
  - conversation analysis
  - discourse analysis
  
- Correlationist (statistical) methods:
  - initiated by Labov
  - applied by Trudgill and many others

## **2. Scenarios (observational and experimental designs)**

- Natural and spontaneous speech corpora
- Surveys:
  - prompted real-world elicitation (Labov)
  - online perception and description (OSCAR)

# ***Labov's 'Correlationist' Approach: Selected Aspects***

# Labov's Surveys

## 1. Dimension 1: *phonetic variables*

for example:

*th* [θ], *ng* [ŋ], *r* [r]

## 2. Dimension 2: *context style variable*

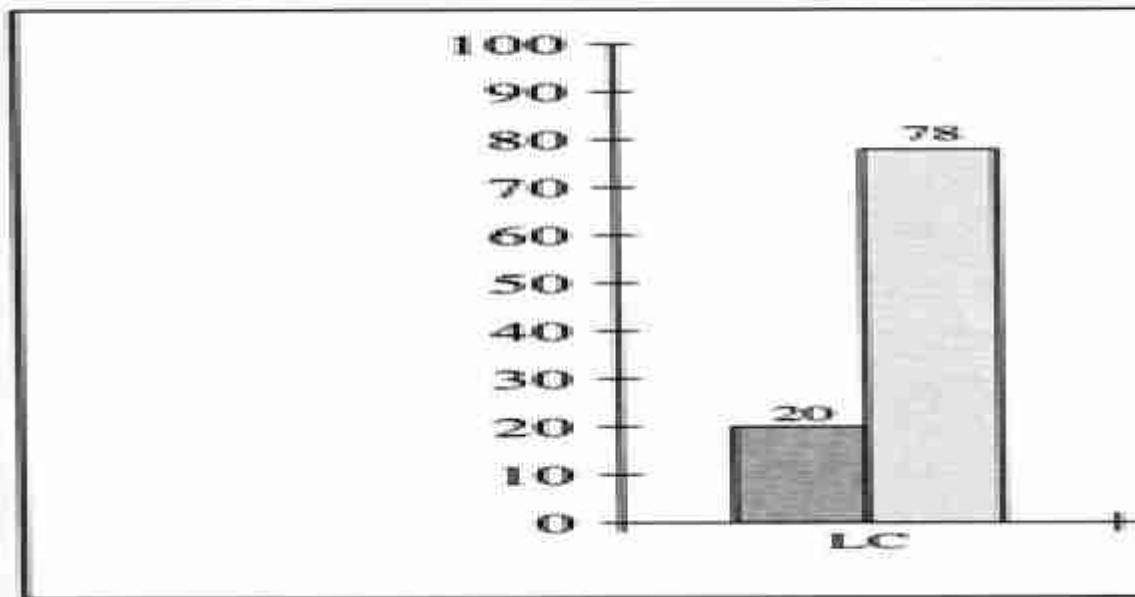
- casual speech
- careful speech
- reading style
- word lists
- minimal pairs

## 3. Dimension 2: *socioeconomic variable*

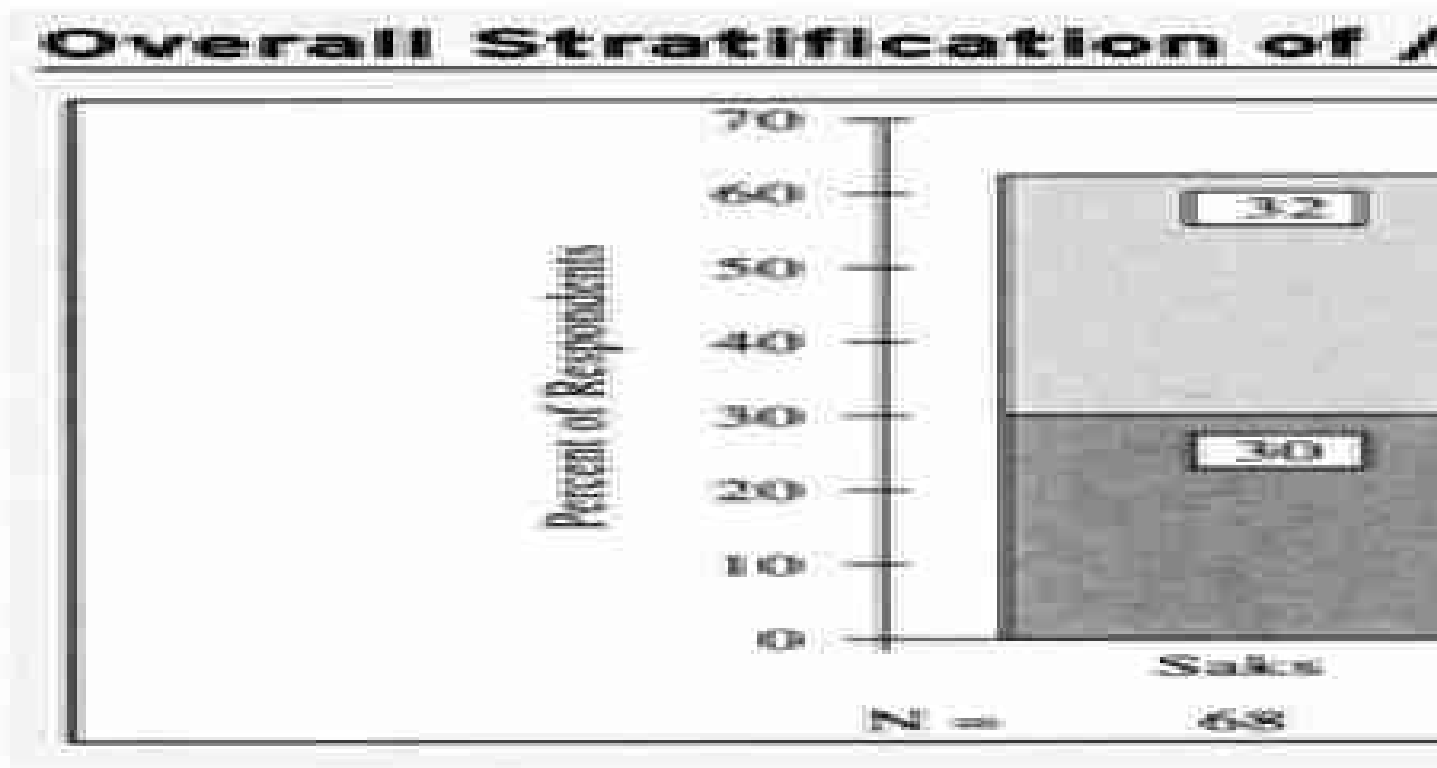
- lower class
- working class
- lower middle class
- upper middle class

# Labov's Surveys

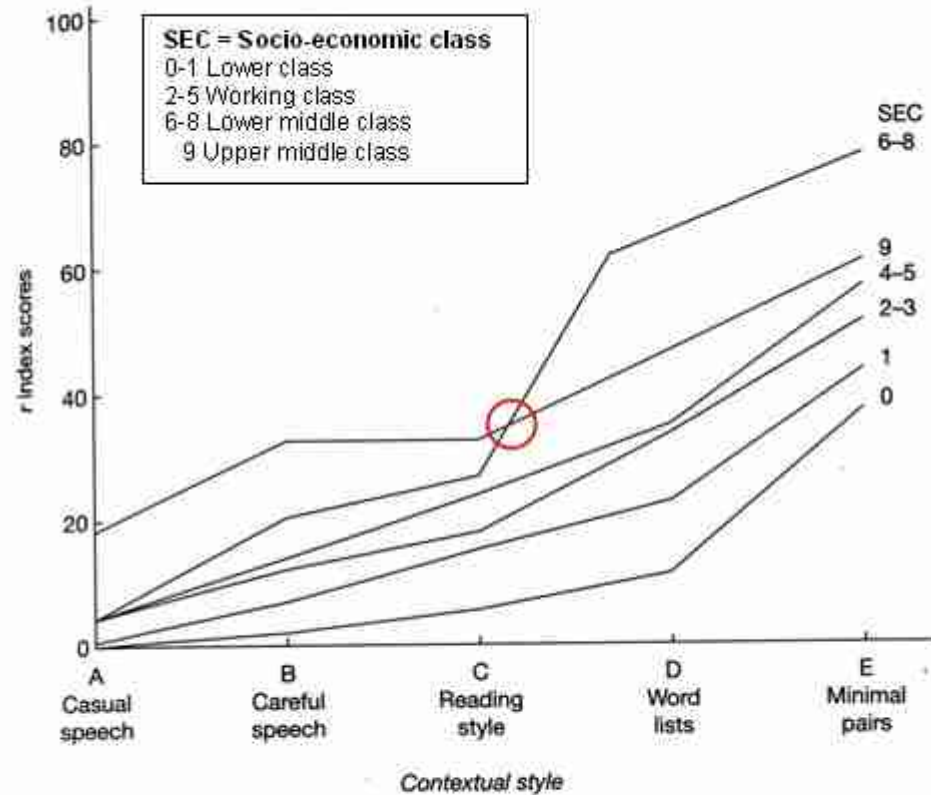
**Percent of -ing Suffix Pro  
New York City**



# Labov's Surveys



# Labov's Surveys



Social and stylistic stratification for the variable (r) in New York City according to Labov (1966)

(Source: Llamas/Mullany/Stockwell, 2007: 55)



# ***An Online Sociolinguistic Survey Method: OSCAR***

***(Online Survey Collation and Reporting)***


## Different approach: Online Opinion Survey

### Task of assigning pitch descriptors to tones

- metalinguistic documentation of perception
  - cf. judgment paradigm of auditory phonetics and phonology
- sociophonetics, ‘folk linguistic’ opinions

### There are many formats for opinion surveys

- open interview
- closed set (standard: Likert scale)
  - for this test:
    - audio input, Likert format response to a statement:

- 
- strongly agree
  - agree
  - don't care
  - disagree
  - strongly disagree



Dr. Rensis Likert  
1903-1981

Likert scale 1932  
(Ph.D. thesis)

## Descriptor assignment

### Task of assigning pitch descriptors to tones

- metalinguistic documentation of perception
  - cf. judgment paradigm of auditory phonetics and phonology
- sociophonetics, 'folk linguistic' opinions

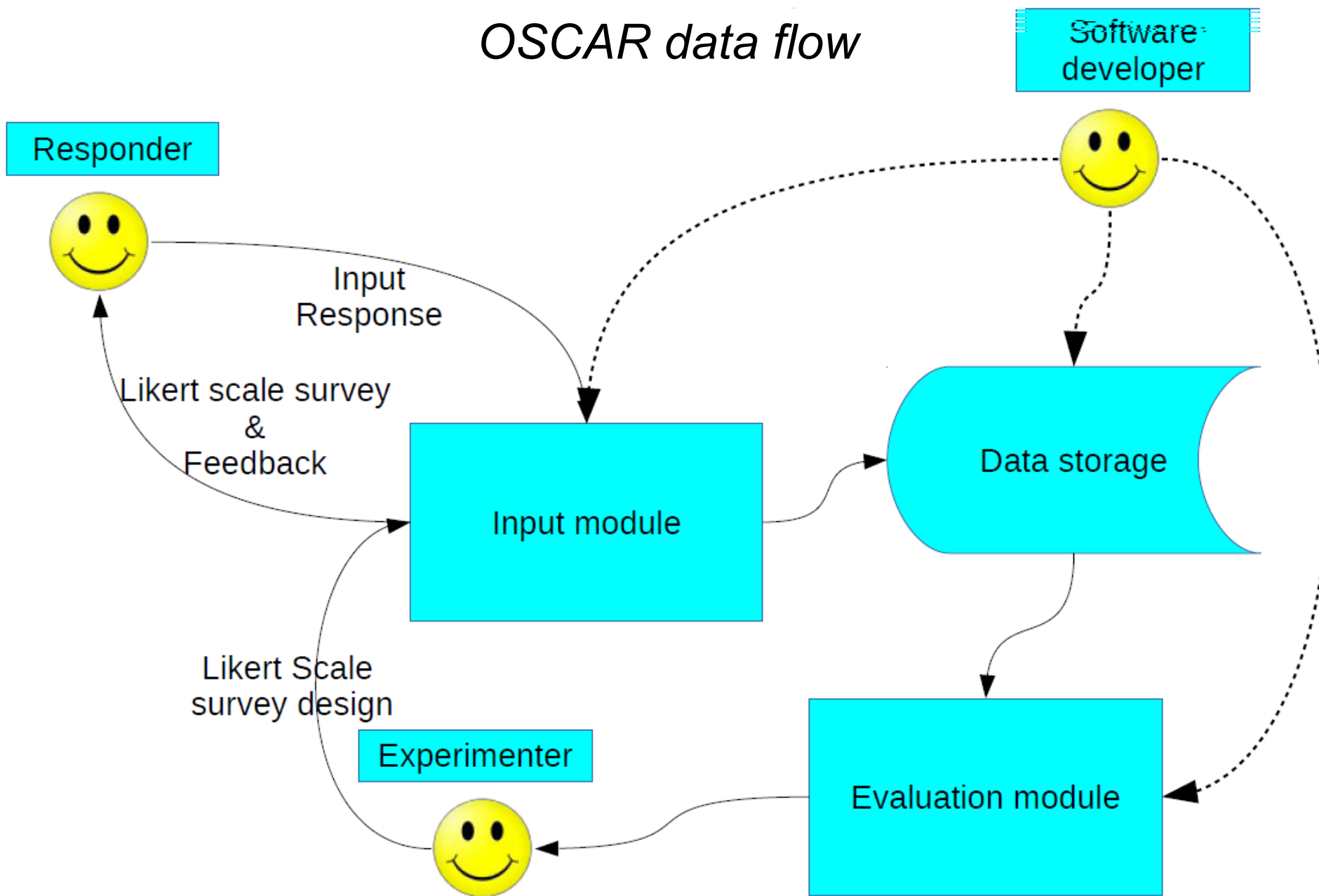
### Custom online tool OSCAR

- “Online Survey Collation And Reporting”
- input:
  - responder metadata: age group, sex, L1, regional variety
  - single-page Likert format survey form
    - list of tones + pitch descriptor choices
- output:
  - for responders: notification of (in-)completeness of responses
  - for experimenter: automatic evaluation



# Variability in Mandarin Tone Perception

## OSCAR data flow



## Acknowledgment for the following section to:

Li Peng. 2018. An Analysis of Impolite Speech Acts in Donald Trump's Speeches with Special Reference to Prosodic Components. M.A. Thesis, Jinan University, Guangzhou.



# *Prosody and Impoliteness*

## 1. Aim

- Examining impolite speech acts through impoliteness strategies:
- previous approaches
  - films, tv series, reality shows, debate
- here: public speech

## 2. Focus on the role of linguistic text

- lexical aspects
- semantic aspects
- prosodic aspects often missed out
- here: both ‘textual’ and prosodic aspects

## 3. Method

- previous approaches
  - mainly qualitative
- here: qualitative and quantitative

# *Prosody and Impoliteness*

**Therefore, new questions are asked:**

- 1) What are the strategies commonly employed by Donald Trump?
- 2) What role does prosody play in the comprehension of impolite speech acts?
- 3) What are respondents' perceptions of Donald Trump's speeches?

Qualitative approaches often said to be 'merely opinion'

Novel twist: validation of 'opinion' by 'opinion survey'

(consensus theory of truth!):

Novel twist:

online questionnaire with sounds and descriptors to characterise the sounds

# *Methodology and Data*

## A combined research approach:

- Typical 5-point Likert scale
- Novel twist: an online questionnaire with sounds/ various attributes

## Procedure:

- Data collection:
  - Prompts:
    - 1)Downloading Donald Trump's presidential election speeches from Youku (> 7 hours)
    - 2)Repeated listening, transcribing, and then cutting out the impolite clips
    - 3)Extracting 42 clips concerning impoliteness within impoliteness model
    - 4)Converting 42 video clips into audio clips (WAV) by Total Video Converter
    - 5)Choosing 10 audio clips concerning marked prosody to design a questionnaire
- Survey:
  - Using OSCAR to distribute, collect and report on the online audio survey results ([http://wwwhomes.uni-bielefeld.de/gibbon/OSCAR\\_a102/](http://wwwhomes.uni-bielefeld.de/gibbon/OSCAR_a102/))



# *Prosodic Aspects of Impoliteness*

## Prosodically Impolite Speech Acts

Pause

Stress

Down-stepping Intonation

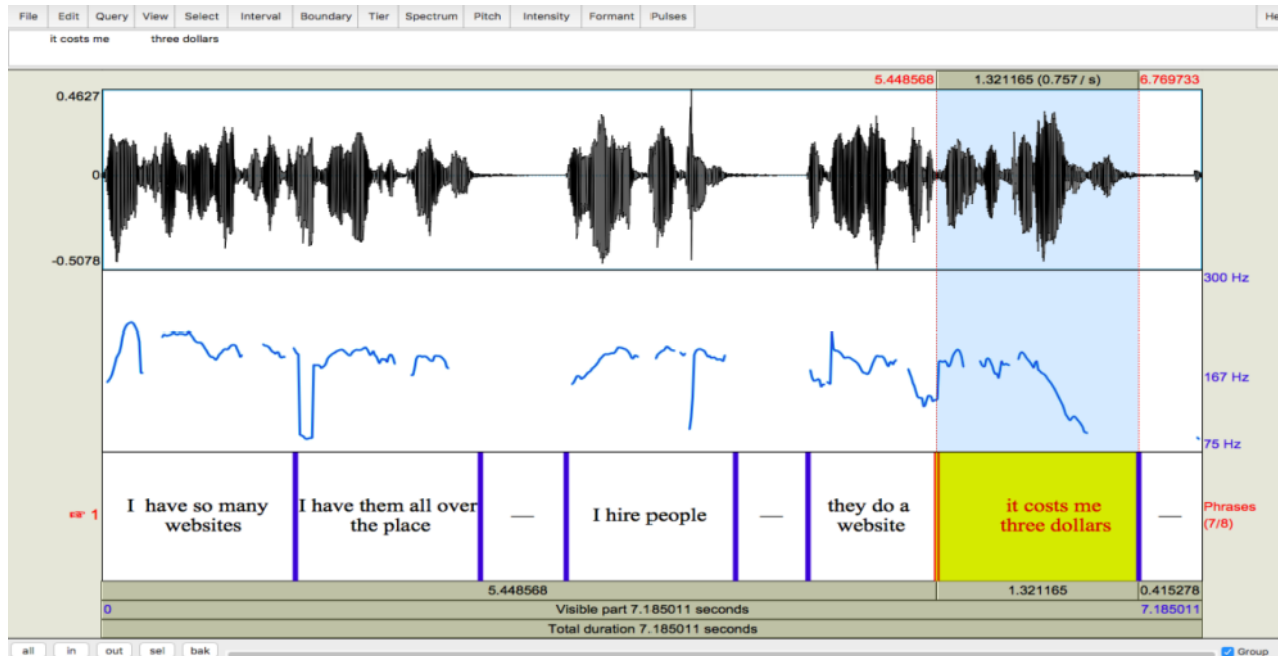
Tempo of Speed

Prosodic Mimicry

## Down-stepping Intonation and Impolite Speech Acts

“Five billion dollar website, I have so many websites, I have them all over the place. I hire people, they do a website, it costs me 3 dollars. Five billion dollar website.” (New York)

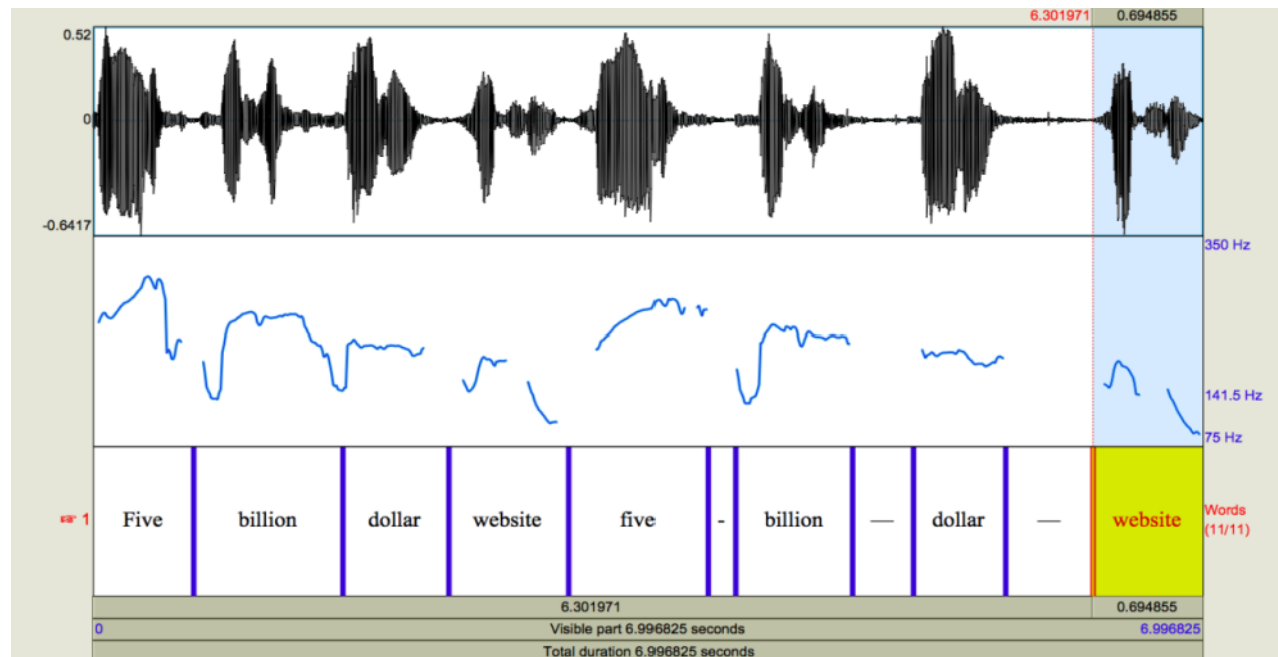
# Prosodic Aspects of Impoliteness



Audio\_5:

Five billion dollar website  
I have so many websites, I have them all over the place.

I hire people, they do a website, it costs me 3 dollars.



Five billion dollar website.  
(Context: Donald Trump is talking about Obama's health care website. Bloomberg Government estimated that Obama's healthcare government cost less than 2.1 billion dollars.)

# Prosodic Aspects of Impoliteness

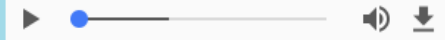


Google Chrome 不是您的默认浏览器 设为默认浏览器

## Section B

- Please listen to each recording at least twice (a transcript is provided).
- Then for each description (sarcastic, appropriate, etc.) click on your impression of whether you *strongly agree*, *agree*, *have no opinion*, *disagree* or *strongly disagree* with the statements "The utterance is sarcastic", "the utterance is appropriate", etc.
- Then please give your ideas about what causes this impression.

**Audio\_1:** They met for thirty-nine minutes, remember, he said: "We talked golf, and we talked about our grandchildren." Three minutes for the grandchildren, two minutes for the golf.


Recording: 

The utterance is

	strongly agree	agree	no opinion	disagree	strongly disagree
sarcastic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
appropriate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
offensive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
polite	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
arrogant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
friendly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please comment on which aspects of language appear to create the effects you perceive (e.g. tone of voice, emphasis, rhythm, pauses, words, grammar, repetitions,...)?

**Audio\_2:** I will be the greatest jobs president that God ever created.

Recording: 

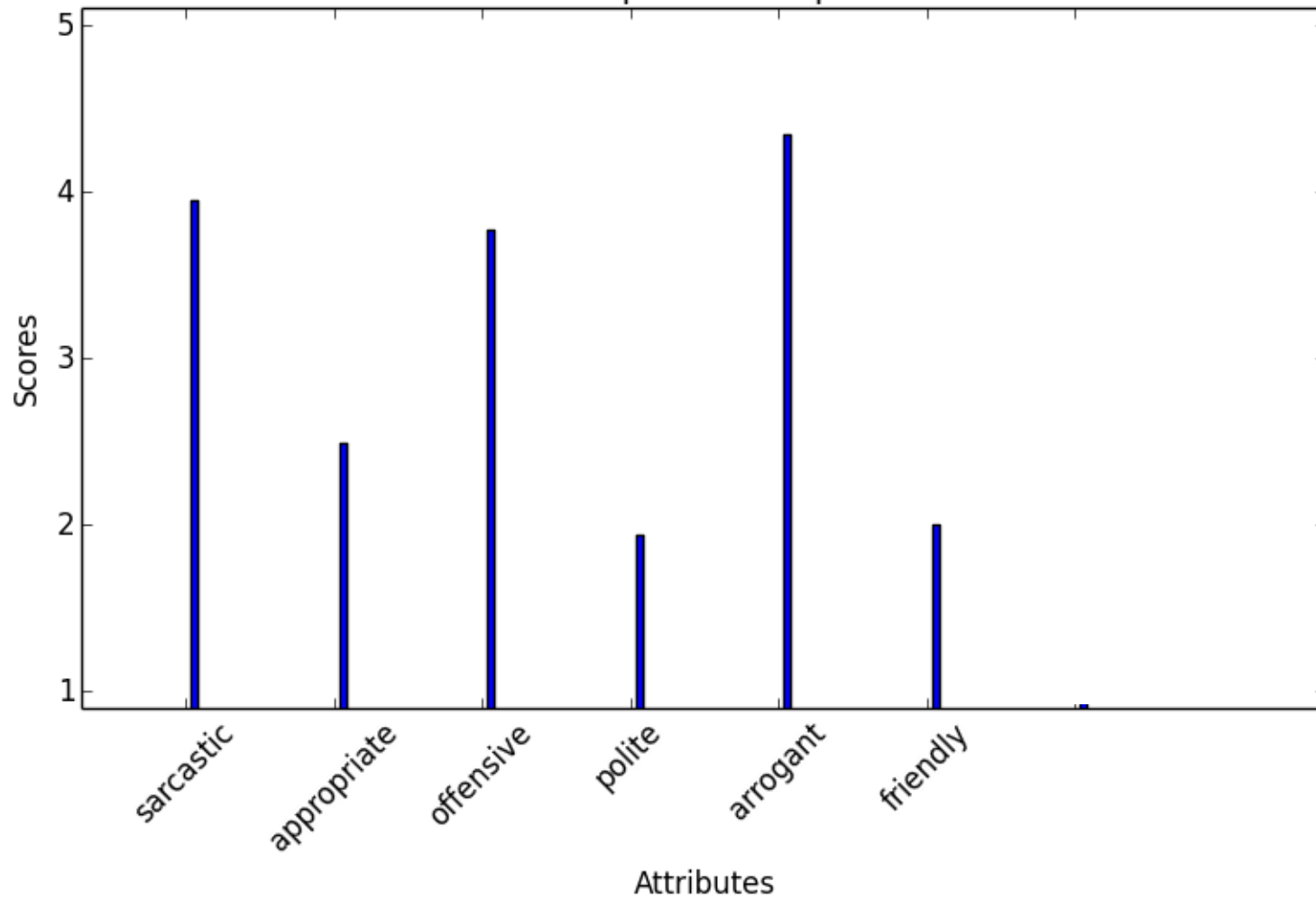
The utterance is

	strongly agree	agree	no opinion	disagree	strongly disagree
sarcastic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
appropriate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
offensive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
polite	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
arrogant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
friendly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

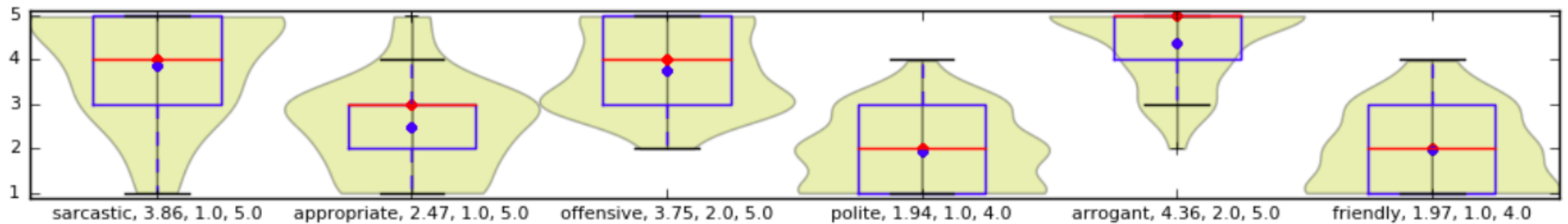
Please comment on which aspects of language appear to create the effects you perceive (e.g. tone of voice, emphasis, rhythm, pauses, words, grammar, repetitions,...)?

# Prosodic Aspects of Impoliteness

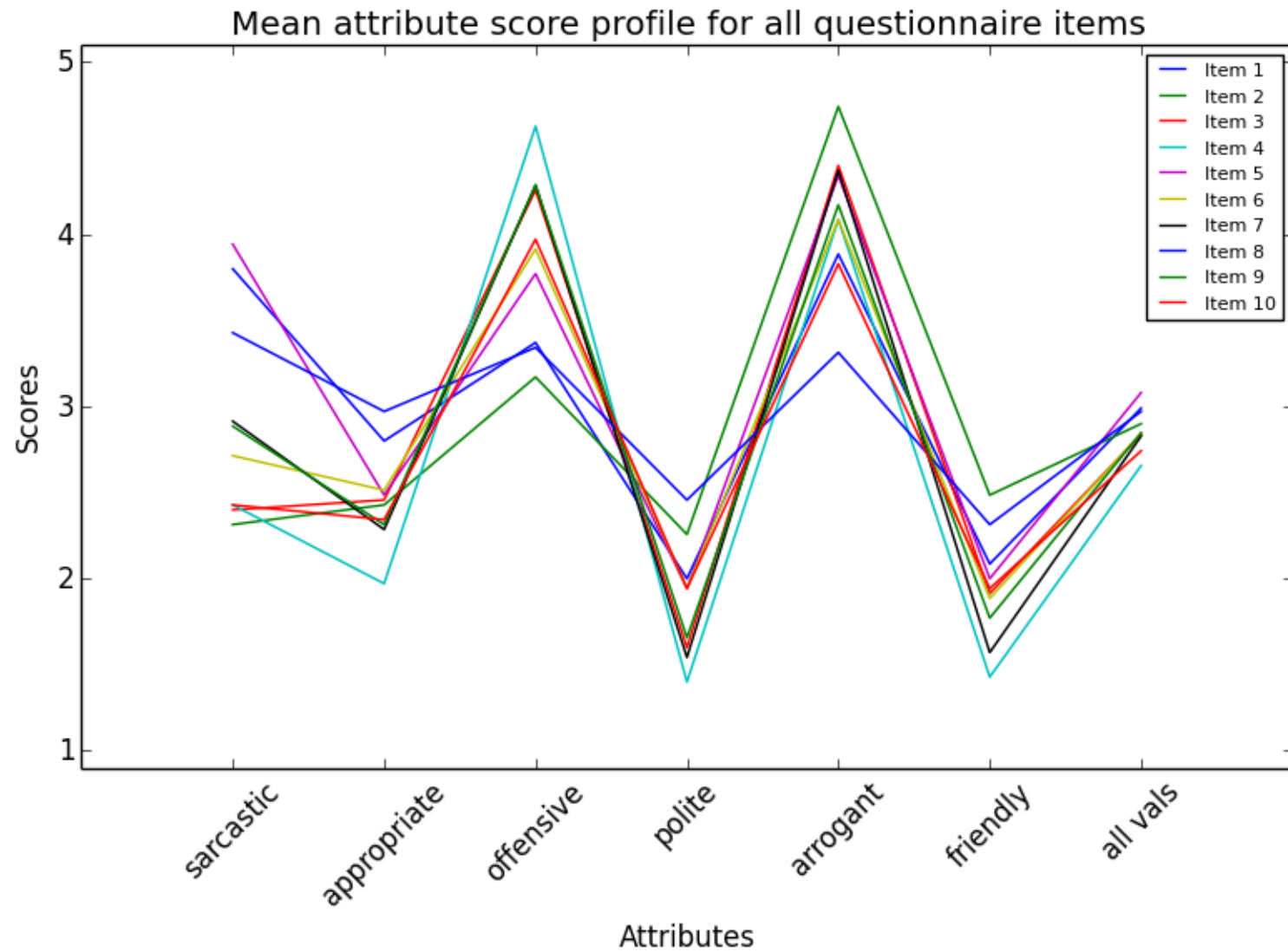
Mean attribute score profile for questionnaire item 5



Score densities: Likert item 5



# Prosodic Aspects of Impoliteness



# *Prosodic Aspects of Impoliteness*

Gender	F=3.113	$p \leq 0.1^\dagger$	$p = 0.0778$
<b>Language</b>	<b>F=3.199</b>	<b><math>p \leq 0.001^{***}</math></b>	<b><math>p = 0.000</math></b>
Language Variety	F=1.554	$p \leq 0.1^\dagger$	$p = 0.091$
<b>Party</b>	<b>F=9.447</b>	<b><math>p \leq 0.001^{***}</math></b>	<b><math>p = 3.355e-06</math></b>
<b>Descriptor</b>	<b>F=298.688</b>	<b><math>p \leq 0.001^{***}</math></b>	<b><math>p = 3.081e-243</math></b>
Prompts	F=1.751	$p \leq 0.1^\dagger$	$p = 0.073$

## Acknowledgment for the following section to:

Gibbon, Dafydd and Huangmei Liu. 2018. Variability in Mandarin Tone Perception. *Proceedings of Speech Prosody 2018, Poznań, Poland* [In the ISCA Proceedings Archive]



## GOALS

### Test of new method

- multidialectal comparison
- contrast with previous bidialectal and bilingual comparisons
- exploratory rather than confirmatory

### Sociophonetic focus

- assignment of descriptors of perceived pitch
  - to standard Mandarin (Pǔtōnghuà) tones
  - by native speaker responders from different regions
- comparison of height and contour descriptors
- focus on inter-rater variability
  - contrast with inter-rater reliability
- preparation for a large-scale multidialectal study
- longer-term goal of relating pitch descriptor assignments
  - to self-ascribed regional dialects
  - to linguistic dialect classification



# Variability in Mandarin Tone Perception

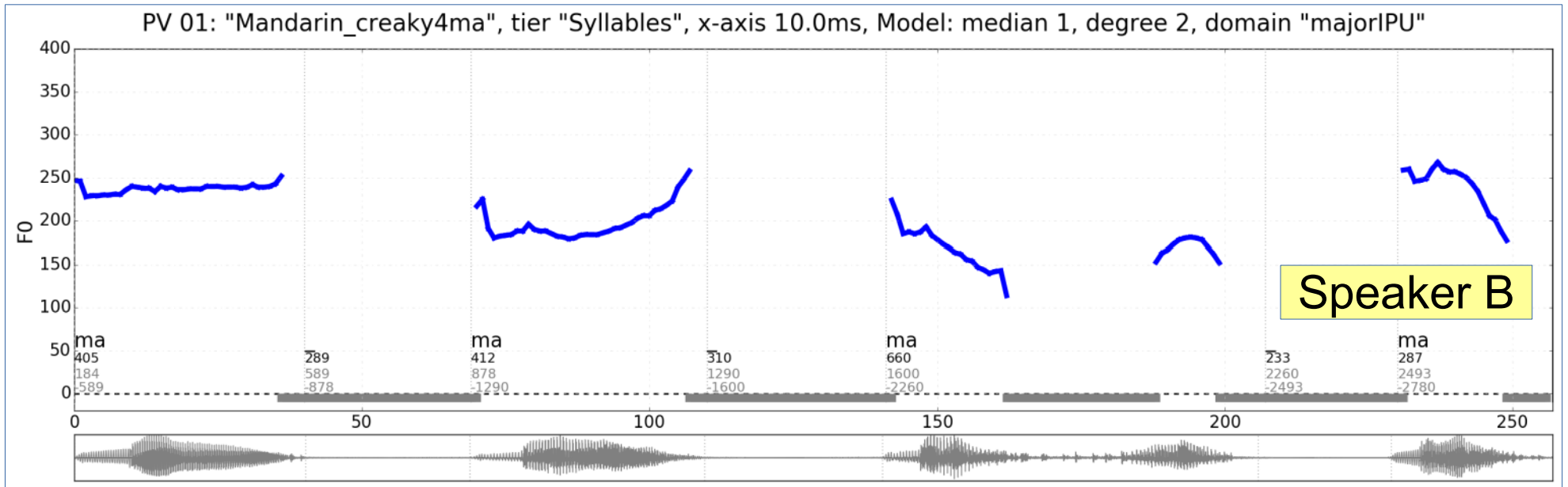
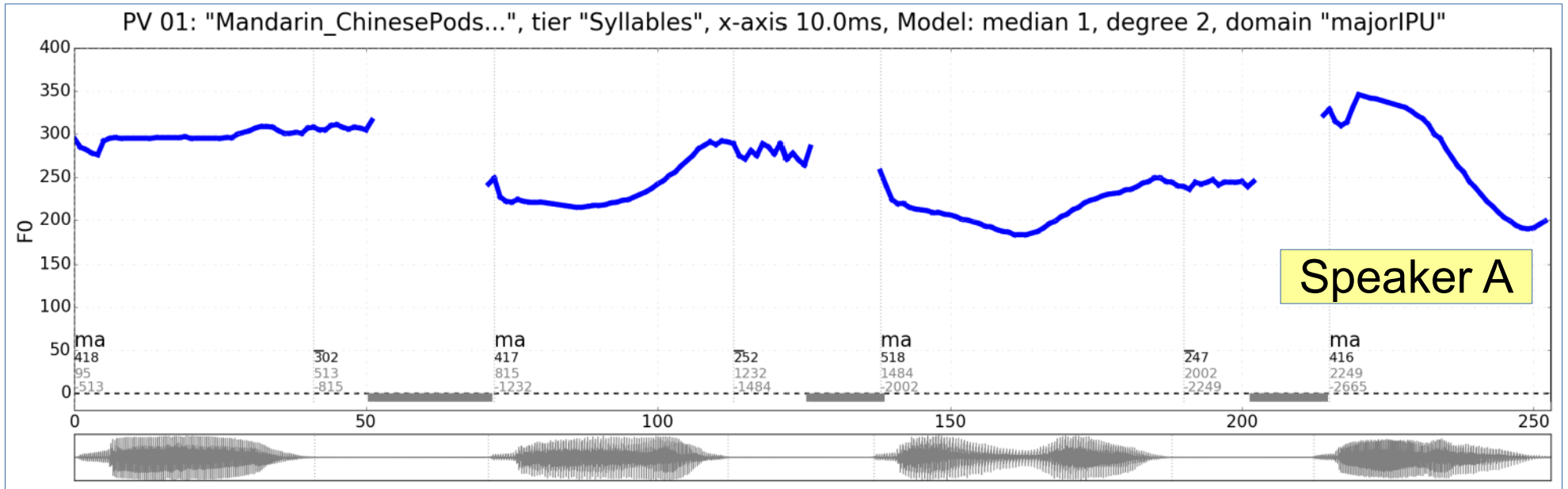
## TASK

## Stimuli

- 16 tone items:
  - 4 tones
  - 2 tokens each
  - 2 female speakers, standard Beijing Mandarin
- for all responders:
  - same randomised token order
  - no adjacent repetitions

# Variability in Mandarin Tone Perception

## Mandarin lexical tones



# Variability in Mandarin Tone Perception

## INPUT TASKS

### Descriptors:

- 8 pitch descriptors
  - contours: *level, rise, fall-rise, rise-fall* (distractor), *fall*
  - heights: *high, mid, low*
- 5-point Likert format input scale
  - yes, maybe, not sure, maybe not, no
- coded for evaluation: 5,4,3,2,1

Audio\_1:  
Listen to the recording at least twice:

The melody of the sample is...

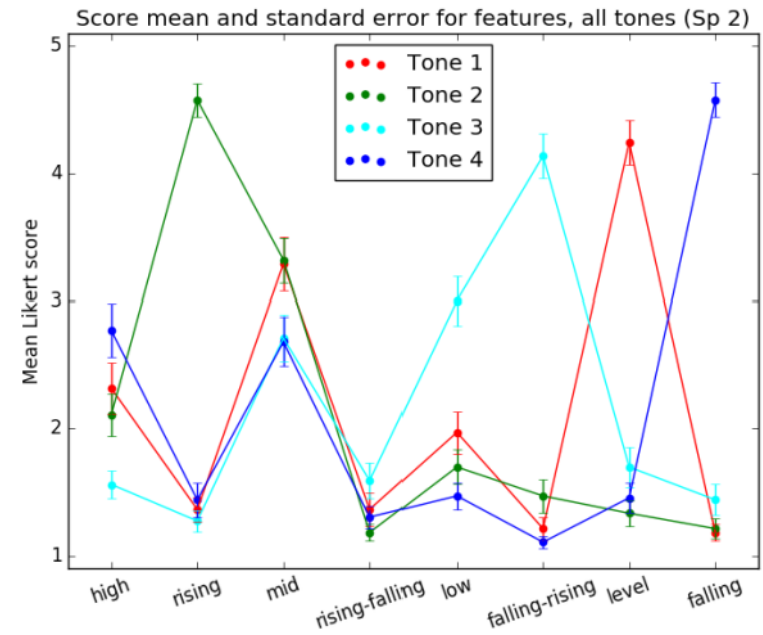
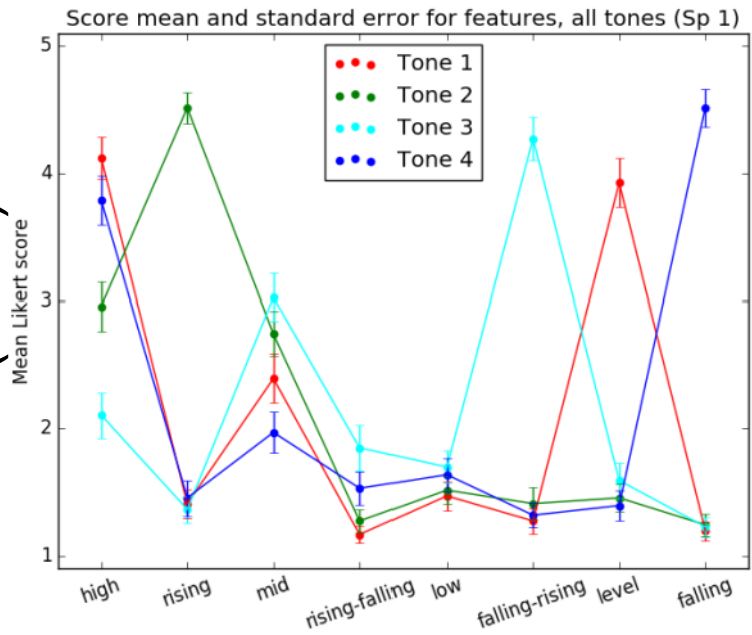
0:00 / 0:03

	yes	maybe	unsure	maybe not	no
high	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
rising	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
mid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
rising-falling	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
falling-rising	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
level	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
falling	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

# Variability in Mandarin Tone Perception

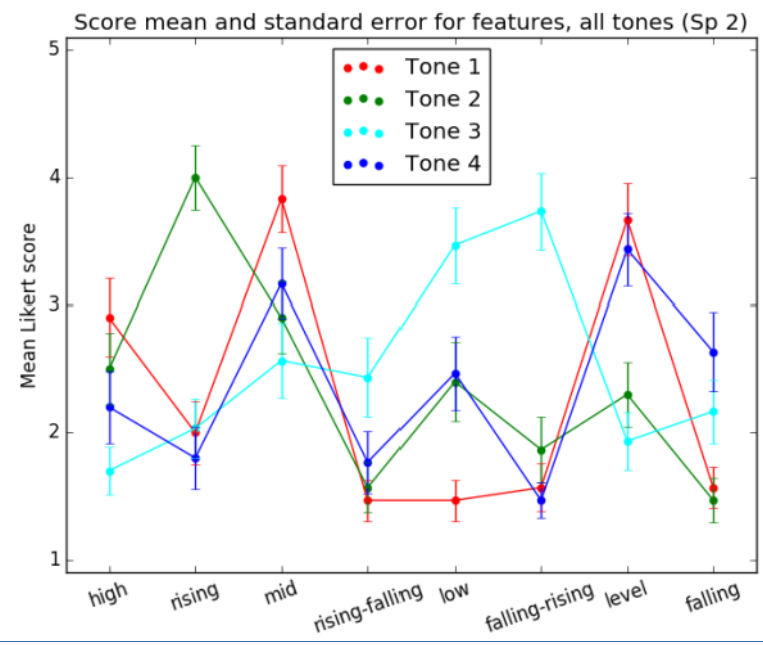
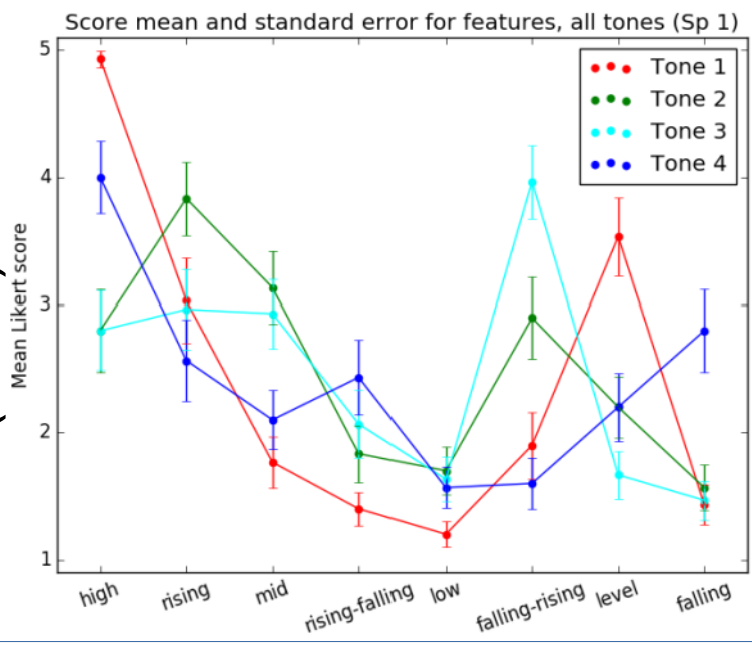
Mandarin  
responders

(n=33)



German (&c.)  
responders

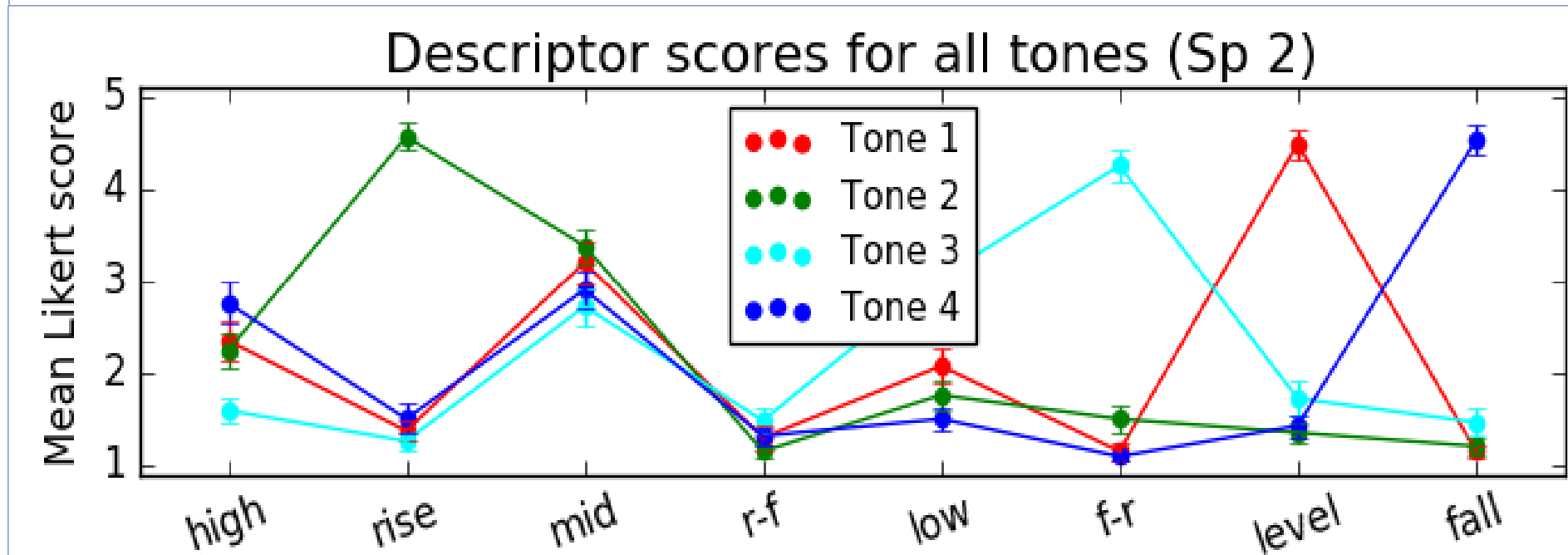
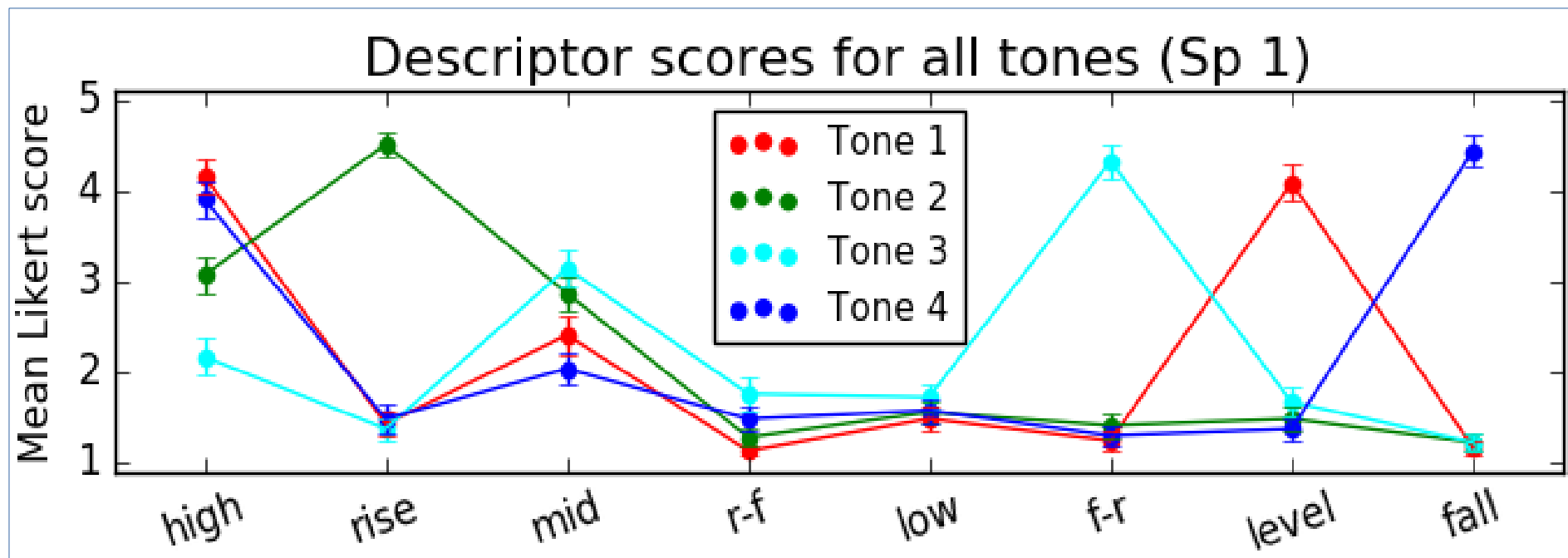
(n=15)



### Descriptor results

- Inter-speaker variability
    - *mid* varies for tones, not so much for speakers
  - low varies strongly
    - for Speaker B
    - not for Speaker A
  - *high* score
    - stronger for Speaker A
    - not for Speaker B
- overall higher pitch for Speaker A than for Speaker B?

## Variability in Mandarin Tone Perception



## Descriptor results

### 1. Contour descriptors:

- ‘canonical tone descriptors’:
  - high skewed distribution for high scores
  - cf. Mandarin tones:
    - Tone 1: *level*, Tone 2: *rise*, Tone 3: *fall-rise*, Tone 4: *fall*
    - Neutral tone not included
    - Distractor tone rise-fall: low

### 2. Height descriptors

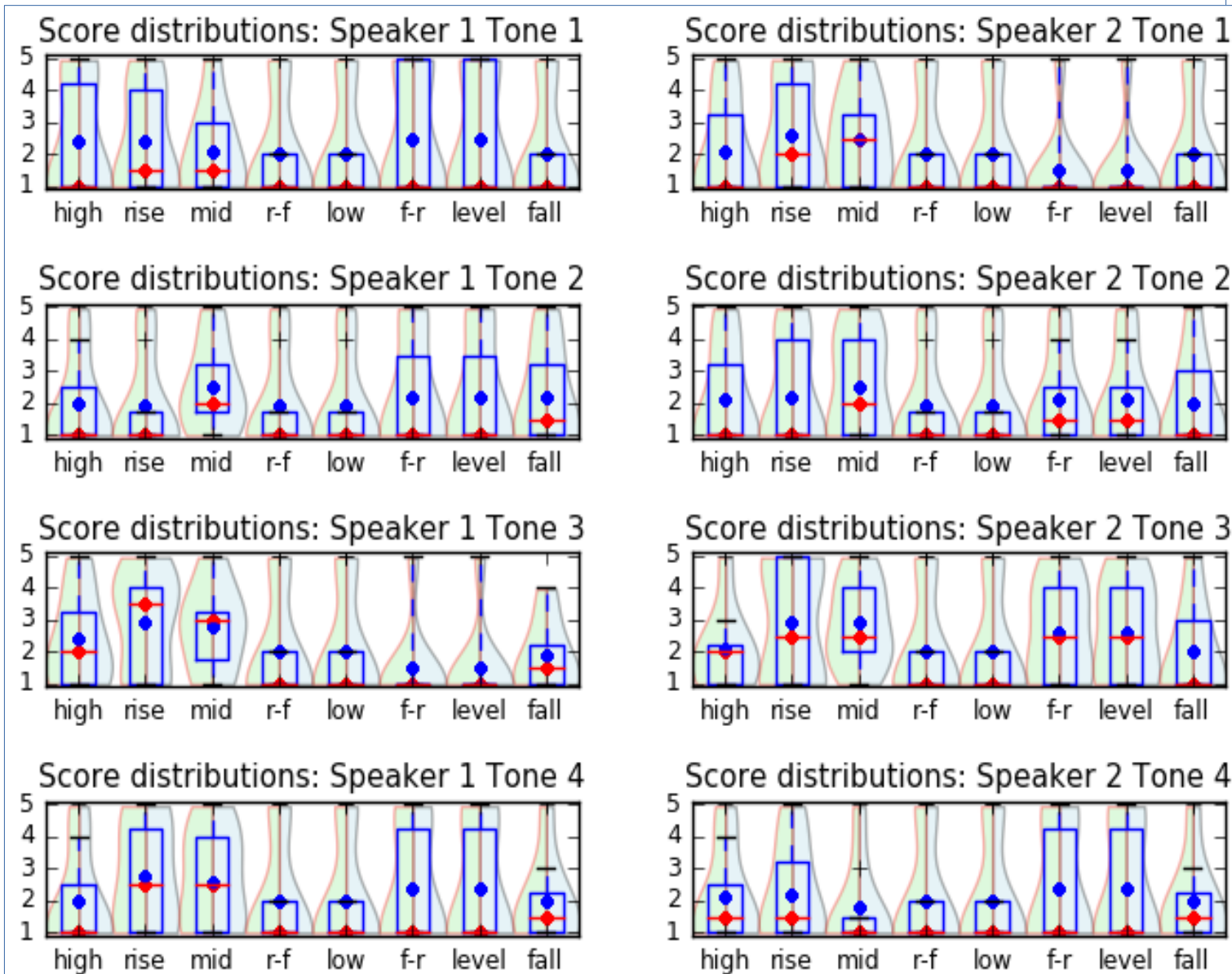
- ‘non-canonical’
  - high, mid, low
  - low scores:
    - skewed distributions
    - bimodal distributions
    - broad distributions

<b>SpA</b>	<b>high</b>	<b>rise</b>	<b>mid</b>	<b>r-f</b>	<b>low</b>	<b>f-r</b>	<b>level</b>	<b>fall</b>
<b>T1</b>	4.12	1.41	2.39	1.17	1.47	1.27	3.92	1.20
<b>T2</b>	2.95	4.52	2.74	1.27	1.52	1.41	1.45	1.24
<b>T3</b>	2.11	1.36	3.03	1.85	1.70	4.27	1.59	1.23
<b>T4</b>	3.79	1.45	1.97	1.53	1.64	1.32	1.39	4.52
<b>SpB</b>	<b>high</b>	<b>rise</b>	<b>mid</b>	<b>r-f</b>	<b>low</b>	<b>f-r</b>	<b>level</b>	<b>fall</b>
<b>T1</b>	2.32	1.36	3.29	1.36	1.97	1.21	4.24	1.18
<b>T2</b>	2.11	4.58	3.32	1.18	1.70	1.47	1.33	1.21
<b>T3</b>	1.56	1.27	2.71	1.59	3.00	4.14	1.70	1.44
<b>T4</b>	2.77	1.44	2.68	1.30	1.47	1.11	1.45	4.58

# Variability in Mandarin Tone Perception

## Descriptor results

Kernel  
density plots  
(violin plots)





## Descriptor results

### 1. Inter-speaker variability:

- some inter-speaker variability
  - pitch height descriptor *mid* varies for tones but not so much for speakers
  - *low* varies strongly for Speaker B but not for Speaker A
  - higher score of *high* for Speaker A: overall higher pitch

### 2. MANOVA

- fixed factors: tone type, pitch descriptor, dialect, speaker, with interactions
- significant effects: dialect region, pitch descriptor
- strong interactions
  - tone + descriptor, speaker + descriptor
  - dialect + tone + shape (multiinteraction)

## Variability in Mandarin Tone Perception

Descriptor results: MANOVA

Fixed factors:

- tone type, pitch descriptor, dialect , speaker, with interactions

Significant effects:

- dialect region, pitch descriptor

Strong interactions:

- tone + descriptor, speaker + descriptor
- multiinteraction: dialect + tone + shape

<b>Factors</b>	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F</b>	<b>p</b>
<b>dial</b>	16	480	30	12.966	<0.001
<b>descr.</b>	1	123	123.19	53.252	<0.001
<b>tone:descr.</b>	1	130	130.39	56.365	<0.001
<b>sp.:descr</b>	1	38	38.26	16.54	<.0001
<b>dial:tone:descr</b>	16	89	5.58	2.413	<0.01

## Hierarchical Clustering

### Method:

- a distance matrix is maintained at each iteration; the  $d[i,j]$  entry corresponds to the distance between cluster and components (here: Pearson distance)
- distance matrix updated to reflect distance of the newly formed cluster with remaining clusters
- classifiers:
  - Nearest Point Algorithm.
  - Farthest Point Algorithm (Voor Hees Algorithm)
  - Unweighted Pair Group Method with Averaging
  - Weighted Pair Group Method with Averaging
  - Unweighted Pair Group Method with Centroid Averaging (Median)
  - Weighted Pair Group Method with Centroid Averaging (Median)
  - Ward variance minimization (incremental)

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html>

## HIERARCHICAL CLUSTERING

Pearson Distance classification, 7 clustering algorithms

→ Ward incremental variance minimisation

## Comparison with geographical location

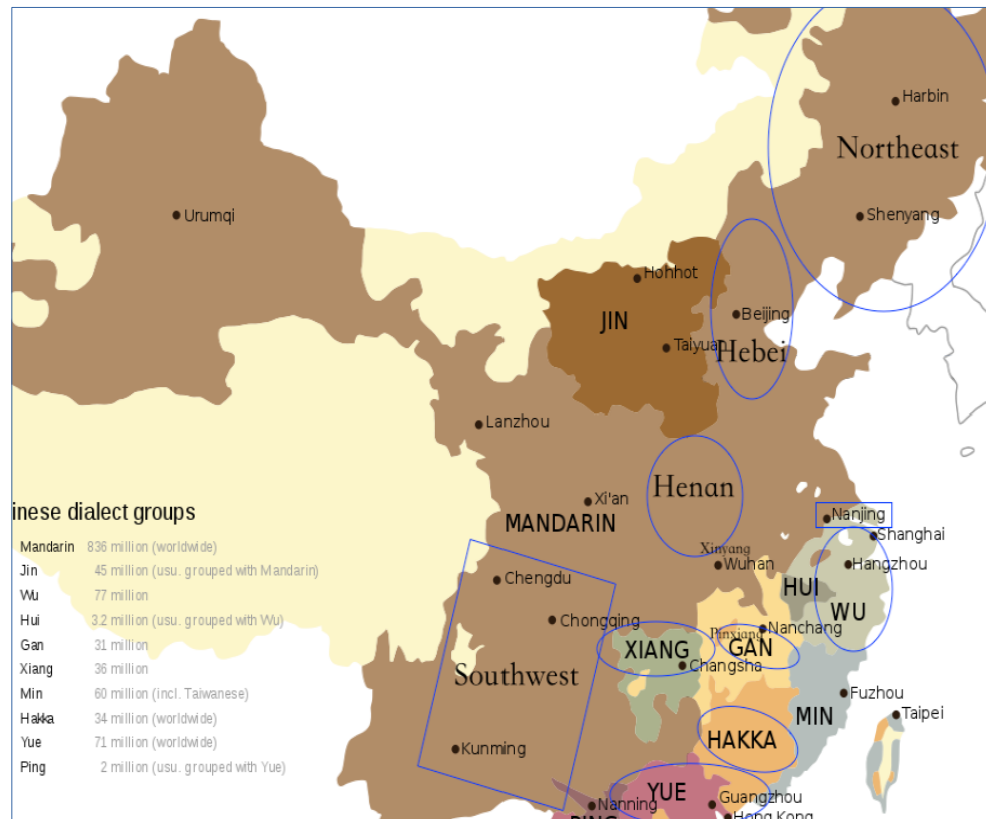
- Shandong+Hebei:
  - geographically close, N. Mandarin
- Hunan, Hakka, Henan:
  - geographically close, historically related
- others:
  - prosodic typology partly plausible, geography and history less so

Noise due small data set with large number of classes

- inaccuracies and normative element in self-ascription
- language graduates, strong influence of standard Mandarin

# Variability in Mandarin Tone Perception

## HIERARCHICAL CLUSTERING

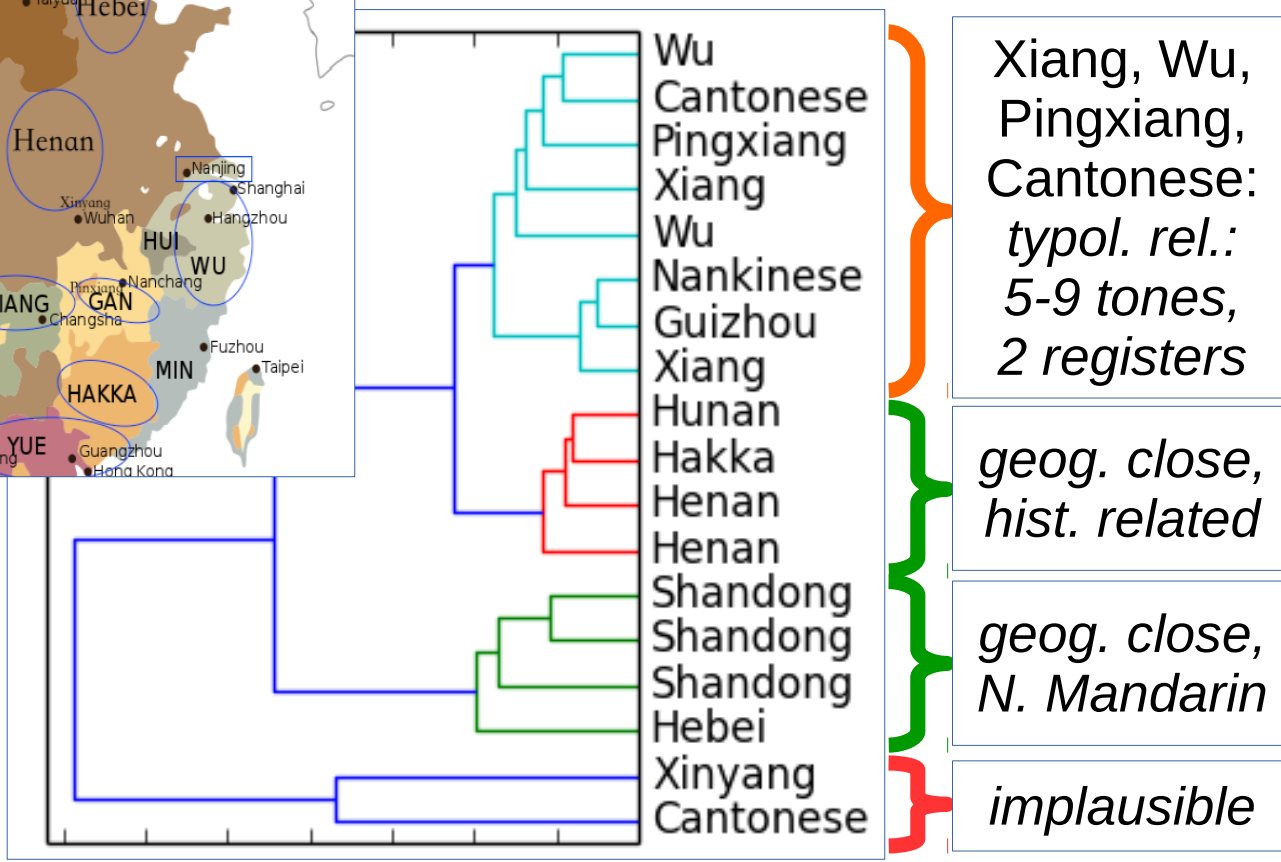


Chinese dialect groups

Mandarin	836 million (worldwide)
Jin	45 million (usu. grouped with Mandarin)
Wu	77 million
Hui	3.2 million (usu. grouped with Wu)
Gan	31 million
Xiang	36 million
Min	60 million (incl. Taiwanese)
Hakka	34 million (worldwide)
Yue	71 million (worldwide)
Ping	2 million (usu. grouped with Yue)

Geographically and/or historically distant pairs:

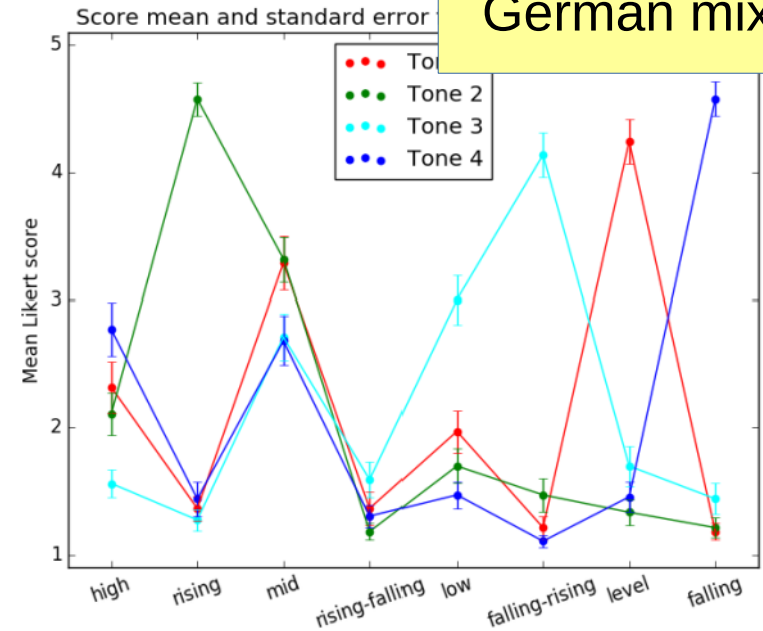
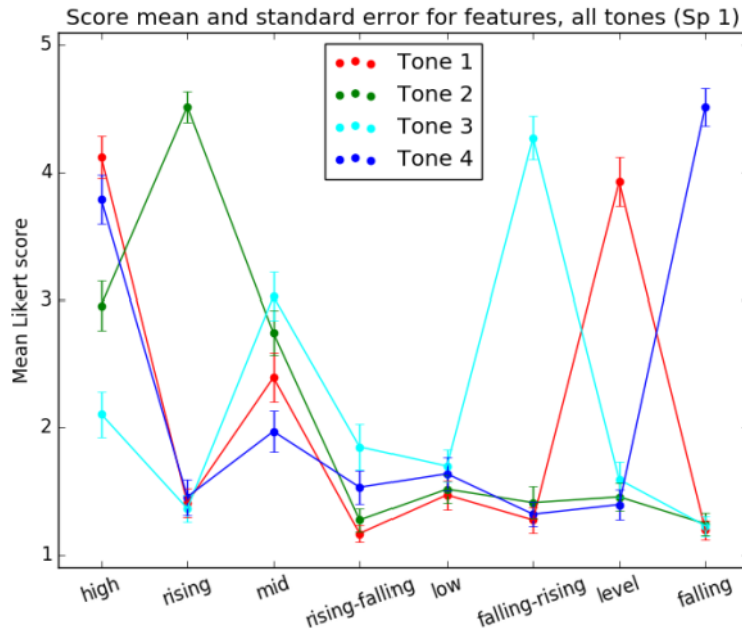
WU-Cantonese, Nankinese-Guizhou, Henan-Hebei-Wu, Henan-Hebei, Xiang-Wu, Xinyang-Cantonese



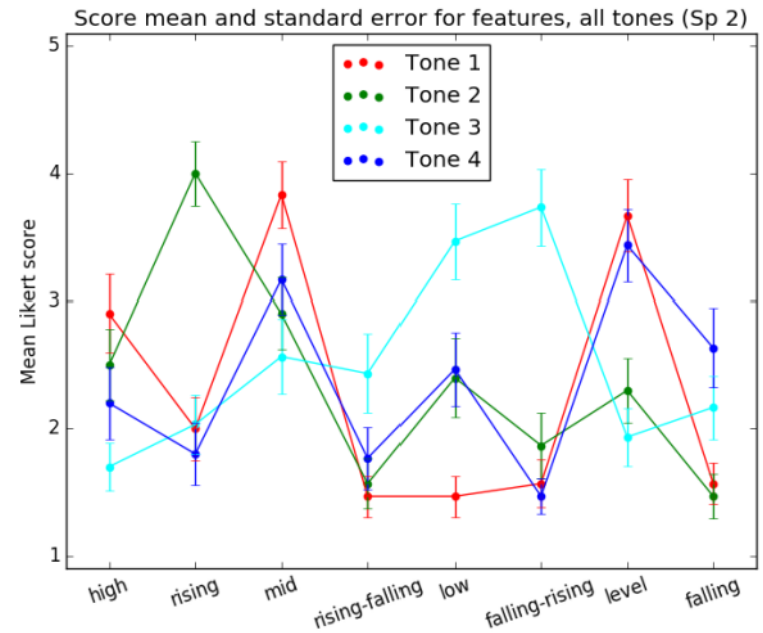
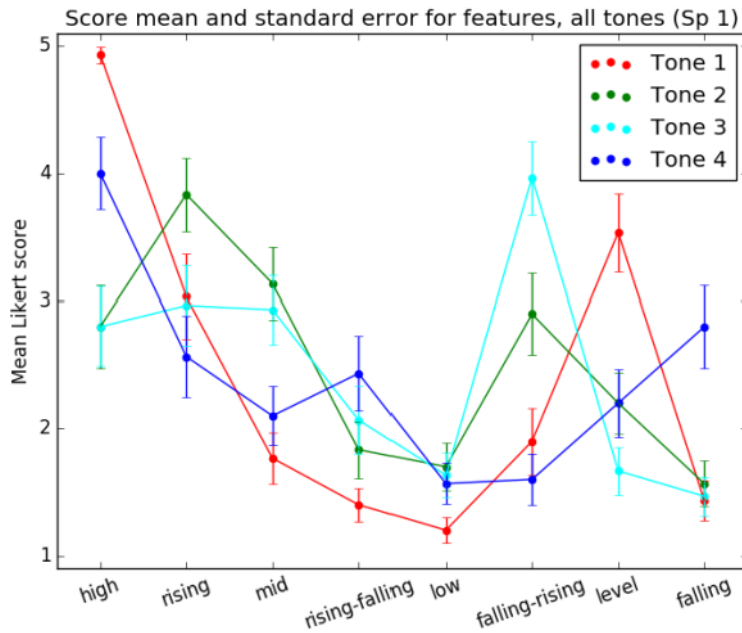
# Variability in Mandarin Tone Perception

Mandarin  
VS.  
German mixed

Mandarin  
responders

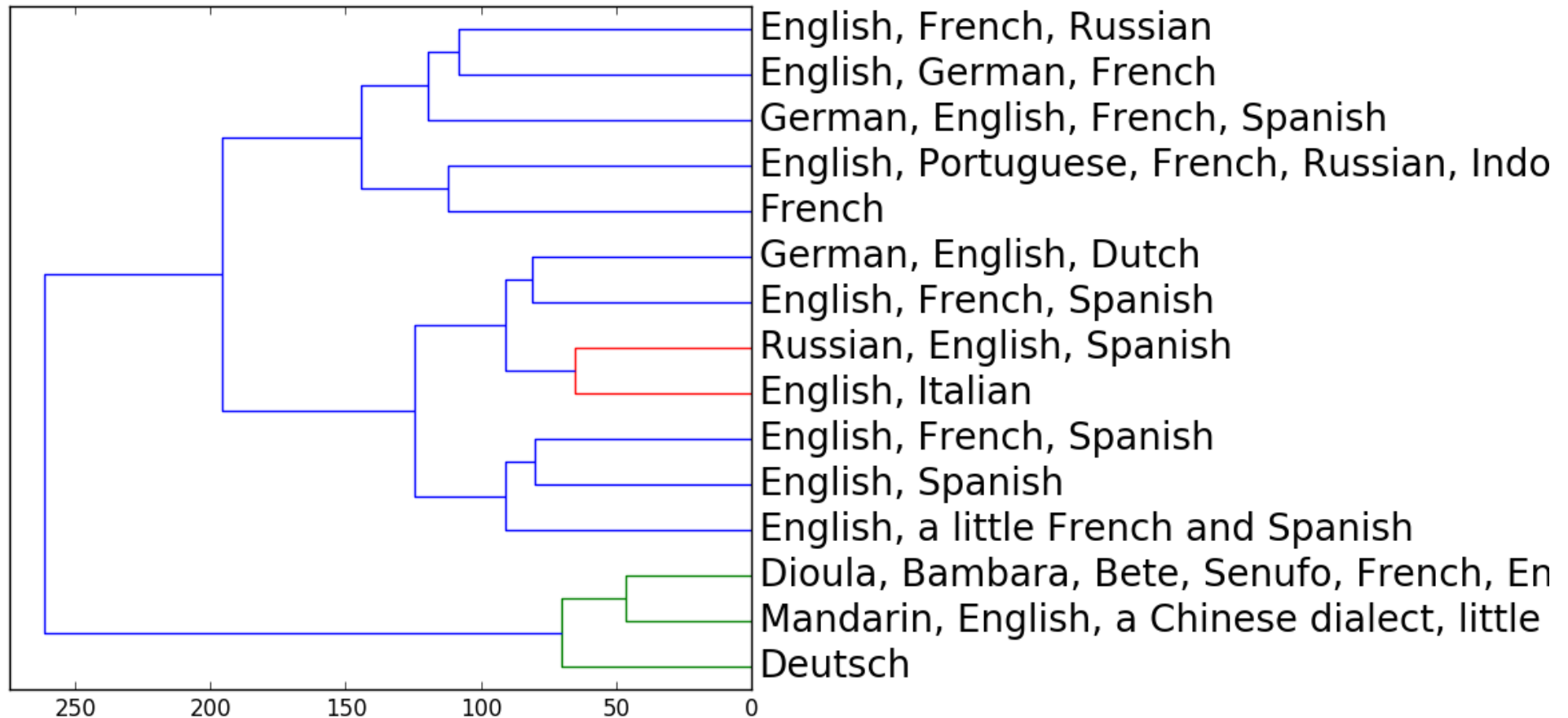


German (&c.)  
responders



## HIERARCHICAL CLUSTERING

An experiment with an *ad hoc* group of German linguistics students, plus



### **Conclusions on tone descriptor assignment**

#### Main descriptive outcomes

- expected: contour unlike height descriptors
  - canonical descriptors more consistent: categorial perception
- significant effects
  - dialect, descriptor; interactions for tone + descriptor, speaker + descriptor, dialect + tone + descriptor

#### Classification

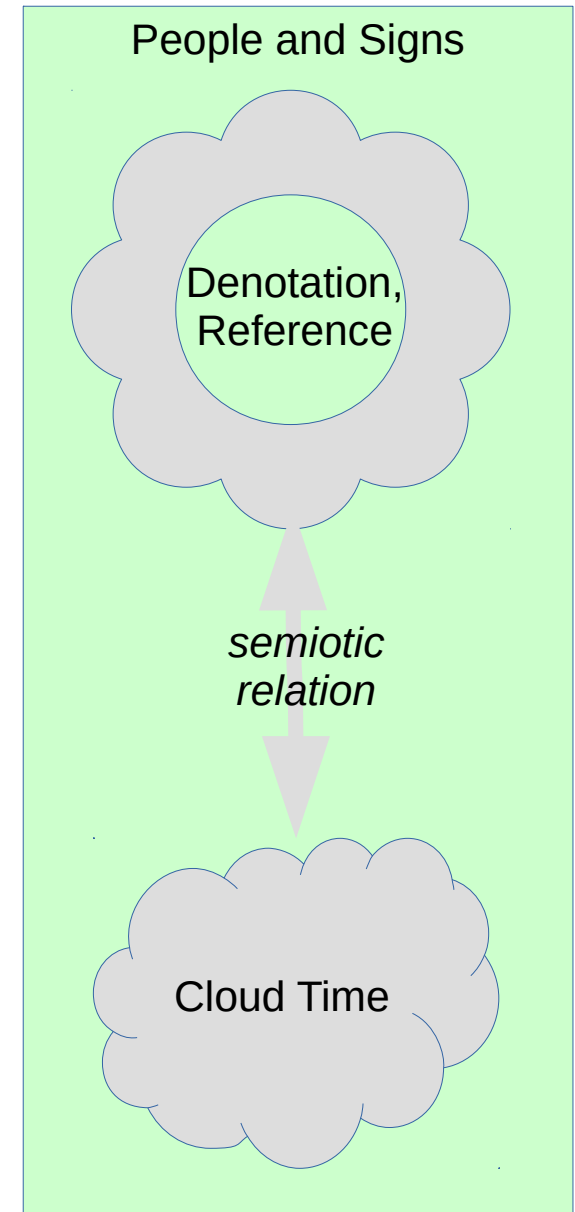
- partly plausible classification results
- despite small dataset – but more data needed

#### Main strategic outcome

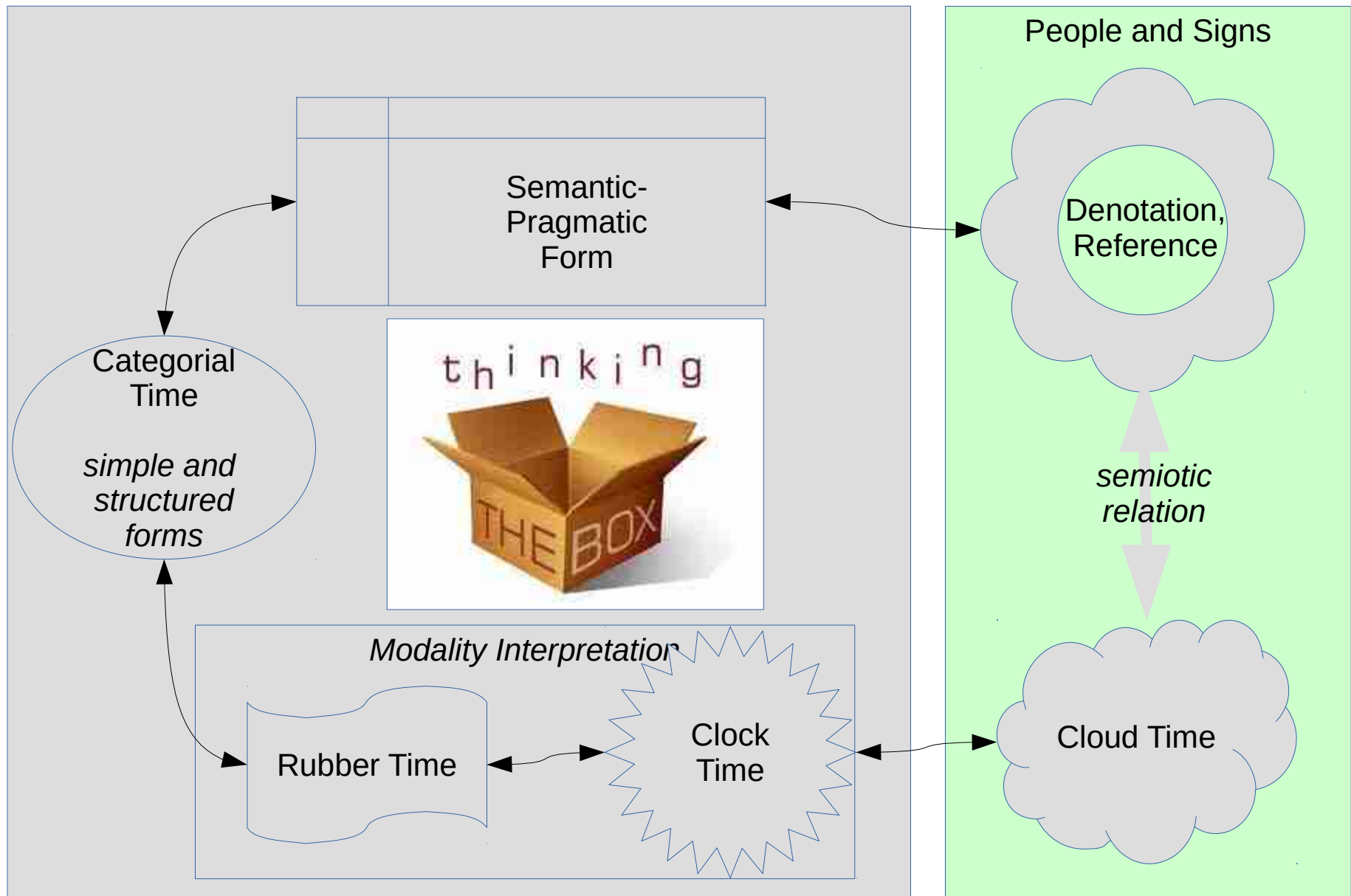
- the novel method is fit for purpose for planning
  - a larger dialect survey
  - more complex contextual data: tone sandhi, accent, intonation
  - more systematic dialect classification for self-ascription
  - more speakers, gender balance, socio-economic information



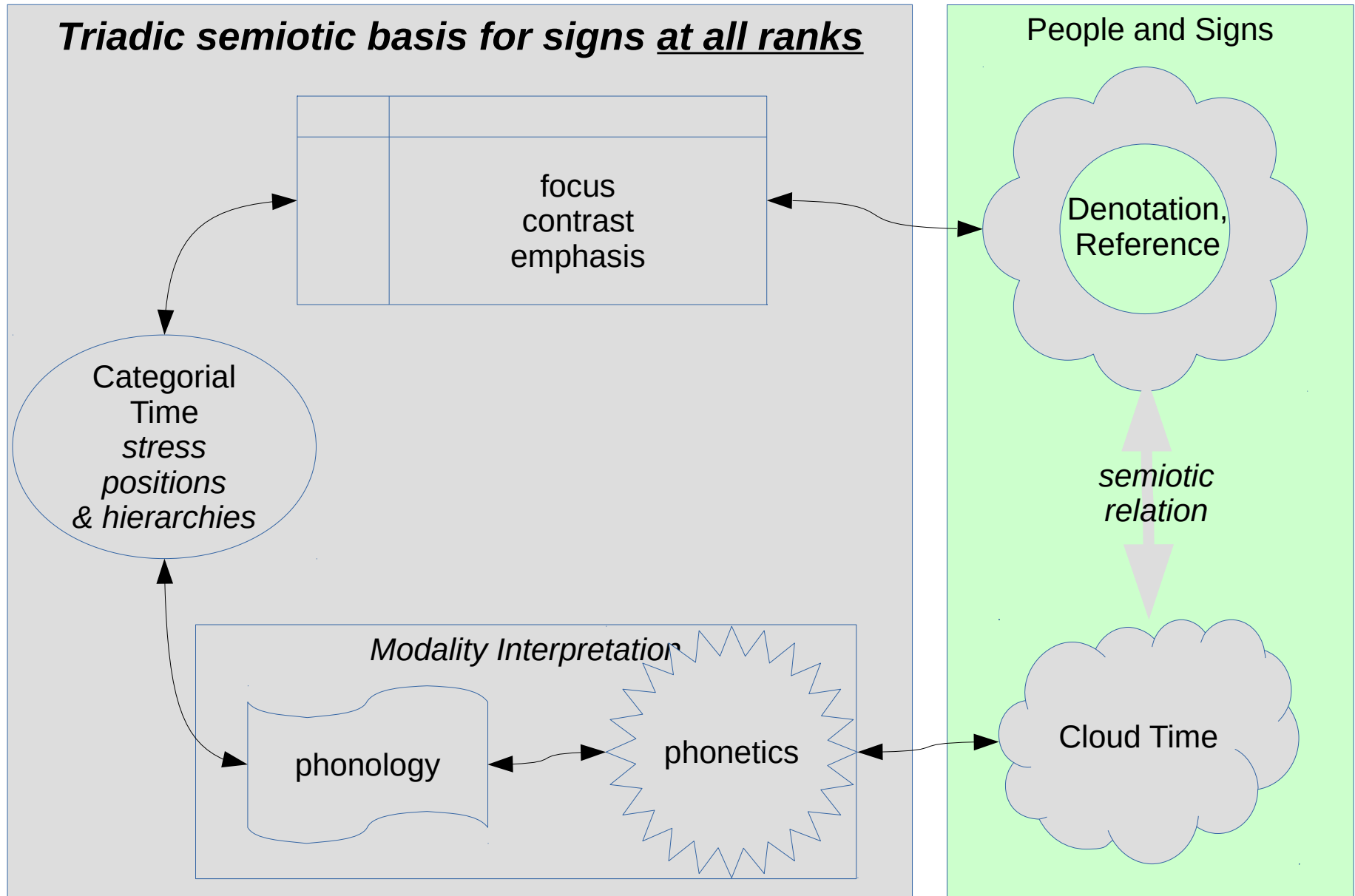
# Rank-Interpretation Model of Language



# Rank-Interpretation Model of Language



# Rank-Interpretation Model of Language



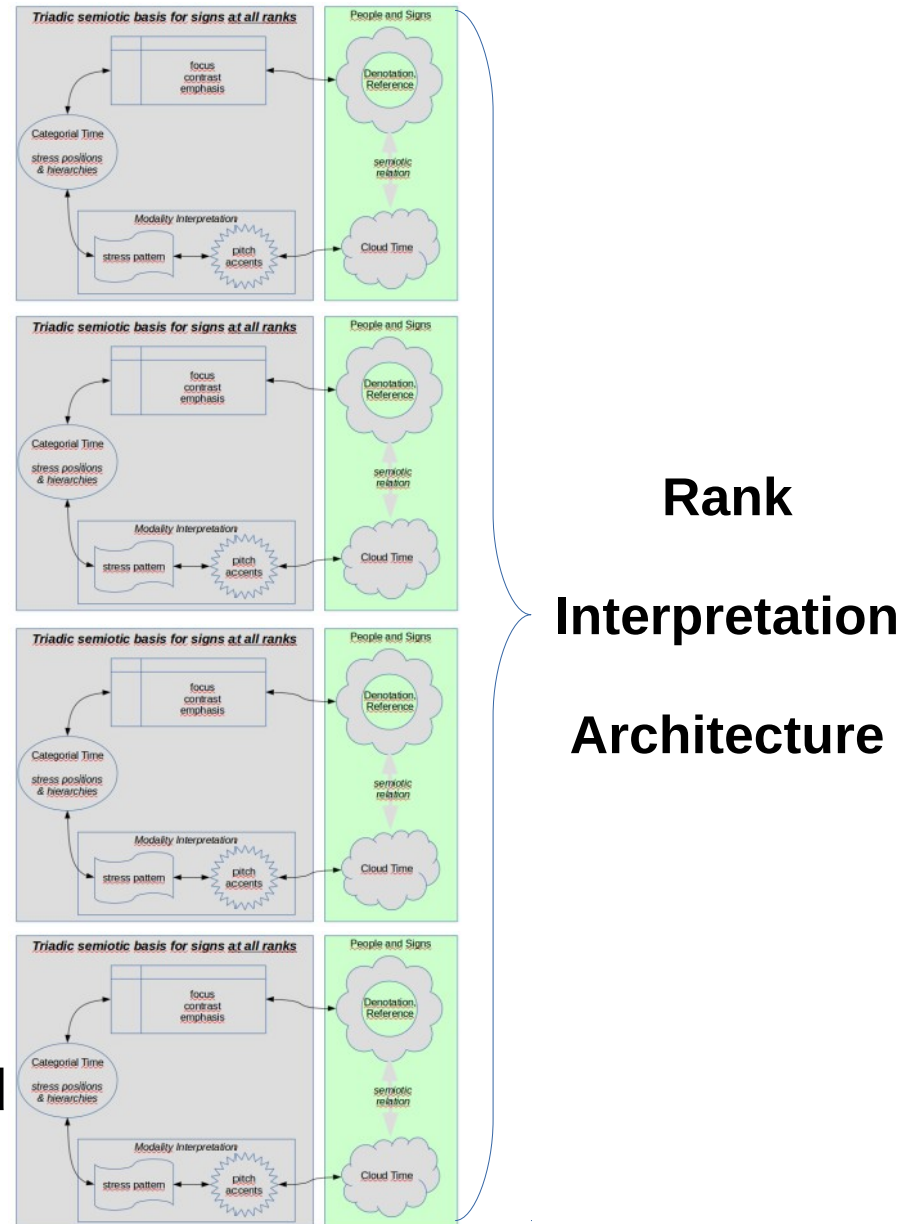
# Rank-Interpretation Model of Language

Discourse: Monologue, Dialogue

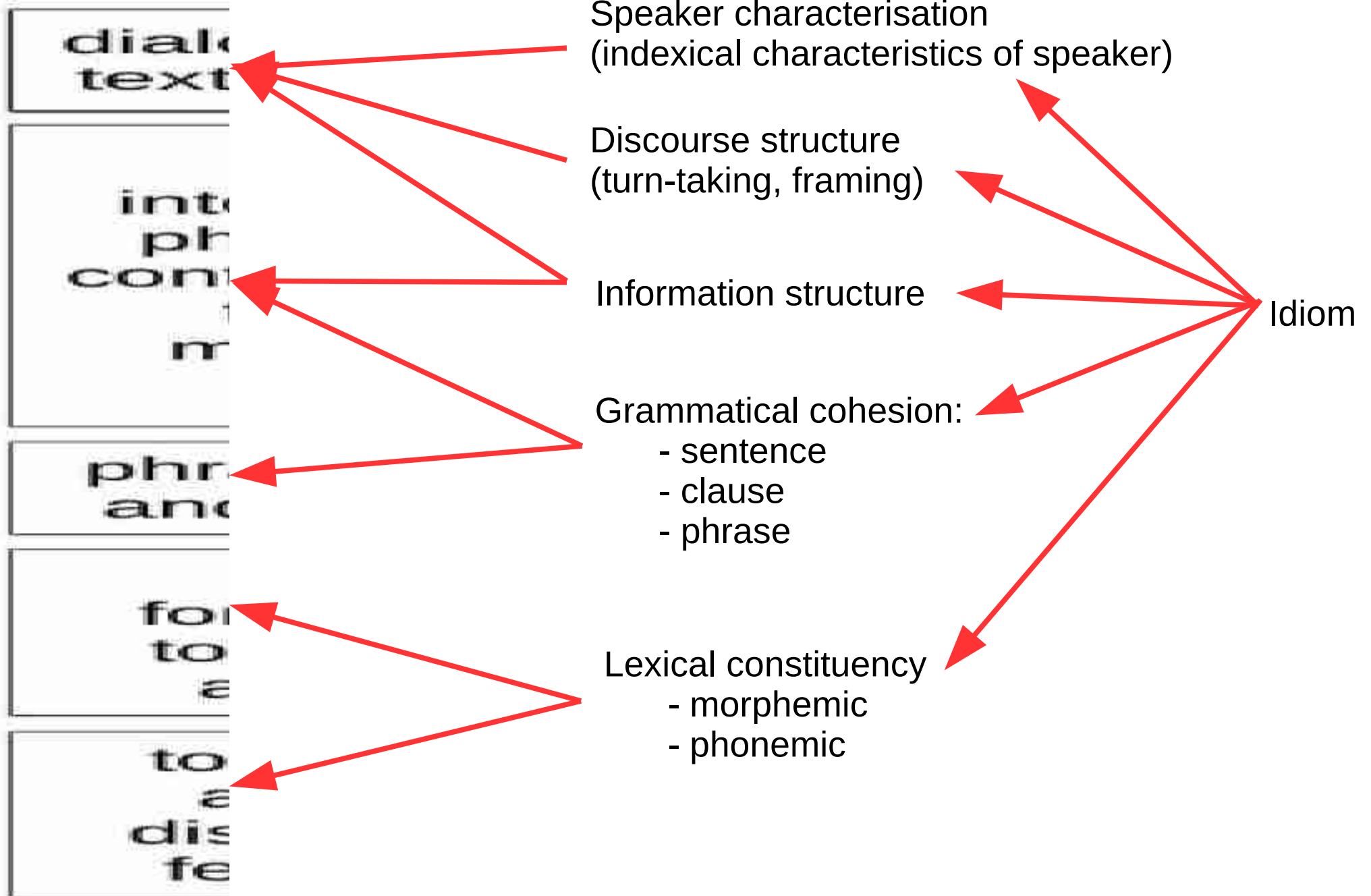
Utterance: turn, IPU, ...

Sentence, phrase, clause

Word: simple, inflected, compound, derived



# The Rank Interpretation Architecture - Prosody



# *Discourse Based Prosodic Analysis*

## **Discourse functions**

- discourse framing
  - “call contours”
  - strategic use of hesitation phenomena, vocalisations
- turn-taking continuity
  - start with high pitch
  - end with low pitch
- dialogue act and speech act marking marking
  - adjacency pair marking
    - e.g. question-answer pattern

## "semantico-pragmatic effects":

- structure:
  - discourse segmentation, topic structure
  - parallelism between mentioned items
  - subordination relationships between propositions salient in the discourse
  - topic shift, digression, interruption
  - turn-taking
- semantics:
  - disambiguation of ambiguous utterances (MW: scope)
  - appropriate choice of referent (reference resolution)
  - distinction between 'given' and 'new' information (information status: given/new, topic/comment, focus/presupposition)
  - conceptual contrast
  - indirect speech acts (MW: other speech acts, too)

**"The central thesis of this work is that there are many ways in which intonation helps to structure discourse."**

## Discourse structure marking

- **linguistic** structure (phrasing, framing)
  - pitch register, pitch range
- **intentional** structure (purposes, speech acts)
  - pitch accent contour type
  - boundary tone type
- **attitudinal** state (objects, properties, relations, and discourse intentions that are most salient at any given point)
  - accent placement, focus, contrast, emphasis
  - given/new, theme/rheme

Grosz, B. J. and C. L. Sidner. 1986. Attention, intentions and the structure of discourse. BBN report.



# *Intentional structure: prosody and speech acts*

1. Grice 1975: conversational maxims

2. Austin 1962 & Searle 1969:

– **locutionary acts:**

- meaning: modality, mood, possibility, predicate & arguments
  - lexical morphemic tone; phrasal intonational meaning

- metallocutionary acts

- marking of properties of locutions (boundary tones, accents, ...)

– **illocutionary acts:**

- interactive creation of new bond between interlocutors
  - question, promise, command; marriage, official appointment, ...

– **perlocutionary acts:**

- creation of an effect by the speaker on the hearer
  - impress, disappoint, interest, excite, bore; praise, insult, ...

## ***Discourse Prosody Case 1: Metalocutionary Framing - Calls***

# *Intonation meaning is 'metallocutionary'*

## 1. Paralinguistic metallocutionary channel

- two aspects:
  - gradient constraints on pitch/intensity/tempo variation
  - affect, sentiment, attitude
- not necessarily automatic: can be imitated

## 2. Linguistic metallocutionary channel

- information marking – a rough correspondence:

<b>Prague school</b>	<b>Halliday</b>	<b>ToBI</b>
delimitative	- tonality	- boundary assignment
culminative	- tonicity	- tone assignment
distinctive	- tone	- tone

## ***Metalocutionary discourse framing:***

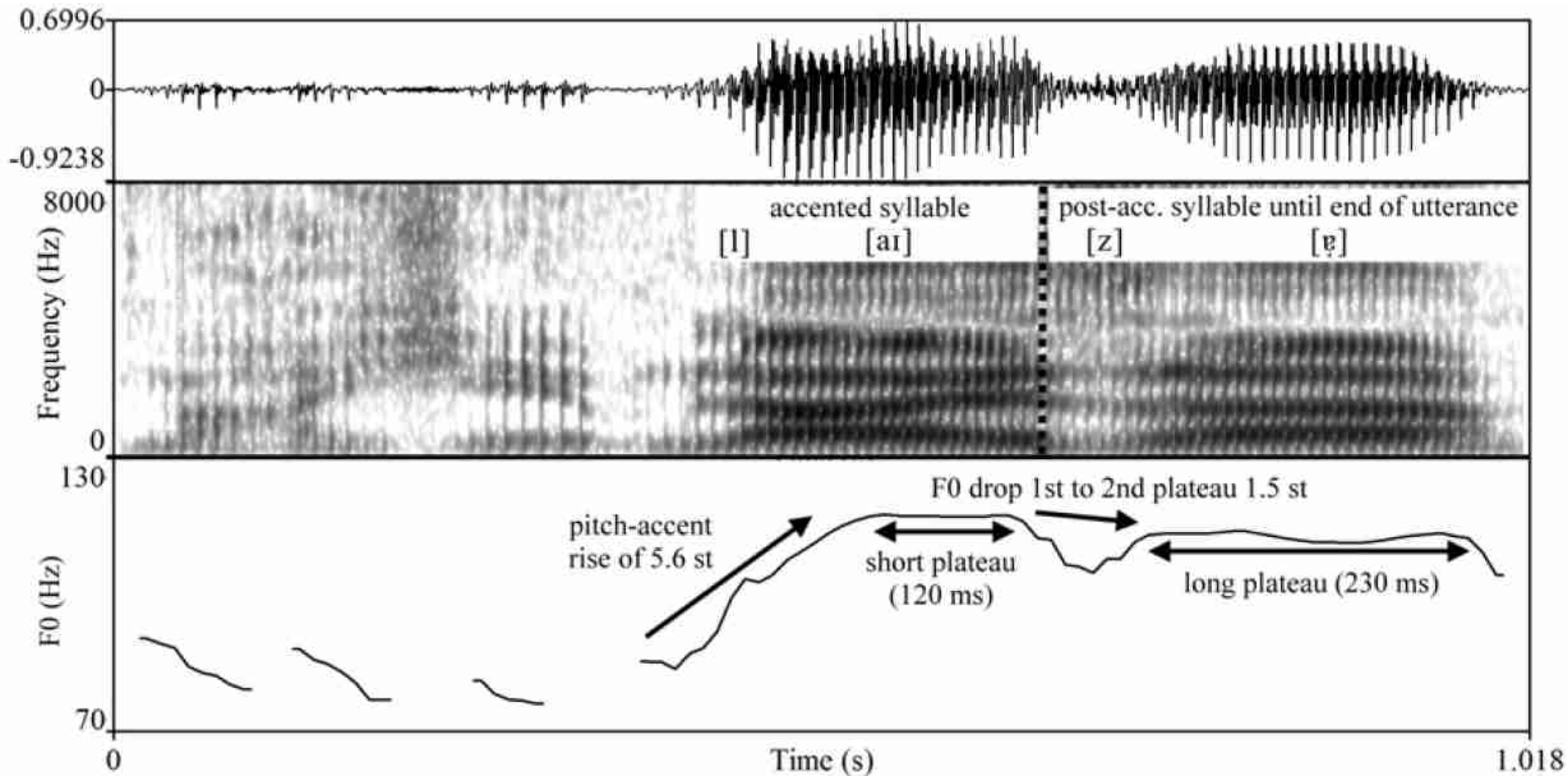
### 3 basic conditions on speech acts:

- Uptake condition / Channel condition
  - Normal input and output conditions obtain
- Essential condition
  - commitment
- Sincerity condition
  - truth – probability – certainty

### Example – “call contours”

- Only discourse framing:
  - Start: “Jooohn-nee!”
  - End: “Byyy-eee!”
  - \* Yesterday I saw Jooohn-nee in town.

# *Metalocutionary discourse framing: German ‘call contour’*



“Dann mach ich eben leiser!”

Niebuhr, O. 2013. Resistance is futile – the difference between continuation rise and falling contour in German. *Interspeech*.



## ***Discourse Prosody Case 2: Discourse Rhythms***

## ***Discourse Prosody Case 2: Discourse Rhythms***

AM and FM spectra:

If a spectrum can be derived from the **AM envelope**, why not derive a spectrum from the **FM track** and see whether they correlate?

Preliminary answer:

**Yes, they do correlate, but not overwhelmingly strongly, and depending on which subspectra are measured.**



## Discourse Prosody Case 2: Discourse Rhythms

### English

A0101B	0.996	0.626	0.645
A1202B	0.993	0.597	0.368
NW048	0.95	0.561	0.549

### Mandarin

jiayan	0.994	0.694	0.632
wuxi	0.991	0.561	0.48

12s: full spectrum  
over the selected  
signal  
suspiciously strong  
correlations

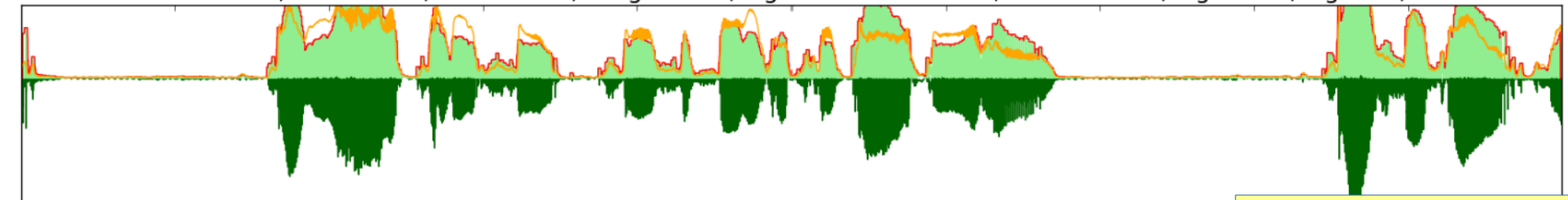
12s: spectrum  
0...20Hz  
(1200...50ms)  
more interesting  
weaker correlations

12s: spectrum  
0...10Hz  
(1200...100ms)  
very interesting  
weaker correlations

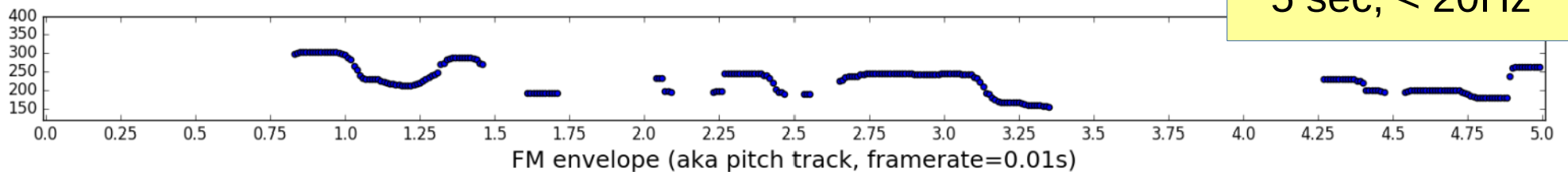
# Discourse Prosody Case 2: Discourse Rhythms

## AM & FM signals and spectra: jiayan

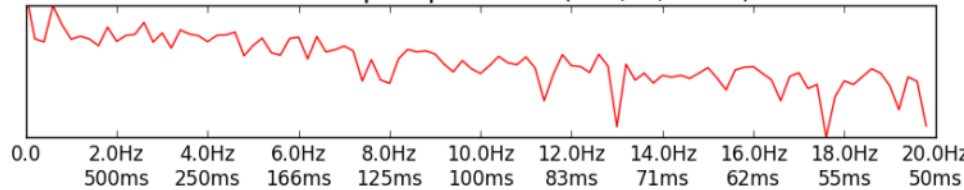
Params: minf0:120, maxf0:400, frame:0.01, weight:0.02, sigmedianfactor:100, f0median:15, sigstart:6, siglen:5, maxhz:20



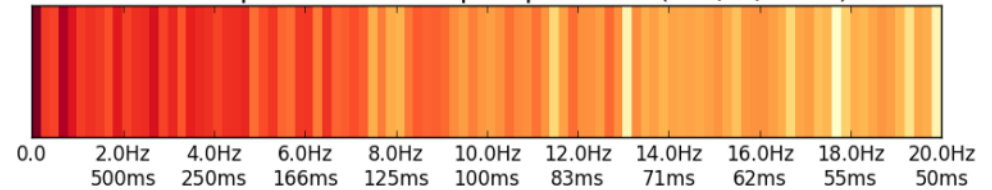
Mandarin, female  
5 sec, < 20Hz



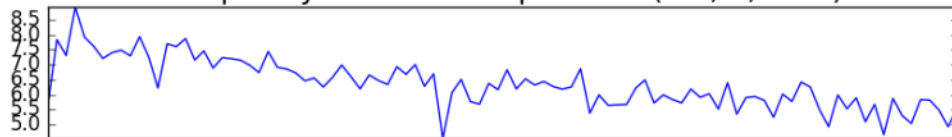
AM Envelope Spectrum (0.0,...,20Hz)



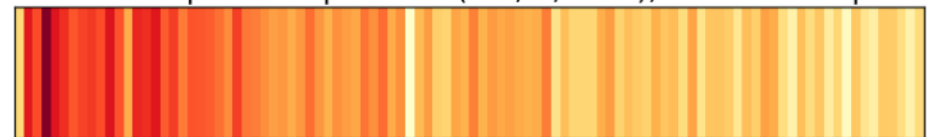
Heatmap of AM Envelope Spectrum (0.0,...,20Hz)



Frequency Modulation Spectrum (0.0,...,20Hz)



Heatmap of FM Spectrum (0.0,...,20Hz), with heatmap



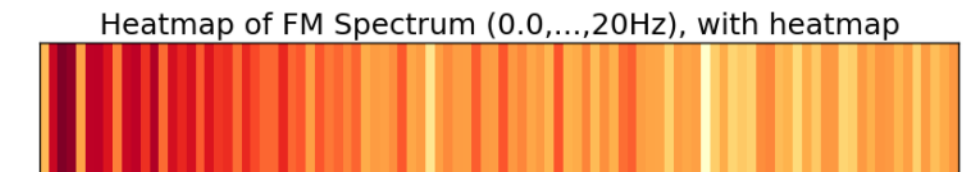
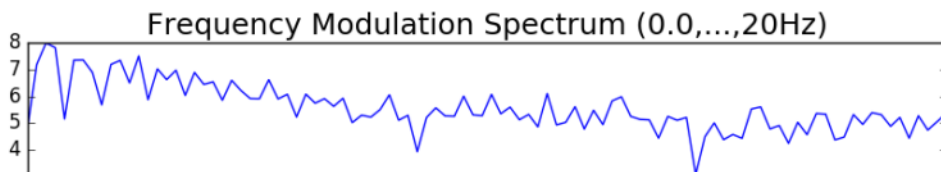
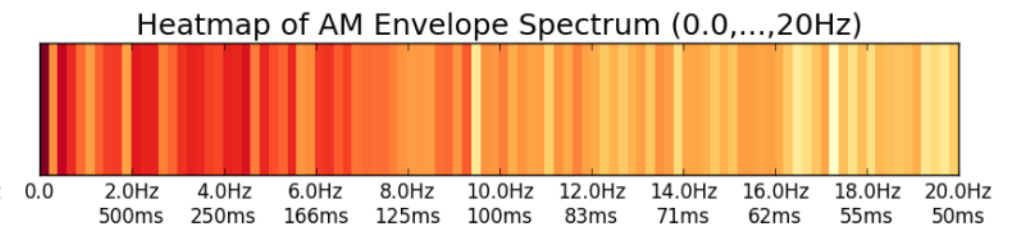
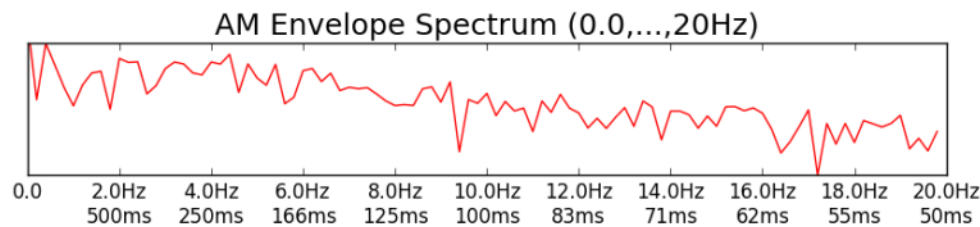
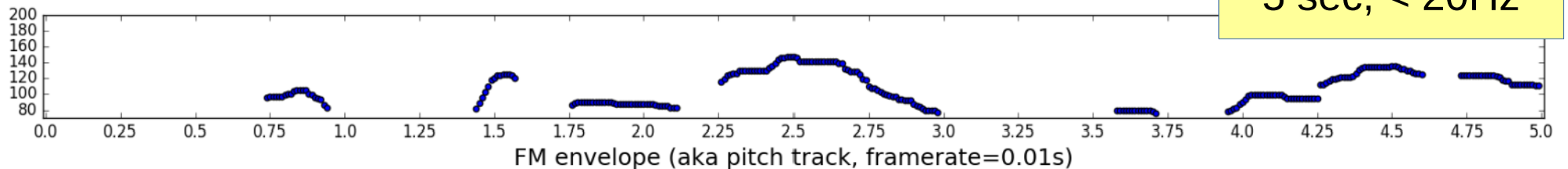
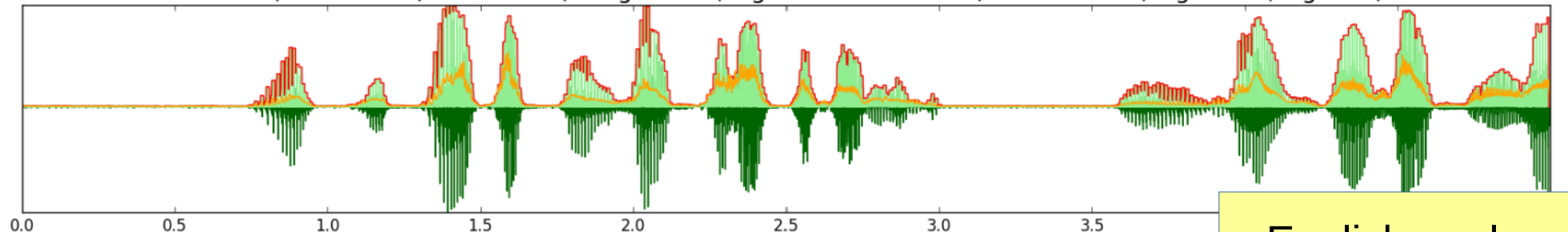
Correlation AME:FME=0.64

Correlation AMS:FMS=0.58

# Discourse Prosody Case 2: Discourse Rhythms

## AM & FM signals and spectra: Abercrombie\_English\_NW048

Params: minf0:70, maxf0:200, frame:0.01, weight:0.02, sigmedianfactor:100, f0median:15, sigstart:6, siglen:5, maxhz:20

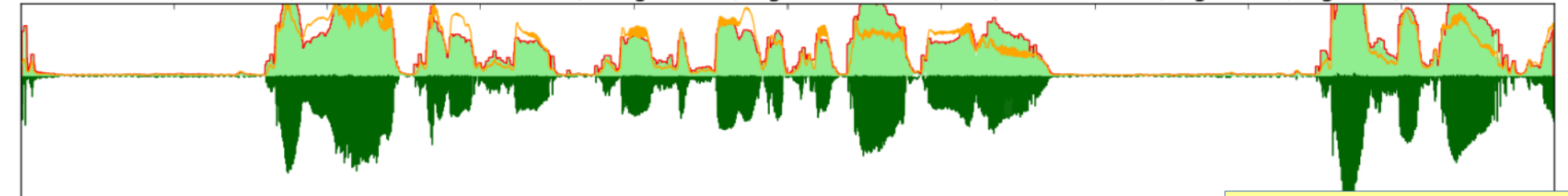


Correlation AME:FME=0.47  
Correlation AMS:FMS=0.56

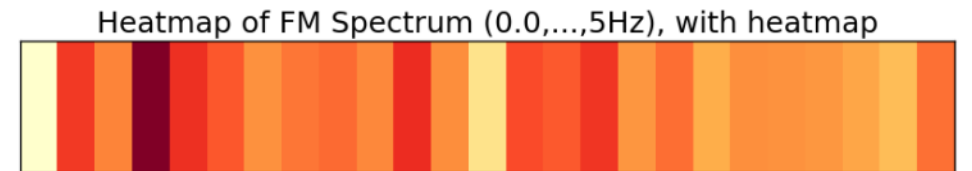
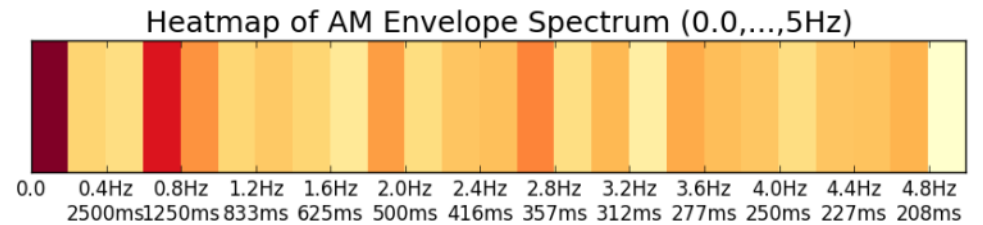
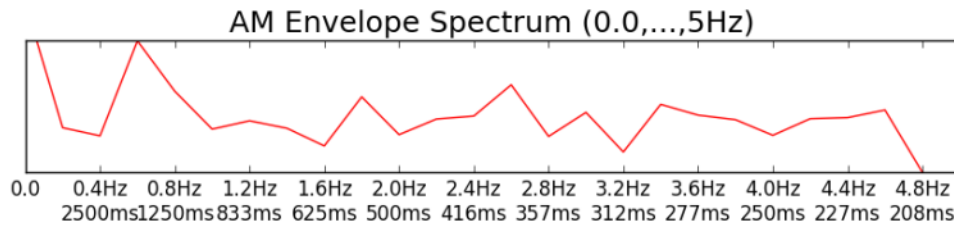
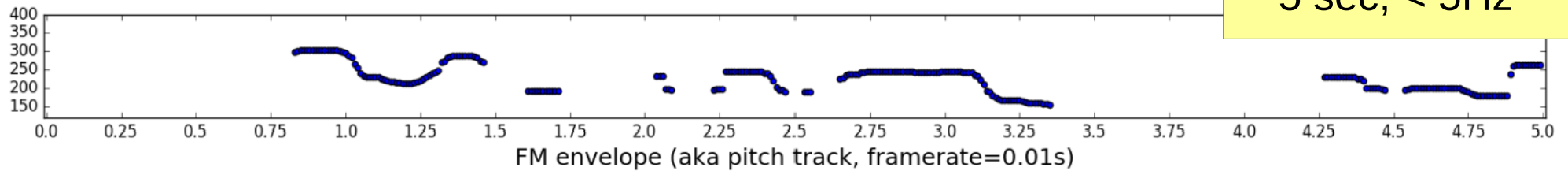
# Discourse Prosody Case 2: Discourse Rhythms

## AM & FM signals and spectra: jiayan

Params: minf0:120, maxf0:400, frame:0.01, weight:0.02, sigmedianfactor:100, f0median:15, sigstart:6, siglen:5, maxhz:5



Mandarin, female  
5 sec, < 5Hz

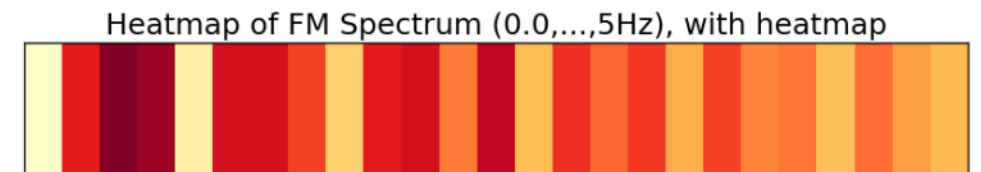
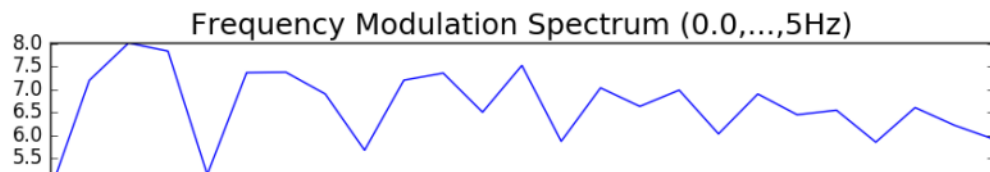
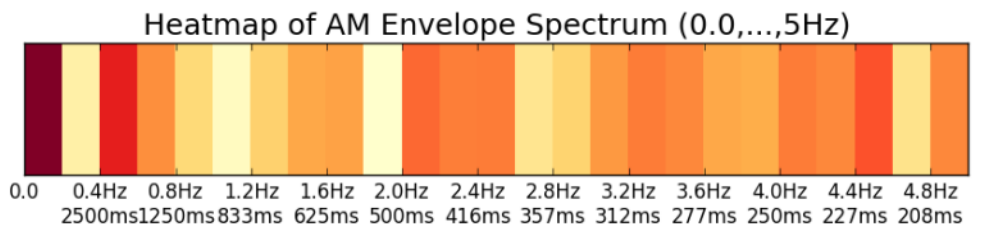
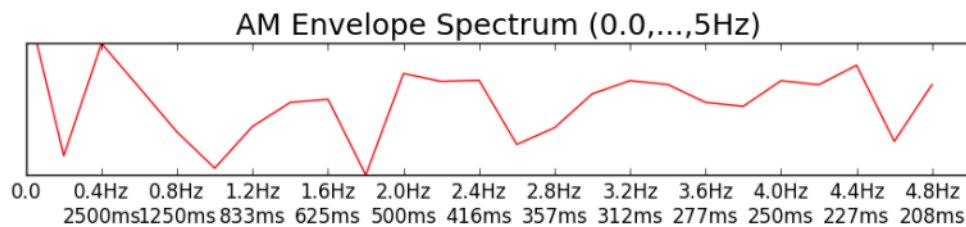
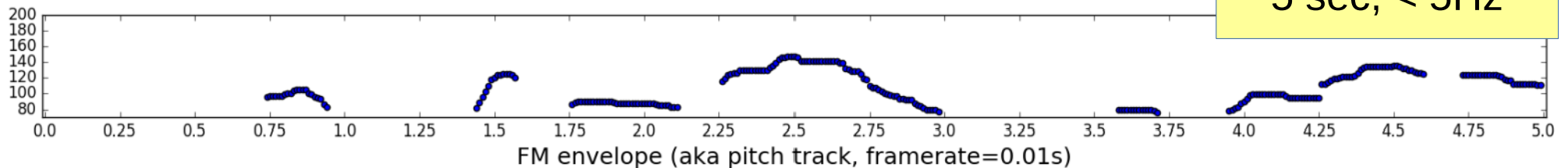
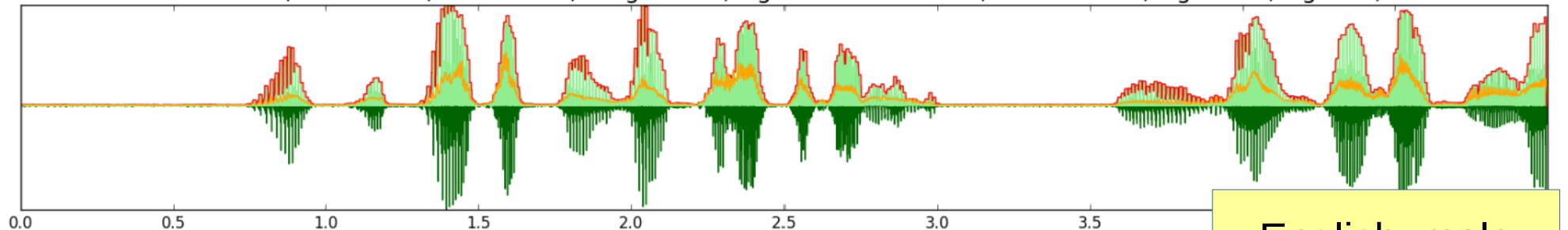


Correlation AME:FME=0.64  
Correlation AMS:FMS=-0.16

# Discourse Prosody Case 2: Discourse Rhythms

## AM & FM signals and spectra: Abercrombie\_English\_NW048

Params: minf0:70, maxf0:200, frame:0.01, weight:0.02, sigmedianfactor:100, f0median:15, sigstart:6, siglen:5, maxhz:5

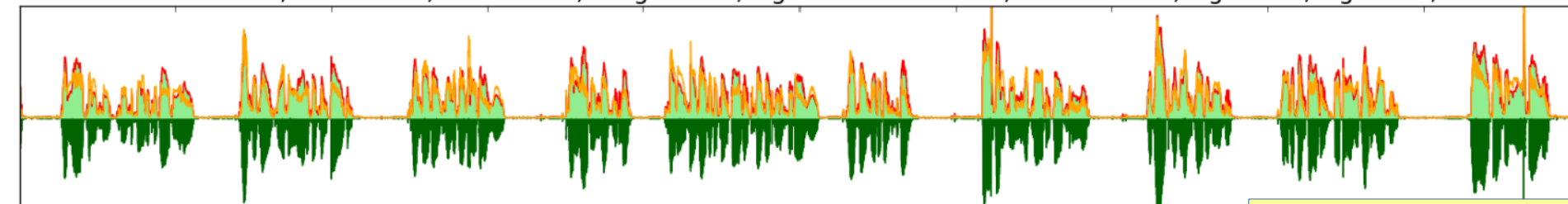


Correlation AME:FME=0.47  
Correlation AMS:FMS=-0.19

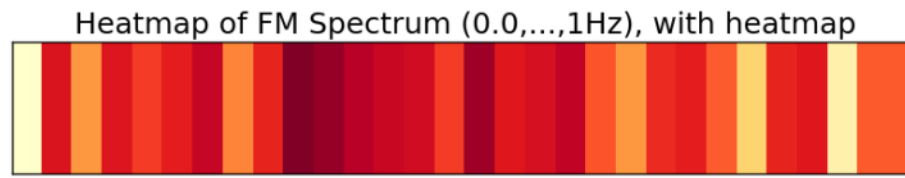
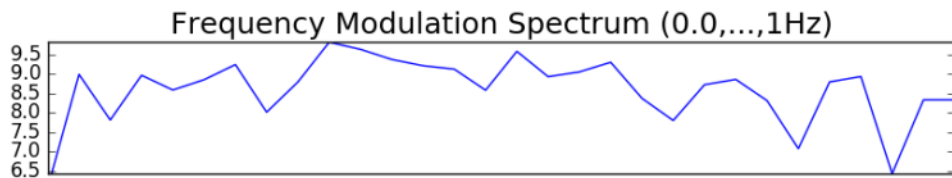
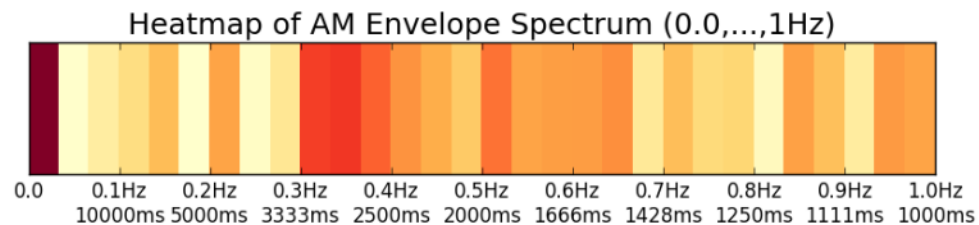
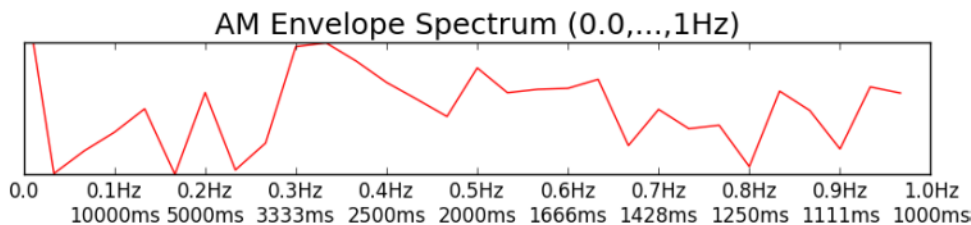
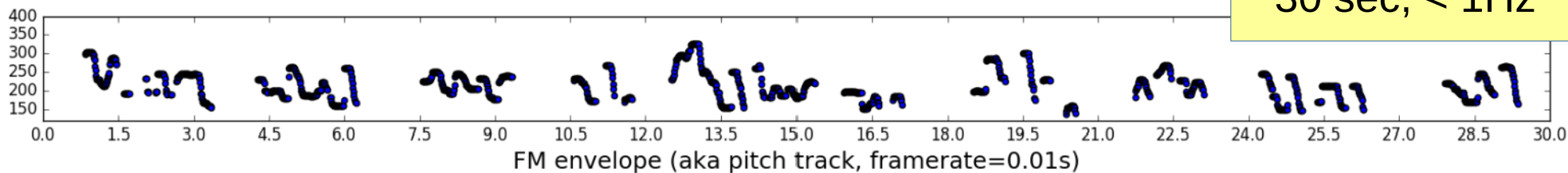
# Discourse Prosody Case 2: Discourse Rhythms

## AM & FM signals and spectra: jiayan

Params: minf0:120, maxf0:400, frame:0.01, weight:0.02, sigmedianfactor:100, f0median:15, sigstart:6, siglen:30, maxhz:1



Mandarin, female  
30 sec, < 1Hz

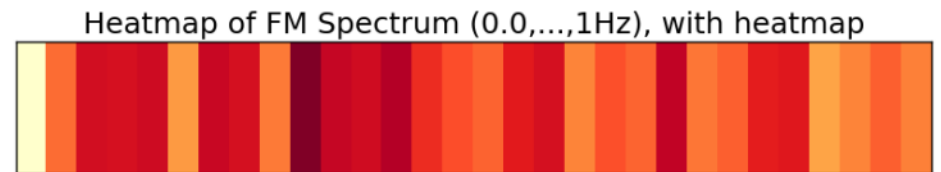
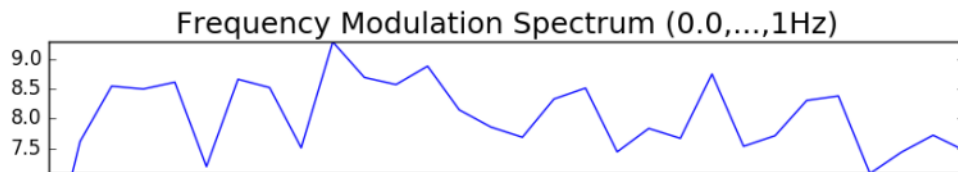
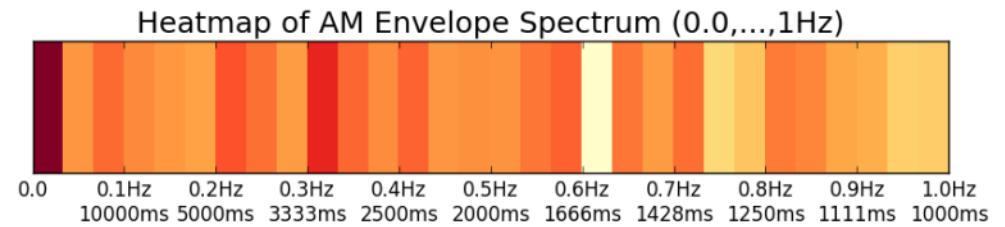
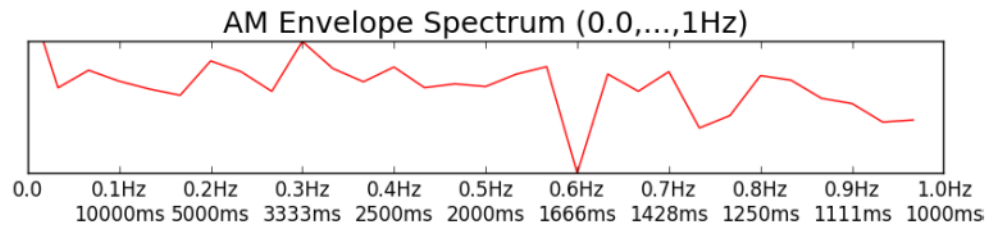
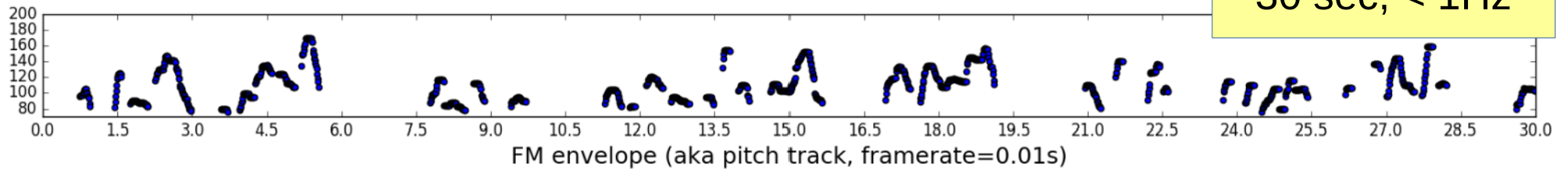
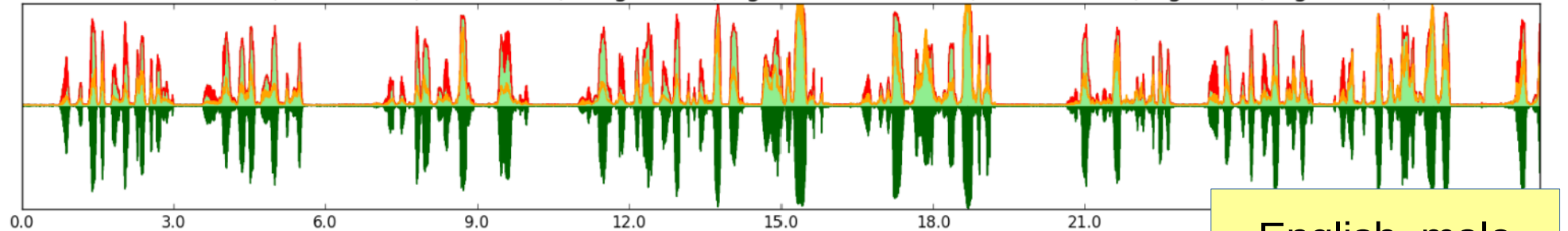


Correlation AME:FME=0.65  
Correlation AMS:FMS=0.15

# Discourse Prosody Case 2: Discourse Rhythms

## AM & FM signals and spectra: Abercrombie\_English\_NW048

Params: minf0:70, maxf0:200, frame:0.01, weight:0.02, sigmedianfactor:100, f0median:15, sigstart:6, siglen:30, maxhz:1



Correlation AME:FME=0.63  
Correlation AMS:FMS=0.07

## **Discourse prosody, Case 2: Accent constraints**

Constraint 1:

**Pitch accents in the same sequence tend to be of the same type.**

Constraint 2:

**Pitch accent sequences tend to match the final phrasal accent:**

- low rising types tend to be followed by a rising final accent
- high rising types tend to be followed by a rising final accent

Constraint 3:

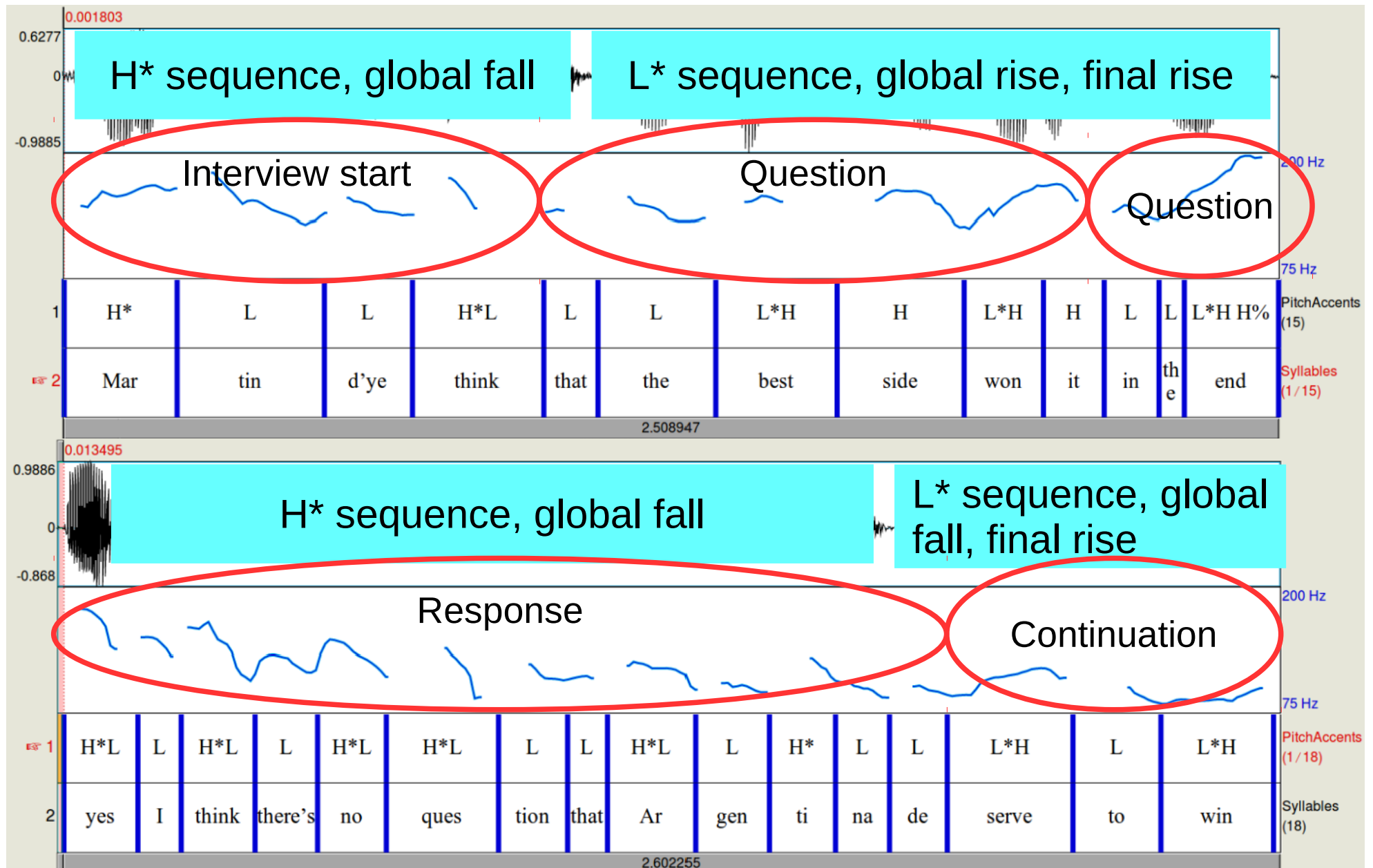
**Pitch accent sequence types tend to match information structure**

- low pitch accent sequences tend to be introductory or questioning
- high pitch accent sequences tend to be closing or stating

*with typologically relevant constraint violations in different languages and dialects*



# Discourse Prosody Case 3: Accent Constraints



# ***Discourse Prosody Case 4: Long FM Contours***

**Thesis: in evolution,**

**– frequency modulation and rhythm came first**

- **emotional cries**
- **turn-taking came before grammar**

Levinson, “Turn-taking in Human Communication – Origins and Implications for Language Processing”, 2015

**Note: in infant speech,**

**– frequency modulation and rhythm also come first**

- **emotional cries**  
Wermke, Sebastian-Galles

- **turn-taking**

Check the infant ‘twin-talk’ videos on YouTube

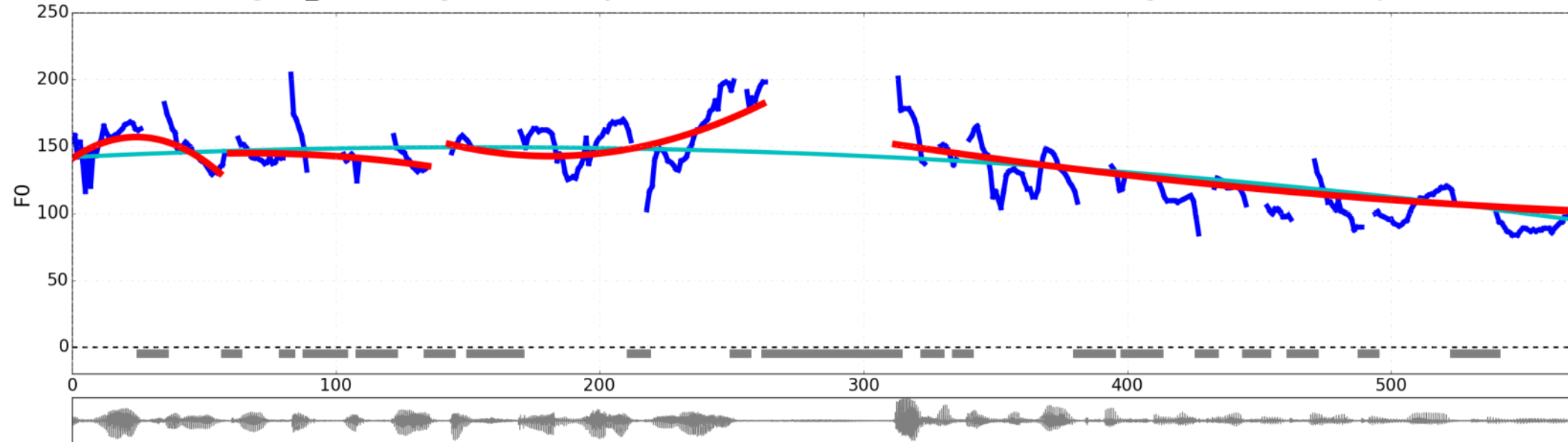


# Discourse Prosody Case 4: Long FM Contours

Question: rising utterance contour

Answer: falling utterance contour

PV 01: "English\_J0104G-Argen...", tier "Syllables", x-axis 10.0ms, Model: median 1, degree 2, domain "majorIPU"



Question+Answer: rising-falling adjacency pair contour  
*syntagmatic entrainment*

# ***Discourse Prosody Case 5: Emotive FM Contours***

Thesis 1:

**In the evolutionary time domain, emotive modulations came before structural modulations**

Thesis 2:

**“Wow!” is among the first emotive utterances**

Thesis 3:

**Or maybe it was the wolf whistle**

Thesis 4:

**In any case, other primates wowed and whistled first – we continued the custom**

***Is this why in some societies whistling is tabooed?***



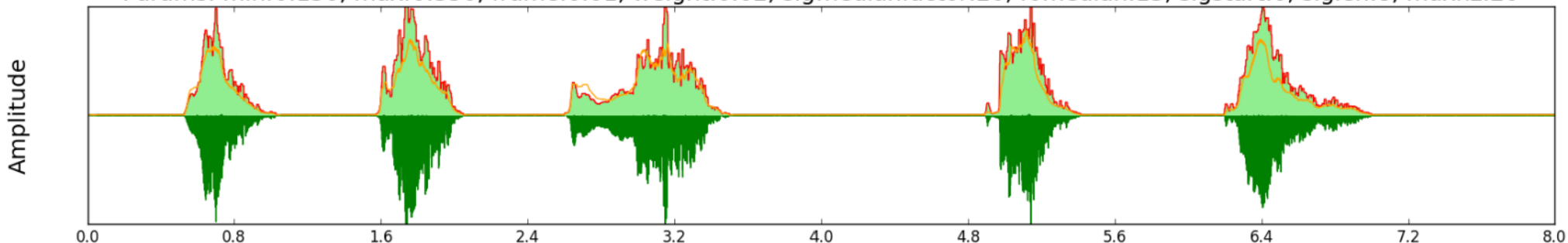
# Discourse Prosody Case 5: Emotive FM Contours

哇

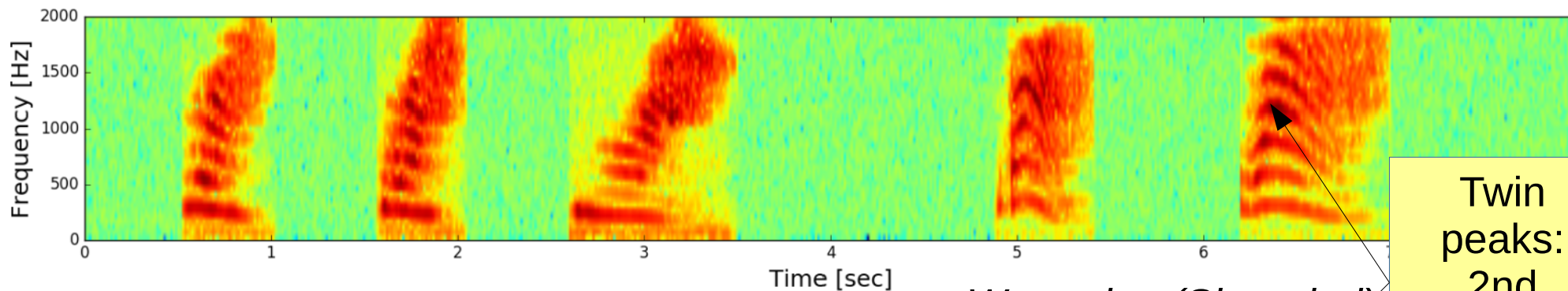
## EMOTIVE EXCLAMATIONS

AM & FM signals and spectra: Ashley-Becky-waa

Params: minf0:150, maxf0:350, frame:0.01, weight:0.02, sigmedianfactor:20, f0median:13, sigstart:0, siglen:8, maxhz:20



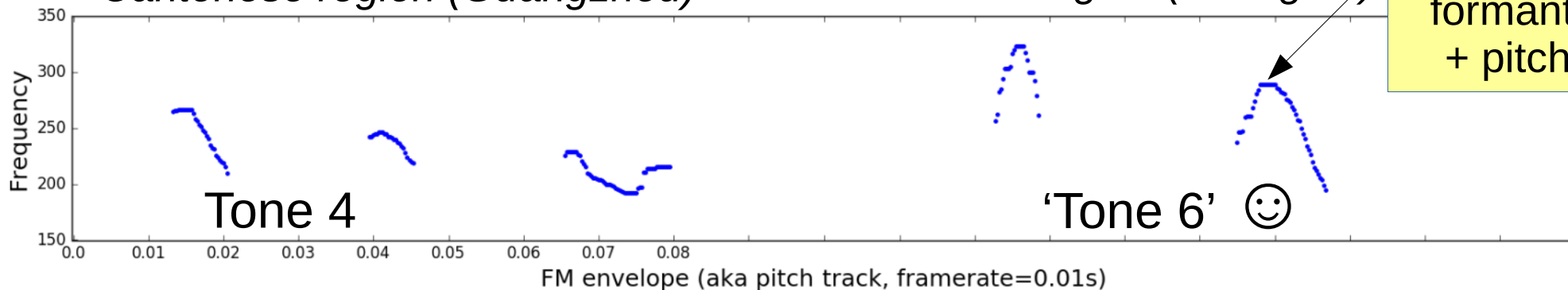
AM carrier with amplitude envelopes (dark green: negative amplitude, light green rectified amplitude)  
red: peak-picked envelope; orange: smoothed absolute Hilbert transform envelope



Cantonese region (Guangzhou)

Wu region (Shanghai)

Twin peaks:  
2nd  
formant  
+ pitch



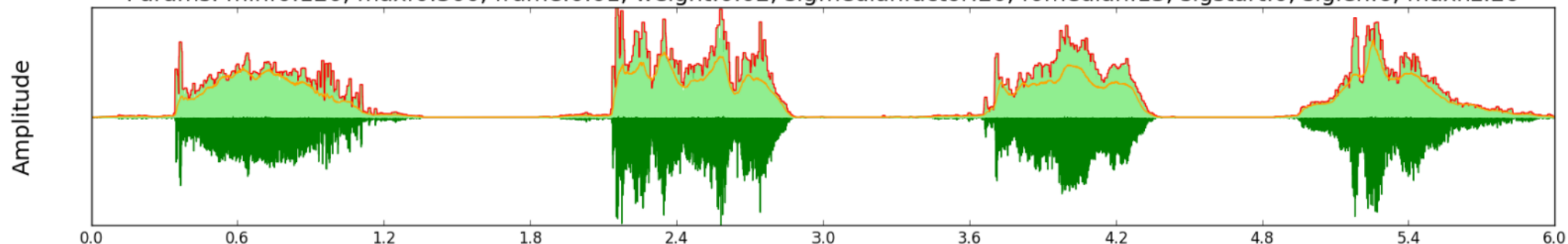
# Discourse Prosody Case 5: Emotive FM Contours

啊

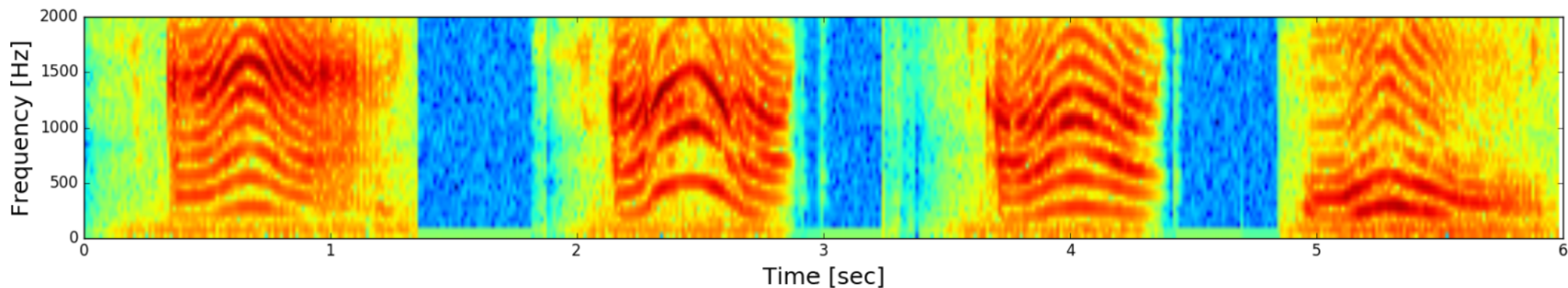
## EMOTIVE EXCLAMATIONS

AM & FM signals and spectra: yilia-aaah-oooh

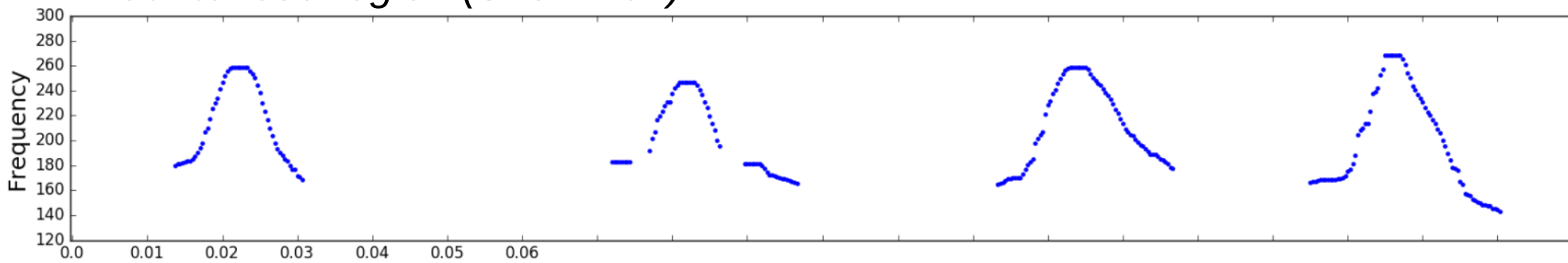
Params: minf0:120, maxf0:300, frame:0.01, weight:0.02, sigmedianfactor:20, f0median:13, sigstart:0, siglen:6, maxhz:20



AM carrier with amplitude envelopes (dark green: negative amplitude, light green rectified amplitude)  
red: peak-picked envelope; orange: smoothed absolute Hilbert transform envelope



*Cantonese region (Shenzhen)*



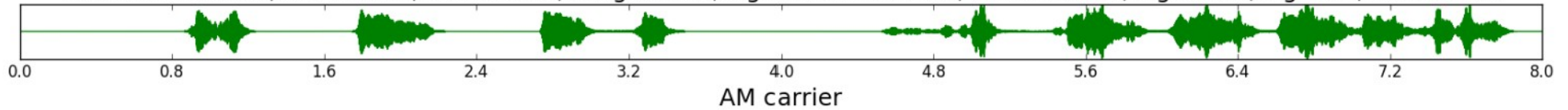
FM envelope (aka pitch track, framerate=0.01s)

# Discourse Prosody Case 5: Emotive FM Contours

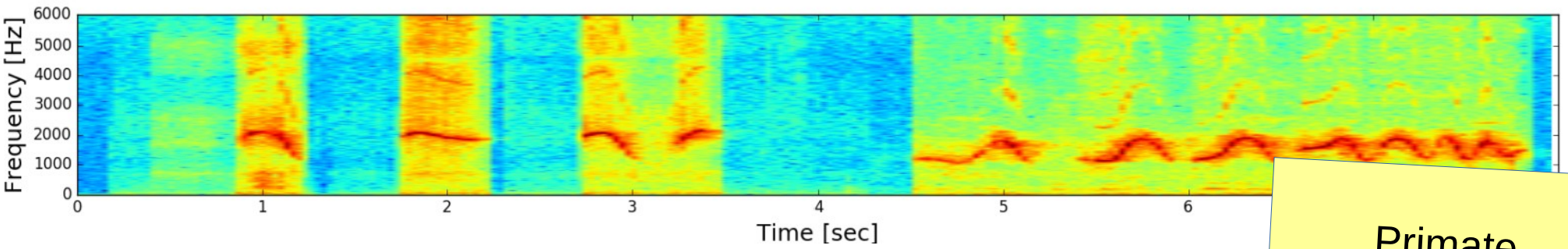
## TELEGLOSSIA

AM & FM signals and spectra: mandarinwhistles-mono

Params: minf0:70, maxf0:200, frame:0.01, weight:0.02, sigmedianfactor:20, f0median:21, sigstart:0, siglen:8, maxhz:100

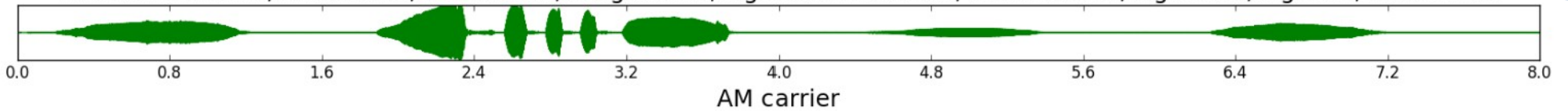


Street whistle  
Cantonese  
shoolboy

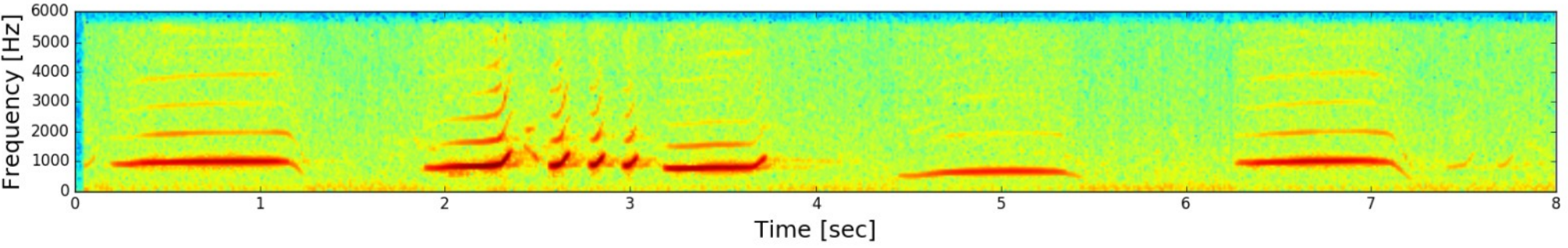


AM & FM signals and spectra: primatecalls-mono

Params: minf0:70, maxf0:200, frame:0.01, weight:0.02, sigmedianfactor:20, f0median:21, sigstart:0, siglen:8, maxhz:100



Primate





# *Discourse Prosody Case 5: Emotive FM Contours*

TELEGLOSSIA

**A Rising Contour: back-channel communication  
from the richest ex-linguist in the world**

**Falling, Rising-Falling and Rising F0 Contours:  
Intonation and Gesture**

## **Summary:**

## **Labov's Sociophonetics**

### **OSCAR: Phonetic Opinion Survey:**

- 1. The Prosody of Impoliteness**
- 2. Description of Mandarin Tones**

### **Phonetic Analysis of Discourse:**

**Case 1: Discourse framing**

**Case 2: AM vs. FM Spectra**

**Case 3: Accent Constraints**

**Case 4: Long FM contours**

**Case 6: Emotive FM contours**

## **Conclusion:**



**... thinking outside the box**

**Summary:**

**Labov's Sociophonetics**

**OSCAR: Phonetic Opinion Survey:**

- 1. The Prosody of Impoliteness**
- 2. Description of Mandarin Tones**

**Phonetic Analysis of Discourse:**

**Case 1: Discourse framing**

**Case 2: AM vs. FM Spectra**

**Case 3: Accent Constraints**

**Case 4: Long FM contours**

**Case 6: Emotive FM contours**



**Conclusion:**

**... thinking outside the box**

Thank you!  
谢谢！