Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

The Automatic Analysis by Synthesis of Speech Prosody
with Preliminary Results on Mandarin Chinese

Daniel Hirst

Laboratoire Parole et Langage, CNRS and Université de Provence
School of Foreign Languages, Tongji University, Shanghai
daniel.hirst@lpl-aix.fr

2012-12-08
8th International Symposium on Chinese Spoken Language Processing
Hong Kong

Analysis of speech prosody

The analysis of prosody is crucial for

- intelligibility "He's not coming back"
- statement? question? order?
- speaker states "Isn't this interesting"
- naturalness
  - - facilitate cognitive processing
  - - cf non-standard, non-native, pathological, or synthetic speech
- limited current use of synthesis for listening tasks but huge potential

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Prosodic Annotation

The explicit characterisation of the length, pitch and loudness of the individual sounds which make up an utterance.

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Prosodic Annotation

The explicit characterisation of the length, pitch and loudness of the individual sounds which make up an utterance.
Analysis by synthesis - synthesis makes it possible to evaluate the analysis.

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Prosodic Annotation

The explicit characterisation of the length, pitch and loudness of the individual sounds which make up an utterance.
Analysis by synthesis - synthesis makes it possible to evaluate the analysis.
Linguists need tools.
Engineers need data.

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Prosodic Annotation

The explicit characterisation of the length, pitch and loudness of the individual sounds which make up an utterance.
Analysis by synthesis - synthesis makes it possible to evaluate the analysis.
Linguists need tools.
Engineers need data.
"Data are cheap"

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Prosodic Annotation

The explicit characterisation of the length, pitch and loudness of the individual sounds which make up an utterance.
Analysis by synthesis - synthesis makes it possible to evaluate the analysis.
Linguists need tools.
Engineers need data.
"Data are cheap" ...but facts are expensive
and we all need facts...

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Looking for prosodic metrics

Linguists propose prosodic typologies

lexical	quantity/tone/stress Korean/Chinese/English (French)
rhythm	stress/syllable/mora timed English/French/Japanese
melody	falling/rising pitch accents English/French

Search for corresponding metrics: objective measurements predicting typological category.

Annotation
Prosodic metrics
Rhythm metrics
Melody metrics
Data
Tools
Melody metrics
Results

Using prosodic metrics

Robust metrics useful for understanding prosodic structure.
Guiding speech recognition.
Evaluating atypical speech:

- Non-standard dialect
- Non-native speech
- Pathological speech
- Synthetic speech

Annotation
Prosodic metrics
Rhythm metrics
Melody metrics
Data
Tools
Melody metrics
Results

Different rhythm metrics

%V	percent duration of vocalic intervals
Δ C, Δ V	standard deviation of duration of consonantal and vocalic intervals
rPVI (c,v)	raw index of variability between duration of successive consonantal and vocalic intervals
nPVI (c,v)	normalised index of variability between duration of successive consonantal and vocalic intervals
VarcoC,VarcoV	coefficient of variation of duration of consonantal and vocalic intervals

Linear discriminant analysis of rhythm metrics

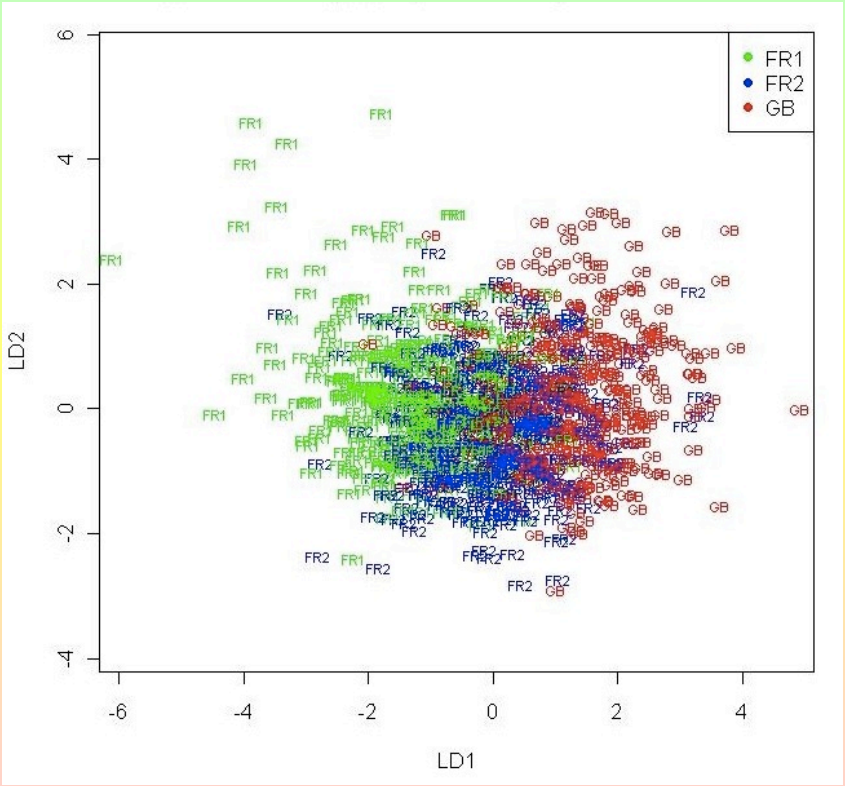


Figure: Tortel & Hirst 2010. Linear discriminant analysis of rhythm metrics for 3 groups reading the same texts.

Annotation
Prosodic metrics
Rhythm metrics
Melody metrics
Data
Tools
Melody metrics
Results

Melody metrics

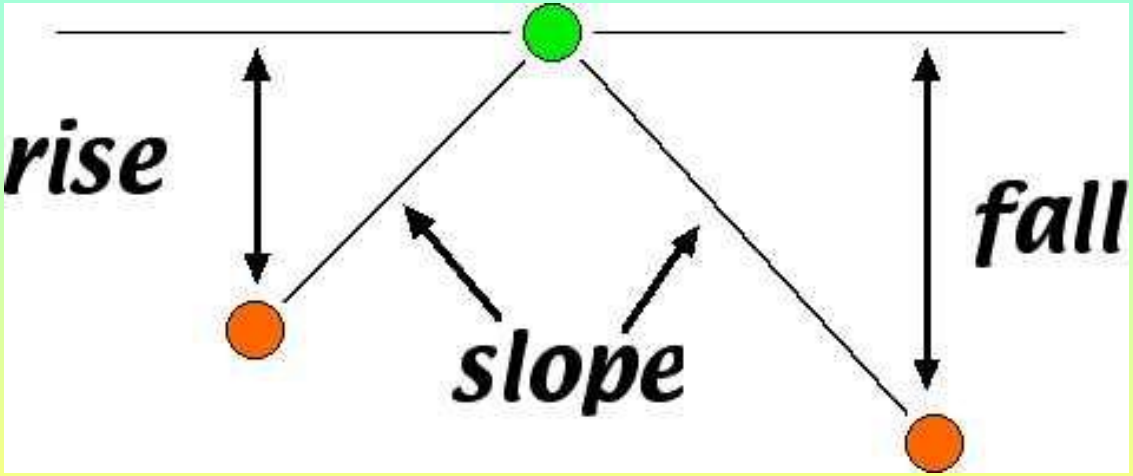


Figure: Melody metrics from a sequence of target points

	Predicted	
	English	French
English	132	18
French	13	87

Table: Classification matrix for discriminant analysis (Hirst 2003).

87.6% correct identification of language from parameters.

Annotation
Prosodic metrics
Rhythm metrics
Melody metrics
Data
Tools
Melody metrics
Results

Building a multilingual prosodic database

The Eurom1 corpus European project SAM 1986
Part of the corpus - 40 five-sentence passages - subsequently used in the project MULTEXT

Last week, my friend had to go to the doctor's to have some injections. She is going to the Far East for a holiday and needs to have an injection against, cholera, typhoid fever, hepatitis A, polio and tetanus. I think she will feel quite ill after all those. She is going to have them all done at once, at one session. I shan't feel sorry for her, though.

Limited number of recordings. Each speaker read only 10 (French) or 15 passages (English).

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

New recordings

New recordings of the same corpus.
10 subjects in each language read all 40 passages (a total of 2000 sentences per language)

Korean	S.Kim, D.J.Hirst, H. Cho, H-Y. Lee, M. Chung (4th International Conference on Speech Prosody, Campinas, Brazil 2008)
English, French	S. Herment, A. Loukina, A. Tortel, D.J. Hirst, B. Bigi (4th International Conference on Corpus Linguistics., Jaèn, Spain, 2012.)
Chinese	H. Ding, D.J.Hirst (8th International Symposium on Chinese Spoken Language Processing, Hong Kong 2012)

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Speech alignment

It takes a linguist several hours to align one minute of speech with a phonetic transcription.

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Speech alignment

It takes a linguist several hours to align one minute of speech with a phonetic transcription.
They have better things to do...

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Tools for alignment

- HTK Toolkit
- Festival
- Julius
- P2FA UPenn aligner (Jiahong Yuan and Mark Liberman)
- EasyAlign (Jean-Philippe Goldman) only Win-Dos...
- SPPAS (Brigitte Bigi)

SPeech Phonetisation Alignment and Syllabification (SPPAS)

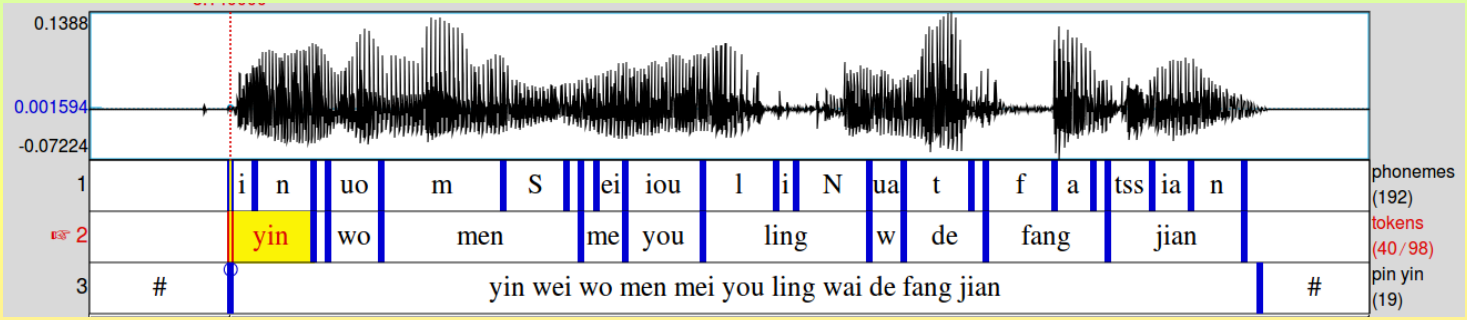


Figure: Sample sentence from Eurom1-ZH corpus ("Because we do not have another room")

Distributed under GPL license and implemented for French, English, Italian and (partially) Chinese
Available from:

http://www.lpl-aix.fr/~bigi/sppas/

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Automatic annotation of pitch

- Momel/INTSINT: Daniel Hirst and Robert Espesser
- RFC and Tilt: Paul Taylor
- Stem-ML: Greg Kochanski and Chilin Shih
- Prosogram: Piet Mertens
- Penta: Santitham Prom-On and Yi Xu
- AuTobi: Andrew Rosenberg

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Problem for modelling f_0

"More news about the Reverend Sun Myung Moon..."

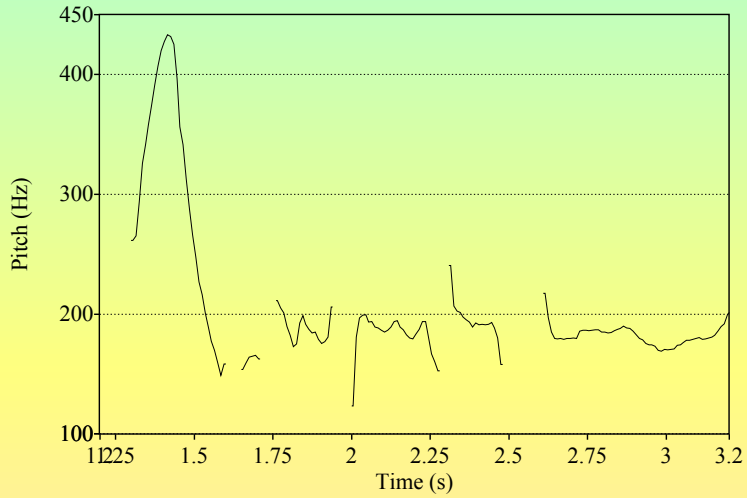


Figure: Two second extract of f_0 curve

Problem for modelling f_0

"More news about the Reverend Sun Myung Moon..."

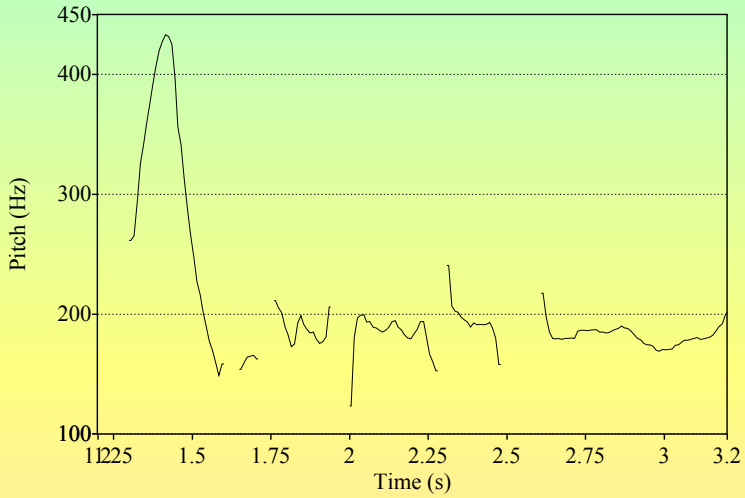


Figure: Two second extract of f_0 curve

- Raw f0 is discontinuous and not smooth.

Problem for modelling f_0

"More news about the Reverend Sun Myung Moon..."

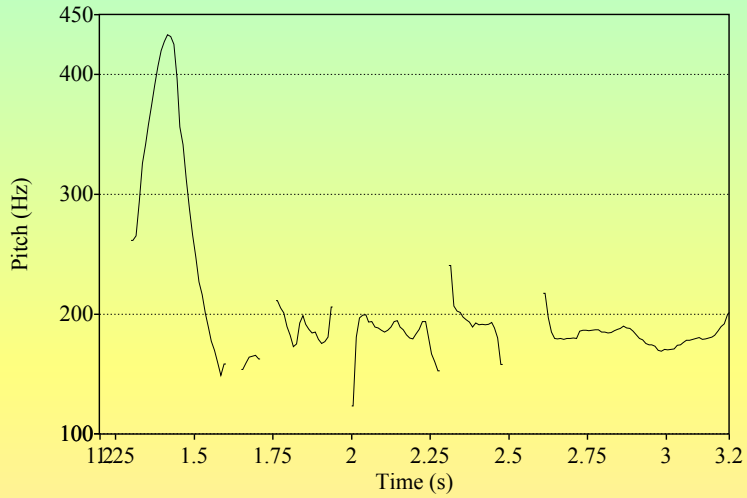


Figure: Two second extract of f_0 curve

- Raw f0 is discontinuous and not smooth.
- Here beginning and end is continuous and smooth

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Problem for modelling f_0

"More news about the Reverend Sun Myung Moon..."

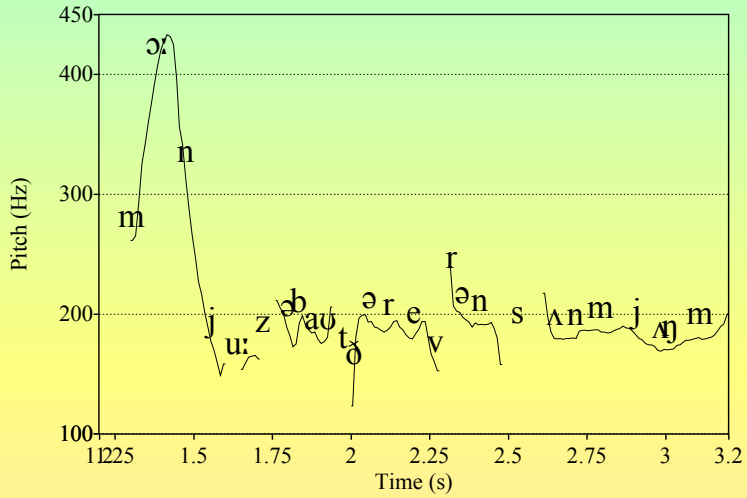


Figure: Two second extract of f_0 curve

- Raw f0 is discontinuous and not smooth.
- Here beginning and end is continuous and smooth
- Discontinuity is due to microprosodic effect of consonants

General model for f_0

Raw f_0 is the combination of two components

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

General model for f_0

Raw f_0 is the combination of two components

- Macromelodic component: smooth and continuous (Underlying intonation pattern)

General model for f_0

Raw f_0 is the combination of two components

- Macromelodic component: smooth and continuous (Underlying intonation pattern)
- Micromelodic component: discontinuous (Surface effect of phonemes)

Macromelodic profile

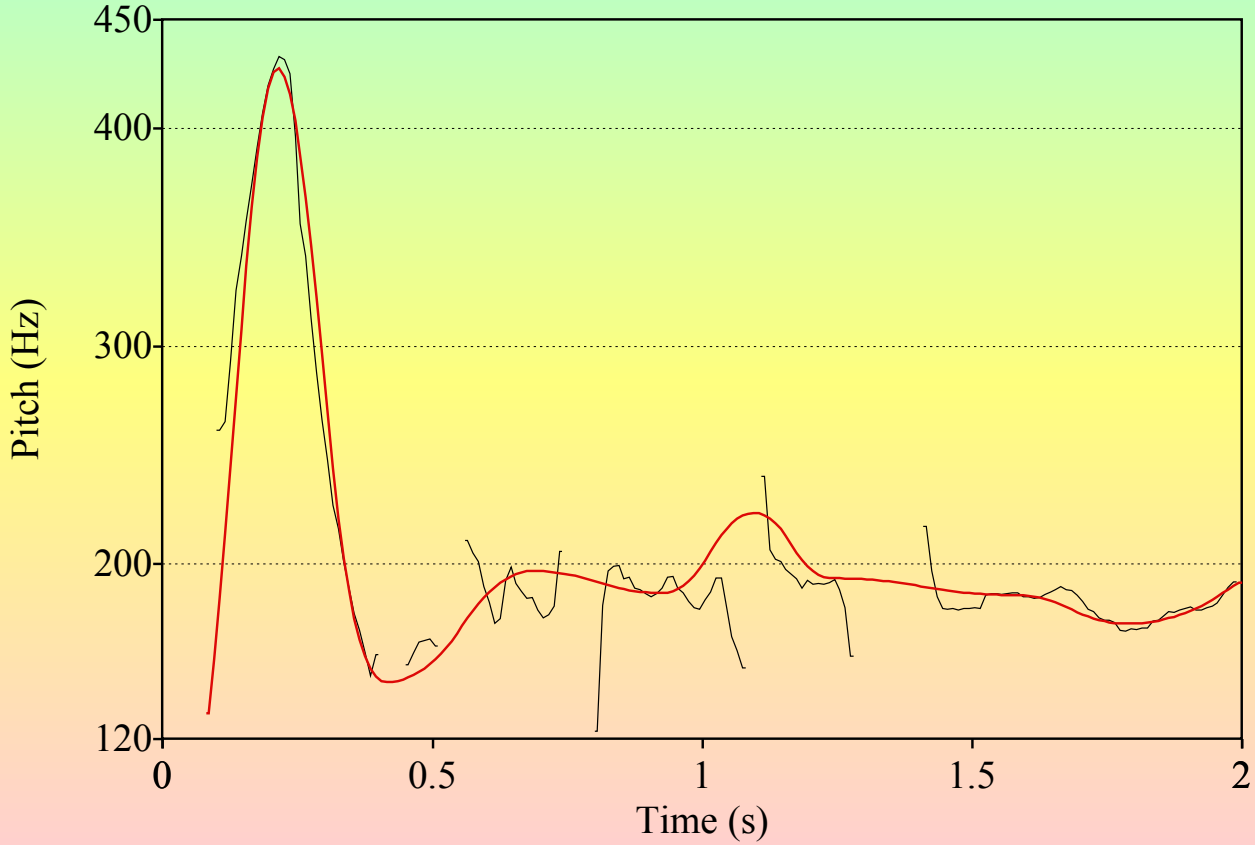


Figure: Macromelodic profile for extract from A01-01

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Macromelodic profile

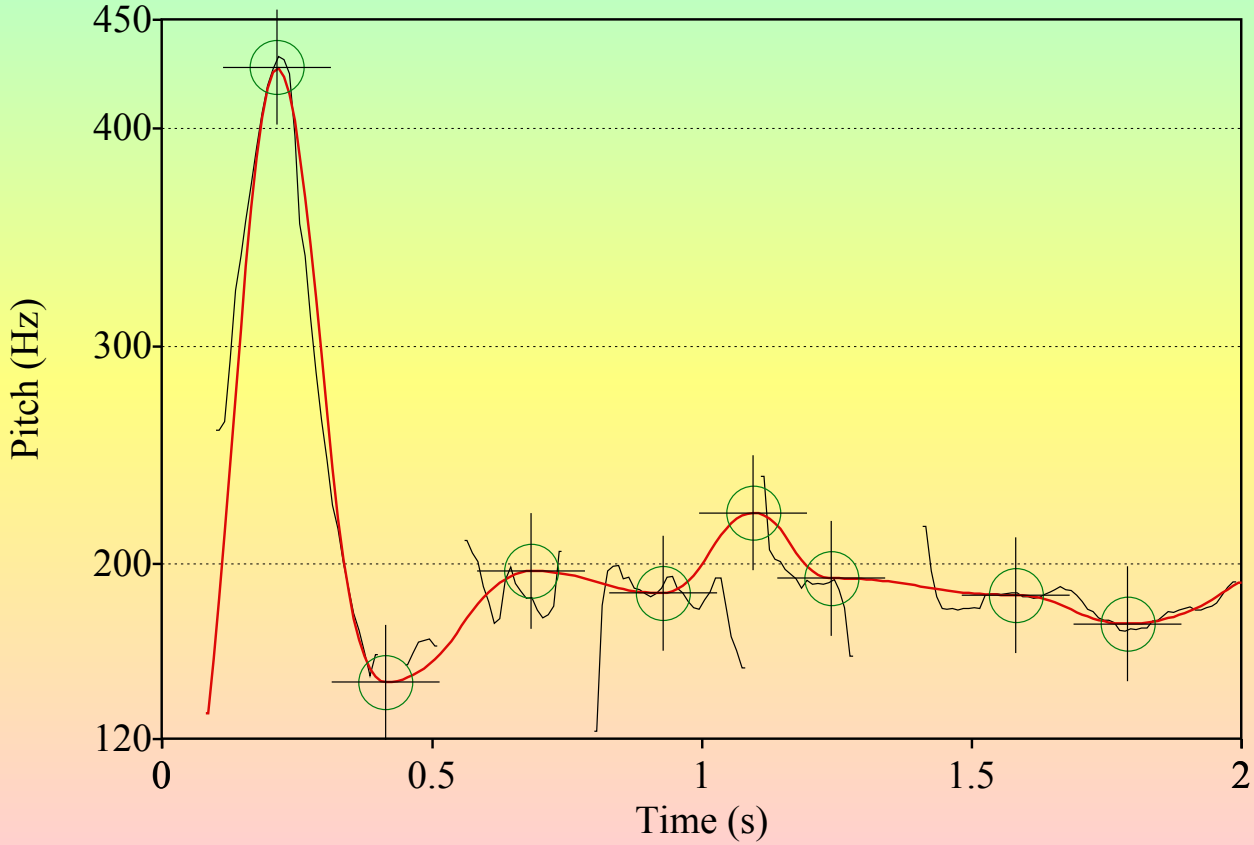


Figure: Macromelodic profile for extract from A01-01

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

MoMel

An algorithm for modelling melody.

Manual momel Used from 1980 on to model melody

Automatic momel Hirst & Espesser (1993)

Asymmetric Modal Quadratic Regression variety of robust regression

- **Quadratic** First derivative is linear
- **Asymmetric** Microprosody is essentially a lowering of f_0
- **Modal** generalisation of mode to function

Hirst (2007) Improved algorithm and Praat plugin.

Annotation
Prosodic metrics
Data
Tools
Analysing Length
Analysing Pitch
Melody metrics
Results

Combining SPPAS and Momel

De Looze & Hirst (JEP 2010) - octave is natural unit for speech.
Speaker independent OMe scale: $log_2(Hz/median)$

Combining SPPAS and Momel

De Looze & Hirst (JEP 2010) - octave is natural unit for speech.
Speaker independent OMe scale: $log_2(Hz/median)$

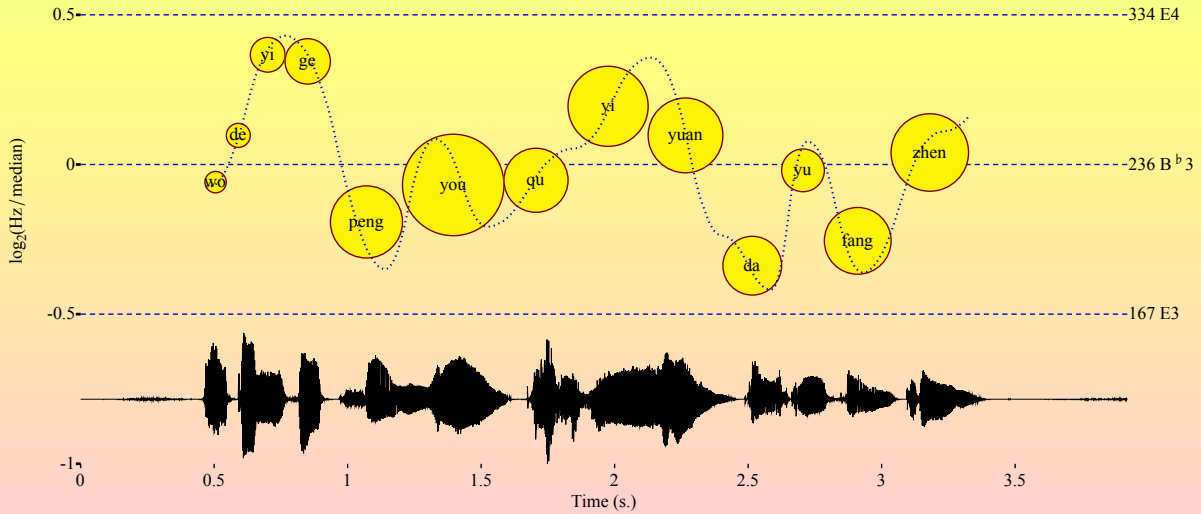


Figure: Automatic analysis of prosody

Data freely available

All these corpora will be analyzed using the automatic annotation facilities.
All the data will be made freely available on the Speech and Language Data Repository

- http://sldr.org

We hope that this will at last lead us to establishing some hard facts about the prosody of these languages

Melody metrics revisited

octave	value of target points on OMe scale
interval	absolute difference from previous target
rise	difference from previous target for rise
fall	difference from previous target for fall
slope	absolute difference from previous target divided by distance in seconds
rise-slope	slope for rise
fall-slope	slope for fall

For each parameter calculate mean and standard deviation.

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Linear Discriminant Analysis : language

	Predicted		
	English	French	Chinese
English	339	148	17
French	84	241	52
Chinese	17	48	328

Table: Classification matrix for discriminant analysis on language.

71% correct prediction of language.
89% correct discrimination of Chinese from English and French.

Linear Discriminant Analysis: language+gender

	Predicted					
	EN-f	EN-m	FR-f	FR-m	ZH-f	ZH-m
EN-f	186	0	49	9	3	1
EN-m	0	172	0	87	0	2
FR-f	44	0	202	26	33	0
FR-m	5	22	0	34	0	11
ZH-f	1	0	27	1	164	0
ZH-m	4	6	1	1	0	183

Table: Classification matrix for discriminant analysis on language and gender.

74% correct prediction of gender and language
76% correct prediction of language.
93% correct discrimination of Chinese from English and French.

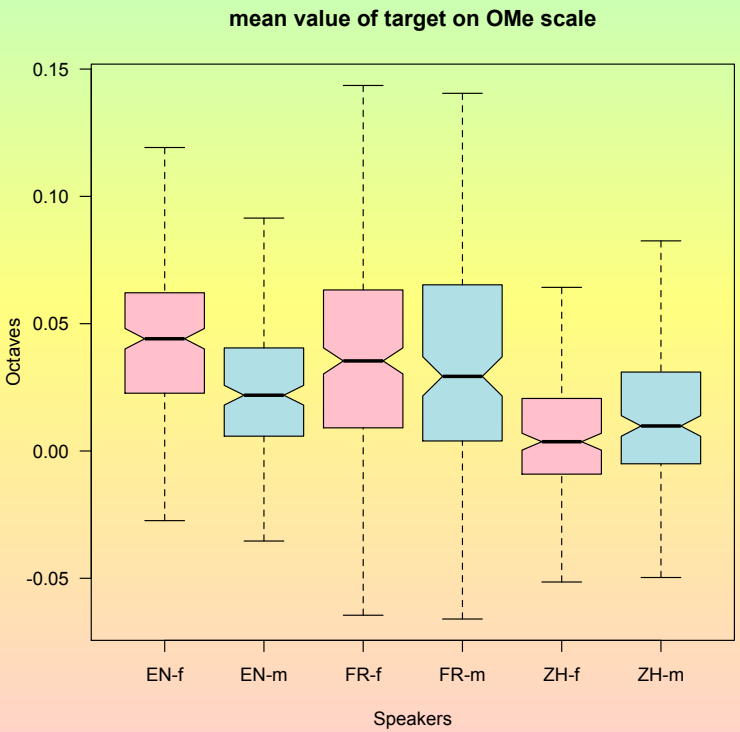
ANOVA

All values in Octave-Median scale = $log2(Hz/median)$

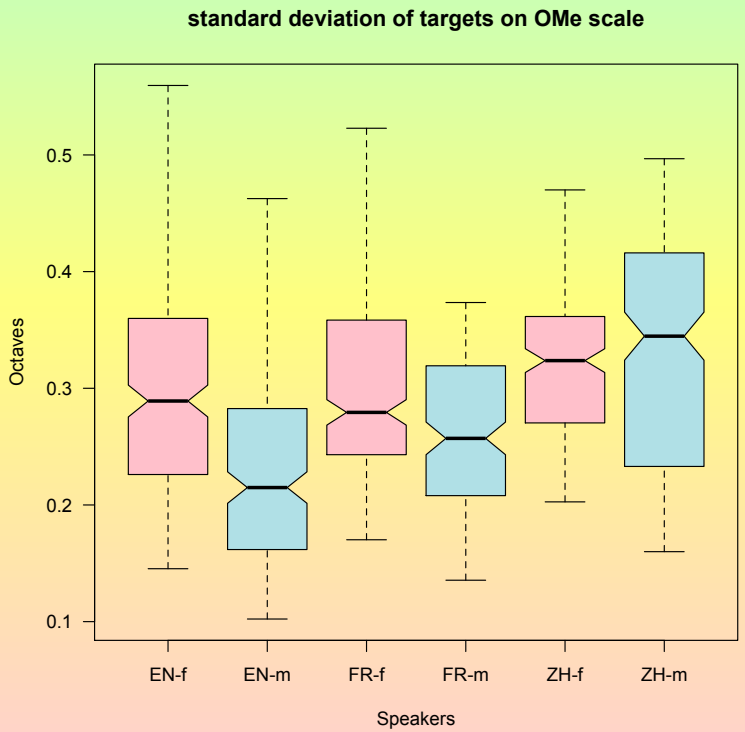
	mean			standard deviation		
	L	G	L*G	L	G	L*G
octave	***	-	***	***	***	***
interval	***	-	*	***	***	***
rise	***	***	***	***	***	***
fall	***	***	***	***	***	***
slope	-	-	-	***	-	-
rise-slope	***	***	***	***	-	-
fall-slope	***	***	***	***	-	-

Table: Significance levels of Anova for each parameter.

ANOVA: Octave - mean and standard deviation

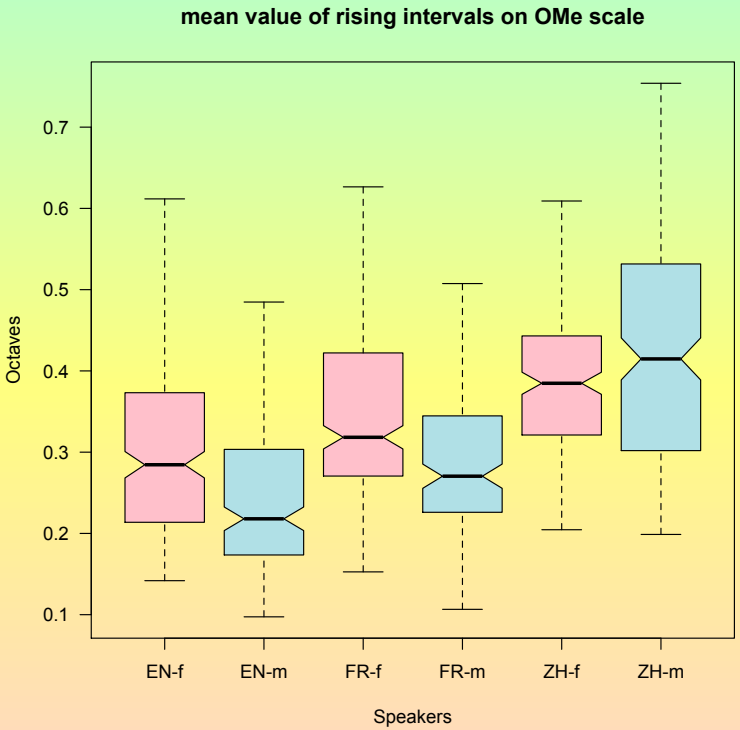


a. Mean value of targets on Octave-Median scale

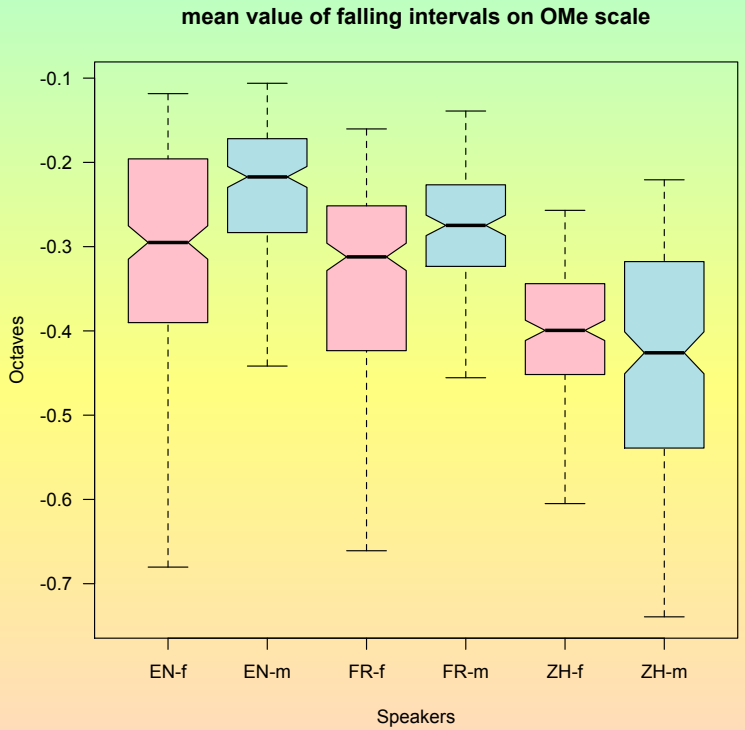


b. Standard deviation of targets on Octave-Median scale

ANOVA: Rise, Fall - mean

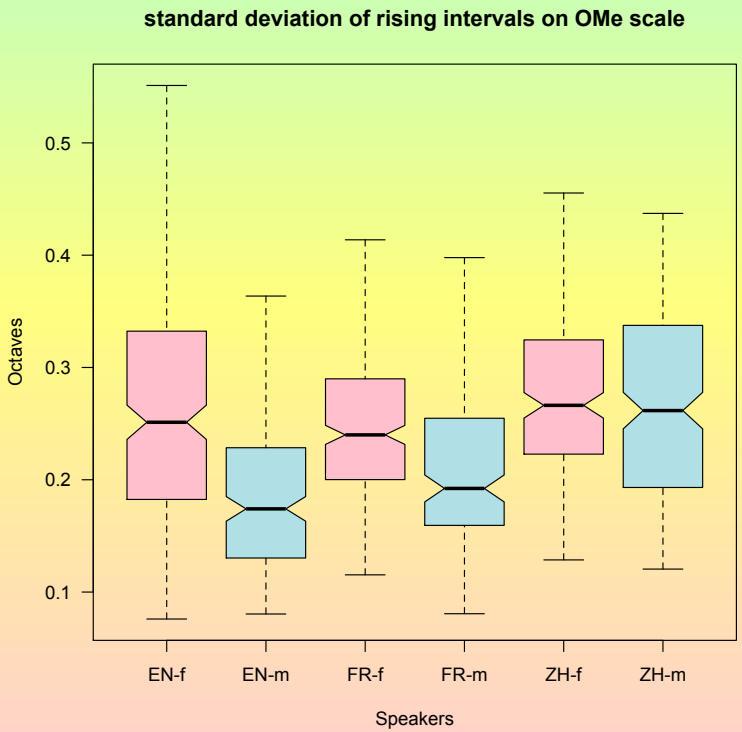


a. Mean value of rising interval on Octave-Median scale

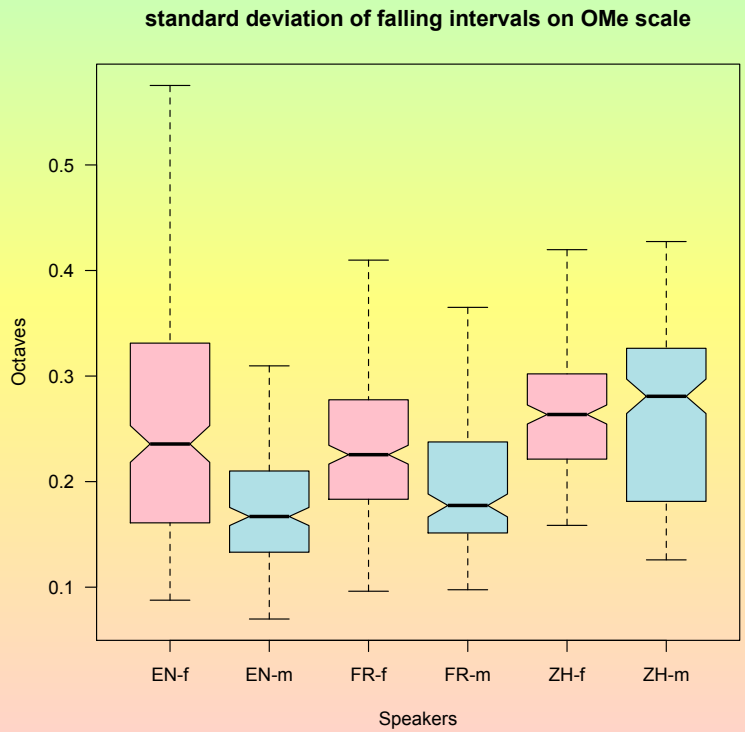


b. Mean value of falling interval on Octave-Median scale

ANOVA: Rise, Fall - standard deviation

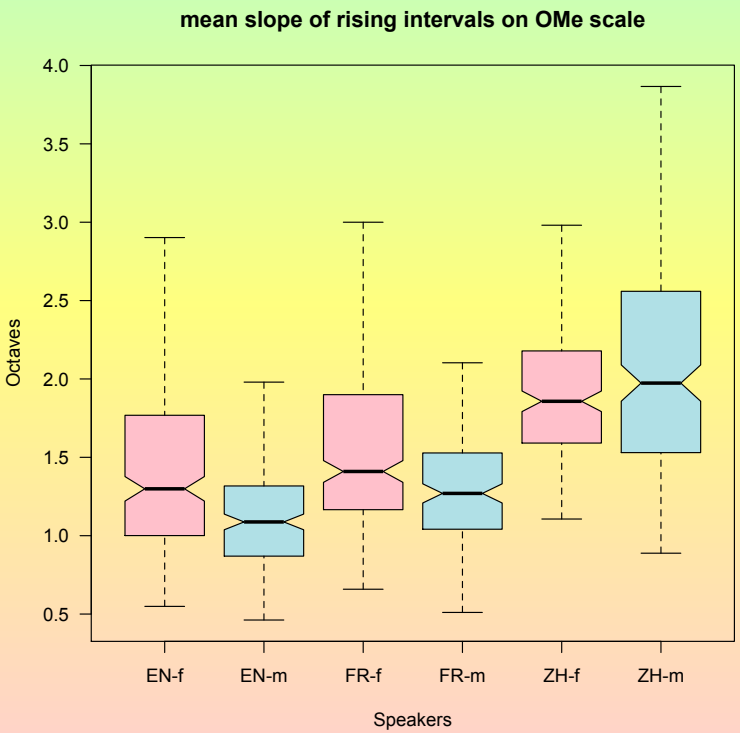


a. Standard deviation of rising interval on Octave-Median scale

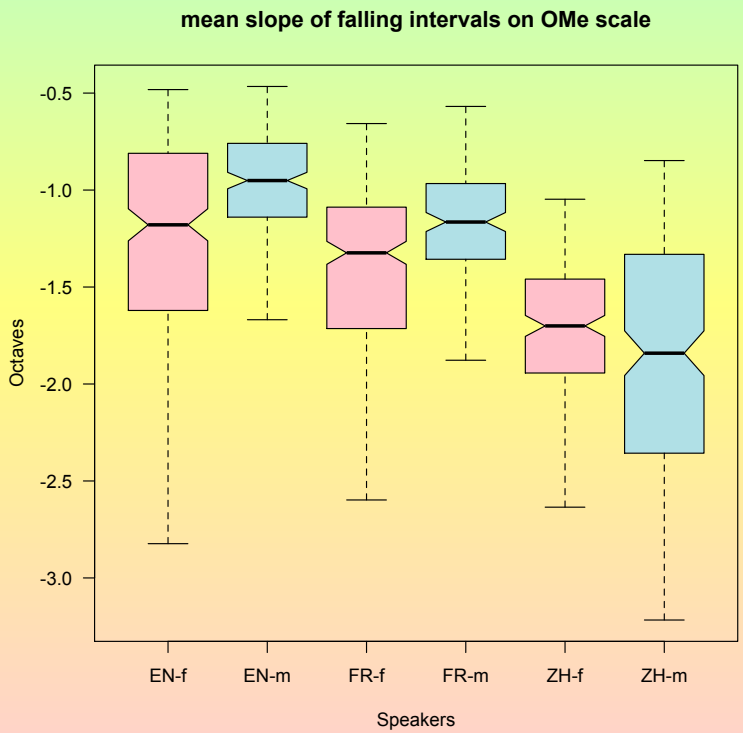


b. Standard deviation of falling interval on Octave-Median scale

ANOVA: Rise-slope, Fall-slope - mean

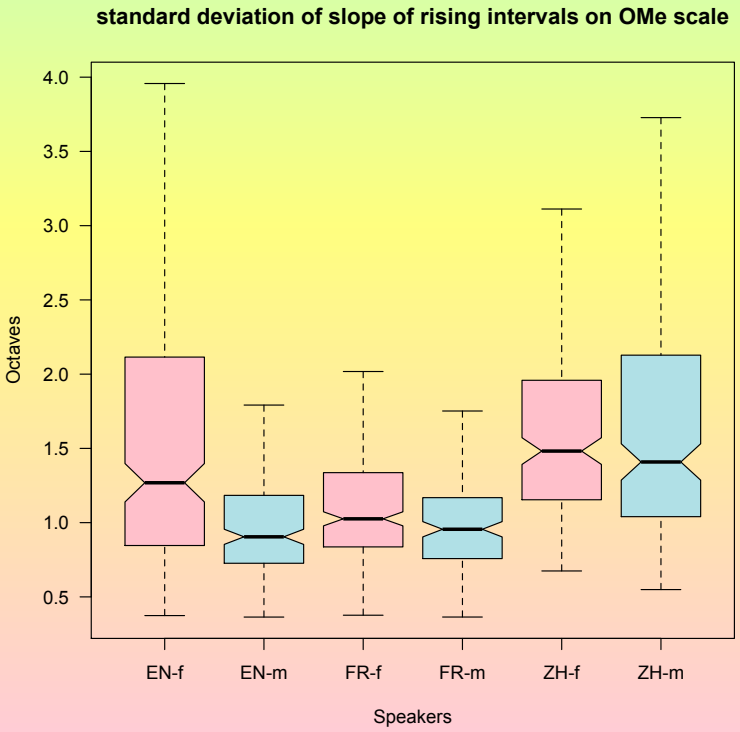


a. Mean slope of rising interval on Octave-Median scale

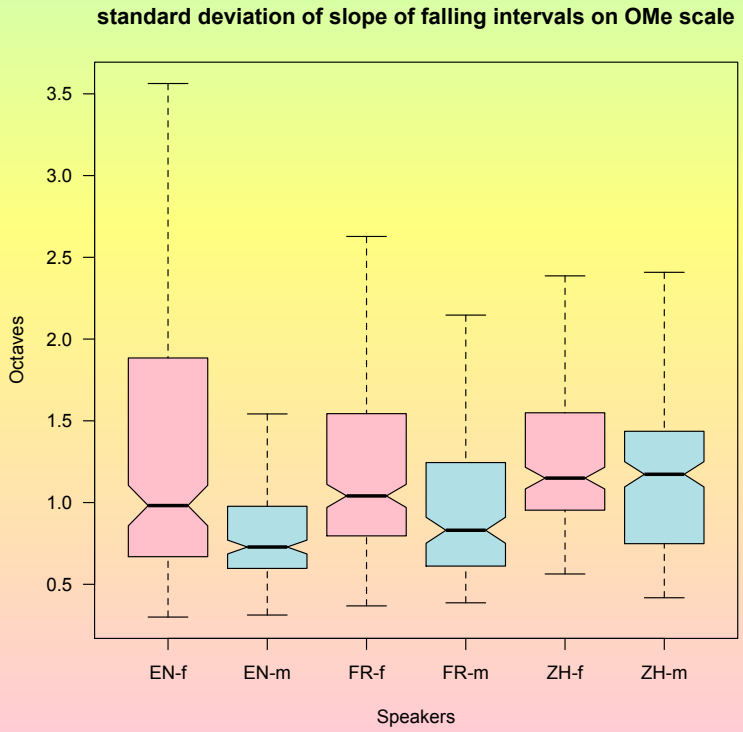


b. Mean slope of falling interval on Octave-Median scale

ANOVA: Rise-slope, Fall-slope - standard deviation



a. Standard deviation of slope of rising interval on Octave-Median scale



b. Standard deviation of slope of falling interval on Octave-Median scale

Summary of results

Chinese seems to be clearly distinct from English and French in making use of pitch movements which are larger (mean interval, fall and rise), with greater variability (sd of interval, fall and rise) and are faster (mean slope, rise-slope, fall-slope).
Furthermore in English and French there is a very significant gender difference (female speakers make larger and faster pitch movements) which is not observed in Chinese
I suggest that this is a result of pressure from the lexical tone that prevents pitch being mobilised for more expressive functions such as gender differences.

What's next?

Align the corpora with transcription and use word boundaries.

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

What's next?

Align the corpora with transcription and use word boundaries.

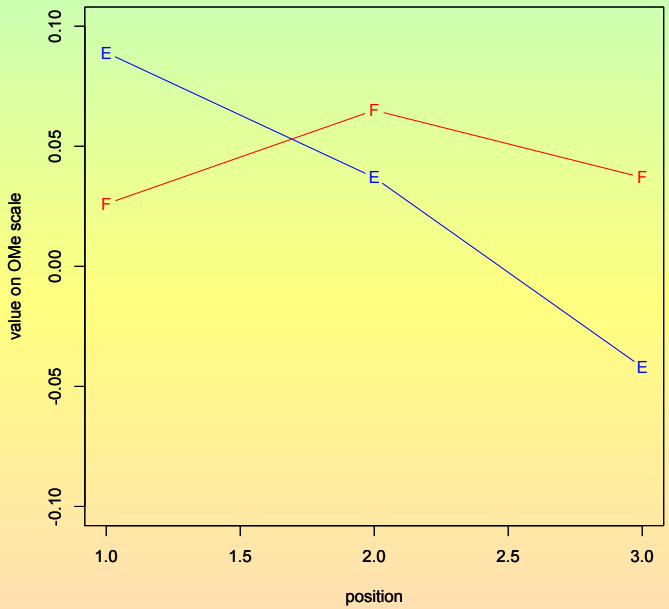


Figure: Mean pitch targets at three positions in word: early, mid and late

and that's just the beginning...

Collaboration

LPL, CNRS Aix-en-Provence

Brigitte Bigi, Sophie Herment

Former doctoral students

Céline De Looze, Anne Tortel, Hyongsil Cho

Oxford University Phonetics Laboratory

Anastassia Loukina, Greg Kochanski

School of Foreign Languages, Tongji University, Shanghai

Hongwei Ding, Ting Wang

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Annotation
Prosodic metrics
Data
Tools
Melody metrics
Results

Thanks for listening!

Questions now or email to daniel.hirst@lpl-aix.fr