Timing analysis with the help of SPPAS and TGA tools

Jue Yu

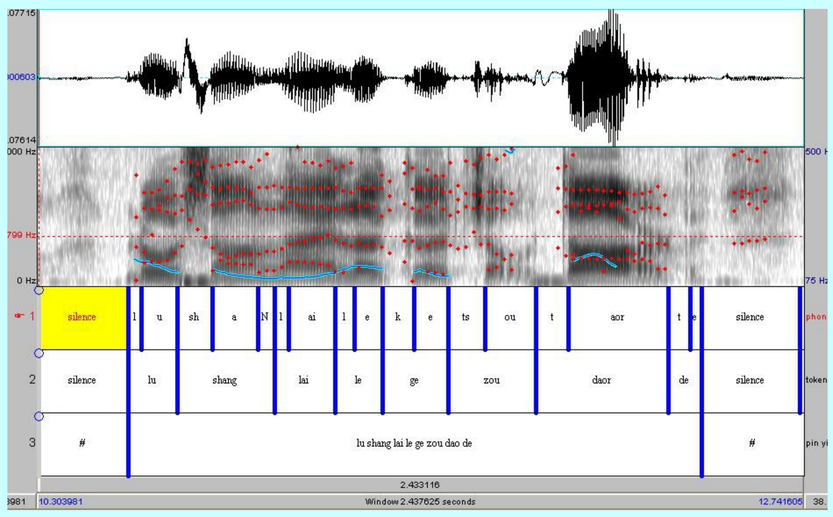
School of Foreign Languages, Tongji University, Shanghai, China

The problem

- **Speech timing:**
  - Domains – speech production, speech synthesis
  - Studies show that an approach based on large corpora is necessary
  - Production and analysis of large numbers of annotations
  - Manual annotation and analysis of speech data is time-consuming: many hours required
- **Objective:**
  - Efficient annotation and analysis of Hangzhou and Beijing Mandarin with web-based computational tools

Methods

- **Efficient syllable annotation: SPPAS**
  - 'SPeech Phonetization Alignment and Syllabification', designed at Laboratorire Parole et Langage, Aix-en-Provence (Brigitte Bigi)
    a) to automatically produce annotations which include utterance, word, syllable and phoneme segmentations and their transcriptions from recorded speech
    b) phonetician-friendly interface, high rate of correct alignment, generation of files in the TextGrid format
  - Illustration of SPPAS output for the Mandarin Chinese utterance “lu4 shang5 lai2 le5 ge4 zou3 daor4 de5” (“on the street came a traveller“) in Pinyin orthography:



- **Timing analysis and visualisation: TGA**
  - 'Time Group Analyser' (TGA, Dafydd Gibbon), designed at Bielefeld University
  - automatic parsing of syllable sequences into Time Groups (TG), e.g. inter-pausal groups, with extensive statistical analyses and visualisations of speech timing in the annotation data
  - illustration of local timing analysis:

Number of Time Groups: 20 ; Total duration (without pauses): 31104 ms

#	n	dur (ms)	rate	mean	median	stdev	npvi	mednpvi	intercept	slope	pattern
1	0	0	0.00	0.00	0.00	0.00	0	0	0.00	0.00	
2	5	1062	4.71	212.40	217.00	27.81	24	27	207.20	2.60	\ / \ = bei3 feng1 gen1 tai4 yang2 189 255 176 225 217

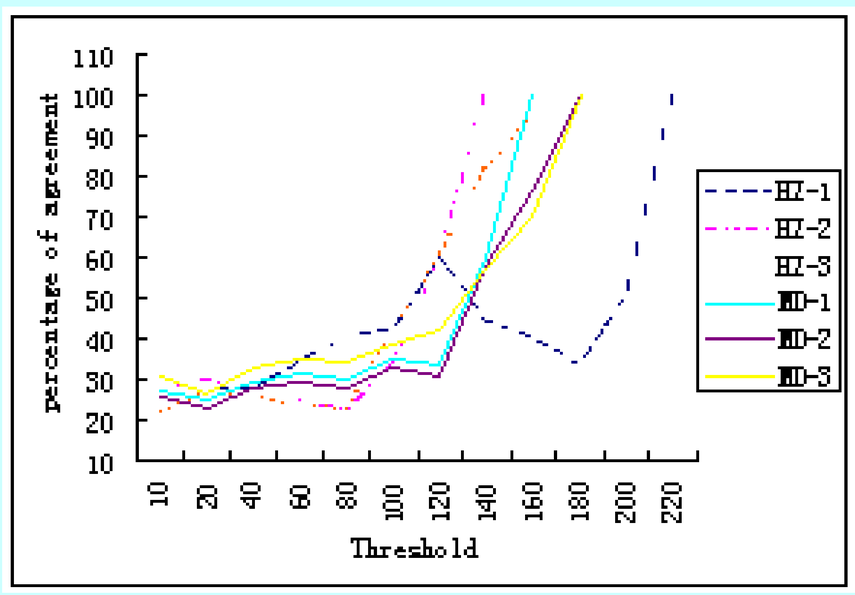
- illustration of comprehensive analysis with TGA:

Summary table of global and accumulated TG duration functions (some do make sense...)
Time Group criterion: pausegroup, local threshold: 10, Min valid TG length: 2
Only inter-pause intervals measured; pauses not included

Overall duration:	31104	Overall raw longer, ms:	5831	Overall raw shorter, ms:	5764
Overall min:	47.00	Overall max:	315.00	Overall range:	268.00
Valid Time Groups:	19	Overall rate/sec:	5.85		
Components: global tendencies					
Overall mean:	170.90	Overall median:	173.00	Overall SD:	52.03
Overall npvi:	40.00	Overall intercept:	171.07	Overall slope:	0.00
Mean of means:	176.27	Median of means:	168.57	SD of means:	18.71
Mean of medians:	174.24	Median of medians:	172.00	SD of medians:	19.57
Mean of SDs:	48.89	Median of SDs:	47.75	SD of SDs:	15.88
Mean of nPVIs:	40.00	Median of nPVIs:	33.00	SD of nPVIs:	14.59
Mean of intercepts:	167.28	Median of intercepts:	162.68	SD of intercepts:	27.27
Mean of slopes:	6.82	Median of slopes:	0.66	SD of slopes:	20.65
Components: correlations					
mean::TGdur:	-0.478	median::TGdur:	-0.235	SD::TGdur:	-0.048
nPVI::TGdur:	0.031	slope::TGdur:	-0.405	intercept::TGdur:	-0.057
nPVI::mean:	-0.506	slope::mean:	0.223	intercept::mean:	0.510
nPVI::median:	-0.733	slope::median:	-0.246	intercept::median:	0.727
nPVI::SD:	0.900	slope::SD:	0.577	intercept::SD:	-0.683

Case study of Hangzhou & Beijing Mandarin

- **Objective:**
  - Comparison of timing patterns in Hangzhou-accented and standard Beijing Mandarin
- **Method:**
  - use of SPPAS and TGA
  - comparison of manual and automatic procedures
- **Results:**
  - Local iambic (decelerating) patterns in inter-pausal groups
  - Agreement of time-tree constituents with words
  - Comparison of Hangzhou and Beijing speakers:



  - correlations of the Time Tree units with multisyllable words in speakers: similar until 50ms threshold
  - speakers HZ-2 & HZ-3 seem to have better Mandarin-like timing (this also matches proficiency evaluation results)
  - above 50ms, correlation of Time Tree units with multisyllable words increases rapidly to a speaker-dependent threshold
  - not significant: speech rate difference ($F(1,5)=0.04$, $p>0.05$), proficiency ($R2=0.028$, $p>0.05$), syllable nPVIs ($F(1,5)=0.444$, $p>0.05$)

Conclusion

- Use of web-based syllable processing tools enables rapid, comprehensive comparison of speech varieties