"An" evaluation of automatic glottal source analysis

John Kane

Thursday June 20th, 2013
NOLISP
Mons, Belgium

Glottal source analysis

Glottal source analysis

Off with her head!

Glottal source in speech technology

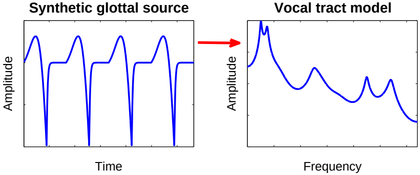
Speech synthesis

Speech recognition

Speaker verification

Evaluating glottal source analysis

Laryngograph/Electroglottograph



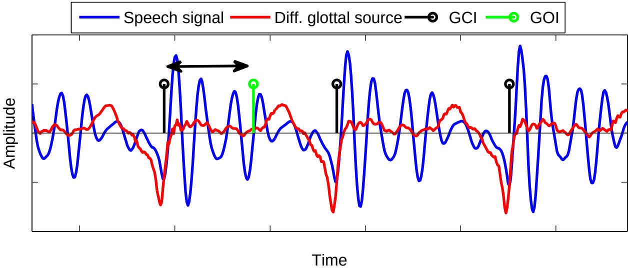
Synthetic speech

Aims

- Quantitative evaluation of glottal inverse filtering and parameterisation methods
- Additionally include model-based glottal parameterisation

Glottal inverse filtering methods -

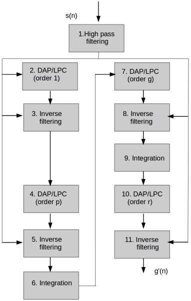
Closed-phase inverse filtering (CPIF)



- Detection glottal opening/closing instants
- Covariance LPC on closed phase (i.e GCI => GOI)

Glottal inverse filtering methods -

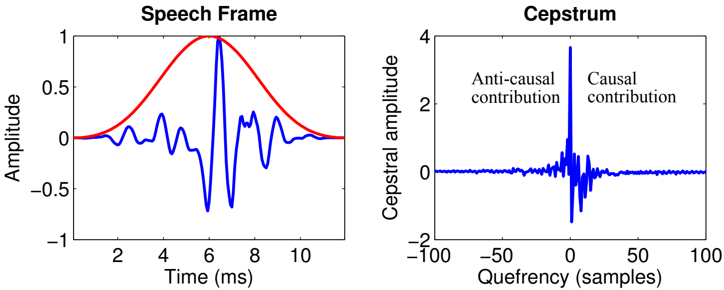
Iterative Adaptive Inverse filtering (IAIF)



- Modelling/inverse filtering of glottal source/vocal tract with increasing prediction order.

Glottal inverse filtering methods -

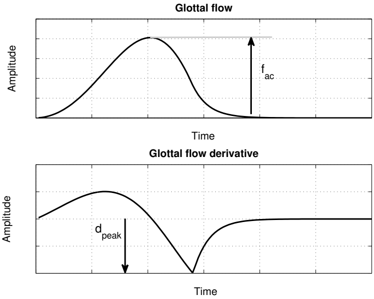
Complex-cepstrum based decomposition (CCEPS)



- Retain only negative quefrency component of the complex cepstrum for glottal source estimation

Parameterisation - Direct measures

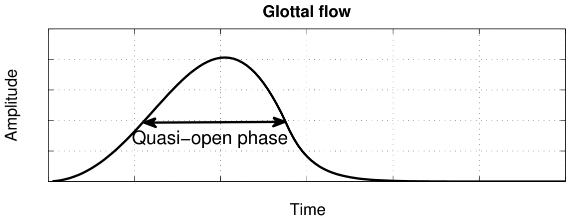
Normalised Amplitude Quotient (NAQ)



- Amplitude based correlate of the Closing Quotient

Parameterisation - Direct measures

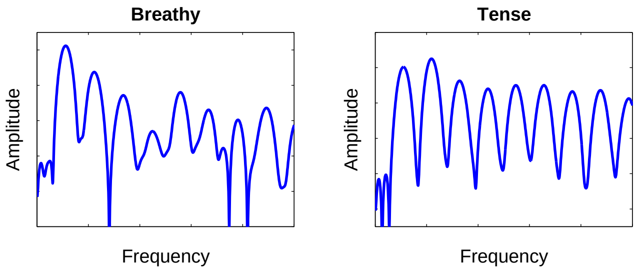
Quasi-Open Quotient (QOQ)



- Amplitude based correlate of the Open Quotient

Parameterisation - Direct measures

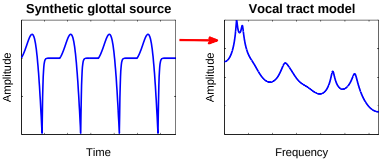
Amplitude difference of first two harmonics (H1-H2)



Parameterisation - Model based methods

- Time domain LF model fitting algorithm (Strik-LF - IAIF)
- Dynamic programming based LF fitting, with time, frequency and transition error criteria (DyProg-LF - IAIF)
- LF model shape parameter (Rd) derived by minimising phase based error criterion (Degott-LF)

Experimental setup - Synthetic stimuli



	GLOTTAL SOURCE				VOCAL TRACT FILTER
	f_0 (Hz)	Ra	Rk	Rg	Vowel
Max	300	0.15	0.5	2.0	8 Vowels settings
Min	80	0.01	0.1	0.6	
Step	10	0.02	0.05	0.1	

=> 198,720 synthetic signals

=> **Relative error:** parameter derived from synthetic source (reference) vs derived following inverse filtering

Experimental setup - Discrimination of phonation types

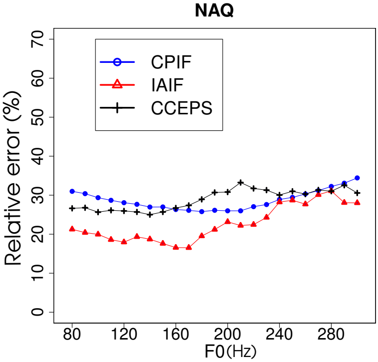
All-voiced sentences

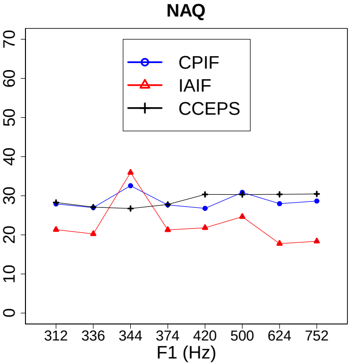
- 6-speakers × 5 -sentences × 3-phonation types
- 3-speakers × 10-sentences × 3-phonation types

Analysis

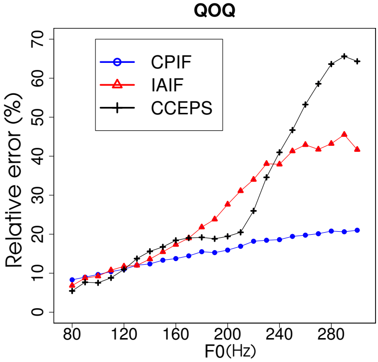
- Parameter contours resampled to 10-samples
- Pearson (R^2): parameter value-*dependent variable*, phonation type-*independent variable*

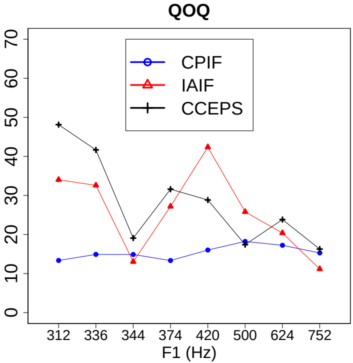
Results - Synthetic stimuli (NAQ)



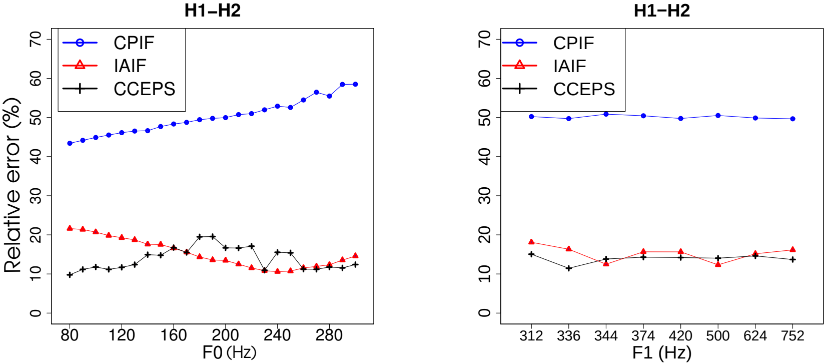


Results - Synthetic stimuli (QOQ)

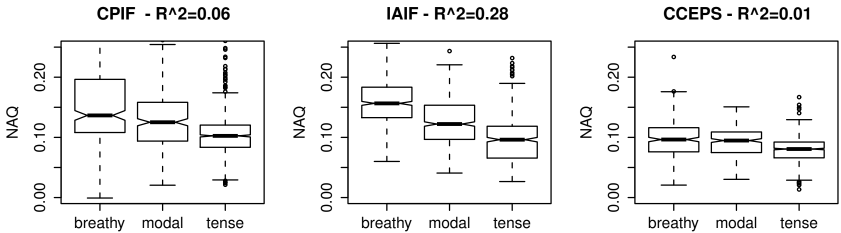




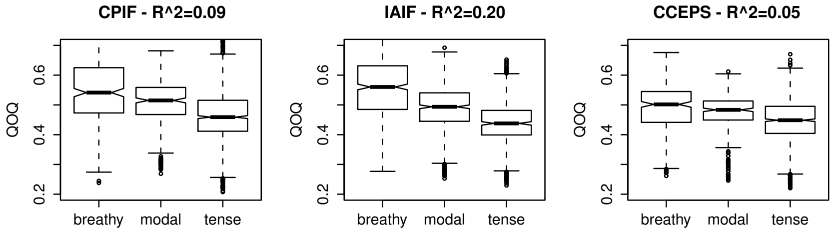
Results - Synthetic stimuli (H1-H2)



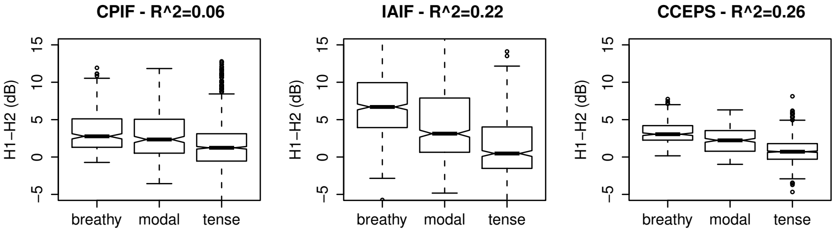
Results - Discrimination of phonation types (NAQ)



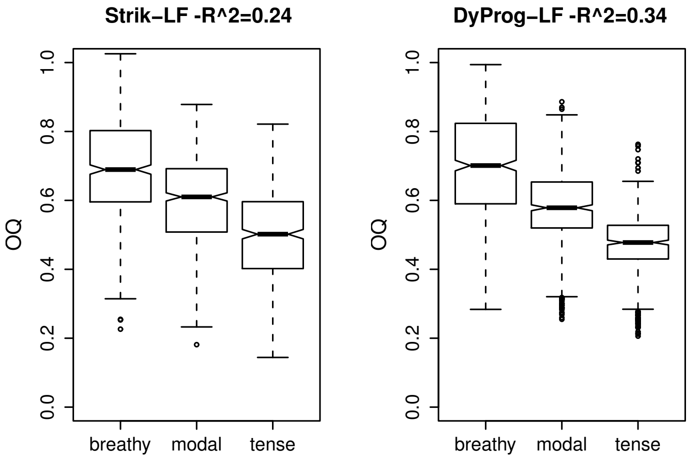
Results - Discrimination of phonation types (QOQ)



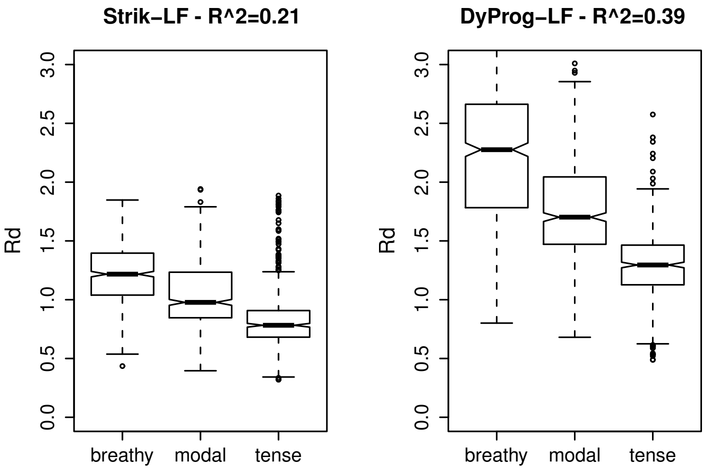
Results - Discrimination of phonation types (H1-H2)



Results - Discrimination of phonation types (model-based-OQ)

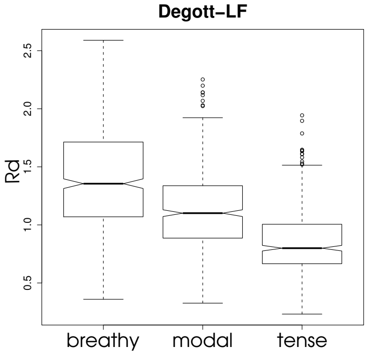


Results - Discrimination of phonation types (model-based-Rd)



Results - Discrimination of phonation types (model-based-Rd)

R^2=0.28



Discussion - Take-home message

- IAIF more robust than in Drugman et al., otherwise results corroborated
- Model based parameterisation better than previously reported (e.g., Airas & Alku, 2007).

Future/ongoing work

- Phonetic feature extraction to determine optimal analysis regions
- Parameterisation and classification of variation in phonation types
- Analysis and synthesis of voice quality variation

Voice Analysis Toolkit

https://github.com/jckane/Voice_Analysis_Toolkit

Thank you!

https://github.com/jckane/Voice_Analysis_Toolkit