Modeling the creaky excitation for parametric speech synthesis.

[1]Thomas Drugman, [2]**John Kane**, [2]Christer Gobl

September 11th, 2012
Interspeech
Portland, Oregon, USA

[1]University of Mons, Belgium
[2]Trinity College Dublin, Ireland

Creaky voice - examples

TTS corpora examples

American Male

Finnish female

Finnish Male

Conversational speech examples

Japanese female

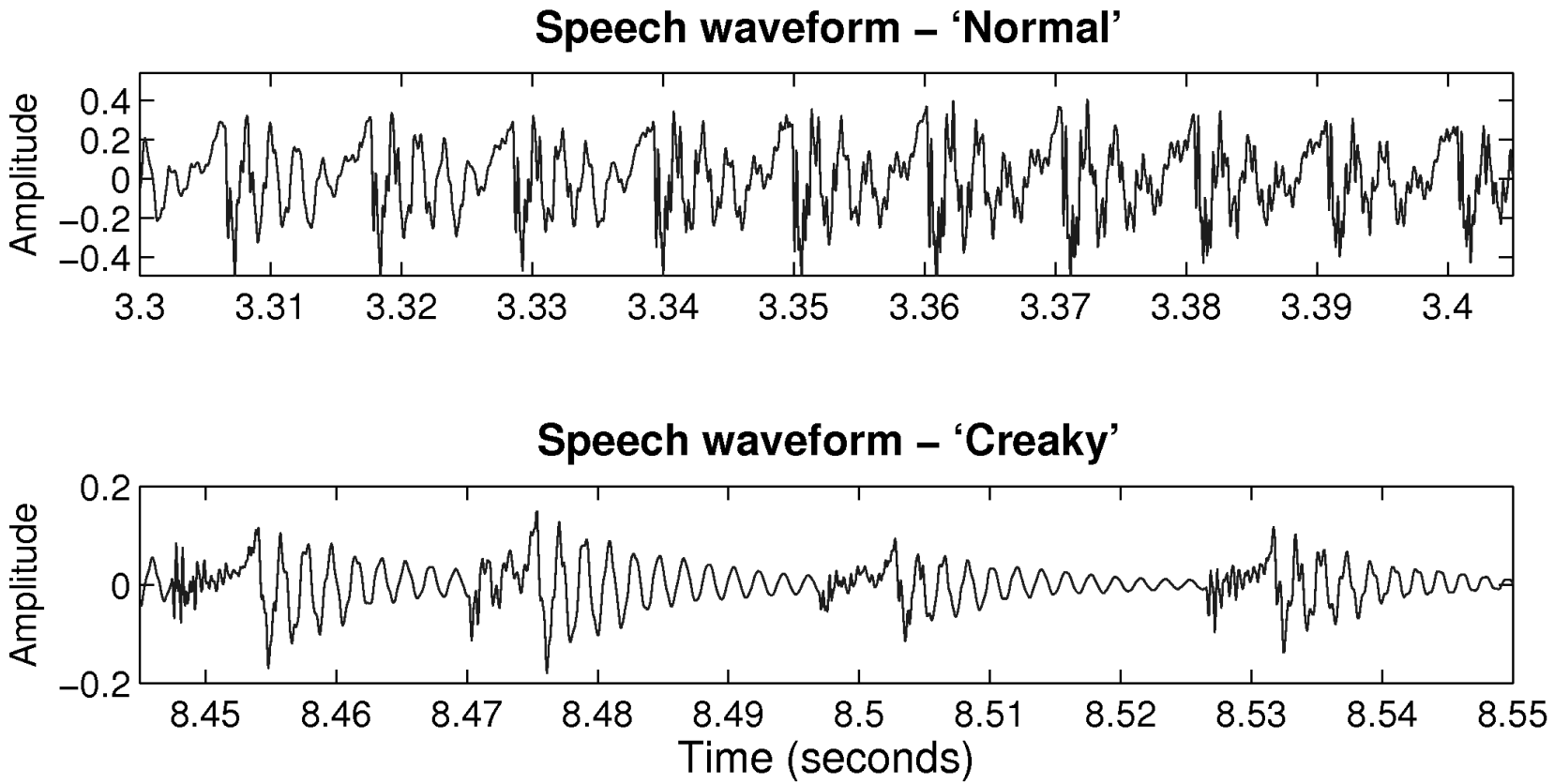
American female

American Male

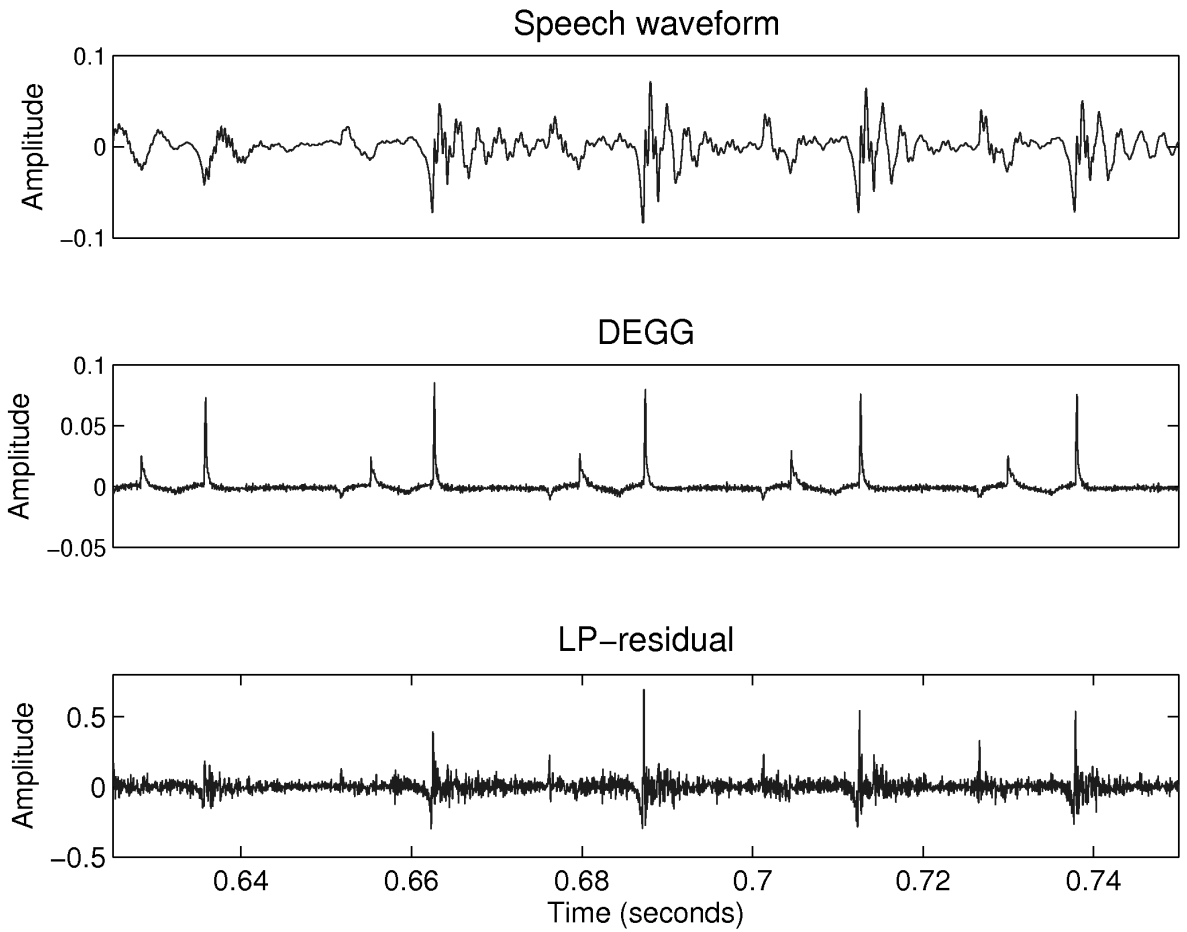
Creaky voice in speech

- Phonetic contrast
  - e.g., Jalapa Mazatec
- Phrase/sentence/turn boundaries
  - Commonly in American English, Finnish etc.
- Interactive speech
  - Turn-taking
  - Hesitations
  - Expression of affective states
  - Stylistic device

Creaky voice - acoustic characteristics



Creaky voice - acoustic characteristics



Problem statement

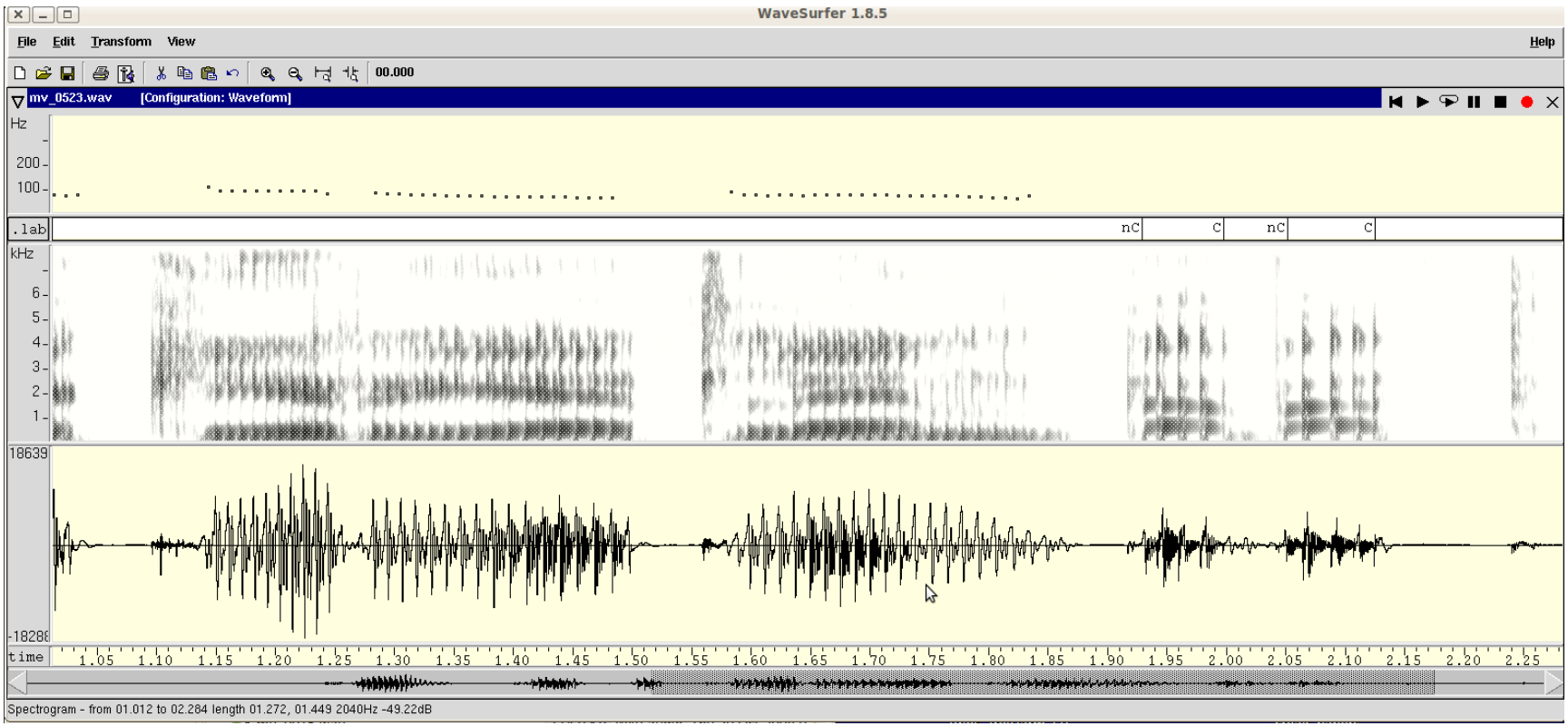
- Unique acoustic characteristics of creak poorly modelled in standard vocoders
- Silen *et al.* (2009) - improved robustness of f_0 and voicing decision
- **Our Aim:** Provide a method for modelling the creaky excitation to improve the timbre of creak in parametric synthesis.

Speech data

- American male (BDL) and Finnish male (MV)
- 100 sentences containing creak

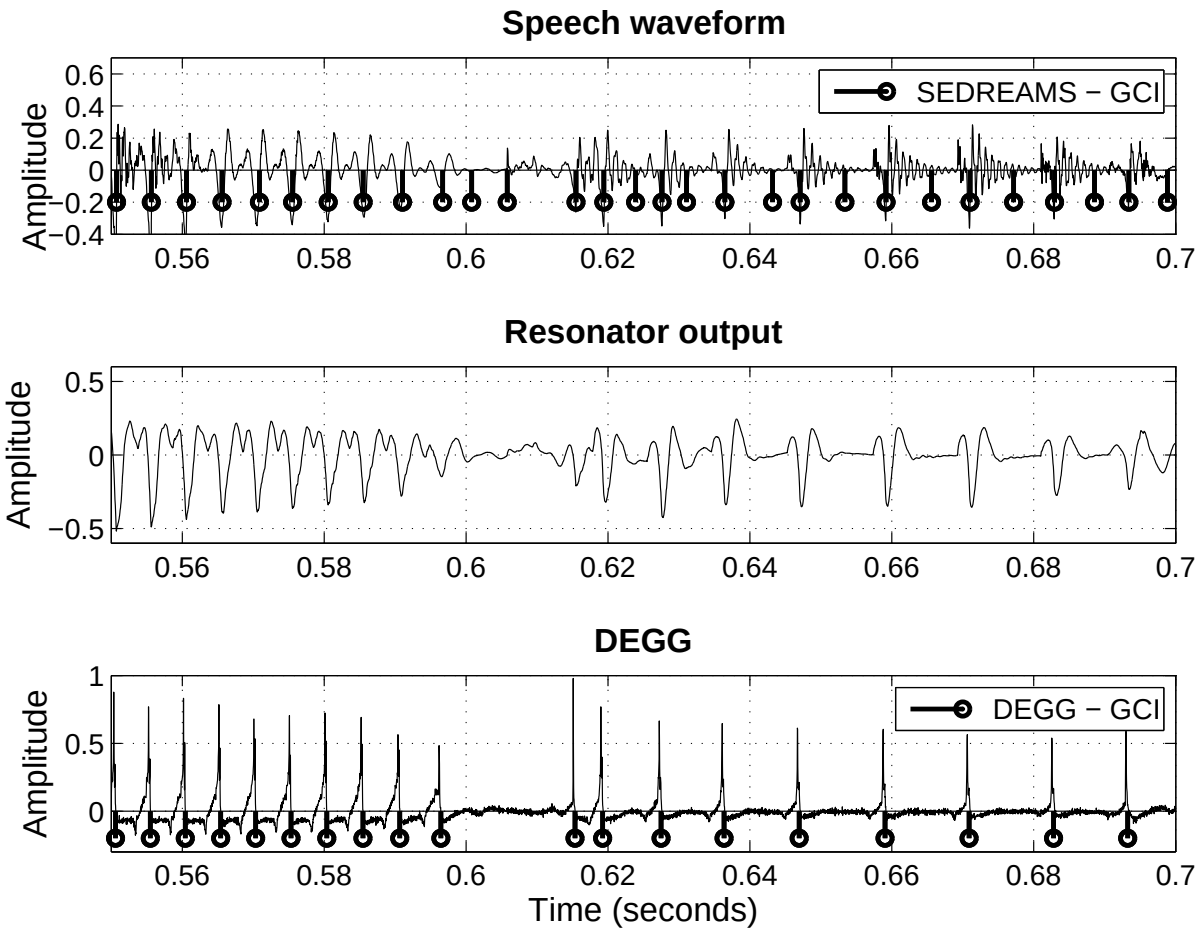
Manual annotation

A rough quality with the sensation of repeating impulses
- Ishi *et al.* (2008)



Glottal closure instants (GCIs)

Newly developed SE-VQ algorithm - Kane & Gobl, In Press

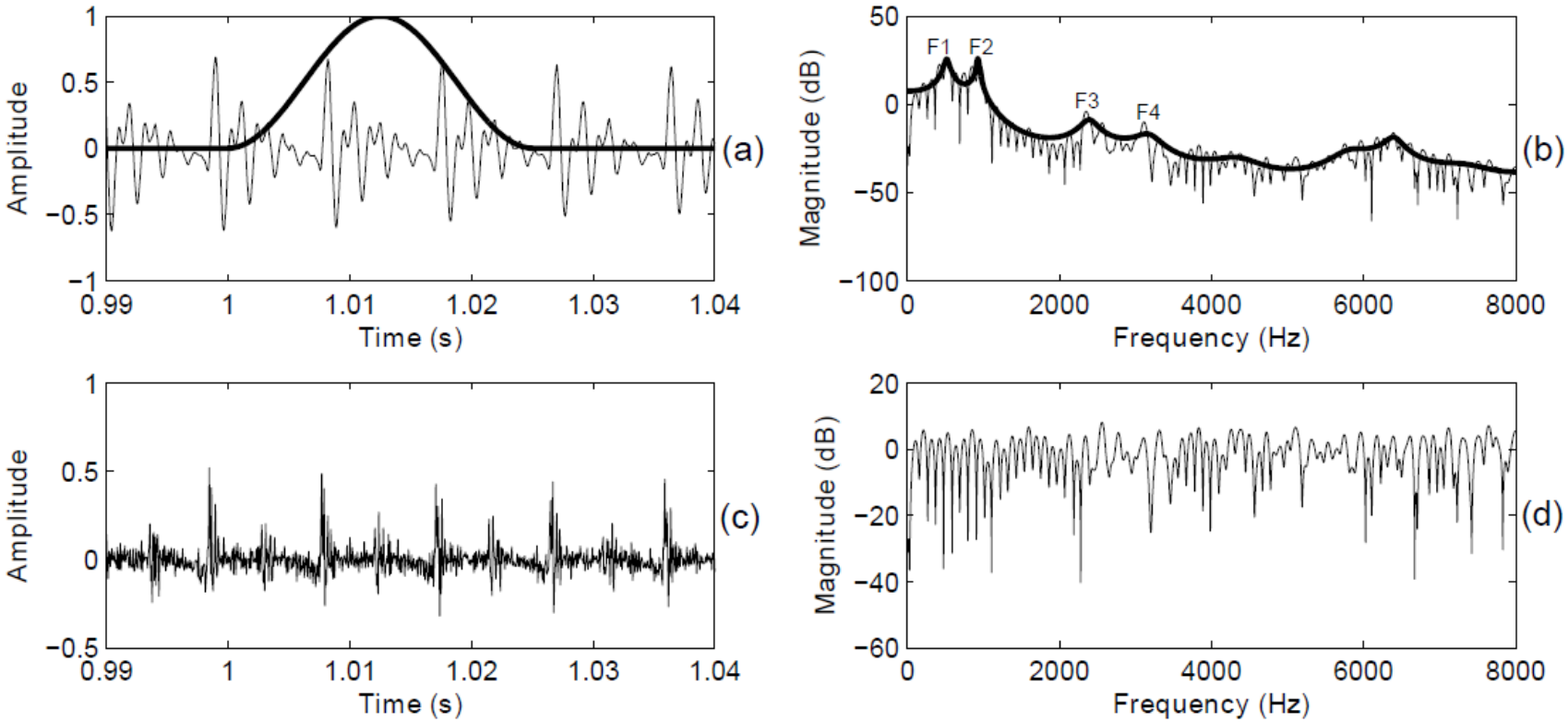


The deterministic plus stochastic model (DSM)

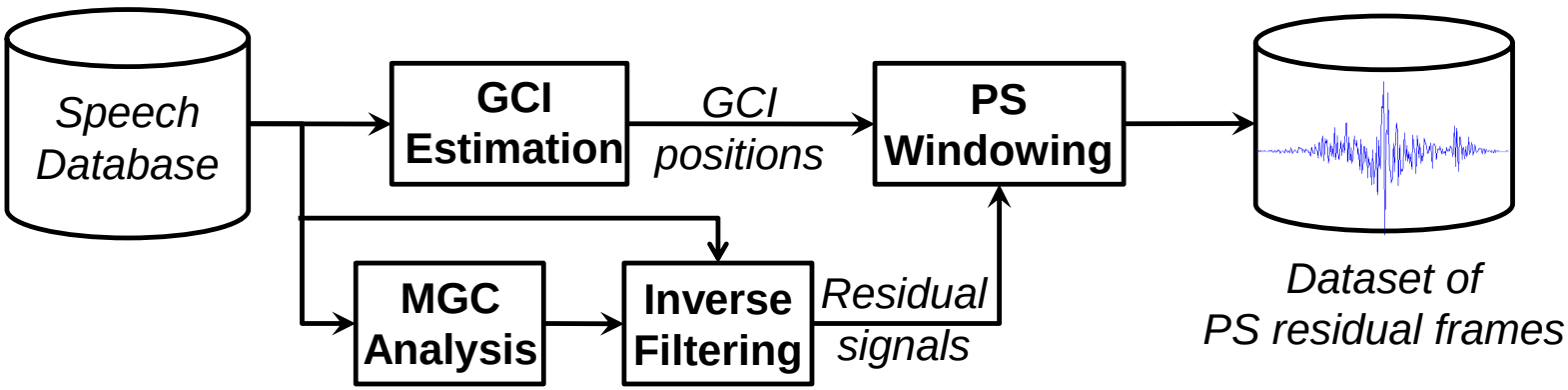
The Deterministic plus Stochastic Model
of the Residual Signal and its Applications

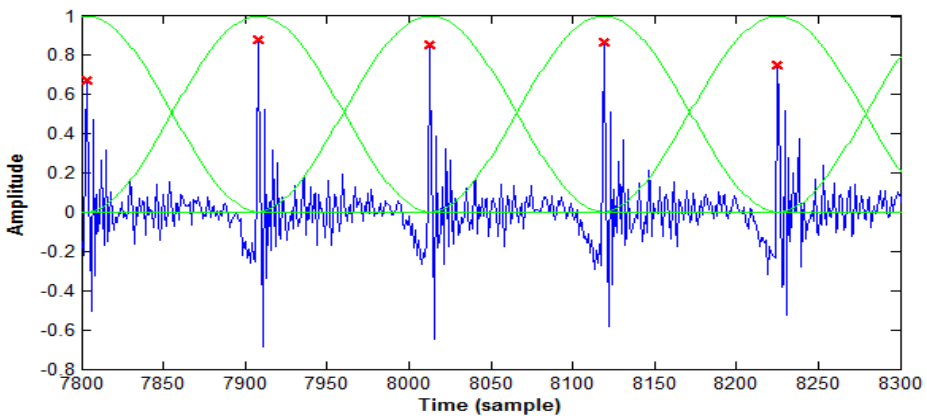
-Drugman & Dutoit (2012), IEEE TASLP

DSM - residual excitation

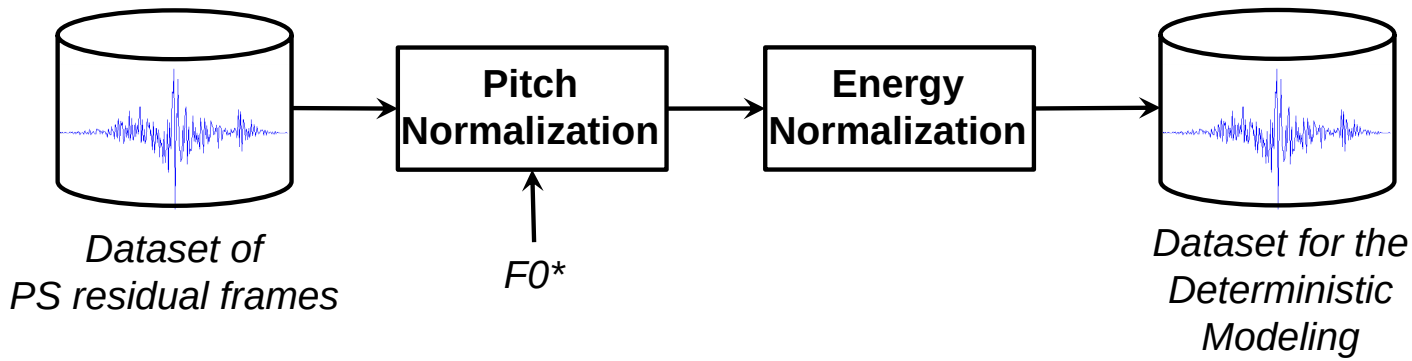


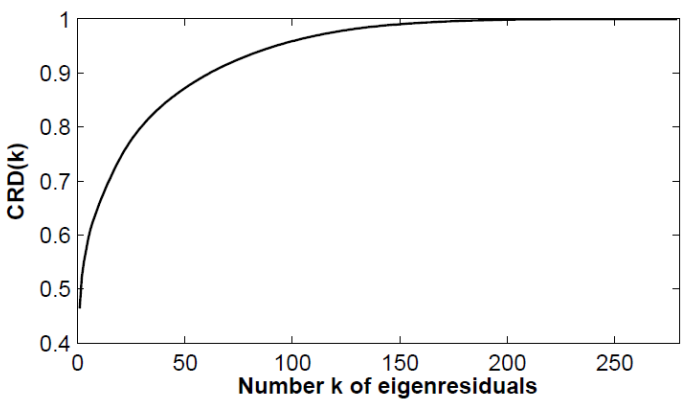
DSM - Residual frames

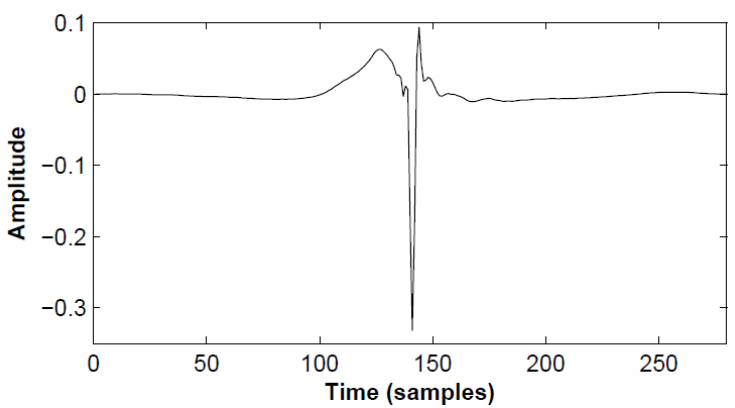




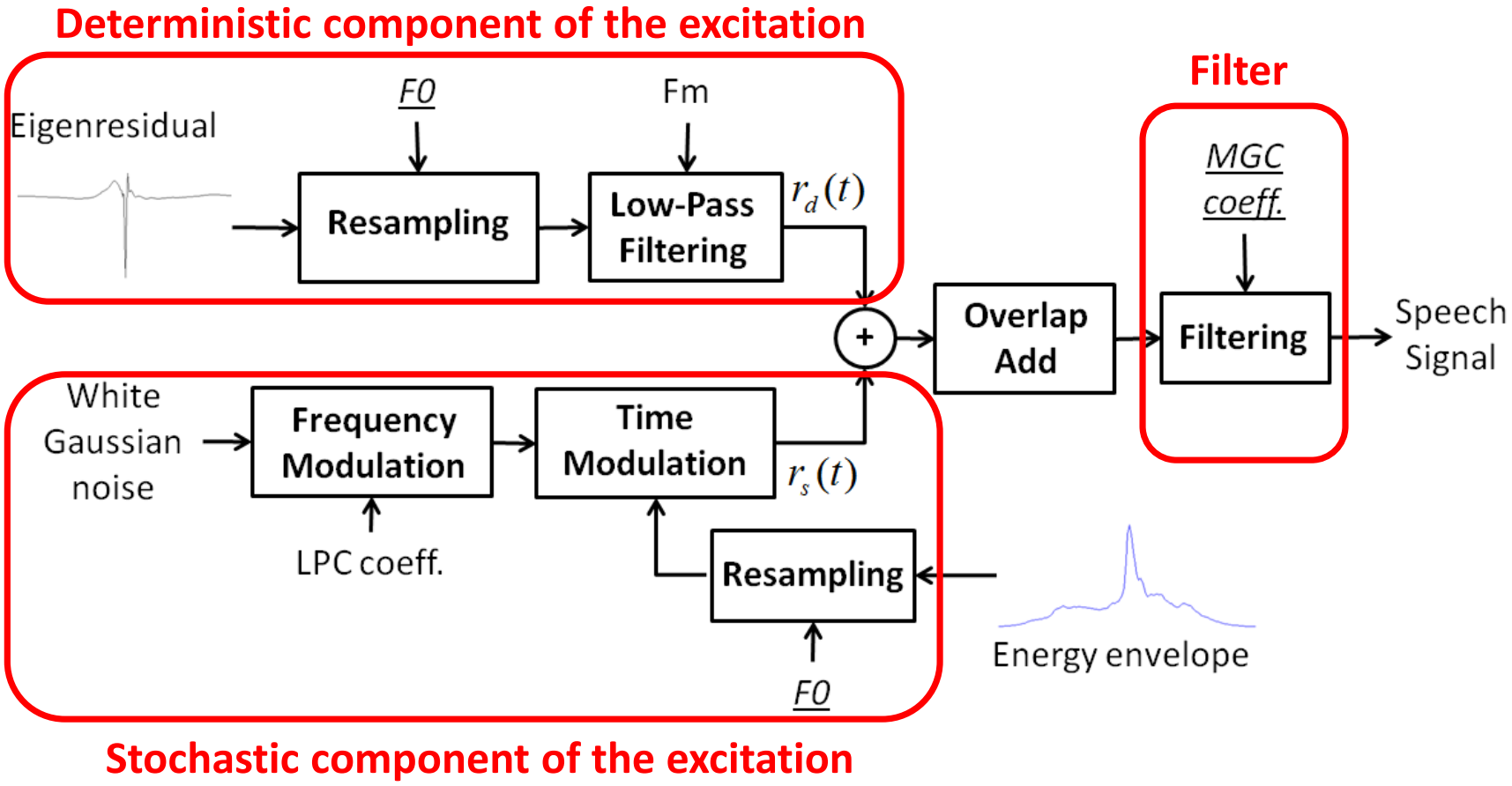
DSM - Deterministic modelling





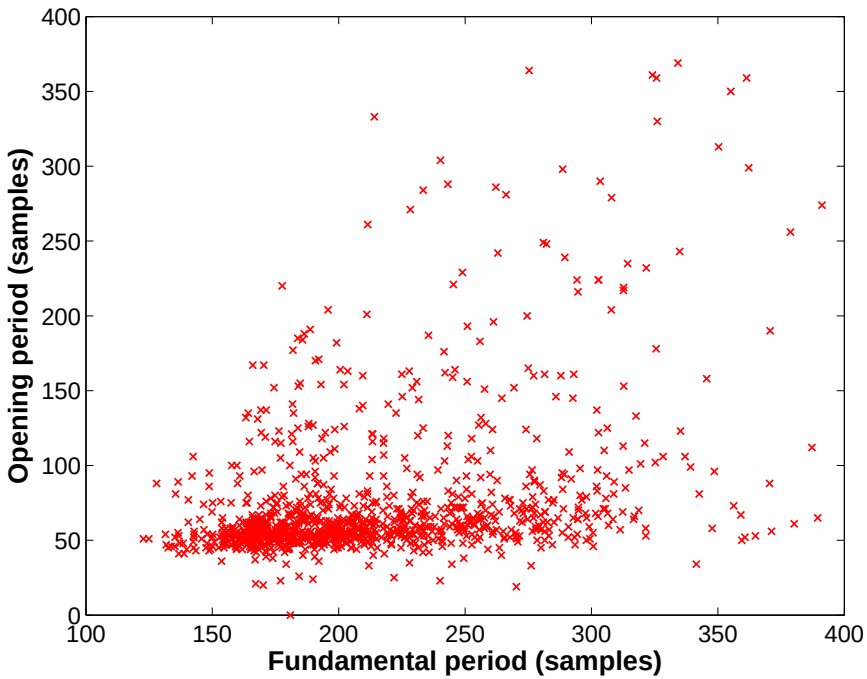


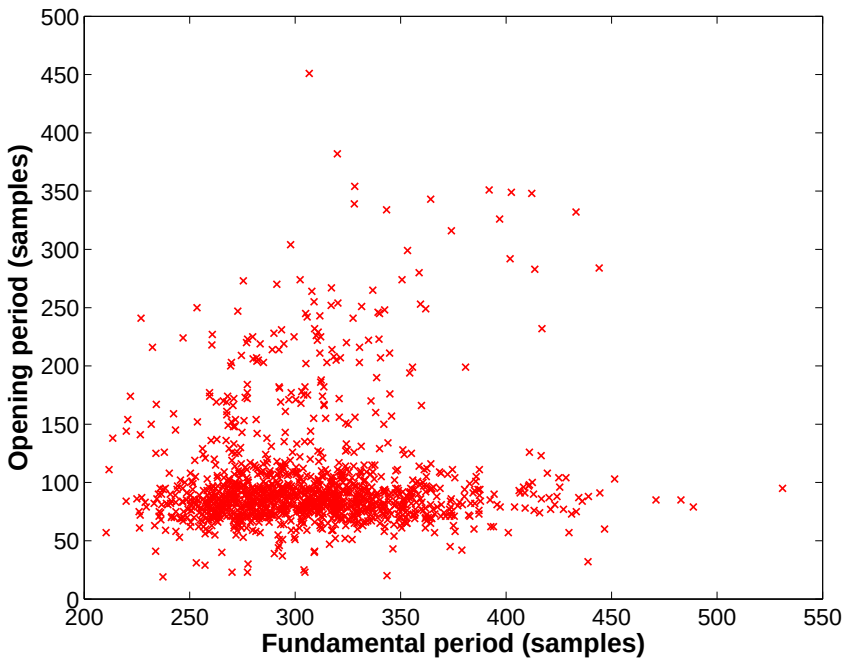
DSM - vocoder


Deterministic component of the excitation
Eigenresidual
F0
Fm
Resampling
Low-Pass Filtering
$r_d(t)$
Filter
MGC coeff.
Filtering
Overlap Add
Speech Signal
+
White Gaussian noise
Frequency Modulation
Time Modulation
$r_s(t)$
LPC coeff.
Resampling
F0
Energy envelope
Stochastic component of the excitation

Extended DSM for creaky voice

DSM (creak) - Fundamental period/opening phase

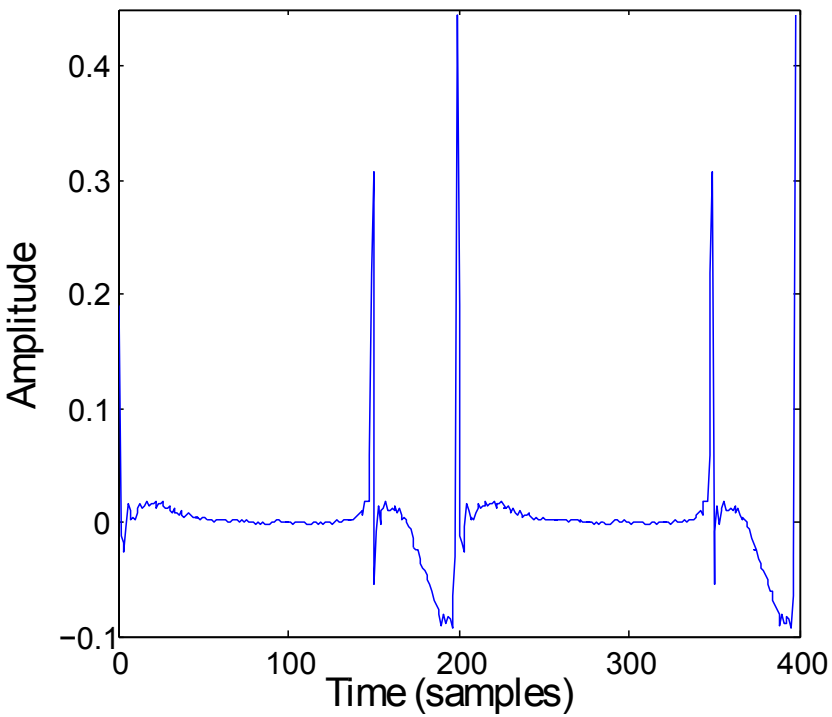


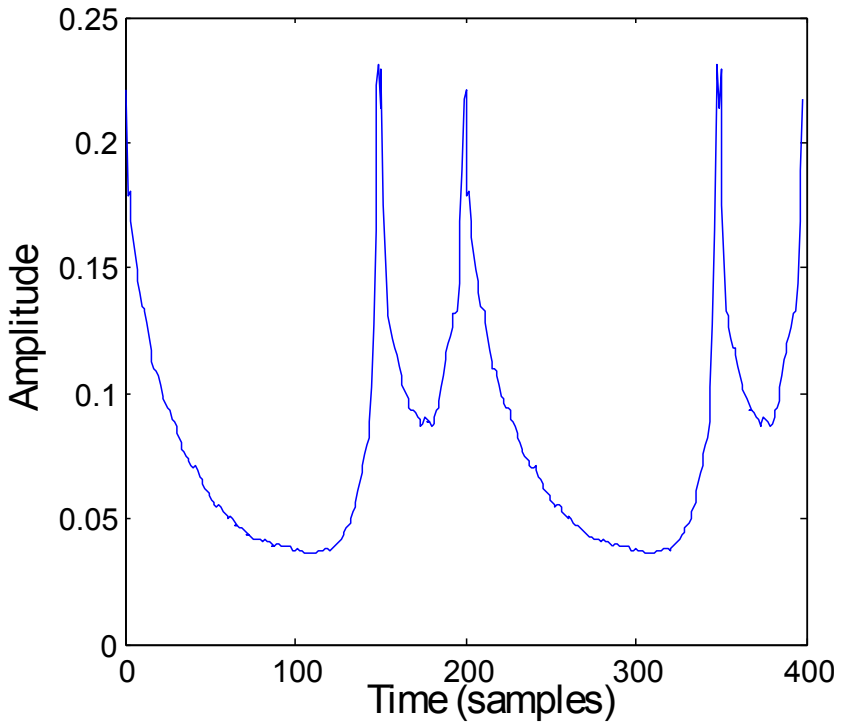


DSM (creak) - Excitation modelling

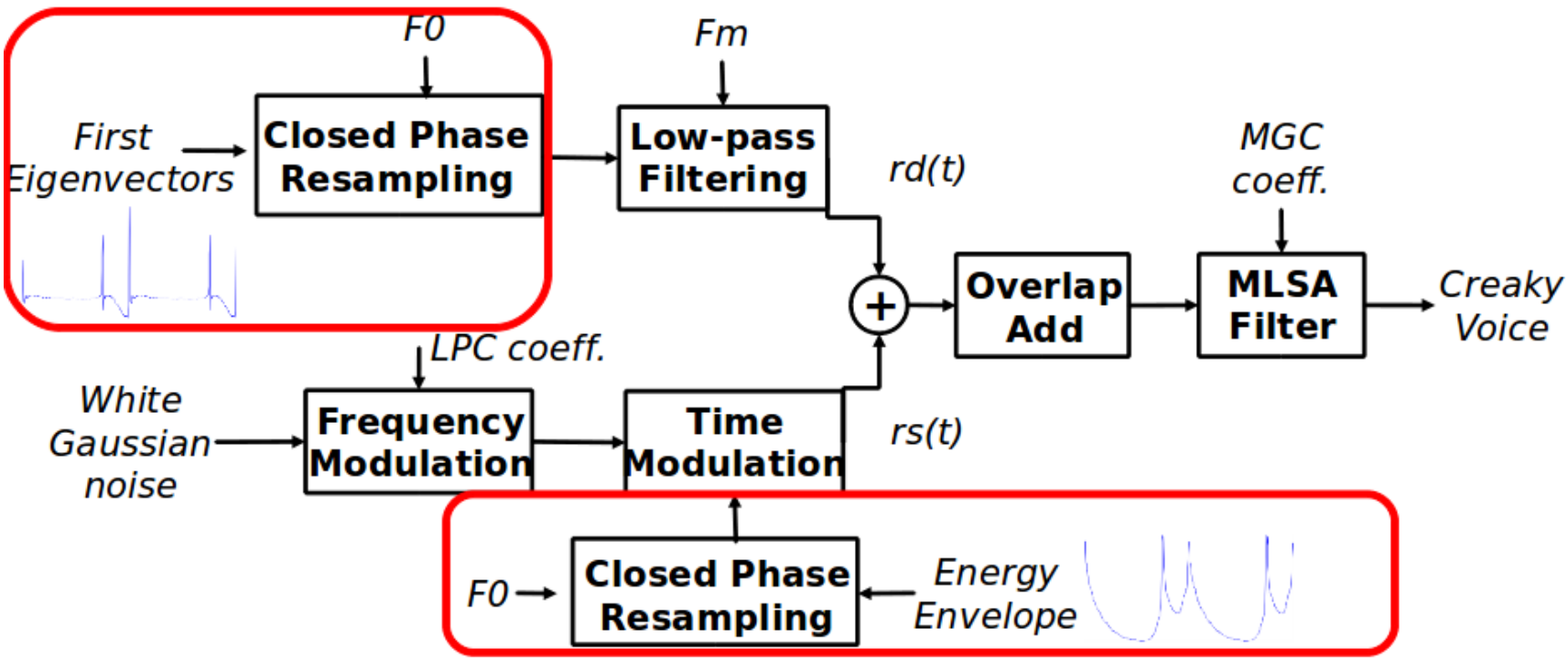
- Separate residual datasets for opening phase (secondary peak => GCI) and closed phase (GCI => secondary peak)
- Principal component analysis of each dataset separately, excitation model combining first eigenvectors for deterministic component.
- Energy envelope also derived for the two datasets separately.

DSM (creak) - Data-driven excitation signal





DSM (creak) - Vocoder

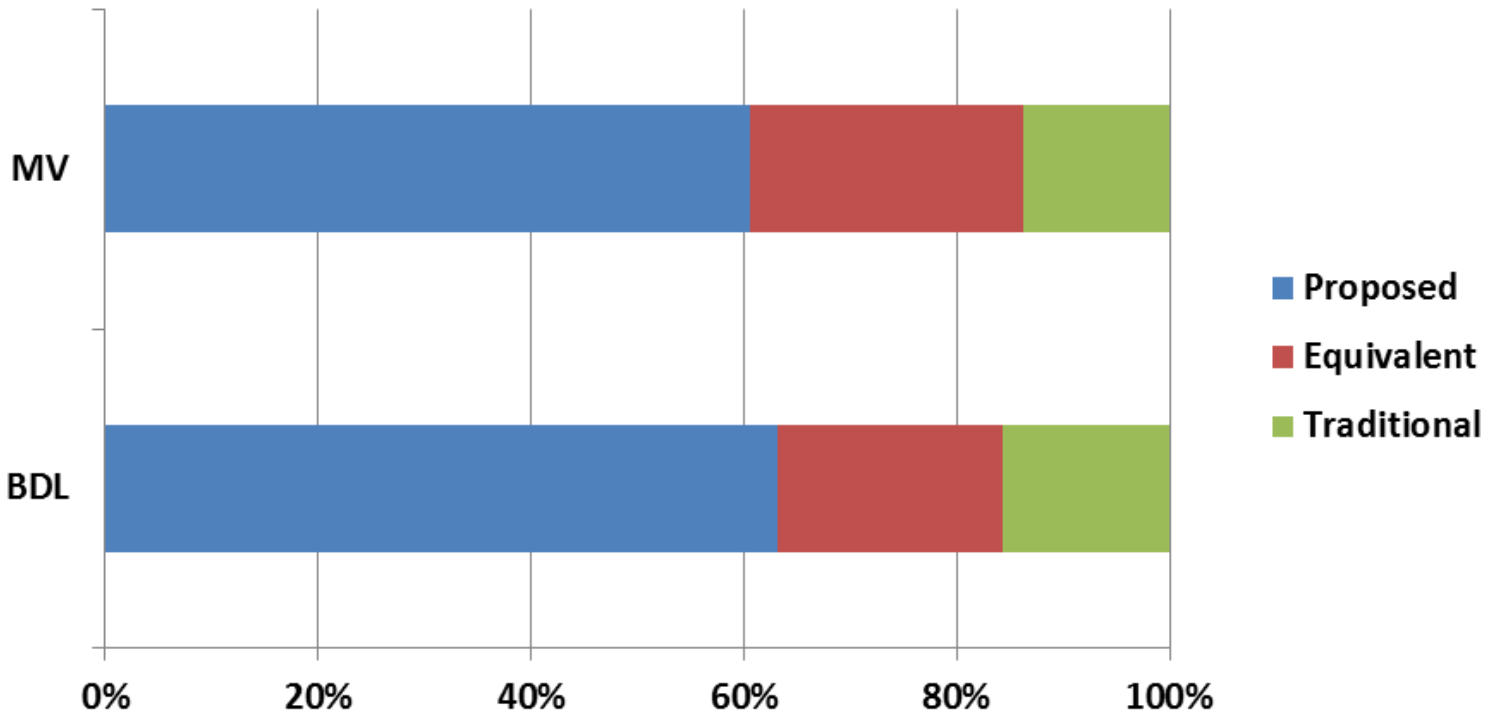


Evaluation

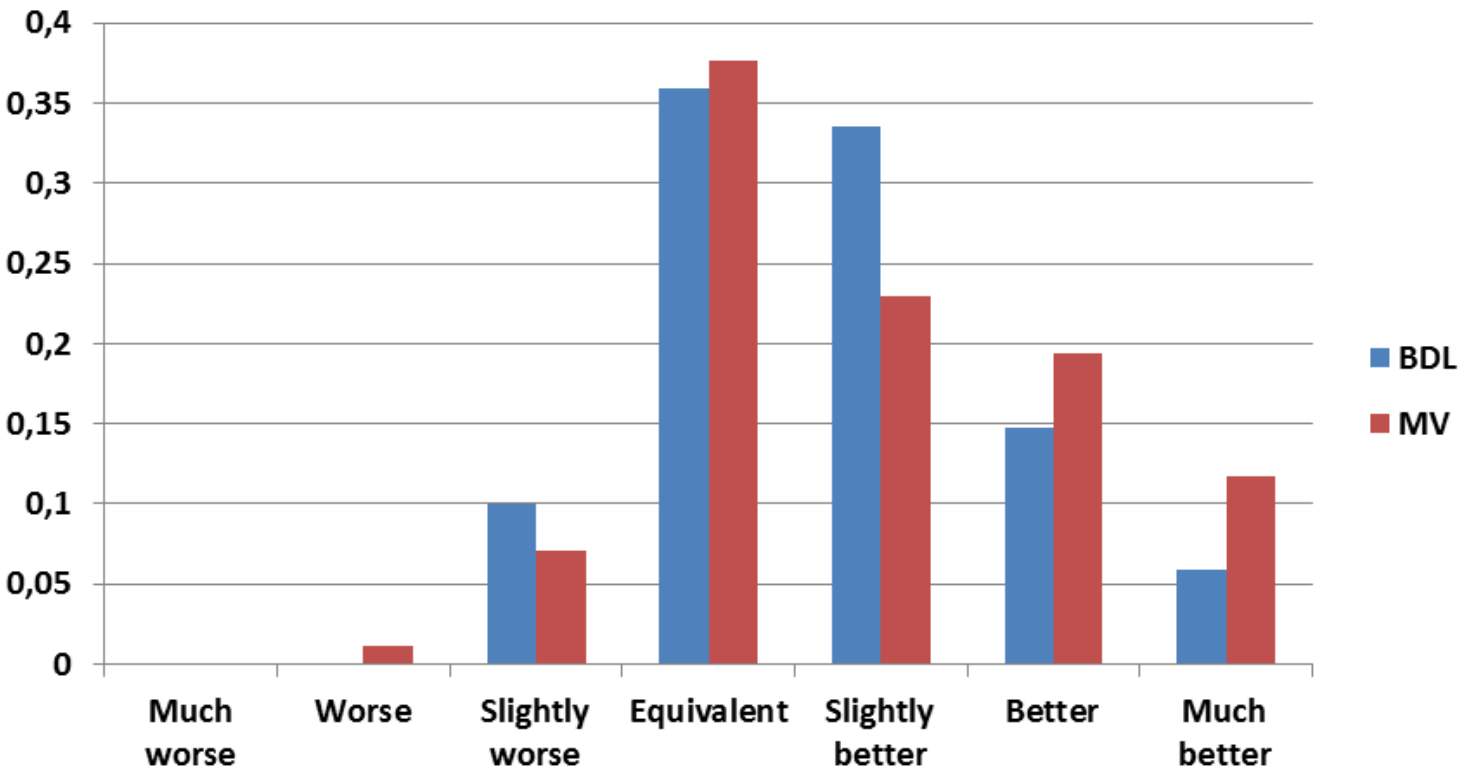
Experimental setup

- Subjective evaluation with 22 participants.
- Copy-synthesis of short utterances by the American and Finnish speaker using the standard DSM vocoder and the proposed method.
- **ABX test**
  - Original utterance (X) and the two copy synthesis versions (A & B). Select most like original
- **Comparative Mean Opinion Score (CMOS) test**
  - Copy synthesis by both vocoders - signal preference on gradual 7 point CMOS scale.

Results - ABX



Results - Comparative Mean Opinion Score (CMOS)



Results - Samples

American Male

1. *Original* *standard HTS vocoder* *DSM vocoder* *DSM-creak*
2. *Original* *standard HTS vocoder* *DSM vocoder* *DSM-creak*
3. *Original* *standard HTS vocoder* *DSM vocoder* *DSM-creak*

Finnish Male

1. *Original* *standard HTS vocoder* *DSM vocoder* *DSM-creak*
2. *Original* *standard HTS vocoder* *DSM vocoder* *DSM-creak*
3. *Original* *standard HTS vocoder* *DSM vocoder* *DSM-creak*

Ongoing/future research directions

- Automate creak segmentation (see our poster at special session - glottal source processing!)
- Prediction of creaky regions from contextual features (e.g., phoneme, word stress, position in sentence, prosodic context etc.)
- Transformation of speakers voice characteristics.

Acknowledgements

- This work was supported by the Science Foundation Ireland, Grant 07 / CE / I 1142 (Centre for Next Generation Localisation, `www.cngl.ie`) and Grant 09 / IN.1 / I 2631 (FASTNET).

Thank you!