A Spectral Glottal Flow Model for the Source-Filter Separation of Speech

PERROTIN Olivier, GIPSA-lab, *olivier.perrotin@gipsa-lab.fr*
MCLOUGHLIN Ian, University of Kent, *I.V.McLouglin@kent.ac.uk*

Context

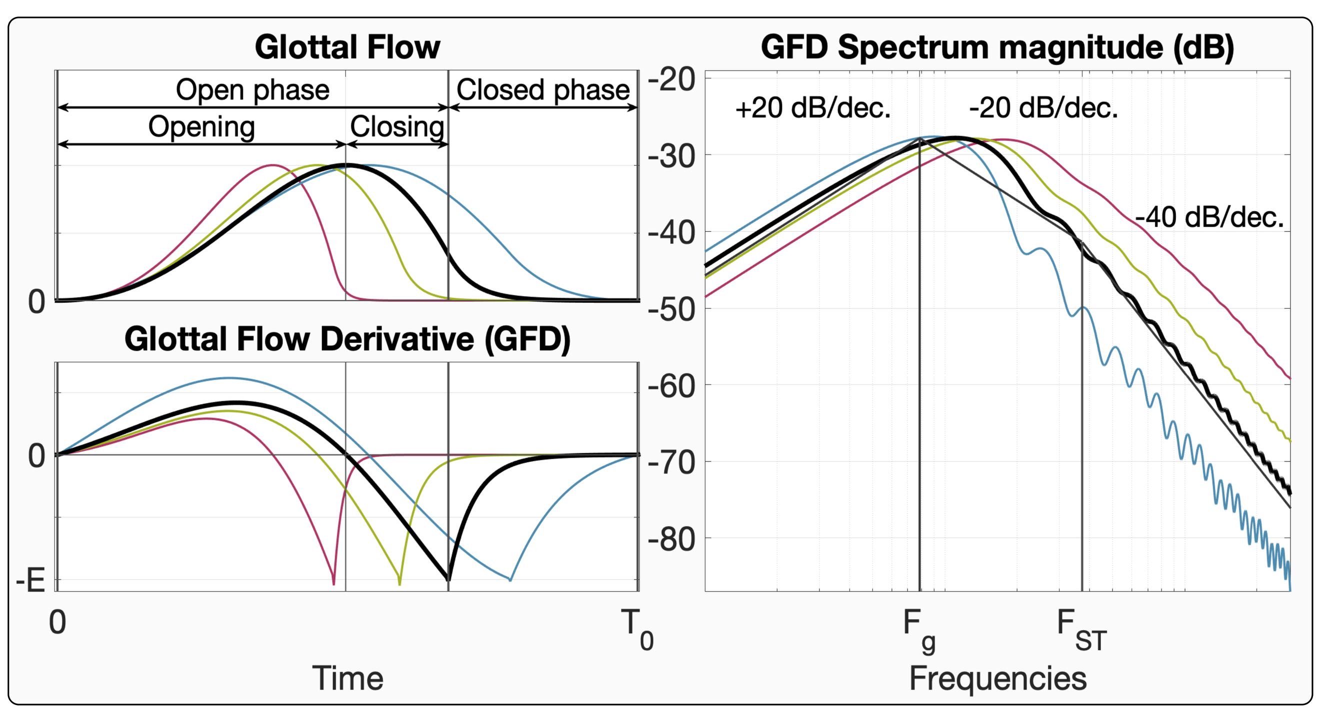
- Estimation of glottis signal from speech is still an unsolved problem [1]
- Issue of modelling the wideband characteristics of the glottal flow: variations of *glottal formant* and *spectral tilt*

➤ Demonstration and evaluation of the **Glottal Flow Model based Iterative Adaptive Inverse Filtering (GFM-IAIF)** method

A spectral glottal flow model

Description of the glottal source signal in the frequency domain as a **third order filter** [2]

Vibration of vocal folds	Spectral correlate
Open phase	Glottal formant
Duration / Asymmetry	*Position (F_g) / Bandwidth (B_g)*
Closing phase	Spectral tilt
Abruptness	*Cutting Frequency (F_{ST})*

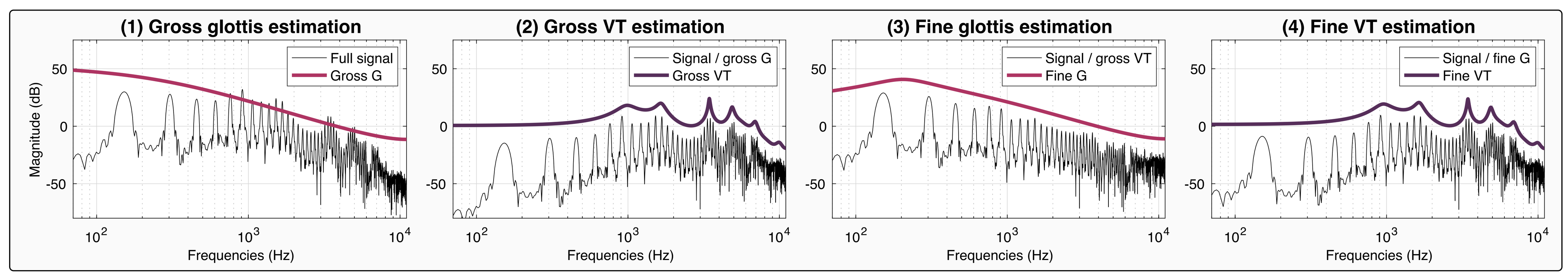


Spectral measurements

- *First-to-second harmonic ratio* (H1H2)
- *Harmonic Richness Factor* (HRF)
  Sum of 2nd to nth harmonics (in dB) over the 1st
- *Spectral Tilt* (ST)
  Linear regression of the nth first harmonics on a log-log scale

IAIF-based algorithm

1. Gross estimation of glottis spectral envelope
2. Remove estimated glottis contribution
   Gross estimation of VT spec. envelope
3. Remove estimated vocal tract (VT)
   Fine estimation of glottis spec. envelope
4. Remove estimated glottis contribution
   Fine estimation of VT spec. envelope

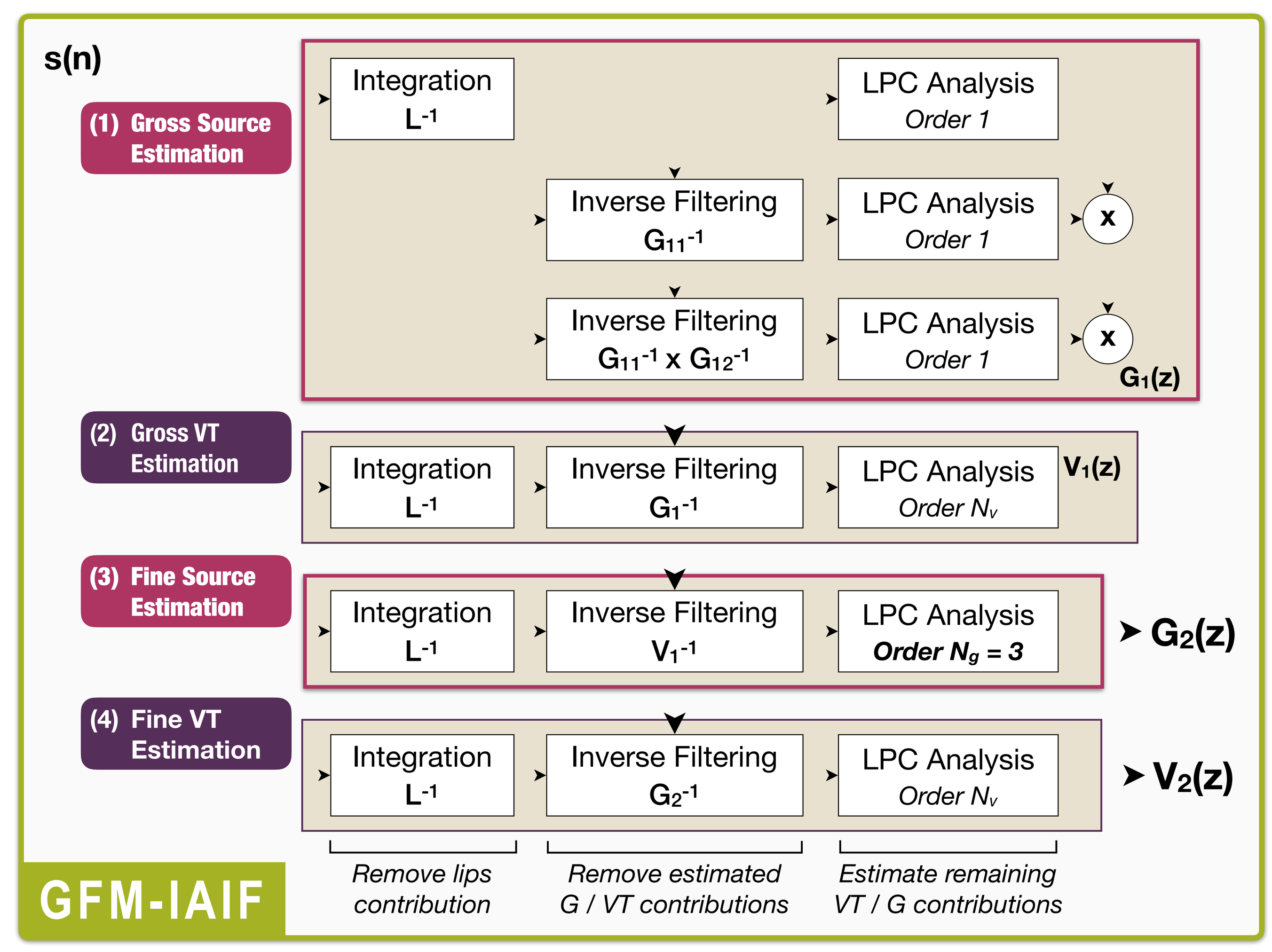


Gross source estimation in IAIF-based methods

Standard IAIF [3]	1st order LPC Pre-emphasis
IOP-IAIF [4]	M x 1st order LPC Until the spectrum is flat
GFM-IAIF	**3 x 1st order LPC** w.r.t. glottal flow model

Fine source estimation in IAIF-based methods

IAIF [3] IOP-IAIF [4]	Nth order LPC Manually optimised
GFM-IAIF	**3rd order LPC** w.r.t. glottal flow model



Evaluation

Corpus

- Synthetic speech: LF waveform + formant filtering
  30 voice qualities (R_d parametrisation) x 29 pitch levels x 10 vowels
- Natural speech: vowels extracted from diphones [5]
  3 voice qualities (Soft, Medium, Loud) x 2 speakers (male and female) x 12 vowels

Methods

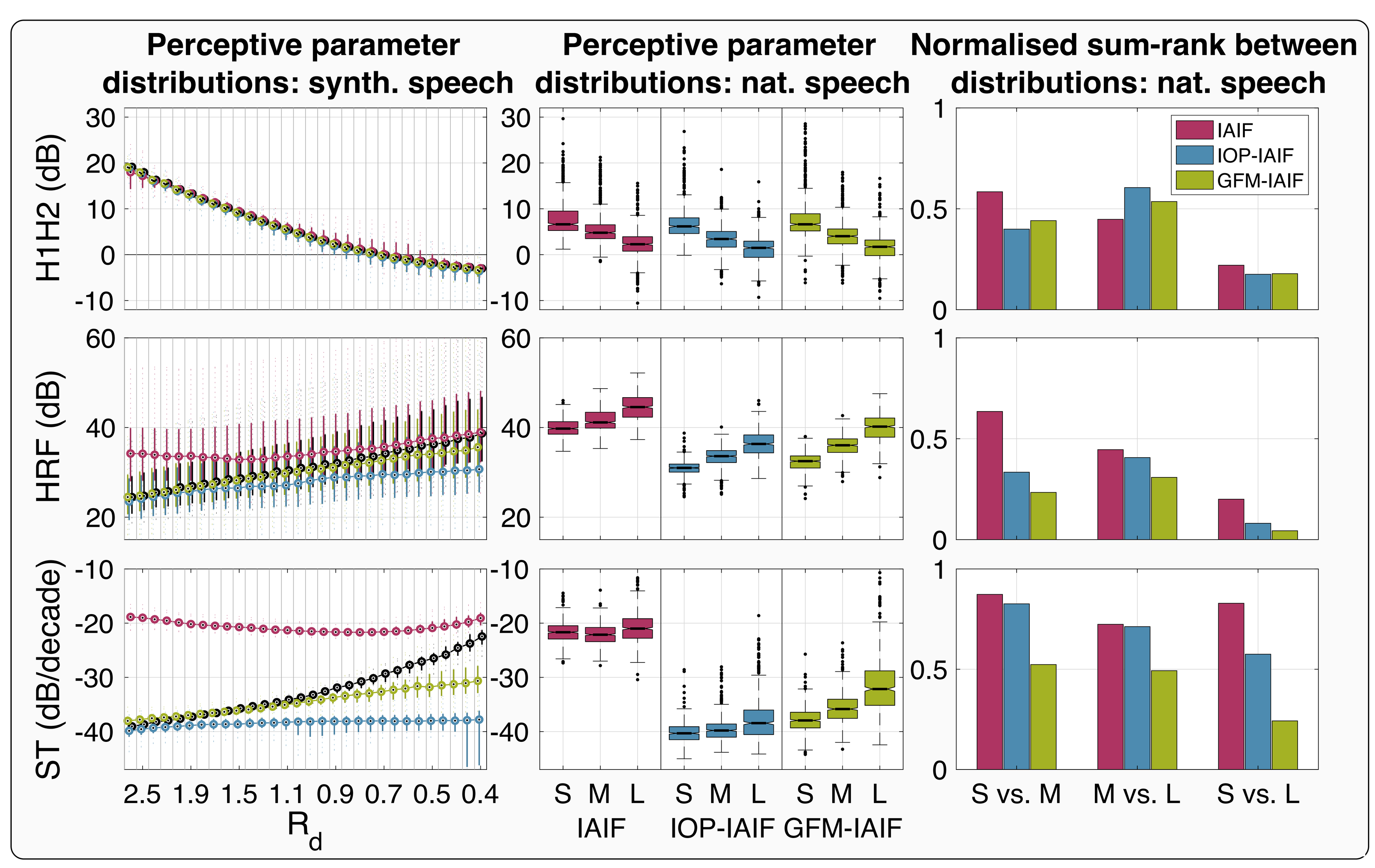
- Comparison of IAIF [3], IOP-IAIF [4], and GFM-IAIF

Analysis

- Extraction of H1H2, HRF, ST, from estimated glottis signals
- Synthetic speech: comparison with extracted parameters from ground truth
- Natural speech: discriminative power of extracted parameters to represent variation in voice quality (Wilcoxon test effect size)

Results

- Consistent variations between synthetic and natural speech
- H1H2 (glottal formant): similar performances
- HRF and ST (high frequency):
  ➤ GFM-IAIF closest to ground truth and more discriminative parameters
  ➤ IOP-IAIF and IAIF tend to attribute too much and not enough spectral tilt, respectively



Conclusion

➤ New source-filter separation method called **GFM-IAIF**
➤ Third order modelling of the glottal flow spectral envelope:
☛ simple parametrisation
➤ Better estimate of spectral tilt compared to other IAIF-based methods

References

[1] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana (2014), "Glottal source processing: From analysis to applications," *Computer Speech & Language*, vol. 28, no. 5, pp. 1117–1138.
[2] B. Doval, C. d'Alessandro, and N. Henrich (2006), "The spectrum of glottal flow models," *Acta Acustica*, vol. 92, no. 6, pp. 1026–1046.
[3] P. Alku (1992), "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Comm.*, vol. 11, no. 2–3, pp. 109–118.
[4] P. Mokhtari, B. Story, P. Alku, and H. Ando (2018), "Estimation of the glottal flow from speech pressure signals: Evaluation of three variants of iterative adaptive inverse filtering using computational physical modelling of voice production," *Speech Comm.*, vol. 104, pp. 24–38.
[5] *http://github.com/numediart/MBROLA-voices/ — de6 and de7 databases*