A comparative study of glottal open quotient estimation techniques

John Kane & Christer Gobl
Phonetics and Speech Lab
Trinity College Dublin, Ireland

Stefan Scherer & Louis-Philippe Morency
Institute for Creative Technologies
University of Southern California, USA

1 Introduction

- Glottal source features **increasingly used in speech technology**, e.g., speech synthesis, speaker identification, emotion classification
- Open Quotient (OQ): **varies significantly** with phonation type
- Time domain OQ extraction: lack of robustness, large computational load
- Frequency domain correlates: confounded by other factors (e.g., skewness of the glottal pulse)
- For emerging speech technology applications (e.g., online vocal behaviour monitoring), **real-time glottal source processing** is required

AIMS

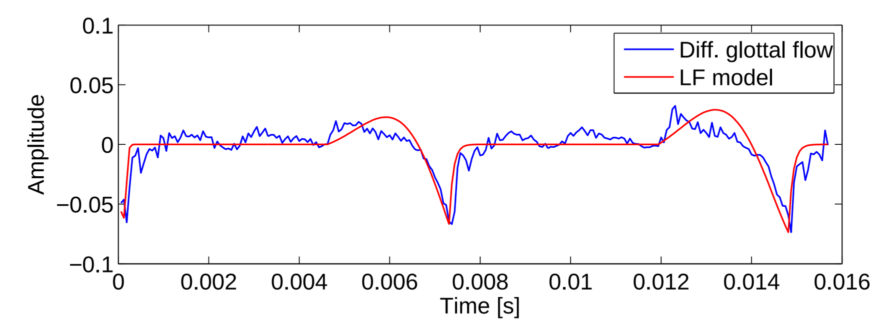
1. Implement an efficient data-driven method for mapping from spectral features to time-domain Open Quotient
2. Evaluate OQ extraction methods:
   (a) Ability to discriminate phonation types
   (b) Error rates (with EGG reference) on a phonetically balanced database
   (c) Computational load

2 Speech data

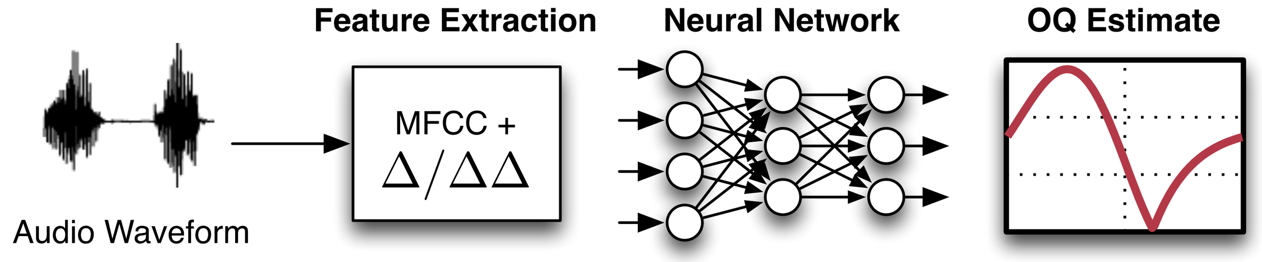
- **Speech data:**
  - Finnish vowels: 7 speakers (4 female), 8 vowels using breathy, modal and tense phonation
  - ARCTIC: 5 speakers: two American male (BDL; 1132 utterances, KED; 450 utterances), an American female (SLT; 1131 utterances), a Canadian male (JMK; 1114 utterances) and a UK male (RAB; 1946 utterances)
  - APLAWD: 5 males and 5 females, 10 repetitions of 5 phonetically balanced sentences. Solely used for training ANN-OQ.

3 OQ estimation techniques

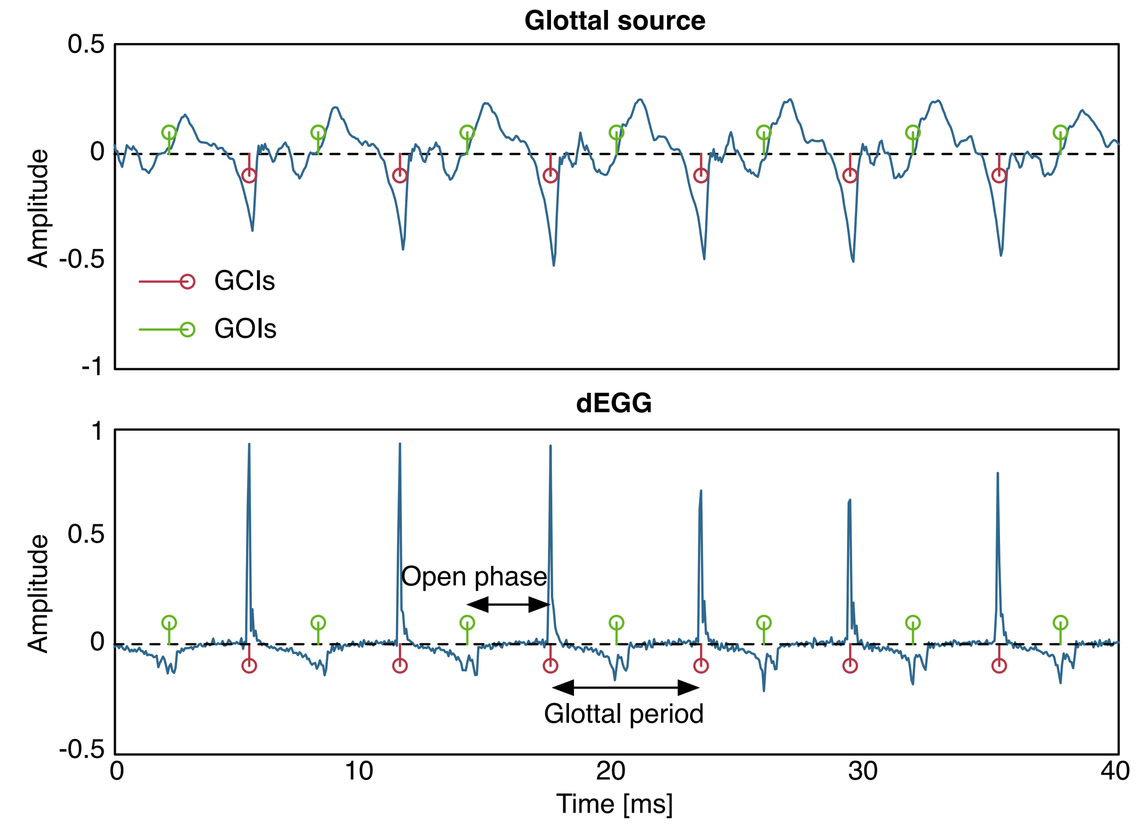
- **Strik-LF [1]:** Glottal inverse filtering (IAIF) and glottal source modelling using Liljencrants-Fant (LF) model



- **YAGA [2]:** Detection of glottal closure/opening instants (GCIs/GOIs): multi-scale product of glottal source with group-delay function. False alarms removed using N-best dynamic programming. OQ(n) $= \frac{\text{GCI}(n) - \text{GOI}(n)}{\text{GCI}(n) - \text{GCI}(n-1)}$
- **ANN-OQ:**
  - Features: 12 Mel-frequency cepstral coefficients (MFCCs): 32 ms Hanning windowed, 10 ms shift, with Δ and Δ^2
  - Learning: Artificial Neural Networks (ANNs), multi-layer perceptron (MLP) with a single hidden layer (contain 100 neurons).

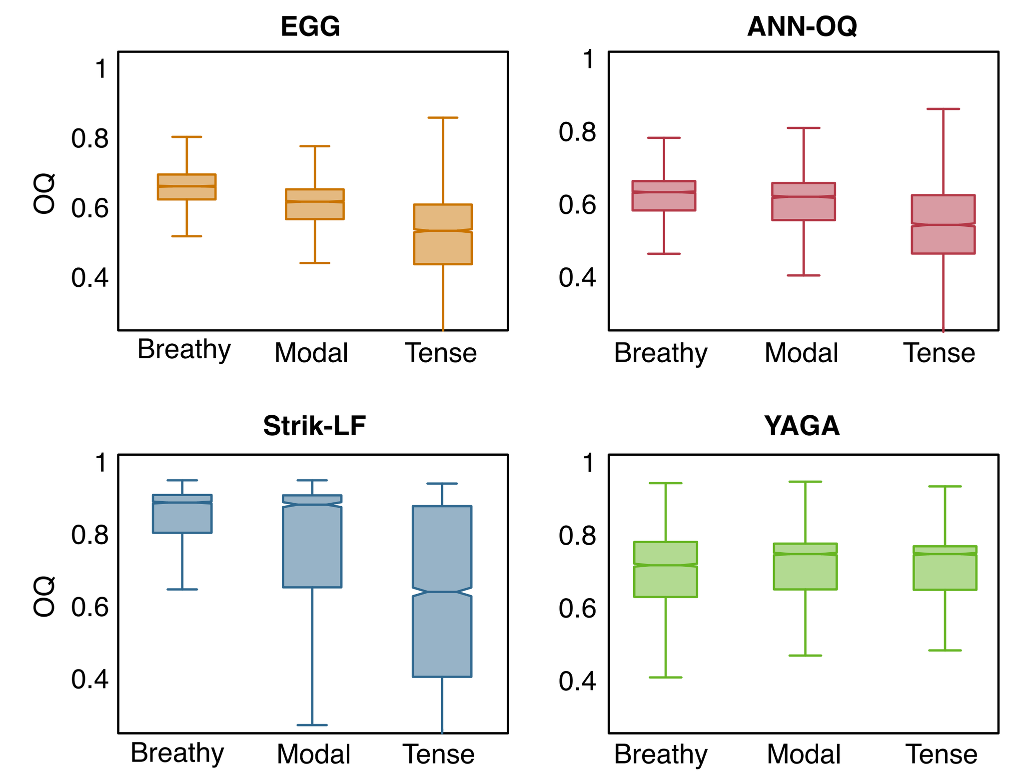


- **Reference OQ values**
  - Detection from Electroglottographic (EGG) signal using SIGMA [3]

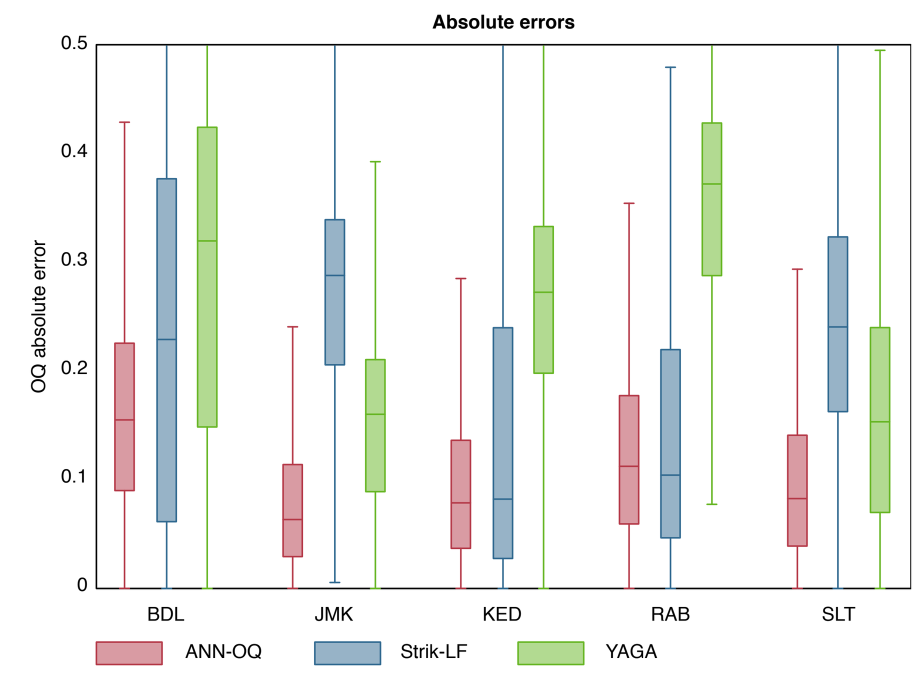


4 Experiments

1. **Discrimination of phonation types**



2. **Error rates on large database**



3. **Computational load (Rel computation time (%))**

Data	Strik-LF	YAGA	ANN-OQ
BDL	178.64	15.17	0.41
SLT	226.45	15.30	0.42

References

[1] Strik, H., (1998) "Automatic parameterization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses" Journal of the Acoustical Society of America, 103(5), pp. 2659-2669.
[2] Thomas, M., Gudnason, J., Naylor, P., (2012) "Estimation of glottal closing and opening instants in voiced speech using the YAGA algorithm" IEEE Transactions on Audio, Speech, and Language Processing, 20(1), pp. 82-91, 2012.
[3] Thomas M., Naylor, P., (2009) "The SIGMA algorithm: A glottal activity detector for electroglottographic signals" IEEE Transactions on Audio, Speech, and Language Processing, 17(8), pp. 1557-1566.

Acknowledgements

This research was supported by the Science Foundation Ireland Grant 09/IN.1/I2631 (FASTNET) and the Irish Department of Arts, Heritage and the Gaeltacht (ABAIR project). We would like to thank Patrick Naylor for providing us with the APLAWD database..

Contact kanejo@tcd.ie, scherer@ict.usc.edu