The Importance of Optimal Parameter Setting for Pitch Extraction

Keelan Evanini[1], Catherine Lai[2]

Educational Testing Service[1], *University of Pennsylvania*[2]

Introduction

- Many studies have compared the performance of different F0 extraction algorithms
- In these studies the pitch extraction parameters may not be given ideal settings
- For example, a recent study showed that SWIPE' and SHS outperformed all other algorithms, but the experiment used unrealistic values for the pitch floor (40 Hz) and pitch ceiling (800 Hz) parameters
- This study compares 5 standard F0 extraction algorithms using optimized values for these two parameters

Speech Corpora

1. FDA: Fundamental Frequency Determination Algorithm Evaluation Database [1]
   - 50 sentences read by one male and one female speaker
   - 37 declaratives and 13 interrogatives (4 yes/no questions and 9 wh-questions)
2. Keele Pitch Database [2]
   - "The North Wind and the Sun" read by 10 speakers
   - 5 females and 5 males

Corpus Statistics

Corpus	Speakers	Total Dur.	Mean Utt. Dur.	# Measurements
FDA	2	5 min 32 sec	3.32 sec	18,098
Keele	10	5 min 37 sec	33.7 sec	11,527

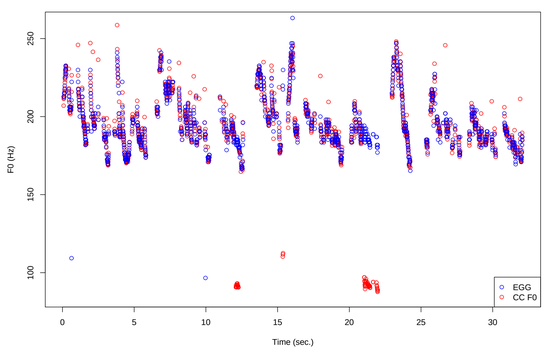
F0 Extraction Methods

Method	Full Name	Source
SWIPE'	Sawtooth Waveform Inspired Pitch Estimator	[3]
SHS	Sub-Harmonic Summation	Praat [4]
AC	Auto-Correlation	Praat [4]
CC	Cross-Correlation	Praat [4]
RAPT	Robust Algorithm for Pitch Tracking	ESPS [5]

Methodology

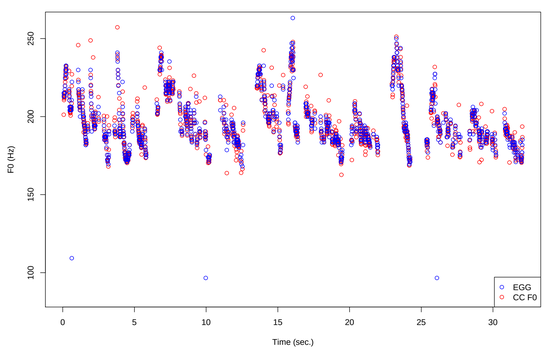
- F0 measurements were extracted from 2 corpora with Electroglottograph (EGG) measurements using 5 standard algorithms
- F0 measurements first extracted using 75 Hz for $pFloor$ and 600 Hz for $pCeiling$
- Then, the optimal pitch floor and ceiling parameters were obtained following the pre-processing procedure in [6]:
  1. Default $pFloor$ and $pCeiling$ values are used to obtain the values of the 35th and 65th quantiles
  2. $pFloor = q35 * 0.72 - 10$
  3. $pCeiling = q65 * 1.90 + 10$
- Performance evaluated using Gross Error Rate, GER, (predicted values that differ from the reference EGG value by $> 20\%$) and RMSE
- Analysis only includes frames that all algorithms predict as voiced

Before Parameter Optimization



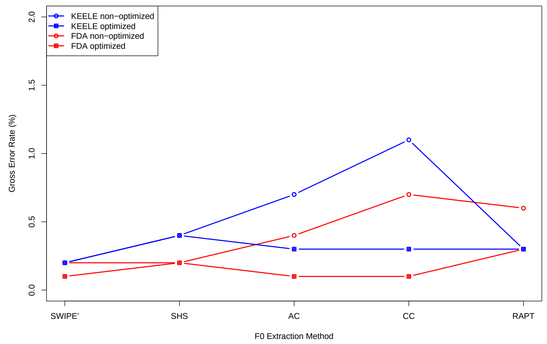
- Example using the CC method for the speaker *f1* from the Keele corpus
- Default pitch floor and ceiling values produce many gross errors

After Parameter Optimization

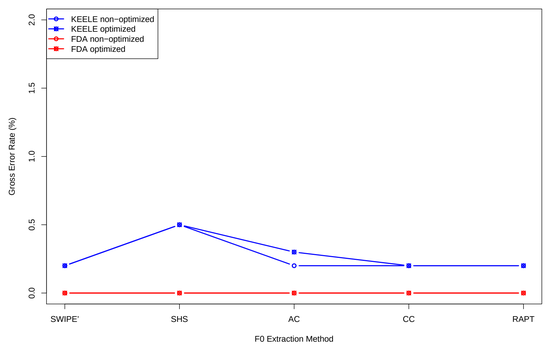


- Parameter optimization sets pitch floor to 125 Hz and pitch ceiling to 390 Hz
- Gross errors are eliminated

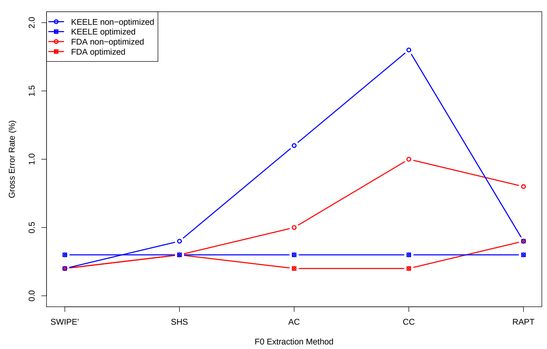
Overall GER Results



Male GER Results



Female GER Results



RMSE Results

FDA corpus:

Method	Overall		Male		Female	
	Def.	Opt.	Def.	Opt.	Def.	Opt.
SWIPE'	8	7.7	3.0	3.0	9.3	8.9
SHS	8.8	9.9	3.1	2.7	10.1	11.7
AC	10.3	7.5	2.5	2.5	12.0	8.8
CC	11.8	7.4	3.1	3.2	13.7	8.6
RAPT	11.9	11.5	3.6	3.5	13.8	13.5

Keele corpus:

Method	Overall		Male		Female	
	Def.	Opt.	Def.	Opt.	Def.	Opt.
SWIPE'	5.2	5.2	3.7	3.7	6.0	6.3
SHS	7.6	6.8	7.1	5.5	7.8	7.8
AC	8.4	5.6	3.6	4.1	10.6	6.7
CC	10.4	5.7	4.3	4.3	13.1	6.7
RAPT	7.3	6.6	4.4	4.0	8.7	8.2

Summary

- All algorithms perform better on male speech than female speech
- Optimization of $pFloor$ and $pCeiling$ parameters improves (or does not change) the overall GER for all algorithms in both corpora
- GER ranges after parameter optimization are 0.1% - 0.3% for FDA and 0.2% - 0.4% for Keele
- *All F0 extraction algorithms perform similarly when parameter optimization is applied*

References

[1] Paul Bagshaw, *Automatic prosodic analysis for computer aided pronunciation teaching*, Ph.D. thesis, University of Edinburgh, 1994.

[2] F. Plante, G.F. Meyer, and W.A. Ainsworth, "A pitch extraction reference database," in *Proc. Eurospeech*, 1995.

[3] Arturo Camacho, *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*, Ph.D. thesis, University of Florida, 2007.

[4] Paul Boersma and David Weenick, "Praat: Doing phonetics by computer, version 5.0.38," *http://www.praat.org*, 2010.

[5] David Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)," in *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal, Eds., pp. 495–518. Elsevier, 1995.

[6] Céline De Looze and Stéphane Rauzy, "Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration," in *Proc. Interspeech*, 2009.