INTERNATIONAL AUDIO LABORATORIES ERLANGEN

A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

Reliability Assessment of Singing Voice F0-Estimates Using Multiple Algorithms

[Paper #3209]

Sebastian Rosenzweig[1], Frank Scherbaum[2], Meinard Müller[1]

[1] International Audio Laboratories Erlangen, Germany [2] University of Potsdam, Germany

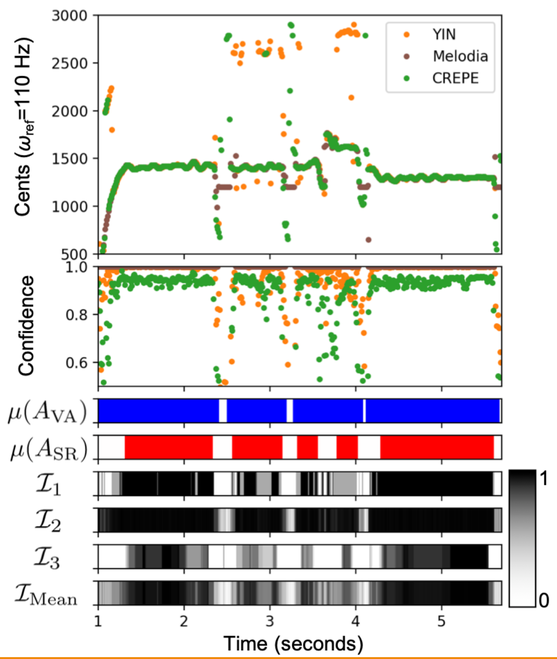
Abstract

Over the last decades, various conceptually different approaches for fundamental frequency (F0) estimation in monophonic audio recordings have been developed. The algorithms' performances vary depending on the acoustical and musical properties of the input audio signal. A common strategy to assess the reliability (correctness) of an estimated F0-trajectory is to evaluate against an annotated reference. However, such annotations may not be available for a particular audio collection and are typically labor intensive to generate. In this work, we consider an approach to automatically assess the reliability of F0-trajectories estimated from monophonic singing voice recordings. As main contribution, we propose three reliability indicators that are based on the outputs of multiple algorithms. Besides providing a mathematical description of the indicators, we analyze the indicators' behavior using a set of annotated vocal F0-trajectories. Furthermore, we show the potential of the proposed indicators for exploring unlabeled audio collections.

Reliability Indicators

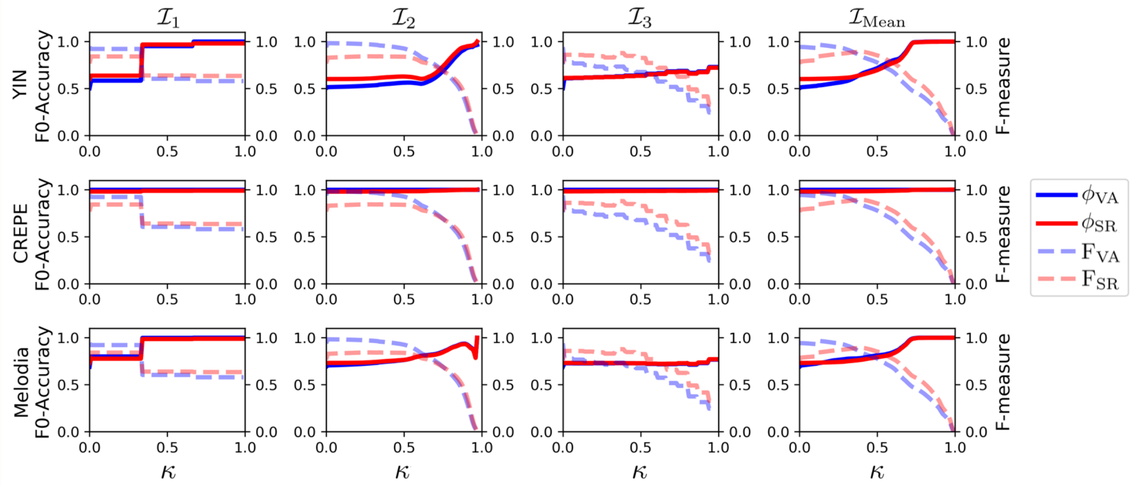
How can we measure the reliability of automatically extracted F0-trajectories?

- Based on the outputs (F0-estimates and confidence) of multiple algorithms
- In our experiments, we use
  - YIN [1]
  - Melodia [2]
  - CREPE [3]
- Two reference annotations for evaluation
  - A_{VA}: F0-annotation for parts where singing **v**oice is **a**ctive (VA)
  - A_{SR}: F0-annotation for roughly **s**table **r**egions (SR)
  - $\mu(A_{\mathrm{VA}})$, $\mu(A_{\mathrm{SR}})$: Activities of annotations
- Three reliability indicators
  - $\mathcal{I}_1$: F0-agreement
  - $\mathcal{I}_2$: Overall Confidence
  - $\mathcal{I}_3$: F0-trajectory stability (based on [4])
- $\mathcal{I}_{\mathrm{Mean}}$: Mean of all indicators (different weights possible)



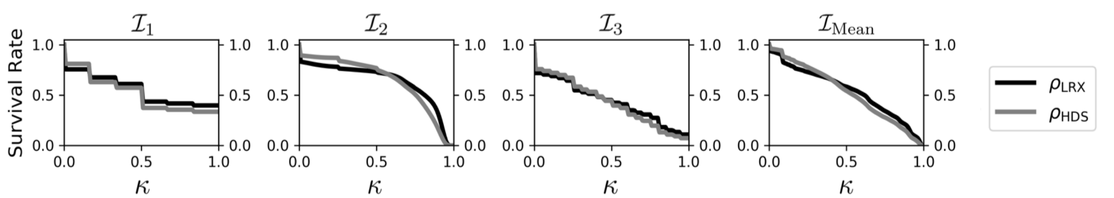
Evaluation Using Labeled Data

- Evaluation using five annotated singing voice tracks from a collection of field recordings of Georgian vocal music [5]
- Different thresholds κ applied to reliability indicators $\mathcal{I}$ (the larger κ, the smaller the number of frames to be evaluated)
- Evaluation measures:
  - F0-Accuracy (ϕ_{VA}, ϕ_{SR}), F0-values of remaining frames vs. F0-values of annotation, tolerance: 10 cents
  - Activity F-Measure (F_{VA}, F_{SR}), activity of remaining frames vs. activity of annotation



Exploring Unlabeled Audio Collections

- 85 multitrack recordings of Georgian vocal music, larynx (LRX) and headset (HDS) microphone signals [5]
- Different thresholds κ applied to reliability indicators $\mathcal{I}$
- Survival rates (ρ_{LRX}, ρ_{HDS}) indicate how many F0-values remain after thresholding



References & Acknowlegments

This work was supported by the German Research Foundation (DFG MU 2686/13-1, SCHE 280/20-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.

[1] Cheveigné et al., "YIN, a fundamental frequency estimator for speech and music.," JASA, vol. 111, no. 4, pp.1917–1930, 2002.
[2] Salamon et al., "Melody extraction from polyphonic music signals using pitch contour characteristics," TASLP, vol. 20, no. 6, pp. 1759–1770, 2012.
[3] Kim et al., "CREPE: A convolutional representation for pitch estimation," ICASSP, 2018, pp. 161–165.
[4] Rosenzweig et al., "Detecting stable regions in frequency trajectories for tonal analysis of traditional Georgian vocal music," ISMIR, 2019, pp. 352–359.
[5] Scherbaum et al., "Multi-media recordings of traditional Georgian vocal music for computational analysis," FMA, 2019, pp. 1–6.