# Efficient Computer-Aided **Pitch Track and Note Estimation** for Scientific Applications

Matthias Mauch
Chris Cannam
György Fazekas

# Problem

Intonation in Unaccompanied Singing:
Accuracy, Drift and a Model of Intonation Memory

Proc. of the National Conference on Communications (NCC), Feb 2012, Kharagpur, India.

**Assessing Vowel Quality for Singing Evaluation**

*Mayank Vibhuti Jha and Preeti Rao*

Department of Electrical Engineering,
Indian Institute of Technology Bombay,
Mumbai 400076, India
Email: {mayankjha, prao} @ee.iitb.ac.in

**Abstract**

The proper pronunciation of lyrics is an important component of vocal music. While automatic vowel classification has been widely studied for speech, a separate investigation of the methods is needed for singing due to the differences in acoustic properties between sung and spoken vowels. Acoustic features combining spectrum envelope and pitch are used with classifiers trained on sung vowels for classification of test vowels segmented from the audio of solo singing. Two different classifiers are tested, viz., Gaussian Mixture Models (GMM) and Linear Regression, and observed to perform well on both male and female sung vowels.

**Keywords:** MFCC; GMM; Linear Regression; Vowel Quality; Singing Voice; Vowel Classification

## 1. Introduction

Singing or vocal music, like instrumental performances, is characterised by musical attributes such as melody and rhythm. However in the case of singing, also important are voice quality and the proper articulation of the lyrics. The automatic assessment of singing ability would therefore require processing the audio signal for the underlying acoustic attributes of pitch (related to melody), onsets (related to rhythm), phoneme quality (related to pronunciation) and timbre (related to voice quality). Such a system for singing assessment and feedback could be very useful both for music education and entertainment. Available systems for singing scoring, including popular karaoke games like SingStar [1] and UltraStar [2], are currently restricted to measuring pitch and timing accuracy with respect to a reference, i.e., only melodic and rhythmic aspects are considered. Our present work builds further on the same essential framework by incorporating new methods for the assessment of phoneme quality in singing.

The scenario under consideration has the singer rendering a known song while listening to the song's karaoke (i.e. background music) track. The acoustic characteristics of uttered phones are then evaluated with respect to the expected phones as provided by the song's lyrics. Our aim is to confirm whether the singer has rendered the lyrics accurately. Our aim is to develop a generalized system which should be text-independent. Once trained on sufficient number of vowel samples, it should be usable for testing vowels on any new song, provided the lyrics are known.

The current task is clearly related to Automatic Speech Recognition (ASR). However singing differs from speech in some important ways as presented in the next section. These differences warrant a separate study on features and classification methods for sung phones. In this paper we focus on sung vowel identification using a standard spectral representation and two different methods of classification. While GMM classifiers are widely applied in speech recognition, we also investigate a linear regression approach to classification that has certain advantages in the singing context [3].

## 2. Singing versus Speech

Singing, compared to speech, has a wider dynamic range in pitch as well as intensity due to the relative importance of expressiveness in singing. Singing tends to be a one-to-many communication at longer distances and hence the need to maintain a loudness balance across sounds [4]. Singing tends to have a higher percentage of sonorants than obstruents so that a singing piece will be largely composed of vowels. In fact, in singing, phonation time can be up to 95%, compared to 60% in normal speech [5]. Hence, restricting phoneme quality assessment to vowels is a reasonable starting point for pronunciation evaluation in singing. Due to the occurrence of high-pitched vowels in singing, it is possible that pitch harmonics do not coincide with the canonical formant locations in some cases. This usually causes singers to modify vowel quality in the interest of maintaining loudness. This dependence of vowel quality on pitch is another distinguishing factor between speech and singing.

## 3. Database

The data sets used in these experiments were chosen from a database of songs sung by various people in sing-along mode at the venue of a technical exhibition. As these songs were recorded in a public place (with moderate noise levels, SNR of the order of 20-30 dB), the database is representative of real-world scenarios. These songs (of about 1 min duration each) were recorded using a directional microphone, sampled at 8 kHz and stored in 16-bit PCM, mono channel, wave format.

Five popular Hindi movie songs each of male and female playback singers were selected for building the database.

- "For use in training, all the vowel tokens in the singer audios were manually labelled in PRAAT [7]…"
(Jha and Rao *Assessing Vowel Quality for Singing Evaluation, 2012*)

- "…onsets and offsets were adjusted manually, and the resulting annotations were fed into customised pitch tracking software…"
(Mauch et al. *Intonation in Unaccompanied Singing…, under review, 2014*)

# Problem

- Ever more research on melody, singing, intonation.
- Still **very cumbersome** to annotate pitch.
  (We have learned the hard way!)
    - using Praat (made for speech)
    - using makeshift, complicated processing chains
- There are **no tools** that allow efficient pitch/note annotation.

# Requirements

# Requirements

| | Melodyne | Praat | Sonic Visualiser |
|---|---|---|---|
| estimate pitch | ✔ | ✔ | ✔ |
| estimate notes | ✔ | ~✔ | ✔ |
| note/pitch correction | ✔ | ✘ | ✘ |
| note/pitch sonification | ✘ | ✘ | ✘ |
| save note/ pitch track | ~✔ | ✘ | ✔ |
| load note/ pitch track | ✘ | ✘ | ✔ |

# Requirements

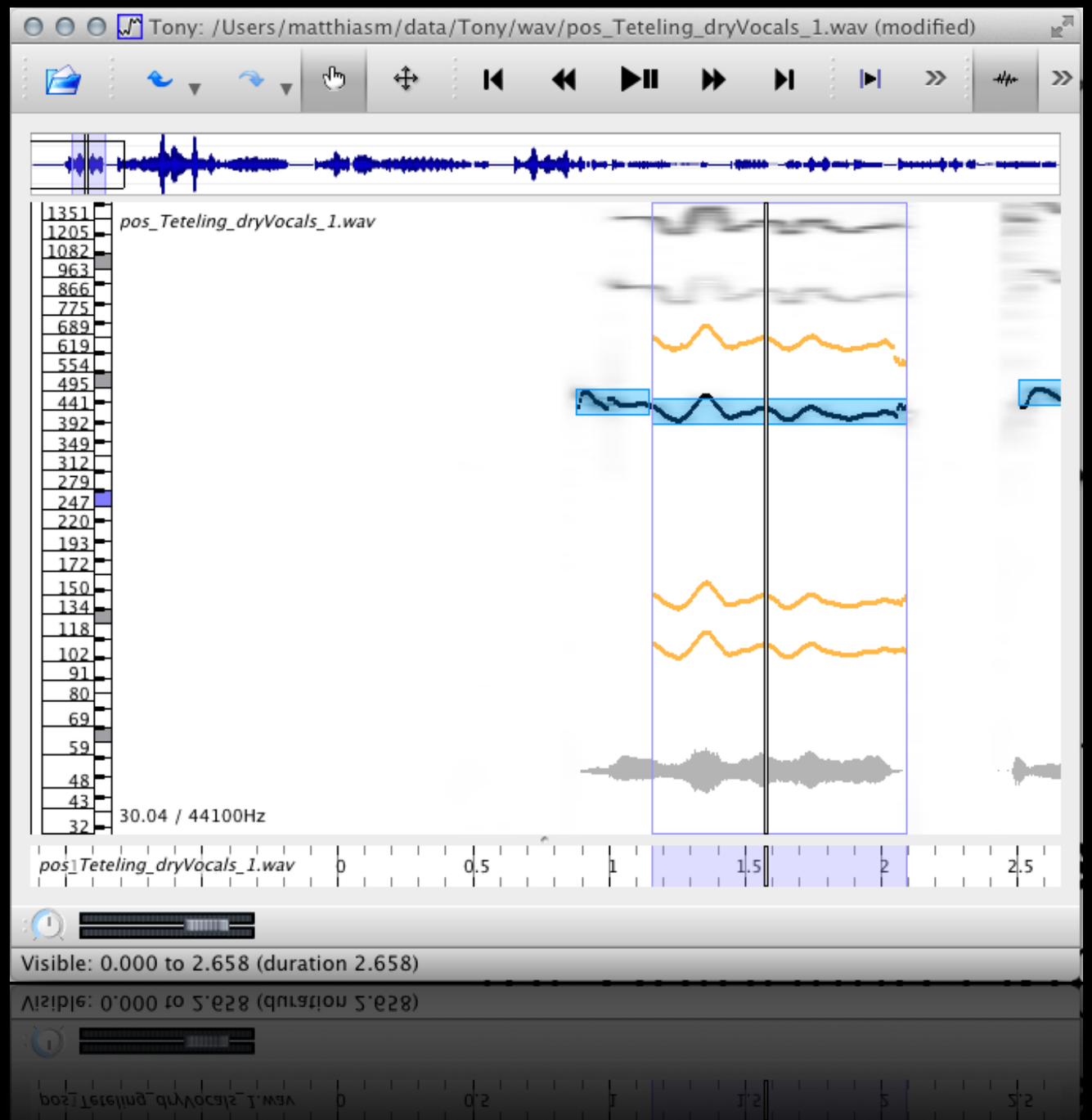
| | Melodyne | Praat | Sonic Visualiser | ? |
|---|---|---|---|---|
| estimate pitch | ✔ | ✔ | ✔ | ✔ |
| estimate notes | ✔ | ~✔ | ✔ | ✔ |
| note/pitch correction | ✔ | ✘ | ✘ | ✔ |
| note/pitch sonification | ✘ | ✘ | ✘ | ✔ |
| save note/ pitch track | ~✔ | ✘ | ✔ | ✔ |
| load note/ pitch track | ✘ | ✘ | ✔ | ✔ |

# Aim

Build a tool that aids researchers investigating melodic data to annotate their recordings!

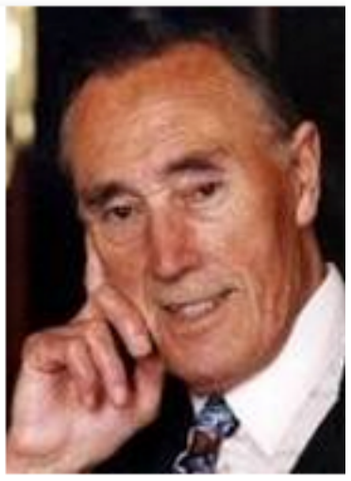
- Automatic pitch and note transcription.
- Sonification of pitch and notes for immediate feedback.
- Fast, efficient correction of auto-transcription errors.
- Versatile import and export for scientific applications.
- Open source for reproducibility.

# Tony


Tony: /Users/matthiasm/data/Tony/wav/pos_Teteling_dryVocals_1.wav (modified)
pos_Teteling_dryVocals_1.wav
1351
1205
1082
963
866
775
689
619
554
495
441
392
349
312
279
247
220
193
172
150
134
118
102
91
80
69
59
48
43
32
30.04 / 44100Hz
pos_Teteling_dryVocals_1.wav
0
0.5
1
1.5
2
2.5
Visible: 0.000 to 2.658 (duration 2.658)

# Tony Melody

Actor



Anthony John "Tony" Melody was an English television actor who appeared in a number of long running comedies and soap operas. He was a prolific character actor with more than 100 television roles. Wikipedia

**Born:** December 18, 1922, London

**Died:** June 26, 2008, Bispham, Blackpool

**Movies and TV shows:** Yanks, Walter, Home and Away (UK), The Nesbitts Are Coming

# Building blocks

- **Pitch Tracking:**
  PYIN — version of widely-used YIN algorithm

  - pitch track smoothing + voiced/unvoiced

  - note track estimation based on pitch track

- **User Interface:**
  Sonic Visualiser libraries

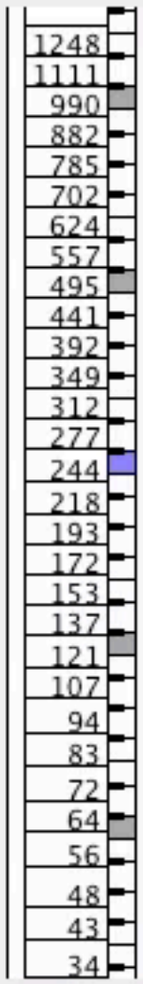
  - simplified interface

  - extended with all the cool stuff we need to doTo

# Basic *Tony* Example

Tony File Edit View Playback Help 99% Wed 2 Apr 15:08

Tony: /Users/matthiasm/data/Tony/wav/oneSampleSilence.wav




*oneSampleSilence.wav*

No local points

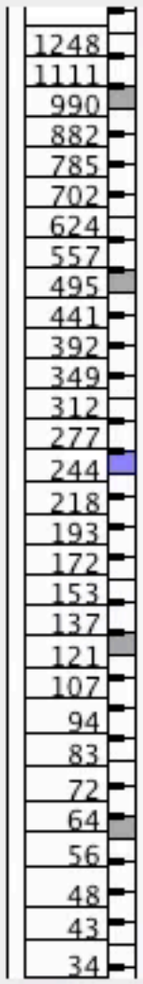
0 / 44100Hz (R)

*oneSampleSilence.wav*

Visible: 0.000 to 0.000 (duration 0.000)

Tony File Edit View Playback Help 99% Wed 2 Apr 15:08

Tony: /Users/matthiasm/data/Tony/wav/oneSampleSilence.wav




*oneSampleSilence.wav*

No local points

0 / 44100Hz (R)

*oneSampleSilence.wav*

Visible: 0.000 to 0.000 (duration 0.000)

# Correcting Notes

# Note correction

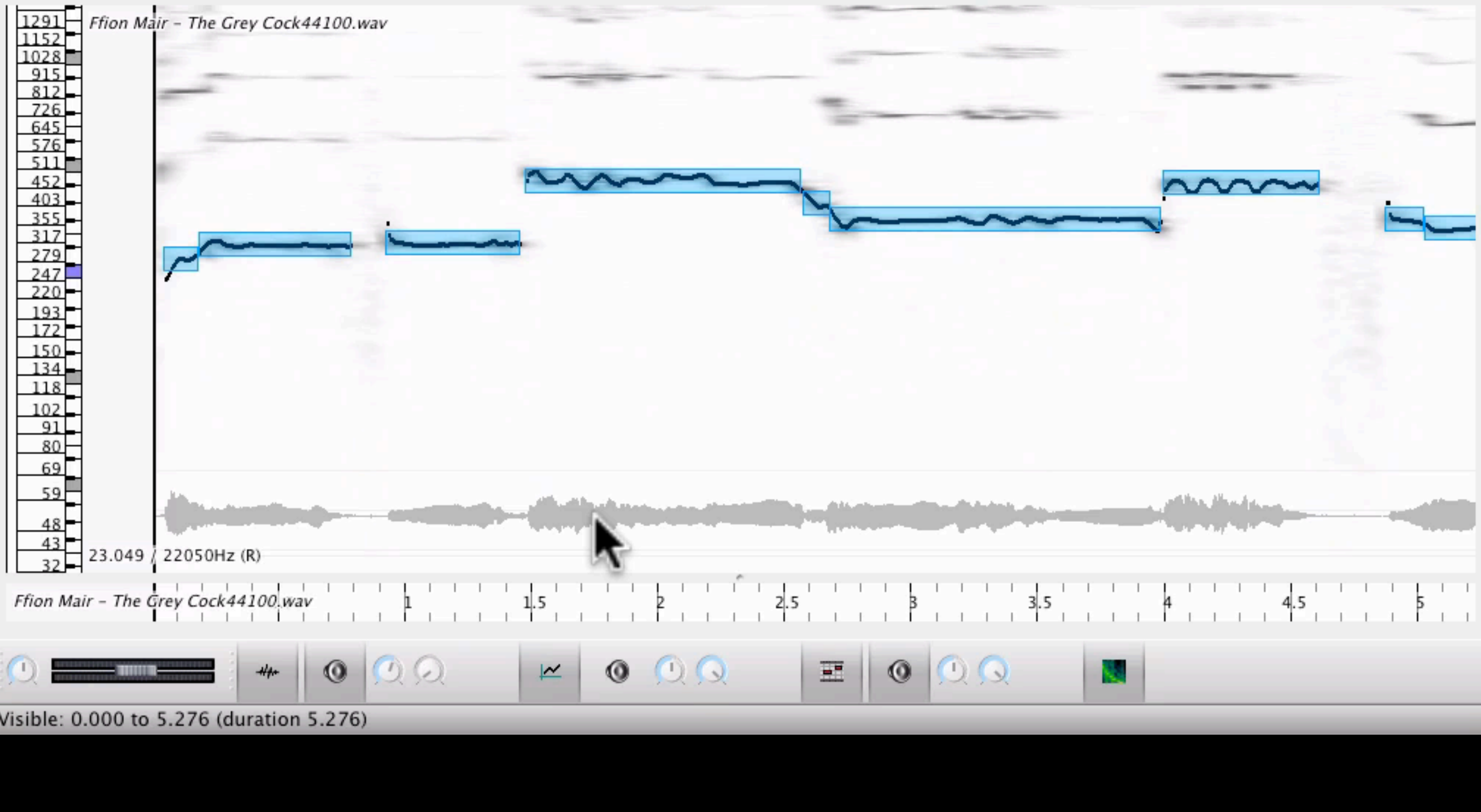
- split notes
- merge notes
- shorten/lengthen notes
- change note pitch
- delete notes

*Example:*

All sorts of note correction

Tony File Edit View Playback Help 100% Wed 2 Apr 18:49

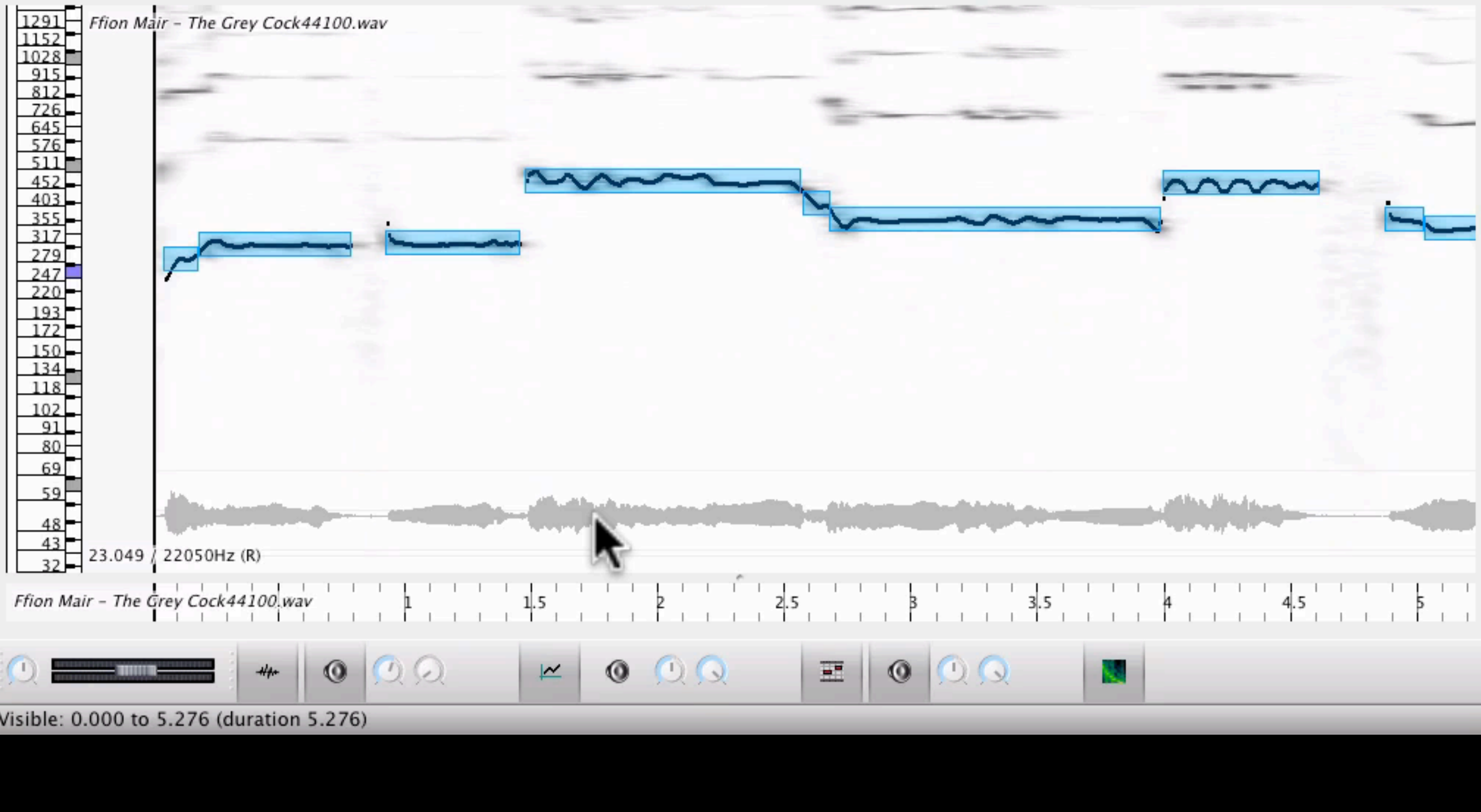
Tony: /Users/matthiasm/data/Tony/wav/Ffion Mair - The Grey Cock44100.wav (modified)




Visible: 0.000 to 5.276 (duration 5.276)

Tony File Edit View Playback Help 100% Wed 2 Apr 18:49

Tony: /Users/matthiasm/data/Tony/wav/Ffion Mair - The Grey Cock44100.wav (modified)




Visible: 0.000 to 5.276 (duration 5.276)

*Example:*

# Note Splitting and Save

Tony File Edit View Playback Help 100% Wed 2 Apr 16:07

Open... ⌘O
Open Location... ⇧⌘O
Open Recent ▶
Save Session ⌘S
Save Session As... ⇧⌘S
Import Pitch Track Data...
Export Pitch Track Data...
Export Note Data...

/Users/matthiasm/data/Tony/wav/oneSampleSilence.wav ⌘R
/Users/matthiasm/data/intonation/contr...3-14-55-01-semitone-down-click.mp3
/Users/matthiasm/data/intonation/contr...013-08-55-01-semitone-up-click.mp3
/Users/matthiasm/data/intonation/controlled_study/pretest_daniel_o/test-1.mp3
/Users/matthiasm/data/intonation/contr...udy/pretest_lucas_thompson/test-2.mp3
/Users/matthiasm/data/intonation/contr...udy/pretest_lucas_thompson/test-1.mp3
/Users/matthiasm/data/intonation/contr...dy/pretest_matthias_mauch/test-1.mp3
/Users/matthiasm/data/intonation/contr...dy/pretest_matthias_mauch/test-2.mp3
/Users/matthiasm/data/Tony/wav/Ffion Mair - The Grey Cock44100.wav
/Users/matthiasm/data/Tony/wav/AhHuuu.wav
/Users/matthiasm/data/Tony/wav/Flute/...nJazz_Premix_04_normalisedExcerpt.wav
file:///Users/matthiasm/data/Tony/wav/Flute/MusicDelta_LatinJazz_Premix_04.wav
file:///Users/matthiasm/data/intonation...n/fnwav/happy_birthday_gs_normal.wav
file:///Users/matthiasm/data/intonation...wav/happy_birthday_dt72_imagined.wav
file:///Users/matthiasm/data/intonation...n/fnwav/happy_birthday_cs_normal.wav
file:///Users/matthiasm/data/intonation.../fnwav/happy_birthday_clm_normal.wav
file:///Users/matthiasm/data/intonation.../fnwav/happy_birthday_phj_normal.wav
file:///Users/matthiasm/data/intonation.../fnwav/happy_birthday_phj_masked.wav
file:///Users/matthiasm/data/intonation...nwav/happy_birthday_phj_imagined.wav
file:///Users/matthiasm/data/intonation...n/fnwav/happy_birthday_kw_normal.wav

ony: /Users/matthiasm/data/Tony/wav/oneSampleSilence.wav

1248 1111 990 882 785 702 624 557 495 441 392 349 312 277 244 218 193 172 153 137 121 107 94 83 72 64 56 48 43 34

0 / 44100Hz (R)

*oneSampleSilence.wav*

Visible: 0.000 to 0.000 (duration 0.000)

Tony File Edit View Playback Help 100% Wed 2 Apr 16:07

Open... ⌘O
Open Location... ⇧⌘O
Open Recent ▶
Save Session ⌘S
Save Session As... ⇧⌘S
Import Pitch Track Data...
Export Pitch Track Data...
Export Note Data...

/Users/matthiasm/data/Tony/wav/oneSampleSilence.wav ⌘R
/Users/matthiasm/data/intonation/contr...3-14-55-01-semitone-down-click.mp3
/Users/matthiasm/data/intonation/contr...013-08-55-01-semitone-up-click.mp3
/Users/matthiasm/data/intonation/controlled_study/pretest_daniel_o/test-1.mp3
/Users/matthiasm/data/intonation/contr...udy/pretest_lucas_thompson/test-2.mp3
/Users/matthiasm/data/intonation/contr...udy/pretest_lucas_thompson/test-1.mp3
/Users/matthiasm/data/intonation/contr...dy/pretest_matthias_mauch/test-1.mp3
/Users/matthiasm/data/intonation/contr...dy/pretest_matthias_mauch/test-2.mp3
/Users/matthiasm/data/Tony/wav/Ffion Mair - The Grey Cock44100.wav
/Users/matthiasm/data/Tony/wav/AhHuuu.wav
/Users/matthiasm/data/Tony/wav/Flute/...nJazz_Premix_04_normalisedExcerpt.wav
file:///Users/matthiasm/data/Tony/wav/Flute/MusicDelta_LatinJazz_Premix_04.wav
file:///Users/matthiasm/data/intonation...n/fnwav/happy_birthday_gs_normal.wav
file:///Users/matthiasm/data/intonation...wav/happy_birthday_dt72_imagined.wav
file:///Users/matthiasm/data/intonation...n/fnwav/happy_birthday_cs_normal.wav
file:///Users/matthiasm/data/intonation.../fnwav/happy_birthday_clm_normal.wav
file:///Users/matthiasm/data/intonation.../fnwav/happy_birthday_phj_normal.wav
file:///Users/matthiasm/data/intonation.../fnwav/happy_birthday_phj_masked.wav
file:///Users/matthiasm/data/intonation...nwav/happy_birthday_phj_imagined.wav
file:///Users/matthiasm/data/intonation...n/fnwav/happy_birthday_kw_normal.wav

ony: /Users/matthiasm/data/Tony/wav/oneSampleSilence.wav

1248 1111 990 882 785 702 624 557 495 441 392 349 312 277 244 218 193 172 153 137 121 107 94 83 72 64 56 48 43 34

0 / 44100Hz (R)

*oneSampleSilence.wav*

Visible: 0.000 to 0.000 (duration 0.000)

*Tony* is already in use

# Two Applications

- **my own research into intonation**
  - ~900 files by two student annotators
  - target: notes
- **large scale project by the Music Technology lab at NYU**
  - music students annotate pitch tracks
  - ~ 10 minutes per 1 minute singing
  - just started 16 tracks (23 minutes)

# Correcting the pitch track

# Pitch track correction

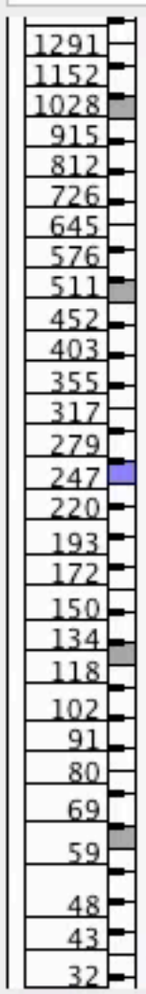
- remove pitches
- alternative pitch candidates
- notes automatically adjust to pitch track

*Example:*

# Pitch Delete/Correct and Save

Tony File Edit View Playback Help 100% Wed 2 Apr 18:25

Tony: /Users/matthiasm/data/Tony/wav/oneSampleSilence.wav




*oneSampleSilence.wav*

No local points

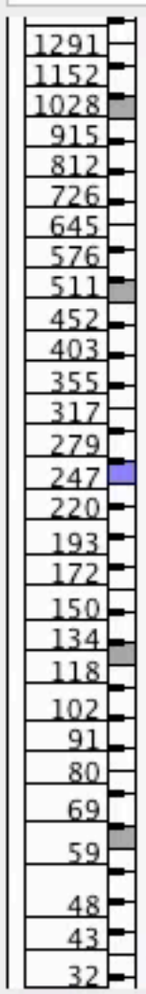
0 / 44100Hz

*oneSampleSilence.wav*

Visible: 0.000 to 0.000 (duration 0.000)

Tony File Edit View Playback Help 100% Wed 2 Apr 18:25

Tony: /Users/matthiasm/data/Tony/wav/oneSampleSilence.wav




*oneSampleSilence.wav*

No local points

0 / 44100Hz

*oneSampleSilence.wav*

Visible: 0.000 to 0.000 (duration 0.000)

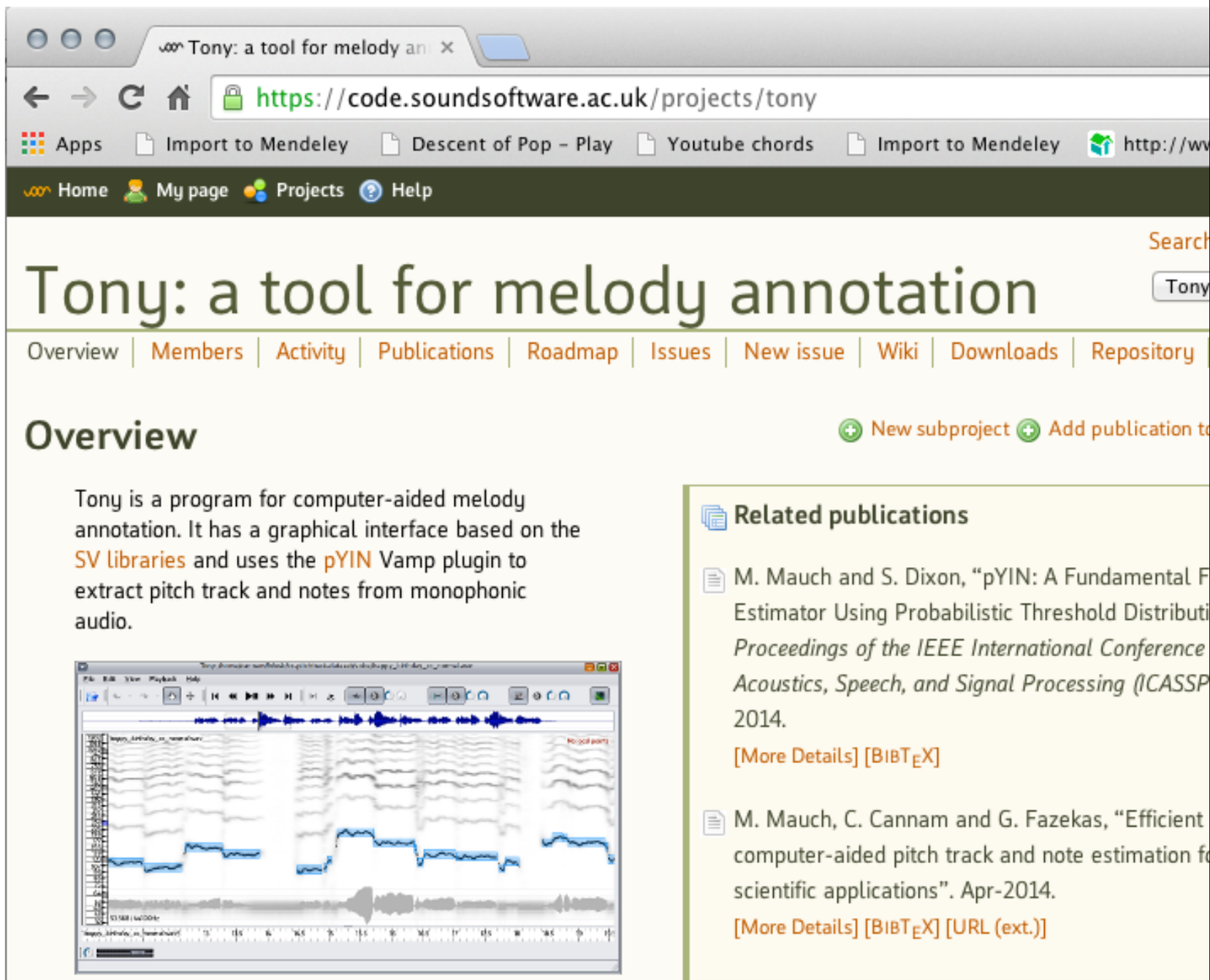
*Tony* is available to all

# Free, Open Source

***http://code.soundsoftware.ac.uk/projects/tony***

- *Tony* is available at SoundSoftware

- Mac
- Windows
- Linux

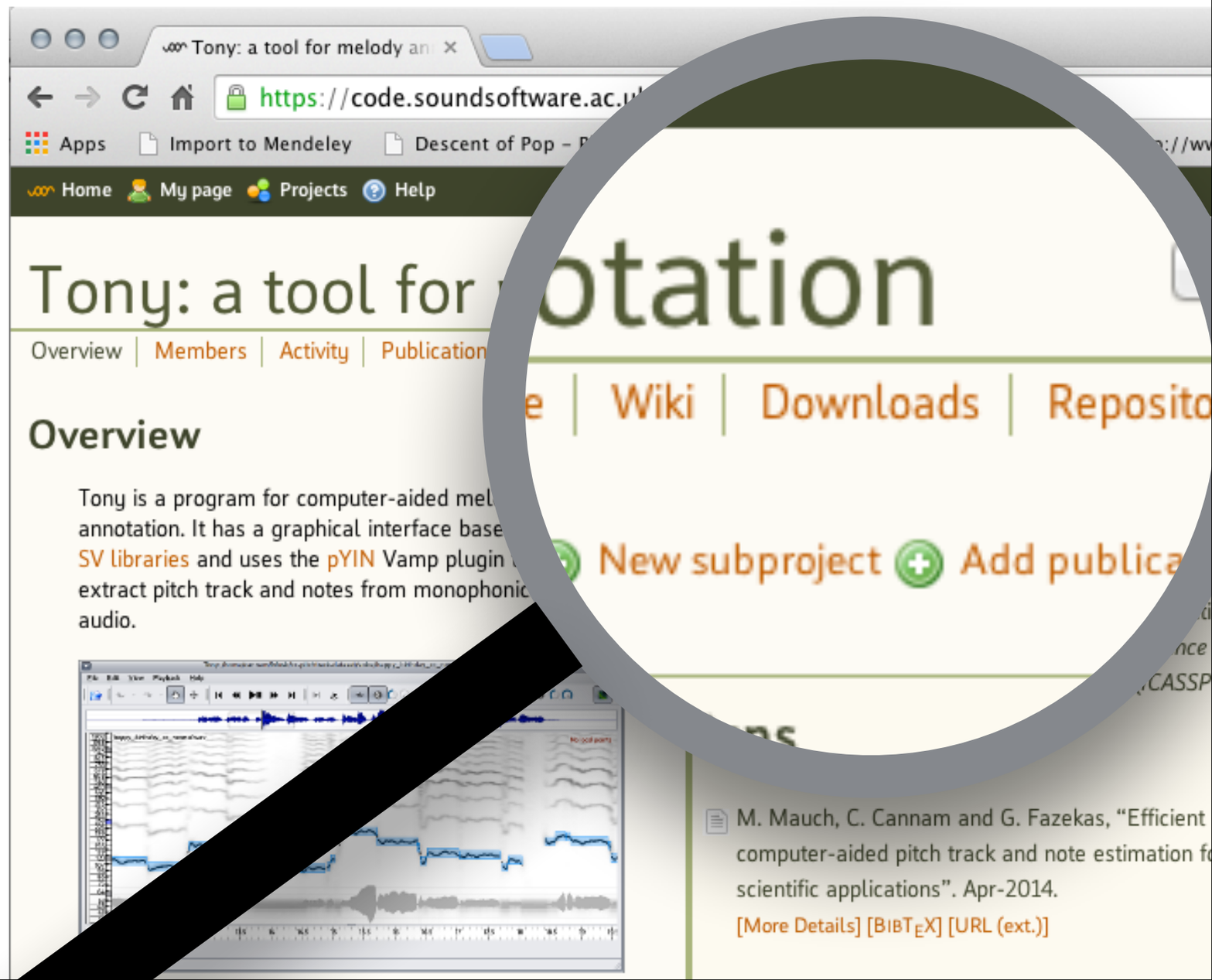


# Free, Open Source

***http://code.soundsoftware.ac.uk/projects/tony***

- *Tony* is available at SoundSoftware

  - Mac
  - Windows
  - Linux



# Conclusions & Outlook

# *Tony*

- Tool for melody annotation for scientific use
- Robust automatic extraction
- Sonification
- Correction
- Export
- Save and continue working another time

# Future work

- Use *Tony* for research on singing intonation
- improve *Tony* interaction using users' feedback
- extend capabilities (**pitch is not everything**)
  - timbre
  - expression
  - predominant frequency estimation

# Thank you.

contact me:
m.mauch@qmul.ac.uk
matthiasmauch.net

contact *Tony*:
http://code.soundsoftware.ac.uk/projects/tony