MELODY EXTRACTION FROM POLYPHONIC AUDIO OF WESTERN OPERA: A METHOD BASED ON DETECTION OF THE SINGER'S FORMANT

Zheng Tang[1] and **Dawn A. A. Black**

University of Washington,
Department of Electrical Engineering
zhtang@uw.edu

Queen Mary University of London, Electronic
Engineering and Computer Science
dawn.black@qmul.ac.uk

Introduction

- Singing voice extraction algorithms are known to perform poorly on multi-track recordings of Western opera.
- The singer's formant is defined [1] as a prominent spectral-envelope peak around 3 kHz and its presence is well documented in the singing of professional male Western opera singers and some female singers.
- This project develops a novel singing voice extraction algorithm based on this feature.

Background

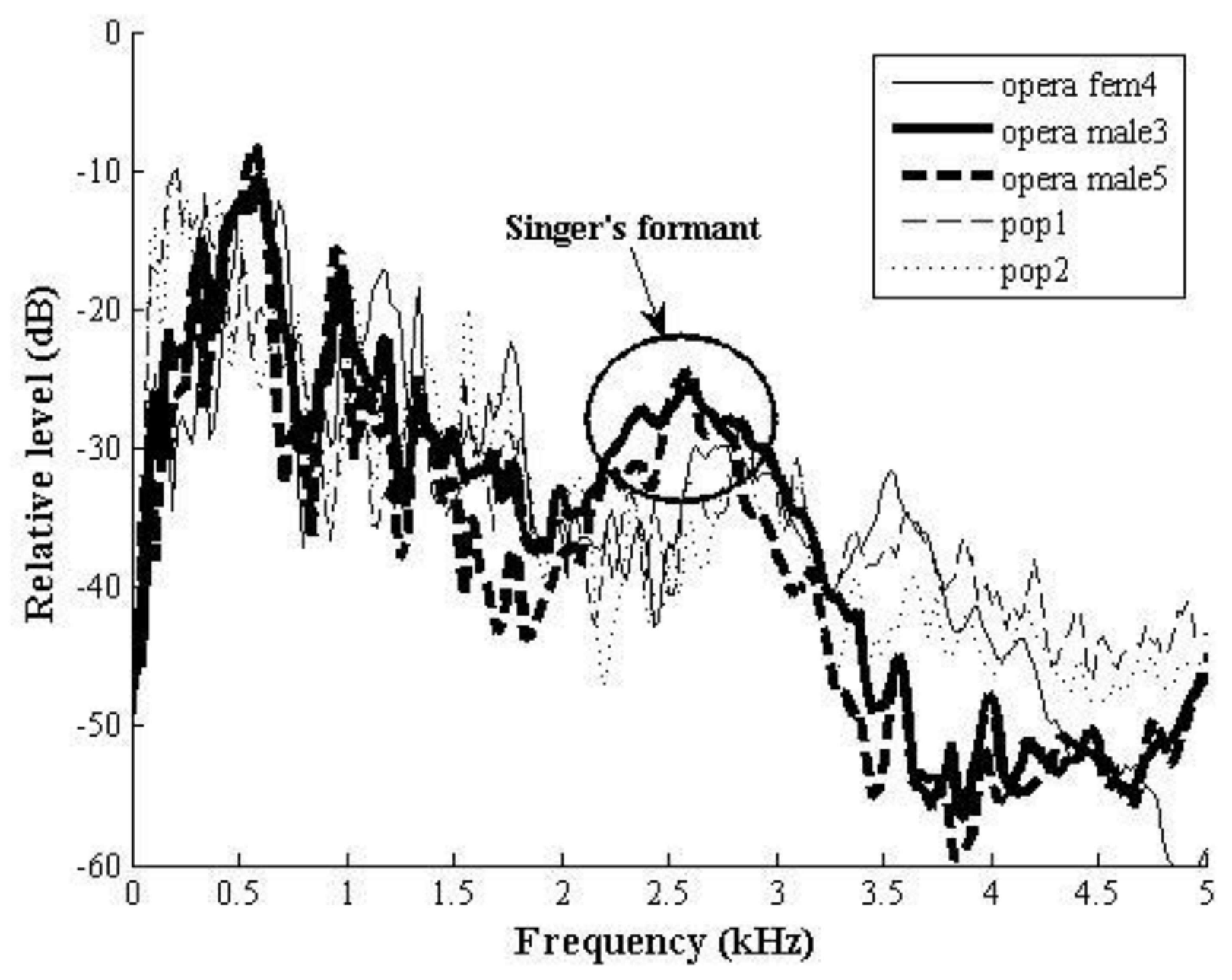

Fig. 1: Normalized LTAS for 5 audio excerpts from the ADC2004 test collection [2]

The singer's formant is known to exhibit the following properties:

1. A spectral peak which has an amplitude greater than 20 dB less than the overall sound pressure level.
2. The peak is located between 2.5 and 3.2 kHz.
3. The peak has a bandwidth of around 500- 800 Hz.

General Workflow

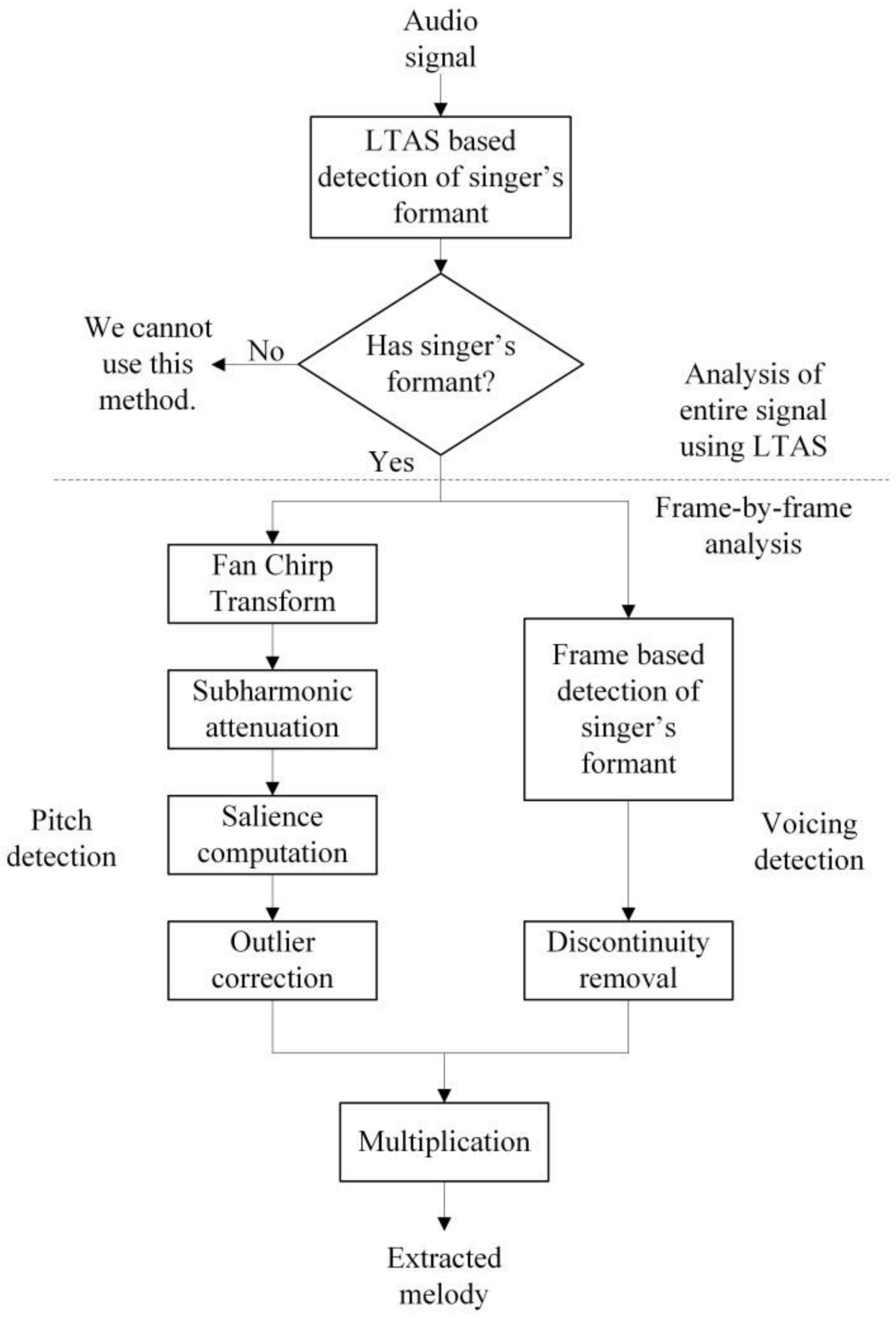

Fig. 2: System overview

Singer's Formant Detection and Voicing Detection

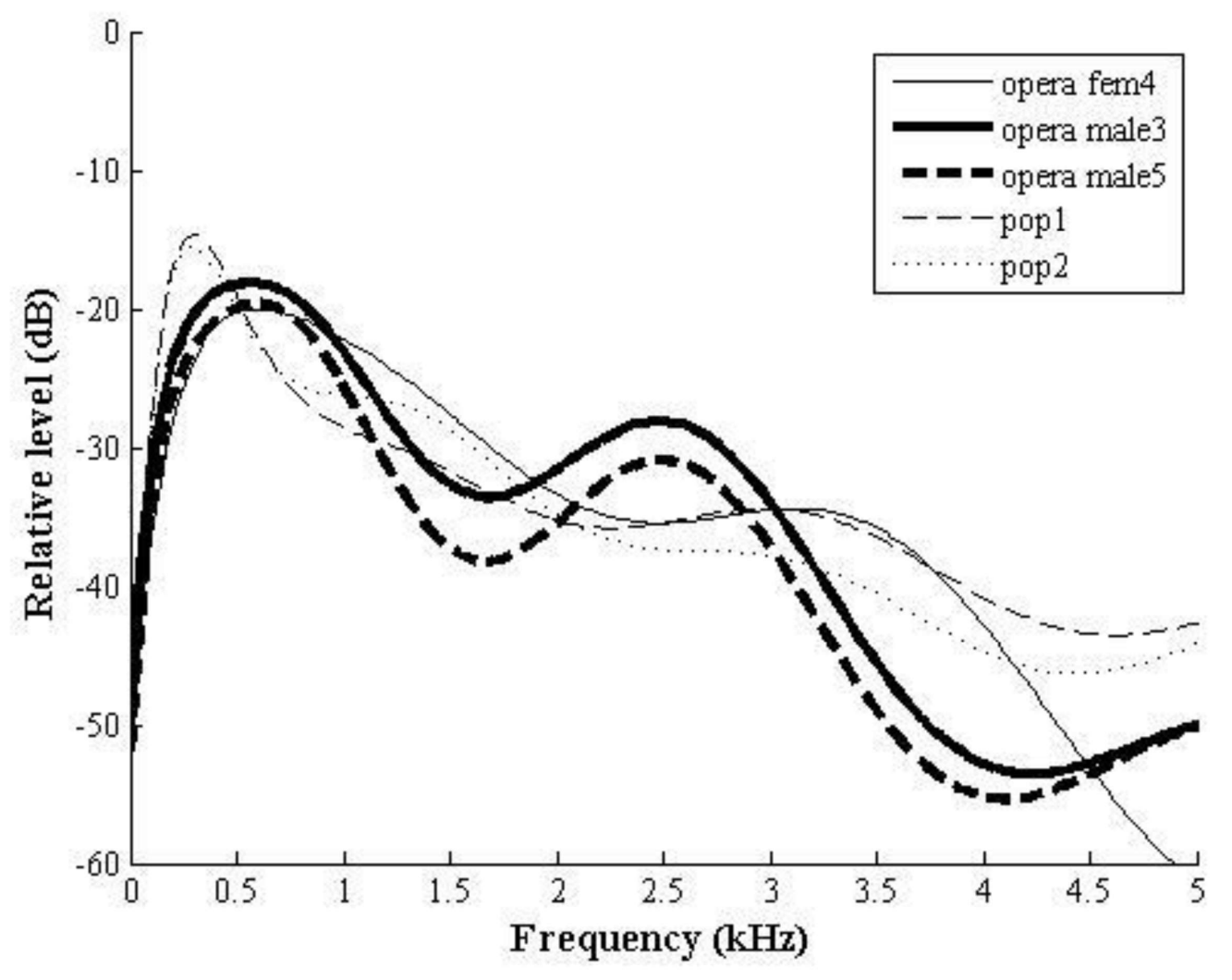

Fig. 3: The fitting polynomials of smoothed LTAS for 5 audio excerpts from ADC2004 [2]

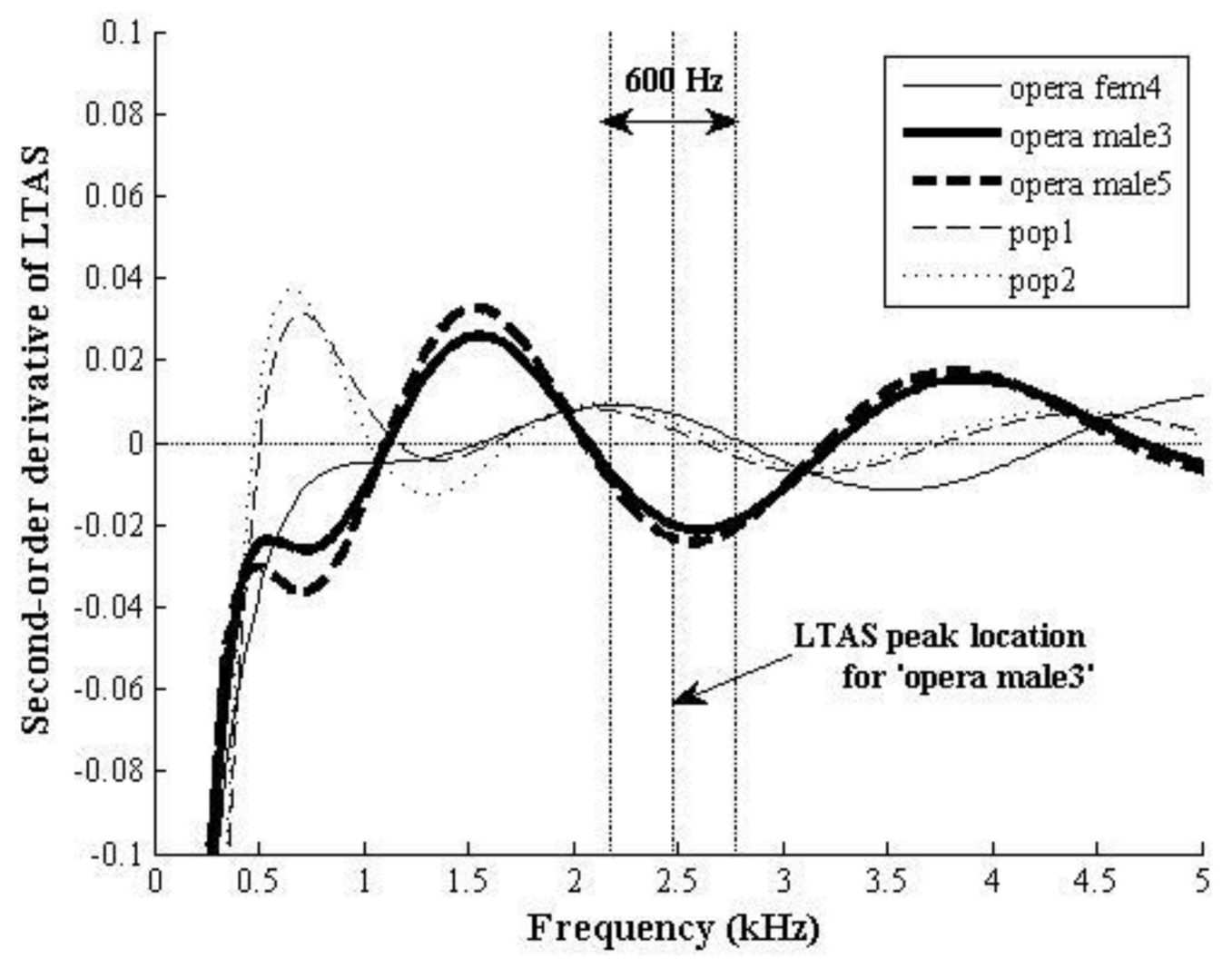

Fig. 4: The second-order derivatives of LTAS for 5 audio excerpts from ADC2004 [2]

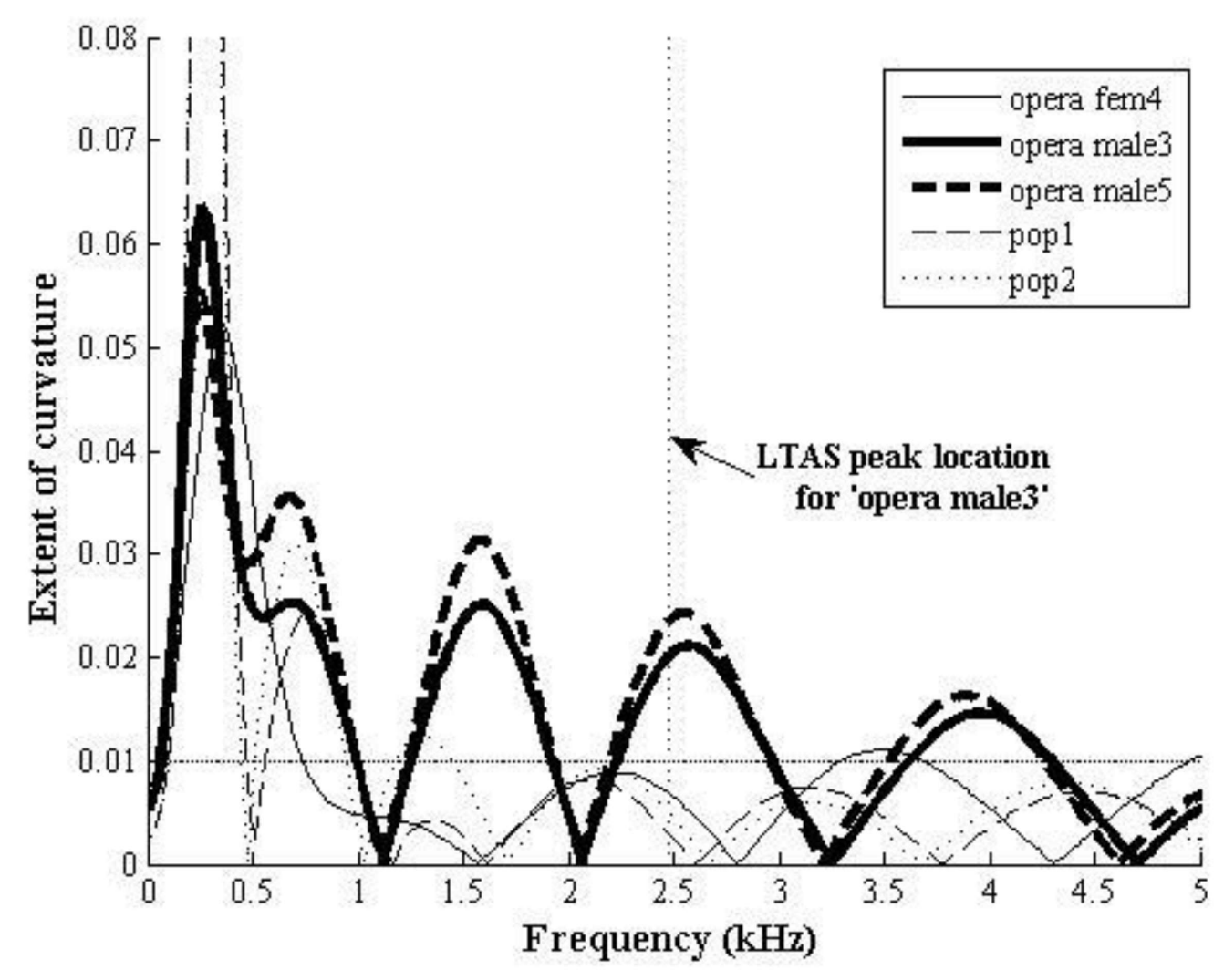

Fig. 5: The curvatures of LTAS for 5 audio excerpts from ADC2004 [2]

Based on the characteristics of the singer's formant we introduce a novel algorithm to automatically detect the presence of a singer's formant and hence the presence of a classically trained singer.

Use of the same criteria to analyse the spectrum of a single audio frame can indicate whether the frame is voiced (contains singing) or not.

Pitch Detection

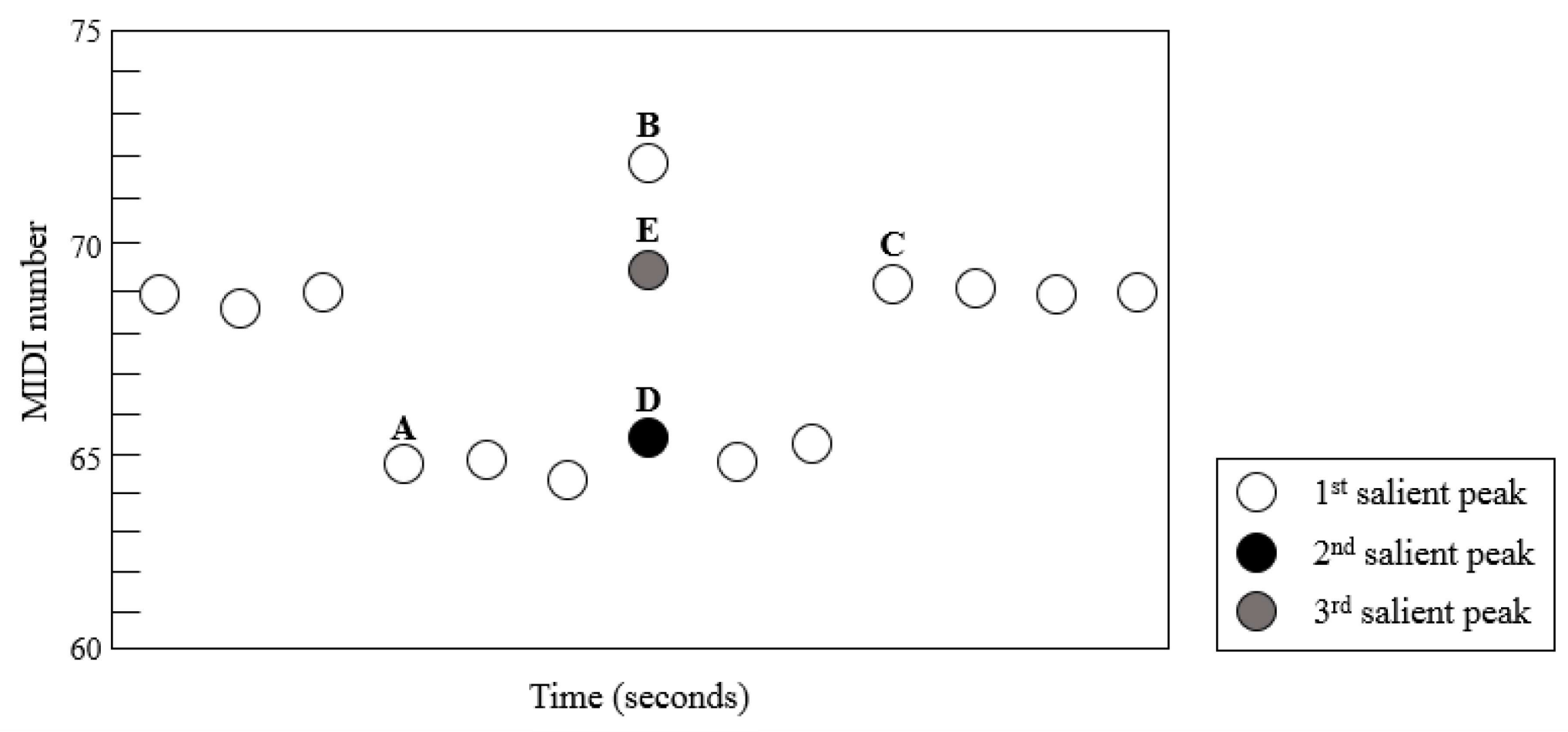

Fig. 6: An example of outlier correction

We adopt Cancela's method [3] to perform FChT since it exhibits optimal time-frequency resolution to solve problems common to vibrato.

In the outlier correction stage, we compute two additional peaks per frame as candidate substitutes for the wrong pitch.

Results and Discussion

Test set	Singing type	No. of songs	Expectation/ detection of singer's formant
ADC2004	Tenor, Western	2	Yes/Yes
	Soprano, Western	2	No/No
	Popular music	4	No/No
The dataset recorded at the Central Academy of Drama	Tenor, Western	16	Yes/Yes
	Soprano, Western	2	No/Yes
	Amateur, Western	2	No/Yes
	Laosheng, Peking	2	No/No
	Qingyi, Peking	2	No/No

Table 1: Test dataset for the evaluations of melody extraction and singer's formant detection.

First author/ completion year	Voicing detection	Voicing false alarm	Raw pitch accuracy	Raw chroma accuracy	Overall accuracy
Vincent (Bayes)/ 2005	N/A	N/A	64.8%	68.6%	N/A
Vincent (YIN)/ 2005	N/A	N/A	69.5%	72.2%	N/A
Sutton/ 2006	89.3%	51.9%	87.0%	87.6%	76.9%
Cancela/ 2008	72.6%	39.3%	83.9%	84.8%	62.4%
Salamon/ 2011	62.3%	21.8%	25.4%	30.1%	31.3%
Tang/ 2014	91.6%	5.3%	84.3%	85.1%	82.3%

Table 2: Results of the audio melody extraction evaluation

Melody extraction evaluation on our dataset confirms that our algorithm provides a clear improvement in voicing detection. Furthermore, its overall accuracy is comparable to state-of-the-art methods when dealing with Western opera signals.

References

[1] E. Gómez, S. Streich, B. Ong, R. P. Paiva, S. Tappert, J. M. Batke, G. Poliner, D. Ellis, and J. P. Bello: "A quantitative comparison of different approaches for melody extraction from polyphonic audio recordings," Univ. Pompeu Fabra, Barcelona, Spain, 2006, Tech. Rep. MTG-TR-2006-01.

[2] J. Sundberg: "Articulatory interpretation of the 'singing formant'," The Journal of the Acoustical Society of America, Vol. 55, No. 4, pp. 838-844, 1974.

[3] P. Cancela, E. López, and M. Rocamora: "Fan chirp transform for music representation," Proceedings of the 13th Int Conference on Digital Audio Effects DAFx10 Graz Austria, 2010.

[1] Zheng Tang is a recipient of the ISMIR2014 Travel Award.