Multi-pitch detection and voice assignment for *a cappella* recordings of multiple singers

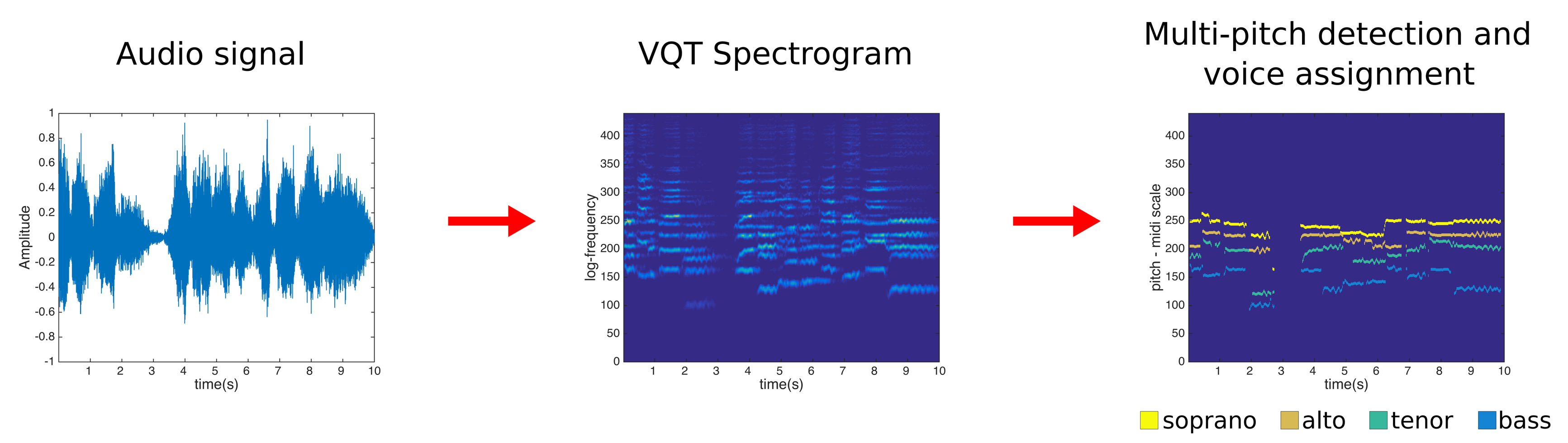
Rodrigo Schramm[1], Andrew McLeod[2], Mark Steedman[2], Emmanouil Benetos[3]

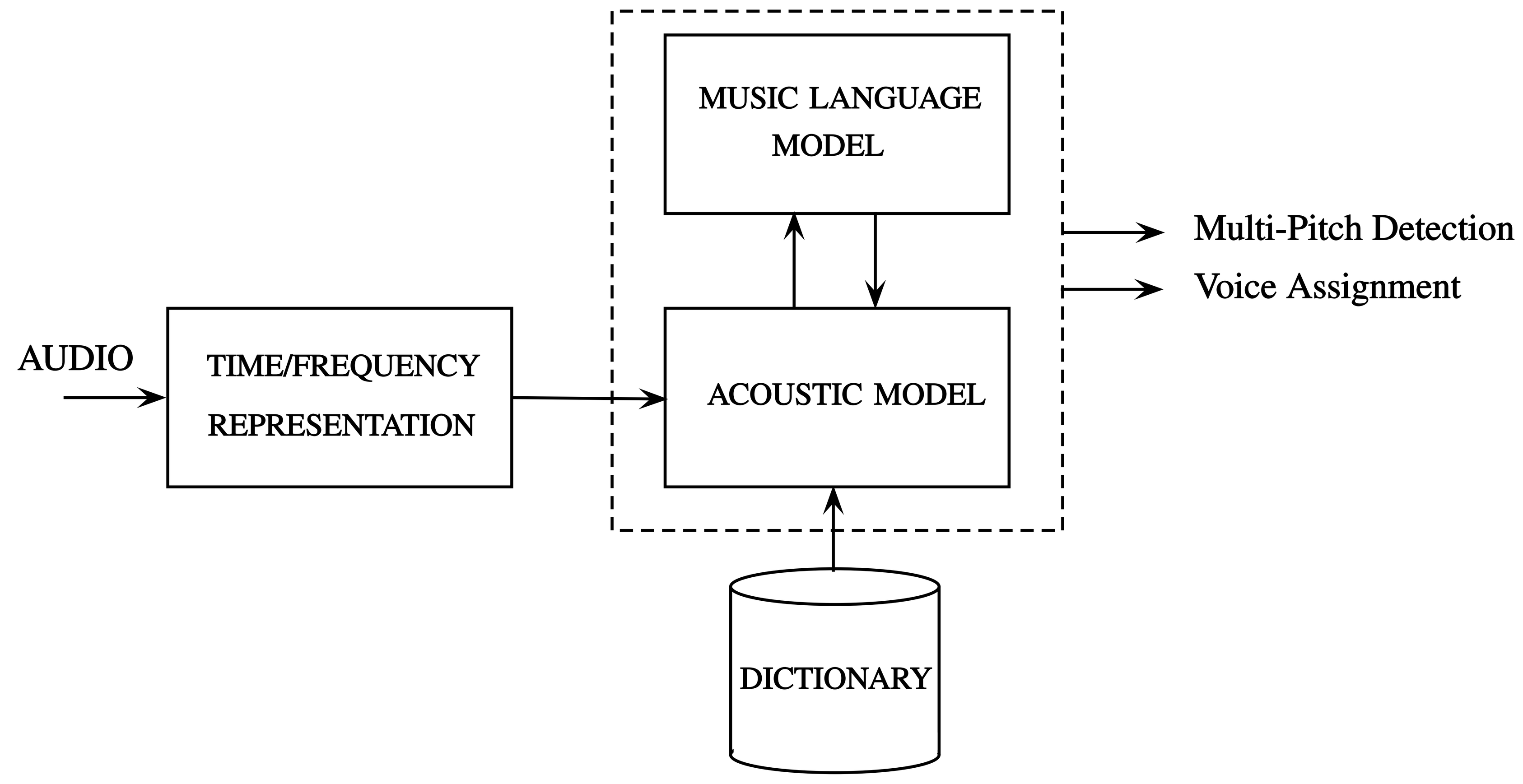
rschramm@ufrgs.br, A.McLeod-5@sms.ed.ac.uk, steedman@inf.ed.ac.uk, emmanouil.benetos@qmul.ac.uk

[1]Department of Music, Universidade Federal do Rio Grande do Sul, Brazil
[2]School of Informatics, University of Edinburgh, UK
[3]Centre for Digital Music, Queen Mary University of London, UK

1 - Introduction

This work presents a multi-pitch detection and voice assignment method applied to audio recordings containing *a cappella* performances with multiple singers. Our approach combines an acoustic model for multi-pitch detection and a music language model for voice assignment.

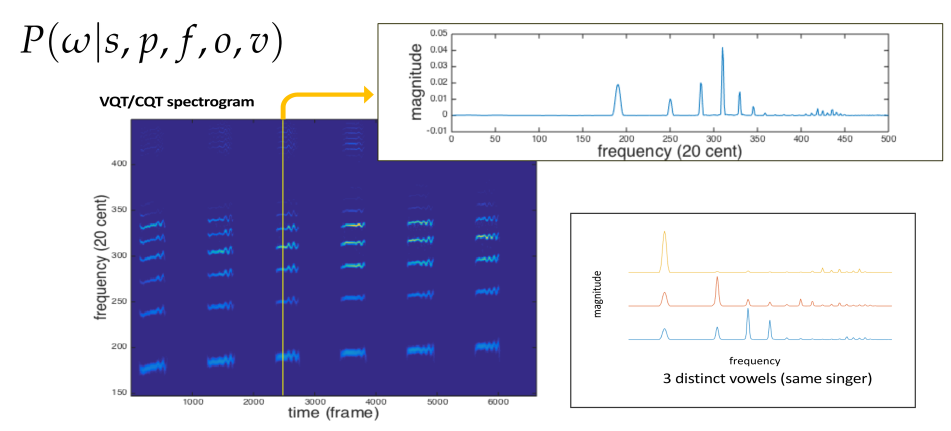




2 - Model

The acoustic model is a spectrogram factorization process based on Probabilistic Latent Component Analysis (PLCA), driven by a 6-dimensional dictionary with pre-learned spectral templates.

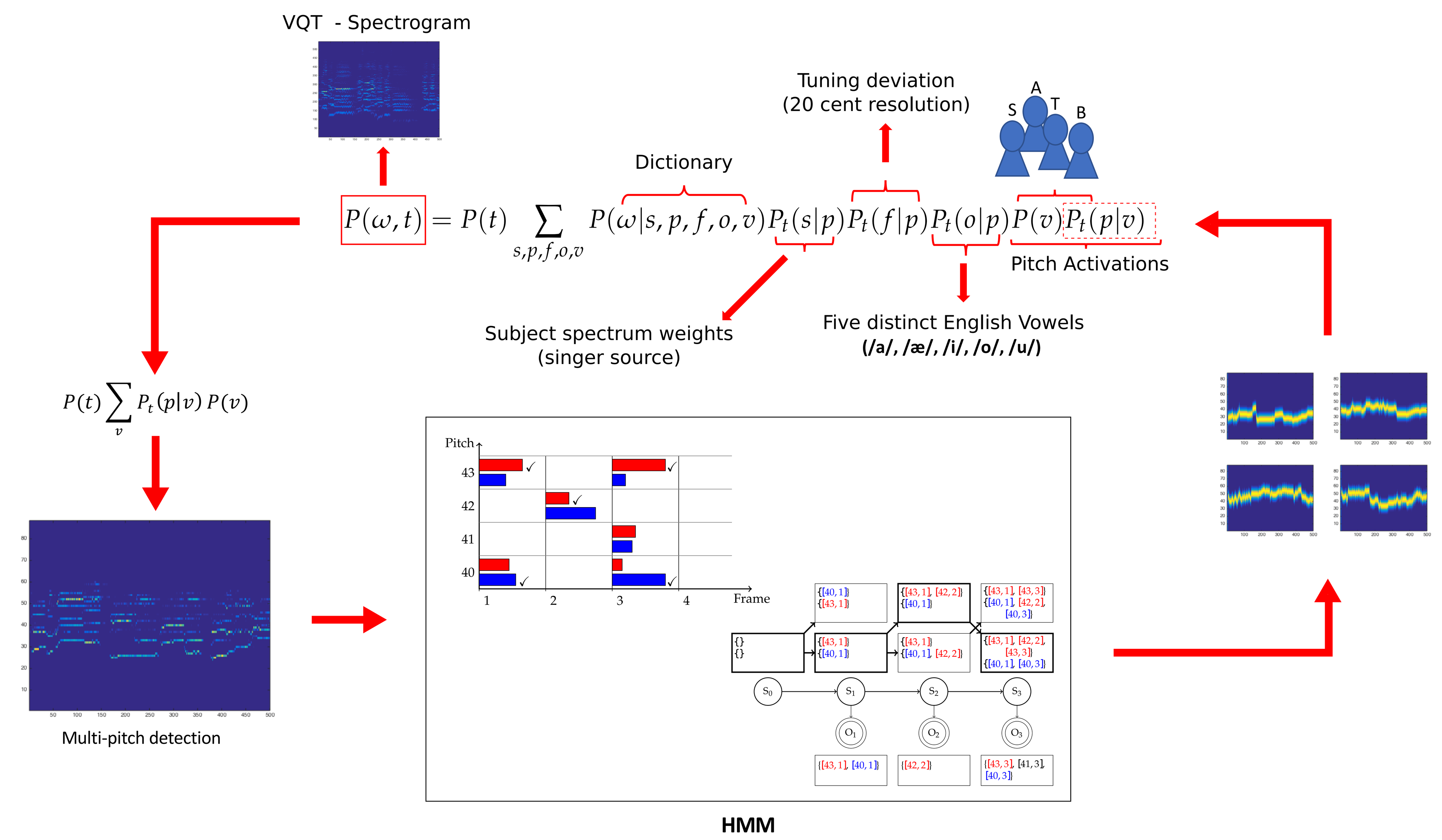
- ω denotes log-frequency
- s denotes the singer index (singer subjects)
- $p \in \{21, ..., 108\}$ denotes pitch in MIDI scale
- $f \in \{1, ..., 5\}$
- o denotes the vowel type
- v denotes the voice type (e.g. soprano, alto, tenor, bass)



$P(\omega|s,p,f,o,v)$

The voice separation component is based on hidden Markov models that use musicological assumptions. By integrating the models, the system can detect multiple concurrent pitches in vocal music and assign each detected pitch to a specific voice corresponding to a voice type such as soprano, alto, tenor or bass (SATB).

$$P(\omega,t) = P(t)\sum_{s,p,f,o,v} P(\omega|s,p,f,o,v)P_t(s|p)P_t(f|p)P_t(o|p)P(v)P_t(p|v)$$

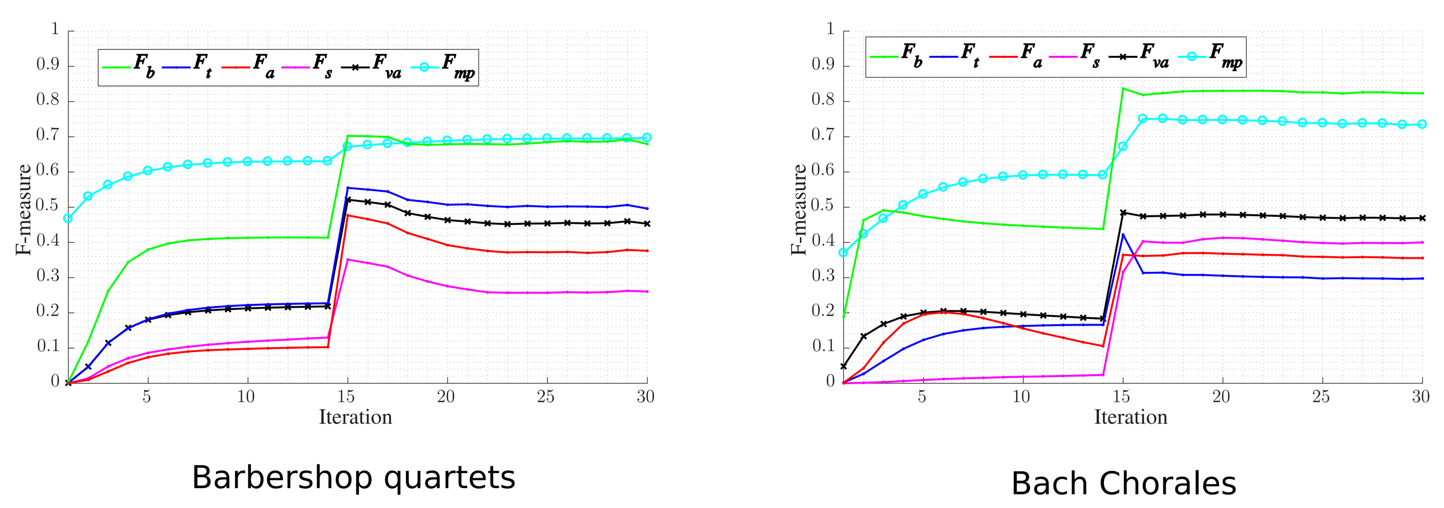


$P(t)\sum_v P_t(p|v)P(v)$

3 - Experiments

Evaluation on two datasets of *a cappella* recordings from http://www.pgmusic.com/ in a total of 104 minutes:
 (a) Bach Chorales = 26 recordings.
 (b) Barbershop Quartets = 22 recordings

The proposed model's F-measures after each Expectation Maximization iteration, averaged across all songs in each dataset:



Barbershop quartets

Bach Chorales

Multi-pitch detection results

Model	Bach Chorales	Barbershop Quartets
Vincent et al., 2010	53.58	51.04
Pertusa et al., 2012	67.19	63.85
Schramm & Benetos, 2017	**70.84**	71.03
Acoustic model only	63.05	59.09
Proposed model	69.66	**73.46**

Voice assignment results

Model	Bach Chorales				
	F_{va}	F_s	F_a	F_t	F_b
Schramm & Benetos, 2017	18.02	15.37	17.59	26.32	12.81
Acoustic model only	21.84	12.99	10.27	22.72	41.37
Proposed model	45.31	26.07	37.63	49.61	67.94

Model	Barbershop Quartets				
	F_{va}	F_s	F_a	F_t	F_b
Schramm & Benetos, 2017	12.29	9.70	14.03	27.93	9.48
Acoustic model only	18.35	2.40	10.56	16.61	43.85
Proposed model	46.92	40.01	35.57	29.76	82.34

4 - Conclusions and Future Work

The proposed acoustic model can not perform reliable voice assignment, but it is notably improved with the integration of the music language model. This integration also improves the multi-pitch detection performance since the acoustic model is guided using musicological principles. Avenues for future work include 1) better handling of overtones in the acoustic model, and 2) better recognition of vibrato in both the acoustic and the music language model.

Queen Mary University of London

UFRGS UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

THE UNIVERSITY of EDINBURGH

SUPPORT: ROYAL ACADEMY OF ENGINEERING