MULTIPITCH ESTIMATION USING A PCLA-BASED MODEL: IMPACT OF PARTIAL USER ANNOTATION

Camila de Andrade Scatolini, Gaël Richard, Benoit Fuentes
Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI
40^{th} IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia, April 2015

1. Introduction

• **Context:** Multipitch estimation in musical recordings based on a probabilistic framework (PLCA).
• **Applications:** Main melody extraction, cover song identification, music transcription.
• **Objective:**
- Can we improve performance with partial annotation of the musical recording?
- How can we better take into account the user inputs?

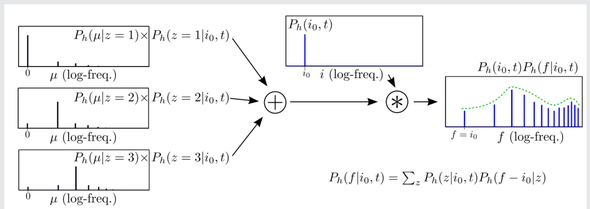
2. The BHAD Model [1]

$|CQT(x)|$: modeled as a probability distribution $P(f,t)$.

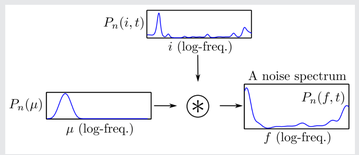
The model:

$$P(f,t) = P(c=h)P_h(f,t) + P(c=n)P_n(f,t)$$

with :
• $P_h(f,t) = \sum_i P_h(i,t)P_h(f|i,t)$,
• $P_h(f|i,t) = \sum_z P_h(z|i,t)P_h(f-i|z)$



Polyphonic harmonic model



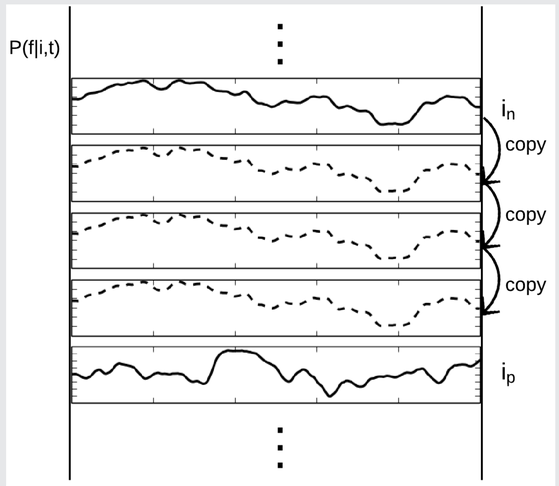
Noise model

Some characteristics:
• Entirely unsupervised
• Control convergence rates by using a "brake"
• Use of priors (*Resemblance prior* on spectral envelopes for each pitch ; *Sparseness prior* on time-frequency activations)

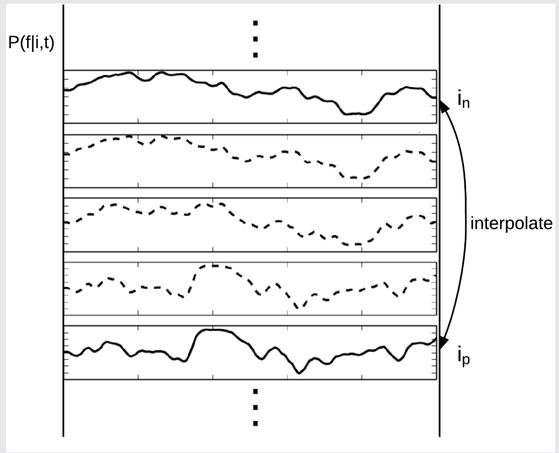
3. Note initialisation

Annotated notes: For each note, the average of all occurrences of this given note.
Three strategies for Non-annotated notes:

- Keep the original slope initialisation,
- Copy previous notes templates,
- Interpolate neighbour notes' templates.



Copy strategy



Interpolate strategy.

4. Convergence rate

Goal: controlling the convergence rate of the well initialised spectral envelopes.

Idea: convergence rate coefficient $\beta_{brake}(n)$ depends on the notes n:

$$\beta_{brake}(n) = \begin{cases} \beta_1, & \text{if } i_n \in \text{learning base} \\ \beta_0, & else \end{cases}, \ \beta_1 > \beta_0$$

5. Results

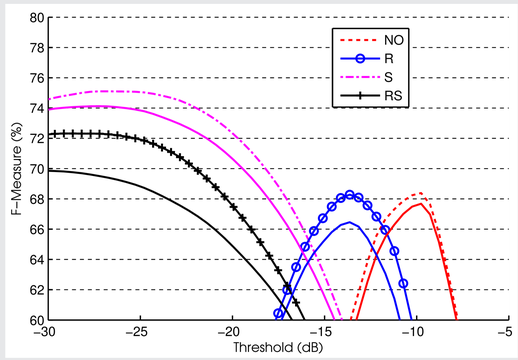
Options	Unsup.	Semi-Guided		
		Slope	Copy	Interp.
NO	67.68	**68.38**	65.00	43.68
S	73.91	**75.07**	71.44	52.07
R	66.46	**68.28**	64.13	47.05
RS	69.51	**72.30**	67.60	52.35
B	77.85	**78.63**	75.88	50.90
SB	74.64	**76.54**	72.47	45.68
RB	77.39	**79.49**	73.89	50.79
RSB	74.85	**77.02**	70.52	47.00

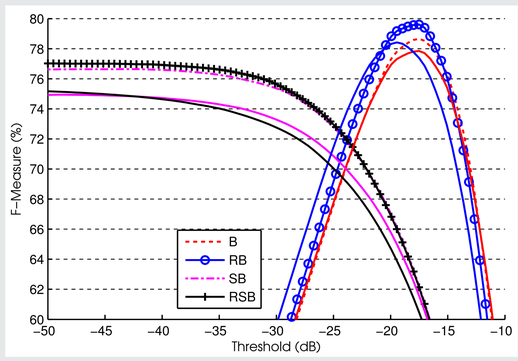
Mean F-Measure

	No brake	Brake annotated notes		
	$\beta_{0,1}=0$	$\beta_1=0.1$	$\beta_1=1$	$\beta_1=10$
B	**68,38**	68,25	66,43	62,93
SB	**75,07**	74,84	74,29	72,74
RB	68,28	68,53	**69,20**	69,16
RSB	72,30	72,35	72,53	**72,66**
	Even	**Brake all notes**		
	$\beta_{0,1}=10$	$\beta_1=10.1$	$\beta_1=11$	$\beta_1=20$
B	78,63	78,63	**78,66**	**78,66**
SB	76,54	76,54	**76,55**	**76,55**
RB	79,49	79,49	79,51	**79,61**
RSB	77,02	77,02	**77,03**	76,98

Mean F-Measure

NO: no prior
S: Sparseness prior
R: Resemblance prior
B: Break





Results for the non-annotated notes (symbols) and the unsupervised approach without and with brake coefficients ($\beta_1 = 20$ and $\beta_0 = 10$).

6. Conclusions

Contributions:
• Better initialisation of the model parameters using partial annotation
• Up to 3% improvement in terms of F-measure in a task of multipitch estimation.

Future work:
• Alternative strategies for the user annotation
• Involve the user in a more interactive way

References:

[1] B.Fuentes, R. Badeau & G. Richard, "Blind harmonic adaptive decomposition applied to supervised source separation", EUSIPCO 2012