Modified Group Delay Feature for Musical Instrument Recognition

Aleksandr Diment[1] Padmanabhan Rajan[2]
Toni Heittola[1] Tuomas Virtanen[1]

[1]Tampere University of Technology
[2]University of Eastern Finland

CMMR 2013, Marseille

18 October 2013

Outline

1 Introduction

2 Methodology

3 Evaluation

4 Conclusions

Introduction

- Music information retrieval: situationally tailored playlisting, personalised radio, social music applications ...
- Musical instrument recognition
- Established set of **features**:
  - **Temporal**: assumption that relevant information is within transient properties of a signal.
  - **Spectral**: those related to harmonic properties preserve the important characteristics of the timbre. Often concentrate only on the magnitude part.

Introduction

- Spectral information: **both magnitude and phase** need to be specified.
- Issue: **wrapping** $\Rightarrow$ direct phase processing spectra is challenging.
- **Phase is informative**: peaks in the spectral envelope (speech: formants)
- Instruments:
  - resonances (**filter** of the body): modelled by MFCCs
  - **source** (also perceptually important) is **neglected**.

Introduction

- A popular solution to direct phase processing issue: **modified group delay function**.
- Previously applied to speech, not yet to instruments.
- **Proposal**: to calculate MODGDF for pitched instrument recognition (primarily or as a complement to MFCCs).
- **Objective**: is MODGDF capable of introducing improvement in recognition accuracy?

1 Introduction

2 Methodology

3 Evaluation

4 Conclusions

Methodology

The *group delay function* is obtained as

$$\tau_g(\omega) = -\mathrm{Im}\left(\frac{d}{d\omega}\log(X(\omega))\right) \quad (1)$$

$$= \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|X(\omega)|^2}, \quad (2)$$

where $Y(\omega)$ is the Fourier transform of $y[n]$, and $y[n] = nx[n]$. No unwrapping needed.

Methodology

- Well-behaved only if zeros of the transfer function are not close to the unit circle.
- Zeros close to the unit circle $\Rightarrow$ high amplitude spikes at corresponding frequencies, resonance structure is masked.

Methodology

Modification[1]: add cepstral smoothing + parameters to control the dynamic range

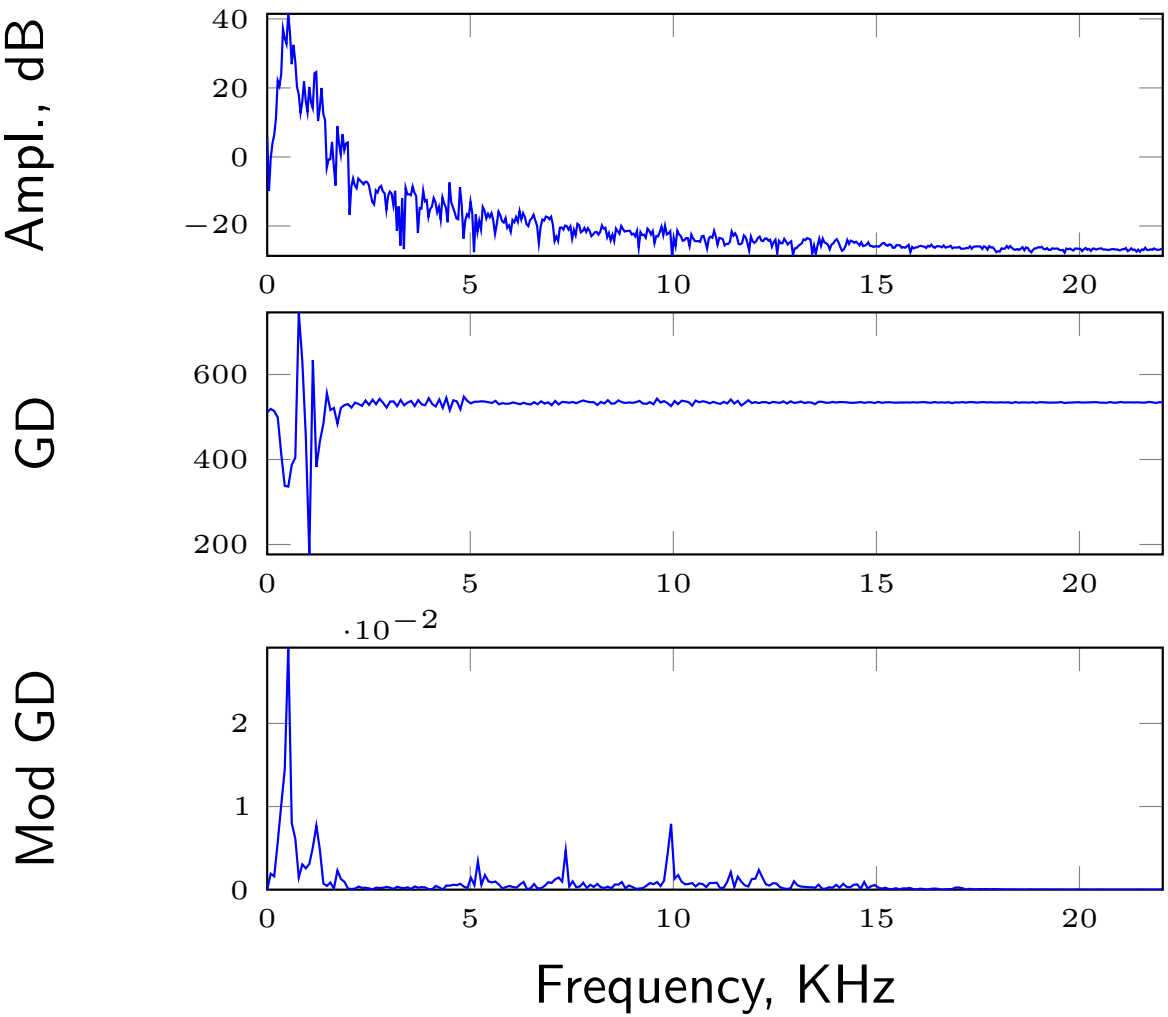
$$\tau_m(\omega) = \mathrm{sign}\big(\tau(\omega)\big)\ \big(|\tau(\omega)|\big)^{\alpha}, \tag{3}$$

where

$$\tau(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|S(\omega)|^{2\gamma}}. \tag{4}$$

[1] H. Murthy and V. Gadde, "The modified group delay function and its application to phoneme recognition", in *Acoustics, Speech, and Signal Processing, 2003. Proc. (ICASSP '03). 2003 IEEE Int. Conf. on*, vol. 1, 2003, pages. DOI: 10.1109/ICASSP.2003.1198718.

E.g. middle-C note played by Bassoon:



1 Introduction

2 Methodology

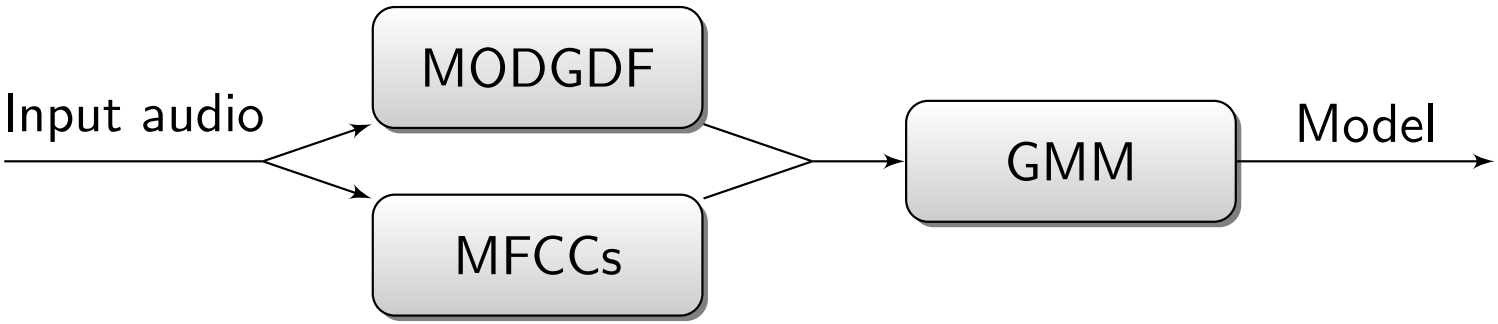
3 Evaluation

4 Conclusions

Evaluation

Set up

- Frame-blocking: 20 ms, 50% overlap.
- MODGDF (16 coefficients) + Δ + $\Delta\Delta$. Parameters from the speech studies[2].
- Baseline: MFCCs (16 coefficients) + Δ + $\Delta\Delta$
- Combination of these by concatenation.



[2] R. Hegde, H. Murthy, and V. Gadde, "Significance of the modified group delay feature in speech recognition", *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 190–202, 2007, ISSN: 1558-7916. DOI: 10.1109/TASL.2006.876858.

Evaluation

Acoustic material

From RWC Music Database. Separate single note-wise recognition scenario.

Instrument set	List of instruments
woodwinds	Oboe, Clarinet, Piccolo, Flute, Recorder
strings	Violin, Viola, Cello, Contrabass
4 various	Acoustic Guitar, Electric Guitar, Tuba, Bassoon
9 various	Piano, Acoustic Guitar, Electric Guitar, Electric Bass, Trombone, Tuba, Bassoon, Clarinet, Banjo.
22 various	"4 various" + "9 various" + "woodwinds" + "strings" + vocals: Soprano, Alto, Tenor, Baritone, Bass

Evaluation

Results

Instrument set	Recognition accuracy, %		
	MFCCs	MODGDF	combined
woodwinds	74.5	66.7	**77.2**
strings	69.7	73.8	**73.6**
4 various	90.9	84.4	**96.0**
9 various	82.6	59.9	**84.9**
22 various	68.8	41.7	**70.7**

1 Introduction

2 Methodology

3 Evaluation

4 Conclusions

Conclusions

- The proposed MODGDF+MFCCs: increase in accuracy with all sets compared to MFCCs. Up to 5.1%.
- MODGDF on its own shows to be a somewhat informative feature for instruments.

Future:

- Dependency on phase-lockness of the instrument groups.
- Parameters of MODGDF, optimal for speech, have been used. To find optimal values for instruments is worthwhile.