DTU

IMM/ISP

On The Relevance of Spectral Features for Instrument Classification

A. B. Nielsen, S. Sigurdsson, L. K. Hansen, and J. Arenas-García
Technical University of Denmark
{abn,siggi,lkh,jag}@imm.dtu.dk

Introduction

In this paper we will:

- Study the validity of two spectral models for instruments.
- Study the suitability of MFCC/HR features to instrument classification.

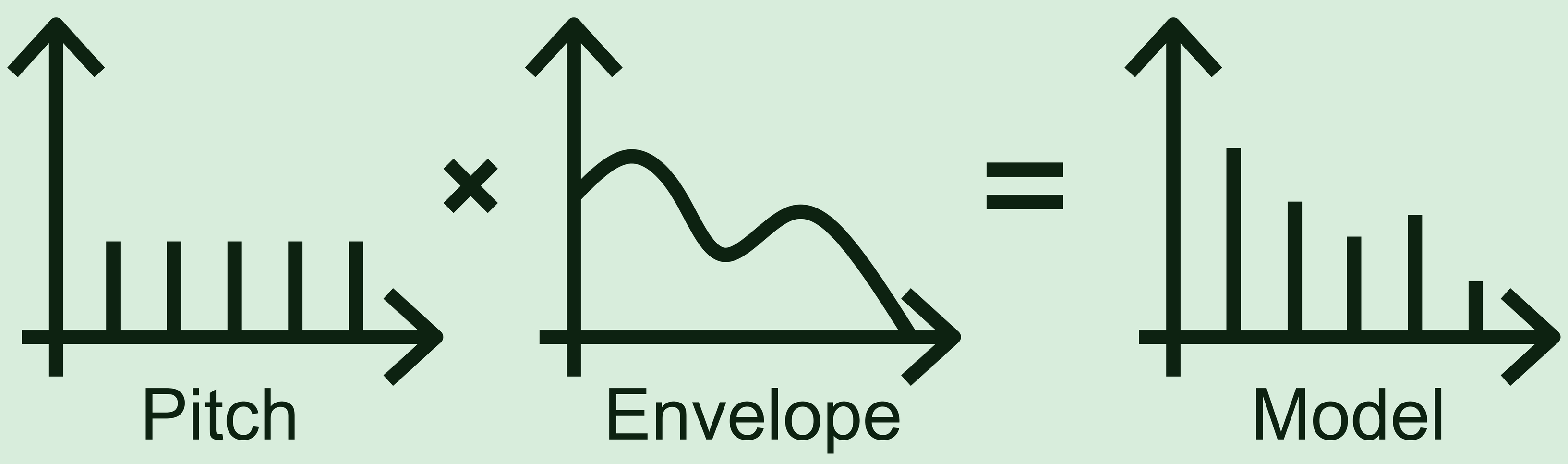
The two models are:

- Fixed envelope.
- Fixed harmonic amplitude.

To increase generality of the results we use two classifiers:

- rKOPLS.
- Multi layer perceptron.

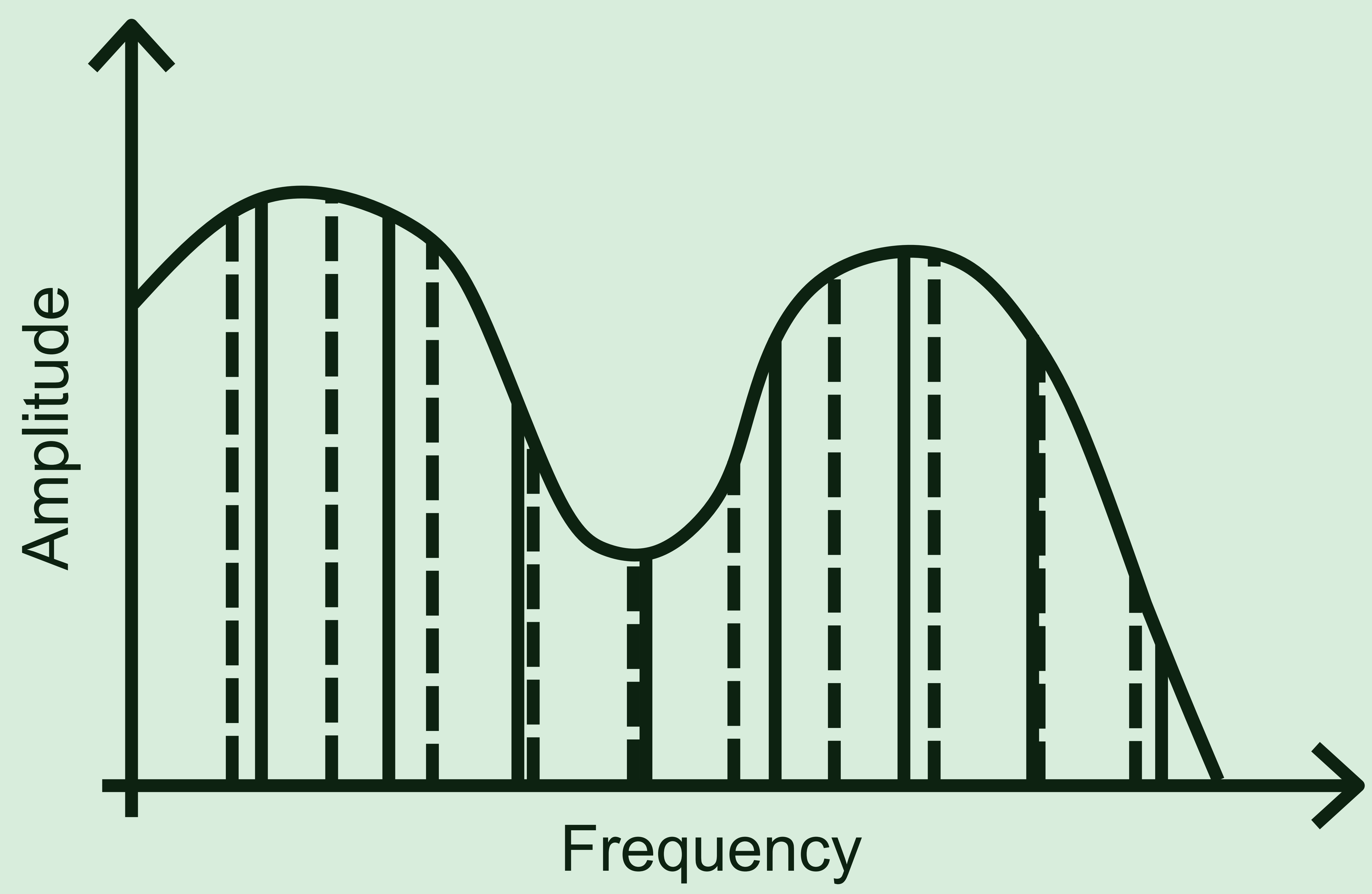
Pitch & Envelope



Pitch - The pitch is what is perceived as the tone, and its value is given by the fundamental frequency, i.e., the frequency of the first harmonic.

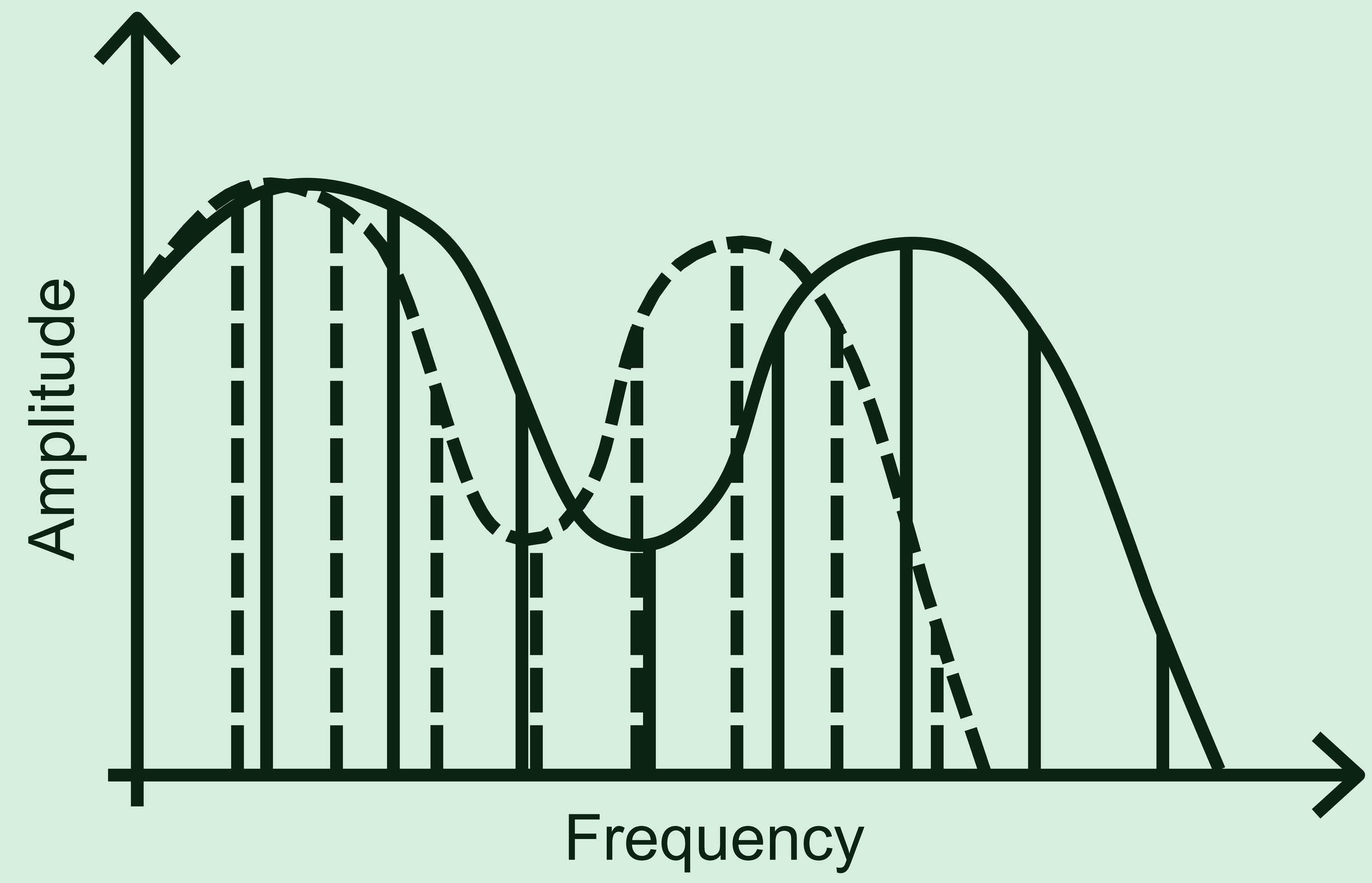
Envelope - The envelope is the amplitude of the harmonics. If two instruments are playing the same note, i.e. the same pitch, only the envelope distinguishes the instruments.

Constant envelope model



MFCC's are used to capture the envelope.

Constant harmonics amplitude



The amplitudes are normalized with the amplitude of the first harmonic. We call these features Harmonic Representation (HR).

rKOPLS

- KOPLS projects data in "Kernel" space, where relevant (non-linear) features are extracted using linear OPLS.
- Assuming a sparse representation of the solution (rKOPLS), complexity of the algorithm can be kept under control.
- The method can easily be extended to compensate for unbalanced classes.
- A single layer perceptron + softmax network is trained on the extracted projections to predict the class.

Multi layer perceptron

- Single layer of hidden units.
- Tanh activation in hidden units.
- Softmax output function
- 30 hidden units.
- Sample size compensated error function,

$$E = -\sum_{i=1}^{N}\sum_{k=1}^{C}\lambda_k y_k^{(i)} \ln \hat{y}_k^{(i)},$$
$$\lambda_k = 1/N_k.$$

.

IMM/ISP

On The Relevance of Spectral Features for Instrument Classification

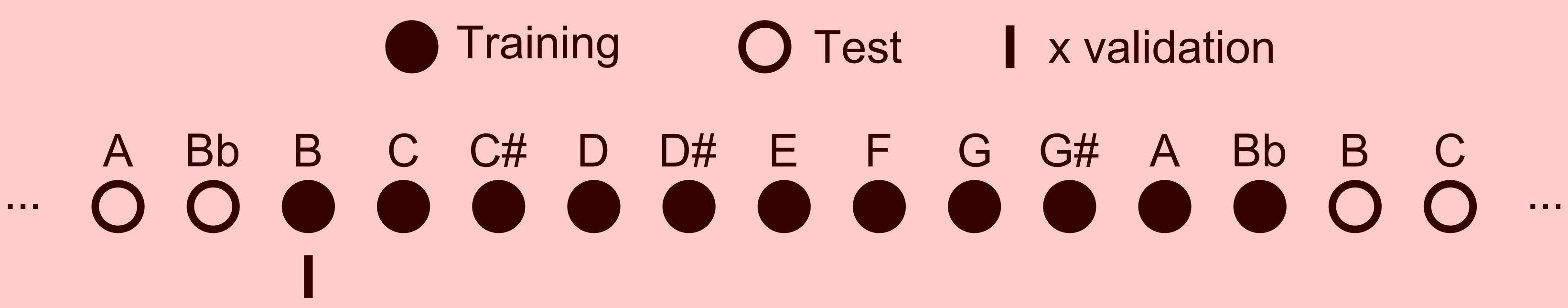
A. B. Nielsen, S. Sigurdsson, L. K. Hansen, and J. Arenas-García
Technical University of Denmark
{abn,siggi,lkh,jag}@imm.dtu.dk

Instrument data set

We use the IOWA Instrument Database.

- 17 different instruments:
  AltoFlute, AltoSax, BassClarinet, BassFlute, Bassoon, BbClar, Cello, EbClar, Flute, Horn, Oboe, Piano, SopSax, TenorTrombone, Trumpet, Viola, Violin.
- Complete note range for each instrument.
- Where possible, different ways of playing instrument was included.
- Fortissimo and mezzoforte (pianissimo was discarded).
- 50 ms time frame, 50 % overlap.
- 282.812 samples.

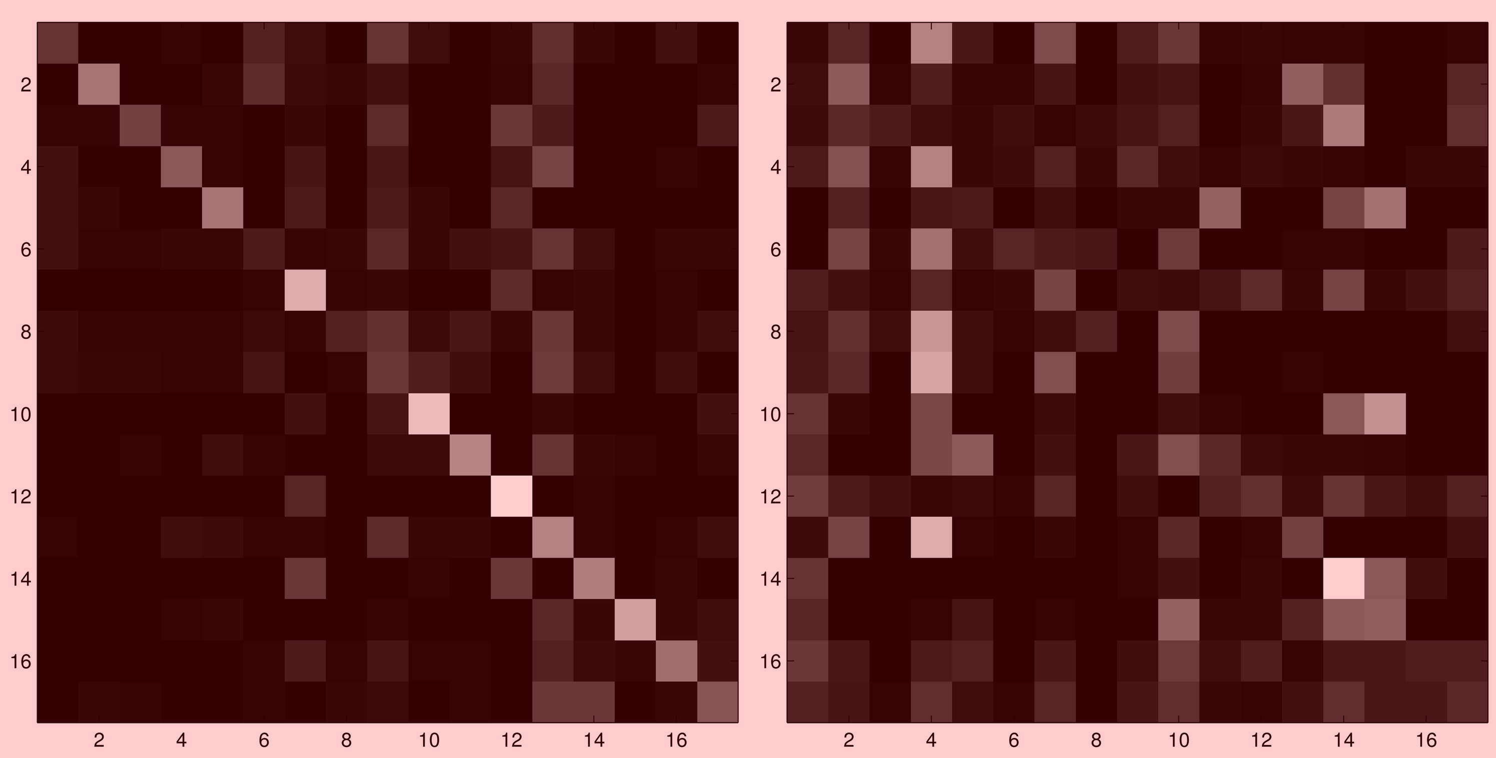
Generalization capability



- Trained on common octave.
- 11 fold cross validation.
- Tested on remaining notes.

		MFCCs	HR
MLP	Tr / Val	91.4 / 79.1	79.3 / 58.8
	Tr / Tst	91.2 / 42.8	78.5 / 12.9
rKOPLS	Tr / Val	89.5 / 80.1	78.2 / 57.7
	Tr / Tst	89.3 / 42.4	77.4 / 14.2

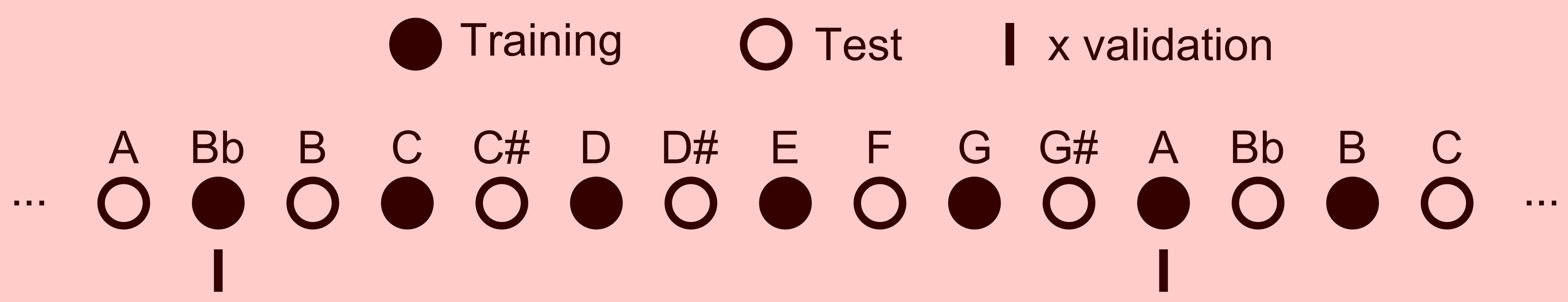
Test set confusion matrices.



MFCC HR

The model based on MFCC features shows a diagonal trend where as the model based on HR does not. The large degradation in performance in the training / test set split shows that, even though the MFCC's perform better, neither of the models gives a complete model of the envelope.

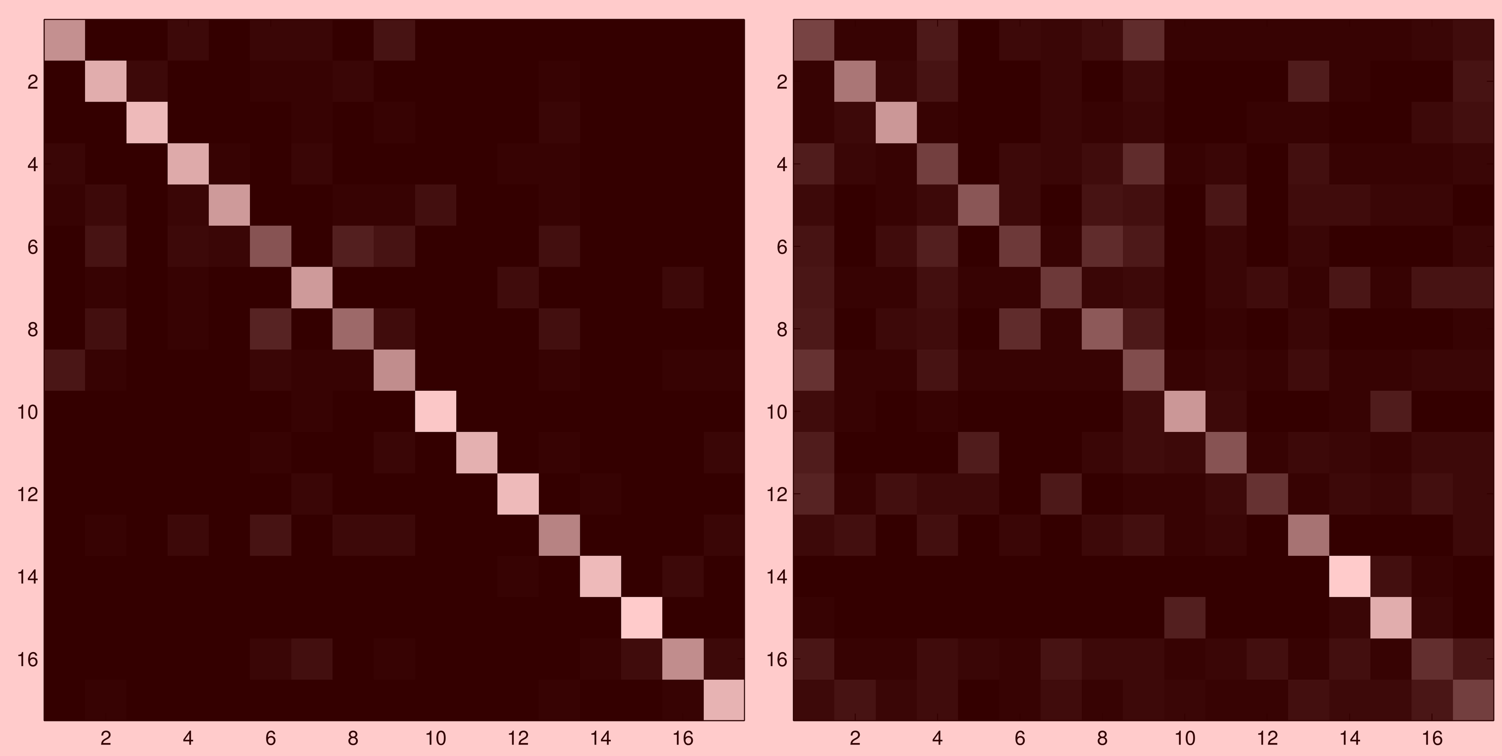
Complete pitch range training



- Trained on every second note.
- 5 fold cross validation.
- Tested on remaining notes.

		MFCCs	HR
MLP	Tr / Val	87.4 / 70.7	52.2 / 29.7
	Tr / Tst	86.1 / 74.7	50.2 / 38.0
rKOPLS	Tr / Val	89.4 / 73.2	63.3 / 32.4
	Tr / Tst	84.4 / 75.9	60.7 / 41.2

Test set confusion matrices.



MFCC HR

Clearly both models perform better and this time both show a diagonal trend. Like before, the model based on MFCC's perform significantly better, especially when looking at the errors presented in the table.

Conclusion

- MFCC coefficients are a better representation than HR.
- Both classification technologies agree on first conclusion.
- High classification performance with a sufficiently rich data set.

.