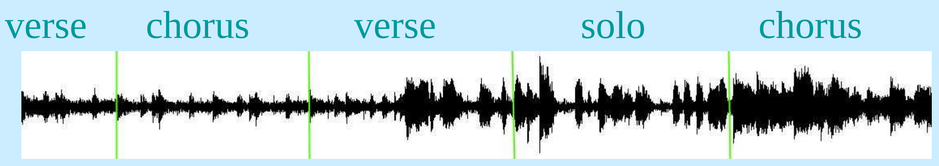
Improving Markov model –based music piece structure labelling with acoustic information

Jouni Paulus
Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany

Music structure analysis from audio signal

1. Divide signal into segment representing musical parts.
2. Group segments that are occurrences of the same musical part.

3. Assign musically meaningful labels to the groups.



Labelling a post-processing step

-Mappings from tags (e.g., *p1, p2*) to labels (e.g., *chorus, verse*).

-Any injective mapping *f* is a valid one, but **which is the best**?

-Define a likelihood model to select the most suitable mapping.

$$f_{\mathrm{OPT}} = \underset{\mathrm{f}}{\operatorname{argmax}} \{ p(f | r_{1:K}, \mathbf{x}_{1:K}) \}, f : R \rightarrow L \text{ injective}$$

$$p(f | r_{1:K}, \mathbf{x}_{1:K}) = p(\mathbf{x}_{1:K} | f(r_{1:K})) p(f(r_{1:K}))$$

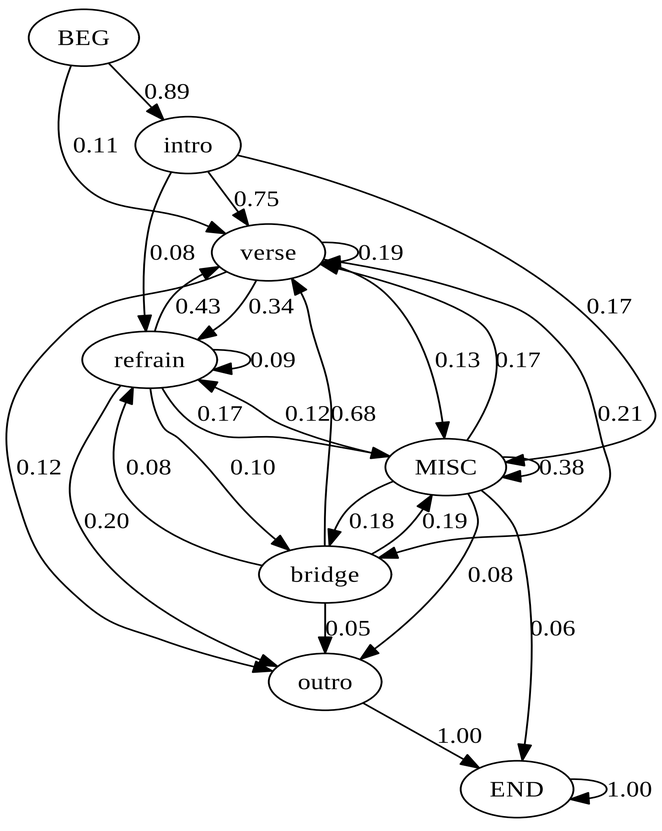
x is acoustic feature vector, *K* is number of parts, *f* is mapping function from tags *r* to labels (R is set of tags and L is set of labels).

Baseline sequence model

-Sequence model: certain part sequences are more likely than others

-E.g., “intro, verse, verse, bridge, verse, bridge, verse, outro”

- N-grams of part labels (baseline system from CMMR08).



Bigram (N=2) transition probabilities between labels in Beatles’ songs. The labels belonging to the most rarely occurring 15% are gathered under “MISC”, and the transitions with probability less than 0.05 are removed for clarity.

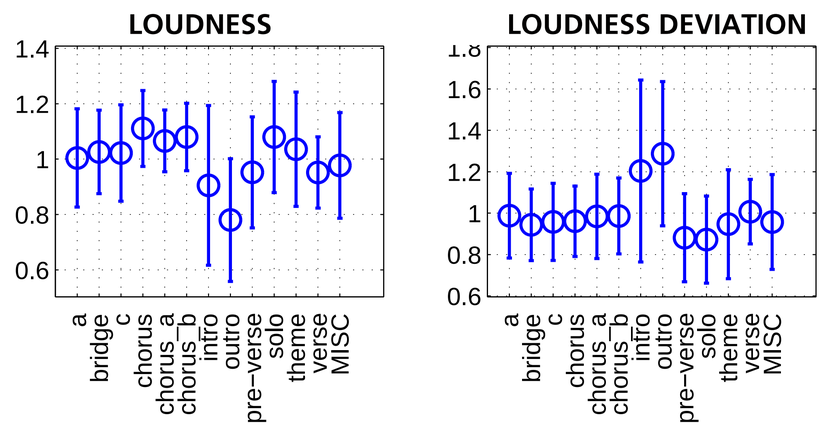
Proposed acoustic model

-Acoustic model: “chorus is the most energetic part in a song”

-Part loudness relative to the mean loudness of the piece.

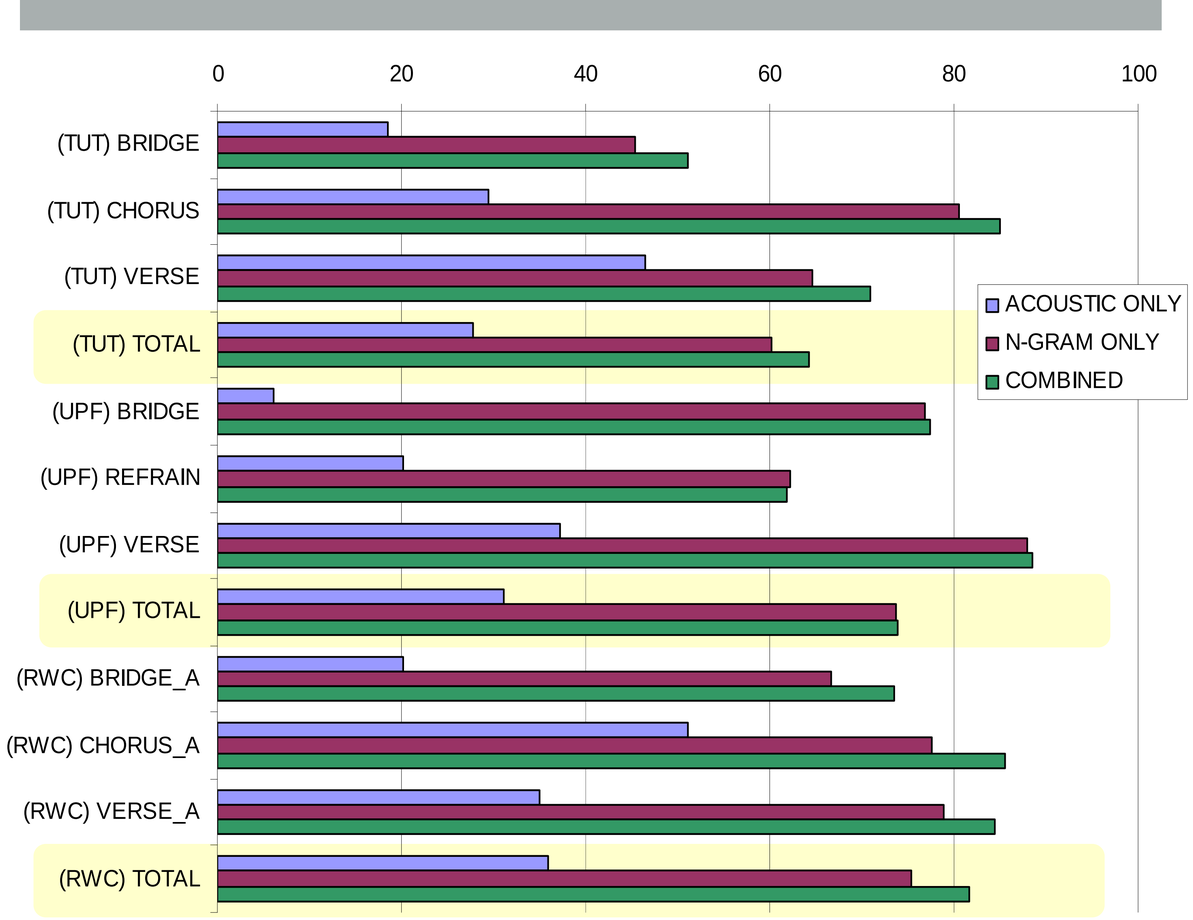
-Loudness variation within part to describe dynamics.

-Parametrised as multivariate Gaussian distributions.



Example of loudness feature statistics in TUTstructure07 data set. The circle denotes the distribution mean, and the error bars denote the standard deviation of the distribution.

Results



Evaluation results on three data sets.of popular music (TUTstructure07, UPF Beatles, and RWC Pop). Only three frequent labels from each set and the overall result are presented here. The presented sequence model uses an N-gram with N=3. Values are % of entire duration of the label recovered correctly. (More results in the paper.)

Conclusions

- The assumption of typical acoustic relationships between parts in a musical piece holds to some extent.
- Using additional acoustic information in labelling provides slight improvement over the sequence modelling baseline method.
- Performance heavily dependent on musical style and part label.
- Proposed addition is easy to integrate to the baseline method.

This work was performed when the author was at the Department of Signal Processing, Tampere University of Technology, Finland, and this work was supported by the Academy of Finland (application number 129657, Finnish Programme for Centres of Excellence in Research 2006-2011).