Project no. FP6-507752

MUSCLE

Network of Excellence
Multimedia Understanding through Semantics, Computation and Learning

DN 4.1: Expanded List of Software Tools for Audio Indexing

Due date of deliverable: 30.11.2007
Actual submission date: 04.02.2008

Start date of project: 1 March 2004 Duration: 48 months

Deliverable Type: PU
Number: DN4.1
Nature: P
Task: WP4

Name of responsible: Andreas Rauber, TU Vienna-IFS

Revision 1.2

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

MUSCLE

Network of Excellence
Multimedia Understanding through Semantics, Computation and Learning

Expanded List of Software Tools for Audio Indexing

Within the MUSCLE Network of Excellence on multimedia understanding, data mining and machine learning researchers have developed a range of tools for audio analysis, speech recognition, sound description and music retrieval. This deliverable (D4.1) of WP4 represents an inventory of audio feature extraction, description and recognition tools.

Table of Contents:

The inventory is available under:
http://www.ifs.tuwien.ac.at/mir/muscle/del/audio_tools.html
also reachable from:
http://www.muscle-noe.org/content/view/112/92/

RP_extract Music Feature Extractor

TU Vienna - IFS, Thomas Lidy

Content-based access to audio files, particularly music, requires the development of feature extraction techniques that capture the acoustic characteristics of the signal, and that allow the computation of similarity between pieces of music. At TU Vienna - IFS three different sets of descriptors were developed:

- **Statistical Spectrum Descriptors**: describe fluctuations by statistical measures on critical frequency bands of a psycho-acoustically transformed Sonogram
- **Rhythm Patterns**: reflect the rhythmical structure in musical pieces by a matrix describing the amplitude of modulation on critical frequency bands for several modulation frequencies
- **Rhythm Histograms**: aggregate the energy of modulation for 60 different modulation frequencies and thus indicate general rhythmic in music

The algorithm considers psycho-acoustics in order to resemble the human auditory system. The feature extractor is currently implemented in Matlab and processes au, wav, mp3 and ogg files. Feature vectors are output in SOMLib format.

Download (V 0.6.21) : http://www.ifs.tuwien.ac.at/mir/muscle/del/rp_extract_0.621.zip

Usage Guide: http://www.ifs.tuwien.ac.at/mir/howto_matlab_fe.html

Audio Feature Extraction Web Service

TU Vienna - IFS, Jakob Frank

The Audio Feature Extraction web service provides users the extraction of the described features above (Statistical Spectrum Descriptors, Rhythm Patterns, Rhythm Histograms) conveniently over the Web.

After submitting audio data, the web service analyses the audio signal, extracts features and in turn provides the extracted descriptors as a file or stream. These can be used either in own local applications for audio and music similarity and the likes (classification, organization), but the Web Service furthermore also provides the possibility to generate Self-Organising Maps, which can be used for organization of musical content and for convenient access to music collections.

Besides allowing an immediate use of the web service with a Java Web Start application, we provide interfaces to the web service for developers, which can integrate the functionality in their own applications.

Details, direct **access to the web service**, and interface description:
http://www.ifs.tuwien.ac.at/mir/webservice/

Sound Description Toolbox

AUTH - AIIA, Emmanouil Benetos

The Sound Description Toolbox extracts a number of MPEG-7 standard descriptors as and other feature sets from WAV audio files. Features covered are:

- Energy: AudioPower
- Harmonic: AudioFundamentalFrequency
- Perceptual: Specific Loudness Sensation Coefficients
- Spectral: AudioSpectrumCentroid, Audio Spectrum Rolloff, AudioSpectrumSpread, MFCCs
- Temporal: Autocorrelation Coefficients, Log-attack Time, TemporalCentroid, Zero-crossing rate
- Various: AudioSpectrumFlatness

Instructions: Freely Distributed MATLAB Source Code to extract 187 features from an audio signal; Main Function: GUI_feature_matrices.m

Download: http://www.ifs.tuwien.ac.at/mir/muscle/del/SoundDescriptionToolbox08-01.zip

GenChords - Automatic Chord Detection

TU Vienna - IFS, Veronika Zenz

We have designed an automatic chord detection algorithm that operates on musical pieces of arbitrary instrumentation and considers music theoretical knowledge. Our detection method incorporates rhythm and tonality of the musical piece the same as knowledge about the common frequencies of chord-changes. An average accuracy rate of 65% has been achieved on a test set of 19 popular songs of the last decades and confirms the strength of this approach.

The GenChords chord-detector consists of 4 modules: The basic chord detection itself, beat tracking, key detection and a chord-sequence optimizer. Beat tracking is used to split the audio data into blocks of sizes that correspond to the computed beat structure. As chord changes usually happen on beat times, beat detection is a good method to enlarge analysis blocks without risking to miss chord-changes. Each of the obtained blocks is passed to an enhanced autocorrelation-algorithm. Its output is then used to compute the intensity of each pitch class, the so called Pitch Class Profile (short PCP). The calculated PCP's are compared to a set of reference chordtype-PCP's using only those reference chords that fit to the key of the song. Finally the smoothing algorithm rates each chord according to the number of chord changes around it.

Details: http://www.ifs.tuwien.ac.at/mir/chorddetection.html

Download (C++ source code):
http://www.ifs.tuwien.ac.at/mir/chords/download/genchords.zip

ARIA - Dynamics Enhancement

CNR-ISTI, Graziano Bertini

ARIA - DDS is a new methodology of sound processing, able to rebuild the dynamic freshness of the recorded music. With ARIA every sound in the mix will be pushed in evidence, reducing the sense of "flatness" we often meet in today's commercial recordings. ARIA allows to perceive the music structure and "groove" also at relative low volumes, where usually standard way of playback fails, sounding poor and uninteresting. ARIA is based on a special algorithm, which recognizes and enhance the dynamic sound variations at the various audio spectrum frequency bands. Only quite fast variations and peaks are altered, following special rules: in this way the overall timbre is not changed as often happens in common exciters or dynamic expanders.

Details: http://www.aria99.com

Download demo: http://www.aria99.com/dsp_aria.zip

Contact: aria_dsp@yahoo.it

Bayesian Extensions to Non-negative Matrix Factorisation for Audio Signal Modelling

Cambridge University, Tuomas Virtanen, Ali-Taylan Cemgil, Simon Godsill

Non-negative matrix factorisation is applied here as a decomposition where spectrogram matrix X is approximated as a product of excitation matrix E and template matrix V. It is an efficient machine learning tool that can be applied in audio signal analysis in supervised or unsupervised manner. Bayesian extensions of the method enable assigning priors for the parameters. In the case of templates V, they allow defining prior information about spectra of sources and when used for excitations E, they can model correlation between adjacent frames, or sparseness of the excitations.

The software package contains non-negative matrix factorisation algorithms targeted for the analysis audio signals. It contains an implementation of the algorithms by Lee and Seung [1], which minimise either the Euclidean distance or the divergence function. It also contains extensions, which include sparseness or temporal continuity terms [2]. Bayesian extensions described in [3] which enable setting priors for the matrices to be estimated are implemented in the package as well. It also includes algorithms for minimising two other reconstruction error measures which are based on the probabilistic models described in [4].

[1] D. D. Lee, H. S. Seung. Algorithms for non-negative matrix factorization, Advances in Neural Information Processing Systems, 2001.

[2] T. Virtanen. Monaural Sound Source Separation by Non-Negative Matrix Factorization with Temporal Continuity and Sparseness Criteria, IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 3, March 2007.

[3] T. Virtanen, A. T. Cemgil, and S. J. Godsill. Bayesian Extensions to Non-negative Matrix Factorisation for Audio Signal Modelling, ICASSP 2008.

[4] T. Virtanen and A. T. Cemgil. Bayesian extensions to Non-negative Matrix Factorisation using Gamma Chain Priors. Technical report, Cambridge University, October 2007.

Instructions: All the algorithms in the package can be used by calling the Matlab function nmffactorize which can be found in the file nmffactorize.m. The help text of the function explains its usage.

Download: http://www.ifs.tuwien.ac.at/mir/muscle/del/UCam_matlab_nmf-audio.zip

Further information about the algorithms in the package can be obtained from Simon Godsill (sjg@eng.cam.ac.uk) or Taylan Cemgil (atc27@eng.cam.ac.uk).

Multi-Object Tracking of Sinusoidal Components in Audio

Cambridge University, Daniel Clark, Ali-Taylan Cemgil, Paul Peeling, Simon Godsill

The tracker identifies individual sinusoidal tracks from audio signals using multi-object stochastic filtering techniques. It can estimate target states when observations are missing and can maintain the identity of these targets between time-frames. The objective to find parameters of a set of damped sinusoidals is achieved through a multi-object generalisation of Bayes filtering. The Probability Hypothesis Density Filter is a recursion of the first-order moment of the multi-object Bayes filter.

This is an example of the tracker, which has been programmed for a two-dimensional constant velocity model. The library is not restricted to this case and can be used for other models and different dimensions. The algorithm is described in the following papers:

D. E. Clark, K. Panta, V. Ba-Ngu (2006) The GM-PHD Filter Multiple Target Tracker. Proc. International Conference on Information Fusion, Florence, Italy, pp. 1-8.

D. Clark, A. T. Cemgil, P. Peeling, and S. Godsill (2007) Multi-object tracking of sinusoidal components in audio with the Gaussian mixture probability hypothesis density filter. Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 339-342.

Instructions: The executable is called 'tracker' and is simply called with:

./tracker

The measurements are in the file test_measurements.txt and parameters in the file parameters.txt. The output files are put in the directory 'tracks'. The file license_check.txt must be in the current directory.

Download: http://www.ifs.tuwien.ac.at/mir/muscle/del/UCam_tracker.zip

Further information about the algorithms in the package can be obtained from Simon Godsill (sjg@eng.cam.ac.uk) or Daniel Clark (danielclark@ieee.org)@

Music Audio Segmentation

TU Vienna - IFS, Ewald Peiszer

Automatic audio segmentation aims at extracting information on a songs structure, i.e., segment boundaries, musical form and semantic labels like verse, chorus, bridge etc. This information can be used to create representative song excerpts or summaries, to facilitate browsing in large music collections or to improve results of subsequent music processing applications like, e.g., query by humming. This algorithm performs 2 phases:

Phase 1: Boundary detection
This phase tries to detect the segment boundaries of a song, i.e., the time points where segments begin and end. The output of this phase is used as the input for the next phase.

Phase 2: Structure detection
This phase tries to detect the form of the song, i.e., a label is assigned to each segment where segments of the same type (verse, chorus, intro, etc.) get the same label. The labels themselves are single characters like A, B, C, and thus not semantically meaningful.

Details: http://www.ifs.tuwien.ac.at/mir/audiosegmentation.html

Download: http://www.ifs.tuwien.ac.at/mir/audiosegmentation/dl/ep_audiosegmentation-2007-07-30.zip

WinSnoori Speech Analysis Software

INRIA-Parole, Yves Laprie

Using tools for investigating speech signals is an invaluable help to teach phonetics and more generally speech sciences. For several years we have undertaken the development of the software WinSnoori which is for both speech scientists as a research tool and teachers in phonetics as an illustration tool. It consists of five types of tools:

- to edit speech signals,
- to annotate phonetically or orthographically speech signals. WinSnoori offers tools to explore annotated corpora automatically,
- to analyse speech with several spectral analyses and monitor spectral peaks along time,
- to study prosody. Besides pitch calculation it is possible to synthesise new signals by modifying the F0 curve and/or the speech rate,
- to generate parameters for the Klatt synthesiser. A user friendly graphic interface together with copy synthesis tools (automatic formant tracking, automatic amplitude adjustment) allows the user to generate files for the Klatt synthesiser easily.

In the context of speech sciences WinSnoori can therefore be exploited for many purposes, among them, illustrating speech phenomena and investigating acoustic cues of speech sounds and prosody.

Download: http://www.ifs.tuwien.ac.at/mir/muscle/del/WinSnoori_1.34_setup.exe

Details and Guide: http://www.loria.fr/~laprie/WinSnoori/index.html

Support Vector Machine Re-Scoring Algorithm of Hidden Markov Models with Applications to Speech Recognition

Tel-Aviv University (TAU – SPEECH), A. Sloin, A. Alfandary, D. Burshtein

According to the statistical approach to automatic speech recognition, each linguistic unit is typically assigned a hidden Markov model (HMM), the parameters of which are estimated using the Maximum Likelihood (ML) approach. However, when the assumed model is not sufficiently accurate or when there is not enough training data, it is possible to significantly improve the recognition rate using discriminative training methods, such as support vector machine (SVM)-type classification. In our work we proposed a new SVM re-scoring algorithm of HMMs with applications to speech recognition. This algorithm utilizes a variable to fixed length transformation of the speech data. The application of the algorithm to isolated and connected digit and word-spotting tasks provides a significant error rate reduction compared to standard ML-trained HMM systems.

Figure 1 shows a comparison between our method and standard ML training (baseline) on a noisy isolated digit task (TIDIGITS database). The error rate reduction is 56.1% compared to the baseline.

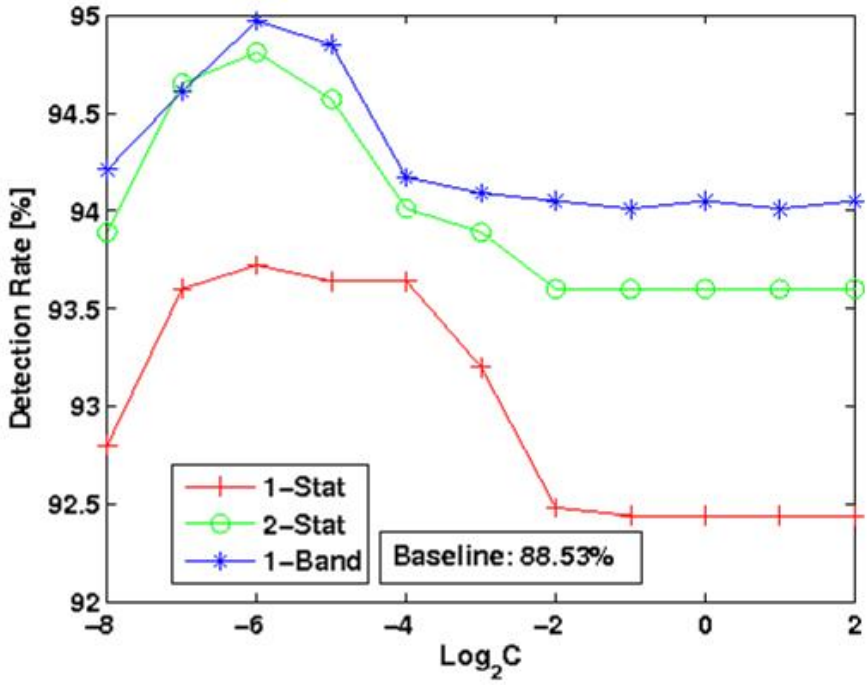

Figure 1

For a connected digit experiment we obtained a reduction of 51% in the error rate. This is shown in the figure 2.

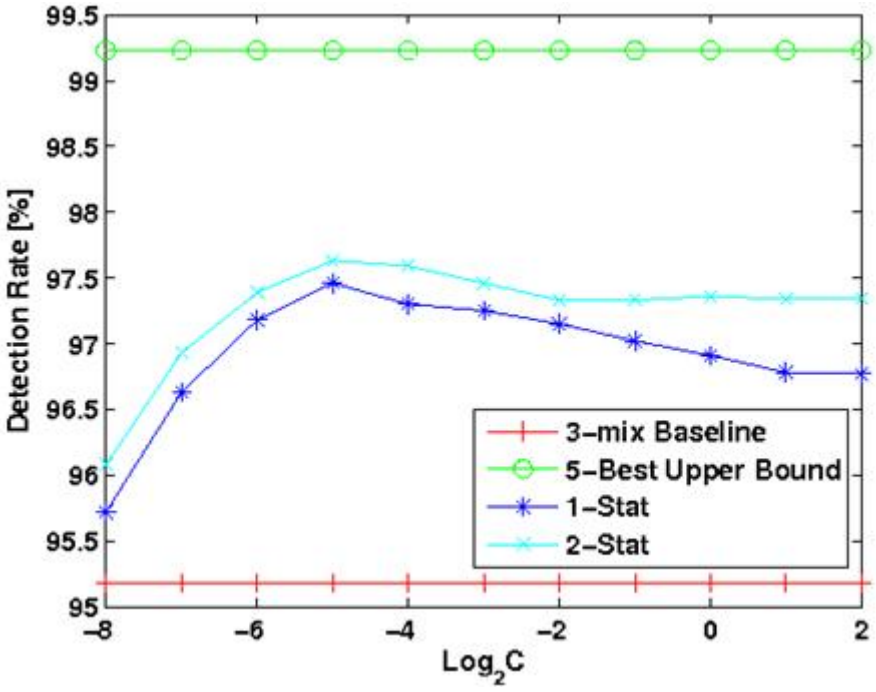

Figure 2

Figure 3 shows the ROC curves of the baseline and new approaches for a keyword spotting task.

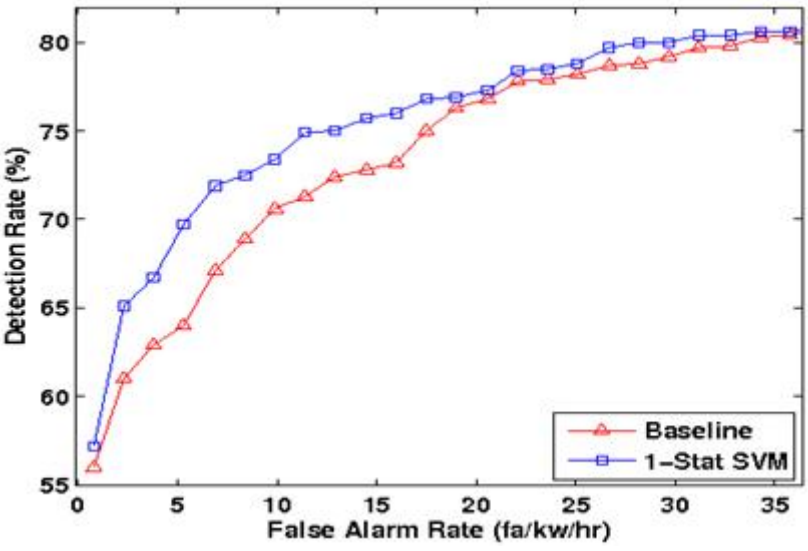

Figure 3

Software: The research programs for the tasks described above (isolated speech, connected speech and word spotting) can be obtained by contacting David Burshtein (burstyn@eng.tau.ac.il).

Related Journal Publications:

- A. Sloin and D. Burshtein, "Support Vector Machine Training for Improved Hidden Markov Modeling," IEEE Transactions on Signal Processing, vol. 56, no. 1, pp. 172-188, January 2008.
- A. Alfandary and D. Burshtein, "Improvements and Generalization of the SVM Re-Scoring Algorithm of Continuous HMMs", to be submitted to IEEE Transactions on Signal Processing.