AUTOMATIC B***** DETECTION

Anna Kruspe (kpe@idmt.fraunhofer.de)
Fraunhofer IDMT, Ilmenau, Germany

Matthias Mauch
Queen Mary University of London, UK

ABSTRACT

Lots of songs contain expletives in their lyrics. These expletives often need to be detected and removed or masked for airplay, which is being done manually. In this demo, we present an approach for automatically detecting such words.

We employ an approach for automatically aligning the textual lyrics to the audio. The lyrics can then be searched for expletives. With this information, these words can then be removed or covered.

We test three acoustic models for alignment. At a tolerance of one second, we obtain an accuracy of 92%.

1. MOTIVATION

Lots of song lyrics contain expletives. There are many scenarios in which it is necessary to know when these words occur, e.g. for airplay and for the protection of minors. In the case of airplay, they are commonly “bleeped” or acoustically removed. In this demonstration, we present an approach for finding such expletives automatically.

2. DATA SET

Our data set consists of 80 popular songs which were collected at Queen Mary University, most of them Hip Hop. 711 instances of 48 expletives were annotated on these songs. In addition, we manually retrieved the matching textual, unaligned lyrics from the internet.

3. APPROACH

We first considered a direct keyword spotting approach, but this did not generate sufficient results since most of the expletives only consist of 2 or 3 phonemes. Keyword spotting becomes notoriously hard for such short keywords [2].

Since textual lyrics are usually easily available on the internet, we therefore employed a new approach that utilizes those:

1. Automatically align textual lyrics to audio (as described in [3])
2. Search for pre-defined expletives in the result
3. If necessary remove those expletives. We tested a stereo subtraction approach, which works adequately for this case since the removed timespans are short. Alternatively, keywords can be asked with a bleep or similar.

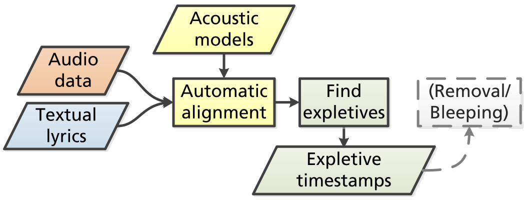
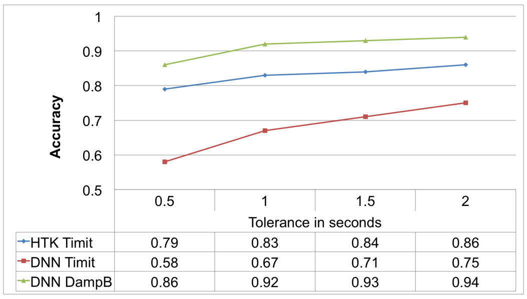


Figure 1. Data flow



	0.5	1	1.5	2
HTK Timit	0.79	0.83	0.84	0.86
DNN Timit	0.58	0.67	0.71	0.75
DNN DampB	0.86	0.92	0.93	0.94

Figure 2. Results at various tolerances.

The data flow is shown in figure 1.

4. EXPERIMENTS AND RESULTS

We tested the alignment using three different acoustic models:

- an Hidden Markov (HMM) acoustic model trained on the *Timit* speech corpus [1]
- a Deep Neural Network (DNN) also trained on *Timit*
- a DNN trained on a subset of the *Damp* singing data set [5] [3]

Accuracy was then calculated by evaluating how many of the annotated expletives were recognized at their correct timestamps (with various tolerances). The results are shown in figure 2.

5. FUTURE WORK

For this first attempt at detecting expletives in songs, we only tested two alignment approaches, but there are others that could provide better results. Whenever the alignment failed, it was mostly due to solo instruments. In order to remedy this, vocal detection (and possibly source separation) could be employed prior to alignment. Additionally, a more sophisticated removal approach could be implemented to remove the expletives.

Lyrics collected from the internet are often incorrect, e.g. because repetitions are not spelled out, because of spelling errors, or because they refer to different versions of a song. Our approach could be expanded to allow for some flexibility in this respect. At the moment, these lyrics need to be provided manually. In the future, those could be retrieved automatically (e.g. using the approach in [4]).

6. REFERENCES

[1] J. S. Garofolo et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus. Technical report, Linguistic Data Consortium, Philadelphia, 1993.

[2] A. M. Kruspe. Keyword spotting in a-capella singing. In *15th International Conference on Music Information Retrieval (ISMIR)*, Taipei, Taiwan, 2014.

[3] A. M. Kruspe. Bootstrapping a system for phoneme recognition and keyword spotting in unaccompanied singing. In *17th International Conference on Music Information Retrieval (ISMIR)*, New York, NY, USA, 2016.

[4] A. M. Kruspe. Retrieval of textual song lyrics from sung inputs. In *INTERSPEECH*, San Francisco, CA, USA, 2016.

[5] J. C. Smith. *Correlation analyses of encoded music performance*. PhD thesis, Stanford University, 2013.