

Fuzzy Audio Similarity Measures Based on Spectrum Histograms and Fluctuation Patterns

Klaas Bosteels, Etienne E. Kerre

*Fuzziness and Uncertainty Modelling Research Group
Department of Applied Mathematics and Computer Science
Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium
{Klaas.Bosteels, Etienne.Kerre}@UGent.be*

Abstract

Spectrum histograms and fluctuation patterns are representations of audio fragments. By comparing these representations, we can determine the similarity between the corresponding fragments. Traditionally, this is done using the Euclidian distance. We propose fuzzy similarity measures as an alternative. First we introduce some well-known fuzzy similarity measures, together with certain properties that can be desirable or useful in practice. In particular, we present several forms of restrictability, which allow to reduce the computation time in practical applications. Next, we show that fuzzy similarity measures can be used to compare spectrum histograms and fluctuation patterns. Finally, we describe some experimental observations for this fuzzy approach of constructing audio similarity measures.

1. Introduction

Recent portable audio players can store about 20 000 songs, and online music stores currently offer millions of tracks. This abundance of music drastically increases the need for applications that automatically analyse, retrieve or organize audio files. Measures that are able to express the similarity between two given audio fragments, are a fundamental component in many of these applications [1, 2, 6, 7, 8]. Such measures are usually constructed by following a feature-based approach. The audio fragments are represented by real-valued feature vectors, and the similarity is calculated by comparing these vectors. We consider two types of feature vectors: spectrum histograms and fluctuation patterns. So far, the Euclidian distance has always been used to compare feature vectors of these types. In this paper, however, we identify the feature vectors with fuzzy sets, such that fuzzy similarity measures can determine the similarity between them.

2. Preliminaries

2.1. Spectrum histograms

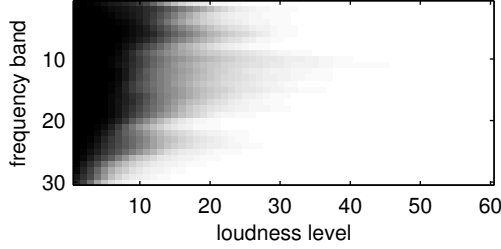
For a given audio segment, the Fourier transform can be used to calculate the amplitude that corresponds with each frequency. By dividing an audio fragment in short subsequent segments and applying the Fourier transform to each of these segments, we get the amplitude for each time-frequency pair. Such a representation of an audio fragment is called a spectrogram. The individual frequencies of a spectrogram are usually consolidated into frequency bands to reduce the computation time.

The amplitudes of the spectrogram can be converted to loudness values. Starting from these values, we can calculate the spectrum histogram (SH) [6, 7] by counting how many times certain loudness levels are reached or exceeded in each frequency band. In this way, we get a simple summarization of the spectral shape of the audio fragment. This summarization is, to some extent, related to the perceived timbre of the audio fragment.

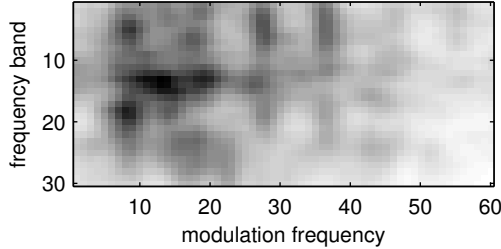
Our implementation of SHs was written in Matlab, using the MA toolbox [5]. Each SH is a matrix with 30 rows (frequency bands) and 60 columns (loudness levels). The bark scale is used for the frequencies, and the sone scale for the loudness values. We use these scales because they are psychoacoustical, i.e., they correspond with human perception. Figure 1(a) shows an example of a SH.

2.2. Fluctuation patterns

By applying the Fourier transform to the subsequent loudness values in each frequency band of the spectrogram, we get the amplitudes that correspond with the loudness modulation frequencies for each frequency band. A fluctuation pattern (FP) [6, 8] consists of weighted versions of these coefficients, based on the psychoacoustic model of



(a)



(b)

Figure 1. The SH (a) and FP (b) for a fragment of “Temps pour nous” by Axelle Red. White depicts zero, and black represents the maximum value.

fluctuation strength. Hence, fluctuation patterns describe the loudness fluctuations for each frequency band. This implies that they are, to some extent, related to the perceived rhythm. Although they are relatively simple, FPs can be very useful in practice. For instance, they were one of the representations used for the audio similarity measure that won the MIREX’06 (Music Information Retrieval eX-change) audio-based music similarity and retrieval task.¹

For implementing FPs, we again used the MA toolbox. The FPs are, like the SHs, 30 by 60 matrices in which the rows correspond with frequency bands. In this case, however, the columns represent modulation frequencies (ranging from 0 to 10 Hz). Figure 1(b) shows an example.

2.3. Fuzzy similarity measures

A fuzzy set A in a universe X is a $X \rightarrow [0, 1]$ mapping that associates with each element x from the universe X a degree of membership $A(x)$. We use the notation $\mathcal{F}(X)$ for the class of fuzzy sets in X . A fuzzy similarity measure M is a fuzzy set in $\mathcal{F}(X) \times \mathcal{F}(X)$ that expresses the similarity between each pair of fuzzy sets in X , i.e., the degree of membership $M(A, B)$ of $(A, B) \in \mathcal{F}(X) \times \mathcal{F}(X)$ denotes the similarity between A and B . If A and B are completely similar then $M(A, B) = 1$, otherwise $M(A, B) < 1$.

¹<http://www.music-ir.org/mirex2006>

We consider 12 well-known fuzzy similarity measures [3]. The first ones are based on the fact that fuzzy sets in a finite universe X can be represented as vectors in $[0, 1]^{|X|}$. The other measures are fuzzy generalizations of cardinality-based similarity measures for crisp sets.

We can determine how similar two vectors are by calculating the Minkowski distance between them: the smaller the distance, the greater the similarity. This observation leads to the following family of fuzzy similarity measures:

$$M_{1,n}(A, B) = 1 - \left(\frac{1}{|X|} \sum_{x \in X} |A(x) - B(x)|^n \right)^{\frac{1}{n}}$$

for all $A, B \in \mathcal{F}(X)$, with $n \in \mathbb{N} \setminus \{0\}$. An alternative approach is to use the cosine between the vectors as similarity measure:

$$M_2(A, B) = \frac{\sum_{x \in X} (A(x) \cdot B(x))}{\sqrt{(\sum_{x \in X} A(x)^2) (\sum_{x \in X} B(x)^2)}}$$

for all $A, B \in \mathcal{F}(X)$.

For two crisp sets A and B in a finite universe X , i.e., $A, B \in \mathcal{P}(X)$, we can use the following measures to determine their similarity:

$$M_3(A, B) = \frac{|A \cap B|}{|A|}$$

$$M_4(A, B) = \frac{|A \cap B|}{|B|}$$

$$M_5(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

$$M_6(A, B) = \frac{|A \cap B|}{\sqrt{|A| \cdot |B|}}$$

$$M_7(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

$$M_8(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

$$M_9(A, B) = \frac{|A \cap B|}{\max(|A|, |B|)}$$

$$M_{10}(A, B) = \frac{\min(|A|, |B|)}{|A \cup B|}$$

$$M_{11}(A, B) = \frac{\max(|A|, |B|)}{|A \cup B|}$$

$$M_{12}(A, B) = \frac{\min(|A|, |B|)}{\max(|A|, |B|)}$$

The classical set-theoretic operations that are used in these measures, can be generalized to fuzzy sets by means of triangular operations (t-operations). An increasing, associative and commutative $[0, 1]^2 \rightarrow [0, 1]$ mapping is called a triangular norm (t-norm) \mathcal{T} if it satisfies $\mathcal{T}(x, 1) = x$ for all

Only the first pages of this paper are presented here because of copyright restrictions. If you would like to obtain the full text, please e-mail to Klaas.Bosteels@UGent.be.