

Mel Frequency Cepstral Coefficient (MFCC)

Metoda MFCC wykorzystywana jest w systemach rozpoznawania mowy oraz klasyfikacji gatunków muzycznych. Jako parametry charakteryzujące dany gatunek stosuje się często średnią oraz odchylenie standardowe MFCC [1].

Wyznaczanie MFCC [2,3]:

1. podział sygnału dźwiękowego na ramki
2. dyskretna transformata Fourier'a (DFT)
3. wyznaczenie logarytmu z widma amplitudy
4. przeliczenie na skalę melową (mel scale)
5. dyskretna transformacja cosinusowa (DTC)

Sygnał dzielony jest na ramki o określonej długości, które częściowo mogą się na siebie nakładać [2]. Aby uniknąć zniekształceń wybierane są różne typy okien czasowych, w tym przypadku najczęściej wykorzystuje się okno Hamminga. Widmo amplitudowe dla każdej ramki wyznaczane jest przy pomocy dyskretniej transformaty Fourier'a.

DFT przekształca ciąg próbek sygnału na w ciąg harmonicznym zgodnie ze wzorem (1).

$$A_k = \sum_{n=0}^{N-1} a_n w_N^{-kn} \quad (1)$$

gdzie:

$$w_N = e^{i\frac{2\pi}{N}}$$

$$k \in [0, N - 1]$$

k - numer harmonicznym

A_k - wartość k -tej harmonicznym

n - numer próbki sygnału

a_n - wartość n -tej próbki sygnału

N - liczba próbek

Następnym krokiem jest wyznaczenie logarytmu. Ze względu na to, że ludzkie ucho jest lepiej przystosowane do rozpoznawania zmian w zakresie niskich niż wysokich częstotliwości [4], wprowadzona została skala melowa, która wydaje się bardziej naturalna. Oparta jest ona na związku pomiędzy częstotliwością czystego tonu harmonicznego, a częstotliwością postrzeganą przez człowieka.

Zależność pomiędzy częstotliwością sygnału, a wartością tej częstotliwości wyrażonej w skali melowej opisana jest wzorem (2).

$$f_m = 2595 \cdot \log_{10} \left(1 + \frac{f}{100} \right) \quad (2) [2]$$

gdzie:

f - częstotliwość sygnału

f_m - częstotliwość sygnału wyrażona w skali melowej

Aby wyznaczyć współczynniki cepstralne należy wykonać odwrotną szybką transformatę Fouriera (IFFT). Ze względu na to, że widmo jest w skali logarytmicznej, symetryczne oraz rzeczywiste zadanie to sprowadza się do zastosowania dyskretnej transformacji cosinusowej. Współczynniki cepstralne wyznaczone przy pomocy DTC opisane są zależnością (3).

$$C_n = \sqrt{\frac{2}{N}} \sum_{i=1}^N \log(S_i) \cdot \cos \left[\frac{\pi n}{N} \left(i - \frac{1}{2} \right) \right] \quad (3) [5]$$

gdzie:

C_n - n -ty współczynnik cepstralny

S_i - i -ty współczynnik otrzymany po przetworzeniu sygnału przez filtry

N - ilość filtrów

Analiza współczynników cepstralnych pozwala na ocenę zależności pomiędzy częstotliwościami składowych spektralnych w sygnale [5].

Najczęściej stosuje się 26 filtrów melowych. Większa ilość nie prowadzi do znaczącej poprawy jakości rozpoznawania [6].

Bibliografia

1. Chang-Hsing Lee, Jau-Ling Shih, Kun-Ming Yu, Jung-Mau Su – *Automatic music genre classification using modulation spectral contrast feature*
2. Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara, Sanjoy Kumar Saha - *Music Classification based on MFCC Variants and Amplitude Variation Pattern: A Hierarchical Approach*
3. Tom LH. Li and Antoni B. Chan – Genre classification and the invariance of MFCC features to Key and Tempo
4. <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>
5. <http://winntbg.bg.agh.edu.pl/rozprawy/9873/full9873.pdf>
6. <http://winntbg.bg.agh.edu.pl/rozprawy2/10009/full10009.pdf>