Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task

Todor Ganchev, Nikos Fakotakis, George Kokkinakis

Wire Communications Laboratory,
University of Patras, 26500 Rion-Patras, Greece
tganchev@wcl.ee.upatras.gr

Abstract

Making no claim of being exhaustive, a review of the most popular MFCC (Mel Frequency Cepstral Coefficients) implementations is made. These differ mainly in the particular approximation of the nonlinear pitch perception of human, the filter bank design, and the compression of the filter bank output. Then, a comparative evaluation of the presented implementations is performed on the task of text-independent speaker verification, by means of the well-known 2001 NIST SRE (speaker recognition evaluation) one-speaker detection database.

1. Introduction

The quest for better speech parameterization led to various speech features, which were reported to provide advantage in specific conditions and applications. Moreover, for some speech features, such as the well-known and widely-used MFCC, multiple implementations were developed. These implementations differ mainly in the number of filters, the shape of the filters, the way the filters are spaced, the bandwidth of the filters, and the manner in which the spectrum is warped. In addition, the frequency range of interest, the selection of actual subset and the number of MFCC coefficients employed in the classification can be also different.

Although there are a number of studies [1÷3] that compare various implementations of the MFCC on the speech recognition task, up to the authors' present knowledge no such study has been performed on the task of speaker recognition. Since the speech and speaker recognition tasks exploit different aspects of the speech signal, we deem it worthy to carry out such a study. Therefore, employing a text-independent speaker verification system, we perform a comparative evaluation of the following implementations:

- MFCC FB-20 – introduced in 1980 by Davis and Mermelstein [4]; Davis and Mermelstein assume sampling frequency of 10 kHz; speech bandwidth [0, 4600] Hz.
- MFCC FB-24 HTK – from the Cambridge HMM Toolkit (HTK) described in Young, 1995 [5]; Young uses a filter bank of 24 filters for speech bandwidth [0, 8000] Hz (sampling rate $\geq$ 16 kHz).
- MFCC FB-40 – from the Auditory Toolbox for MATLAB [6] written by Slaney in 1998; Slaney assumes sampling rate of 16 kHz, and speech bandwidth [133, 6854] Hz.
- HFCC-E FB-29 (Human Factor Cepstral Coefficients) of Skowronski and Harris, 2004 [3]; Skowronski and Harris assume sampling rate of 12.5 kHz and speech bandwidth [0, 6250] Hz.

The abbreviation FB-nn (Filter Bank), which we stick after the designation MFCC (HFCC), provides information about the number of filters in the filter bank as described by the corresponding authors. Since these implementations assume different sampling rates (and different bandwidth of the speech signal) they are not directly comparable. To solve that discrepancy, keeping the filter spacing and filter bandwidth as proposed in the original description of these implementations, we reduce the number of filters to adapt it to sampling frequency of 8 kHz. Sampling frequency of 8 kHz is common for all telephone driven services, and thus, it is default for the contemporary real-world speaker recognition corpora (for instance: Switchboard, NIST SRE data [9], etc).

2. Implementations of the MFCC parameters

Following the introduction of the MFCC [4], numerous variations and improvements of the original idea were proposed. One of the main reasons for such diversity of implementations is the desire of researchers to follow the progress made in the area of psychoacoustics during the years. For instance, let's consider the various approximations of the nonlinear pitch perception by the human auditory system. An early approximation, referred to as Koenig scale is exactly linear below 1000 Hz and logarithmic above 1000 Hz. It provides a computationally inexpensive representation of the Mel scale, which however is not very precise and significantly deviates from the original scale for frequencies both lower and higher than 1000 Hz. A more precise approximation, suggested by Fant, is:

$$\hat{f}_{mel} = k_{const} \cdot \log_n \left(1 + \frac{f_{lin}}{F_b} \right), \tag{1}$$

where $F_b = 1000.$ A specific form of (1), presented in [7]:

$$\hat{f}_{mel} = \frac{1000}{\log_n 2} \cdot \log_n \left(1 + \frac{f_{lin}}{1000} \right) \tag{2}$$

was found to provide a more close approximation of the Mel scale (only for the frequency range of [0, 5] kHz), when compared with the approximation offered by the Koenig scale. In addition, the formulation (2) is particularly interesting since the values of $\hat{f}_{mel}$ remain unaffected by the choice of the base n of the logarithm. Other approximations of the Mel scale that were derived from (1) make use of natural or decimal logarithm, which leads to different choice of the constant k_{const}. The following two representations:

$$\hat{f}_{mel} = 2595 \cdot \log_{10} \left(1 + \frac{f_{lin}}{700} \right) \tag{3}$$

$$\hat{f}_{mel} = 1127 \cdot \ln \left(1 + \frac{f_{lin}}{700} \right) \tag{4}$$

are widely used in the various implementations of the MFCC. The formulae (3) and (4), when compared to (2), provide a closer approximation of the Mel scale for frequencies below 1000 Hz, at the price of higher inaccuracy for frequencies higher than 1000 Hz.

2.1. The original MFCC FB-20

In the paradigm introduced by Davis and Mermelstein, 1980, [4] the novel MFCC were designed as a set of discrete cosine transform decorrelated parameters, which were computed through a transformation of the logarithmically compressed filter-output energies. These energies were derived through a perceptually spaced bank of twenty equal height triangular filters that are applied on the Discrete Fourier Transform (DFT)-ed speech signal. In brief, given N-point DFT of the discrete input signal $x(n)$,

$$X(k)=\sum_{n=0}^{N-1} x(n)\cdot\exp\left(\frac{-j2\pi nk}{N}\right),\quad k=0,1,\ldots,N-1,\quad (5)$$

a filter bank with M equal height triangular filters is constructed. Each of these M equal height filters is defined as:

$$H_i(k)=\begin{cases} 0 & \text{for} \quad k<f_{b_{i-1}} \\ \dfrac{\left(k-f_{b_{i-1}}\right)}{\left(f_{b_i}-f_{b_{i-1}}\right)} & \text{for} \quad f_{b_{i-1}}\le k\le f_{b_i} \\ \dfrac{\left(f_{b_{i+1}}-k\right)}{\left(f_{b_{i+1}}-f_{b_i}\right)} & \text{for} \quad f_{b_i}\le k\le f_{b_{i+1}} \\ 0 & \text{for} \quad k>f_{b_{i+1}} \end{cases},\quad i=1,2,\ldots,M \quad (6)$$

where i stands for the i-th filter, f_{b_i} are the boundary points of the filters, and $k=1,2,\ldots,N$ corresponds to the k-th coefficient of the N-point DFT. The boundary points f_{b_i} are expressed in terms of position, which depends on the sampling frequency F_s and the number of points N in the DFT:

$$f_{b_i}=\left(\frac{N}{F_s}\right)\cdot\hat{f}_{mel}^{-1}\left(\hat{f}_{mel}\left(f_{low}\right)+i\cdot\frac{\hat{f}_{mel}\left(f_{high}\right)-\hat{f}_{mel}\left(f_{low}\right)}{M+1}\right).\quad (7)$$

Here, the function $\hat{f}_{mel}(.)$ states the transformation (4), f_{low} and f_{high} are respectively the low and high boundary frequencies for the entire filter bank, M is the number of filters, and $\hat{f}_{mel}^{-1}$ is the inverse to (4) transformation, formulated as:

$$\hat{f}_{mel}^{-1}=f_{lin}=700\cdot\left[\exp\left(\frac{\hat{f}_{mel}}{1127}\right)-1\right].\quad (8)$$

Here, and everywhere next, the sampling frequency F_s, and the frequencies f_{low}, f_{high}, and f_{lin}, are in Hz, and the $\hat{f}_{mel}$ is in mels. Equation (7) guarantees that the boundary points of the filters are uniformly spaced in the Mel scale. The endpoints of each one of the triangular filters are determined by the centre frequencies of its adjacent filters. Therefore, the bandwidth of the filters is not an independent variable.

The filter bank of Davis and Mermelstein is comprised of twenty equal height filters which cover the frequency range $[0,4600]$ Hz. Unlike (7), the centre frequencies of the first ten filters are linearly spaced between 100 Hz and 1000 Hz, and the next ten have centre frequencies logarithmically spaced between 1000 Hz and 4000 Hz. The choice of centre frequency f_{c_i} for the i-th filter can be approximated [3] as:

$$f_{c_i}=\begin{cases}100\cdot i, & i=1,\ldots,10 \\ f_{c_{10}}\cdot 2^{0.2(i-10)}, & i=11,\ldots,20\end{cases}\quad (9)$$

where the centre frequency f_{c_i} is assumed in Hz.

Having the filter bank constructed, the MFCC parameters are computed [4], as:

$$C_j=\sum_{i=1}^{M} X_i\cdot\cos\left(j\cdot(i-1/2)\cdot\frac{\pi}{M}\right),\ \text{with}\ j=1,2,\ldots,J,\quad (10)$$

where M is the number of filters in the filter bank, J is the number of cepstral coefficients which are computed (usually $J<M$), and X_i is formulated as the "log-energy output of the i-th filter" [4]. Here, the "log-energy output of the i-th filter" is understood as:

$$X_i=\log_{10}\left(\sum_{k=0}^{N-1}\left|X(k)\right|\cdot H_i(k)\right),\qquad i=1,2,\ldots,M.\quad (11)$$

The log-energy output X_i of each filter is derived through the magnitude spectrum (5) and filter bank (6). It has to be specified here that since X_i is derived through the magnitude spectrum, and not through the power spectrum, it does not comply with the Parseval's definition of energy as sum of squared terms. Nevertheless, this definition of energy is used in most of the MFCC implementations.

2.2. The HTK MFCC-FB24

Another widely-used implementation of the MFCC was provided in the framework of the Cambridge Hidden Markov Models (HMM) Toolkit [5], known as HTK. The designation HTK MFCC FB-24 reflects the number of filters M=24 recommended by Young for speech bandwidth of 8 kHz.

The HTK MFCC FB-24 makes use of the definition (3) of the Mel frequency. In this implementation, the limits of the frequency range are the parameters that define the basis for the filter bank design. Specifically, the lower and the higher boundaries of the frequency range of the entire filter bank, $\hat{f}_{low}$ and $\hat{f}_{high}$ respectively, determine the computation of the unit interval $\Delta\hat{f}$:

$$\Delta\hat{f}=\frac{\hat{f}_{high}-\hat{f}_{low}}{M+1},\quad (12)$$

which serves as footstep in the definition of the centre frequencies of the individual filters. The centre frequency $\hat{f}_{c_i}$ of the i-th filter is given by:

$$\hat{f}_{c_i}=\hat{f}_{low}+i\cdot\Delta\hat{f},\quad i=1,\ldots,M-1,\quad (13)$$

where M is the total number of filters in the filter bank. The conversion of the centre frequencies of the filters to linear frequency (Hz) is given by:

$$f_{c_i}=700\cdot\left(10^{\hat{f}_{c_i}/2595}-1\right).\quad (14)$$

In HTK, similarly to the filter bank of the original MFCC FB-20 [4], a filter bank of equal height filters is used. The shape of the individual triangular filters is defined by (6).

The HTK MFCC FB-24 parameters are computed as follows: The DFT $X(k)$ (5), computed for the discrete input signal $x(n)$, is used for computing the magnitude spectrum $\left|X(k)\right|$, which acts as input for the filter bank $H_i(k)$ (6). Next, the filter bank output is logarithmically compressed:

$$X_i=\ln\left(\sum_{k=0}^{N-1}\left|X(k)\right|\cdot H_i(k)\right),\quad (15)$$

and then decorrelated by the DCT (10) to provide the HTK MFCC FB-24 parameters.

2.3. The MFCC FB-40

The MFCC FB-40 speech features were described in the Slaney's Auditory Toolbox [6]. Assuming sampling frequency 16 kHz, Slaney implemented a filter bank of 40 equal area filters, which cover the frequency range $[133,\ 6854]\,\mathrm{Hz}$. The centre frequencies of the first 13 of them are linearly spaced in the range $[200,\ 1000]$ Hz with a step of 66.67 Hz and the ones of the next 27 are logarithmically spaced in the range [1071, 6400] Hz with a step $logStep = 1.0711703$, computed as:

$$logStep = \exp\left(\ln\left(\frac{f_{c_{40}}}{1000}\right)\bigg/ numLogFilt\right). \qquad (16)$$

Here $f_{c_{40}} = 6400\ \mathrm{Hz}$ is the centre frequency of the last of the logarithmically spaced filters, and $numLogFilt = 27$ is the number of logarithmically spaced filters. Each one of these equal area triangular filters is defined as:

$$H_i(k) = \begin{cases} 0 & \text{for} \quad k < f_{b_{i-1}} \\ \dfrac{2\left(k - f_{b_{i-1}}\right)}{\left(f_{b_i} - f_{b_{i-1}}\right)\left(f_{b_{i+1}} - f_{b_{i-1}}\right)} & \text{for} \quad f_{b_{i-1}} \le k \le f_{b_i} \\ \dfrac{2\left(f_{b_{i+1}} - k\right)}{\left(f_{b_{i+1}} - f_{b_i}\right)\left(f_{b_{i+1}} - f_{b_{i-1}}\right)} & \text{for} \quad f_{b_i} \le k \le f_{b_{i+1}} \\ 0 & \text{for} \quad k > f_{b_{i+1}} \end{cases}, \qquad (17)$$

where $i = 1,2,\ldots,M$ stands for the i-th filter, f_{b_i} are $M+2$ boundary points that specify the M filters, and $k = 1,2,\ldots,N$ corresponds to the k-th coefficient of the N-point DFT. The boundary points f_{b_i} are expressed in terms of position, as specified above. The key to equalization of the area below the filters (17) lies in the term:

$$\frac{2}{\left(f_{b_{i+1}} - f_{b_{i-1}}\right)}. \qquad (18)$$

Due to the term (18), the filter bank (17) is normalized in such a way that the sum of coefficients for every filter equals one. Thus, the i-th filter satisfies:

$$\sum\nolimits_{k=1}^{N} H_i(k) = 1, \ \text{for} \ \ i = 1,2,\ldots,M \qquad (19)$$

Next, the equal area filter bank (17) is employed in the computation of the log-energy output (11). Finally, the DCT (10) provides the MFCC-FB40 parameters.

2.4. The HFCC-E FB-29

The Human Factor Cepstral Coefficients (HFCC) introduced in 2004 by Skowronski and Harris [3], provide the most recent update of the MFCC filter bank. Assuming sampling frequency of 12.5 kHz Skowronski and Harris proposed the HFCC-E filter bank composed of 29 Mel-warped equal height filters, which cover the frequency range [0, 6250] Hz. The most significant difference between the HFCC, and the earlier MFCC, is that in HFCC-E the filter bandwidth is decoupled from the filter spacing. Specifically, the filter bandwidth in the HFCC-E is derived from the equivalent rectangular bandwidth (ERB) introduced by Moore and Glasberg [8]:

$$ERB = 6.23 \cdot 10^{-6} \cdot f_c^2 + 93.39 \cdot 10^{-3} \cdot f_c + 28.52, \quad (20)$$

where f_c is the centre frequency of the individual filters in Hz. The filter bandwidth (20) is further scaled by a constant, which Skowronski and Harris labeled as E-factor.

In brief, the HFCC filter bank design [3] consists of the following steps: First the low f_{low} and high f_{high} boundaries of the entire filter bank and the number M of filters are chosen. The centre frequencies f_{c_1} and f_{c_M} of the first and the last of the filters, respectively, are computed as:

$$f_{c_i} = \frac{1}{2}\cdot\left(-\bar{b} + \sqrt{\bar{b}^2 - 4\cdot\bar{c}}\right), \qquad (21)$$

where the index i is either 1 or M, and $\bar{b}, \bar{c}$ defined as:

$$\bar{b} = \frac{b - \hat{b}}{a - \hat{a}} \quad \text{and} \quad \bar{c} = \frac{c - \hat{c}}{a - \hat{a}} \qquad (22)$$

receive different values for the two cases. The values $a,\ b,\ c$ are these from (20): $6.23\cdot10^{-6}$, $93.39\cdot10^{-3}$, 28.52, respectively. For the first filter, the values of the coefficients $\hat{a},\ \hat{b},\ \hat{c}$ are computed as:

$$\hat{a} = \frac{1}{2}\cdot\frac{1}{700+f_{low}},\ \hat{b} = \frac{700}{700+f_{low}},\ \hat{c} = -\frac{f_{low}}{2}\cdot\left(1+\frac{700}{700+f_{low}}\right). (23)$$

For the last filter these are:

$$\hat{a} = -\frac{1}{2}\cdot\frac{1}{700+f_{high}},\ \hat{b} = -\frac{700}{700+f_{high}},\ \hat{c} = \frac{f_{high}}{2}\cdot\left(1+\frac{700}{700+f_{high}}\right) (24)$$

Once the centre frequencies of the first and the last filter are computed, the centre frequencies of the filters situated between them are easily calculated since they are equidistant on the Mel-scale. The step $\Delta\hat{f}$ between the centre frequencies of adjacent filters is computed as:

$$\Delta\hat{f} = \frac{\hat{f}_{c_M} - \hat{f}_{c_1}}{M-1} \qquad (25)$$

where all the frequencies are in mels. The conversions $f_{c_1} \to \hat{f}_{c_1}$ and $f_{c_M} \to \hat{f}_{c_M}$ are given by (3). Having $\Delta\hat{f}$, the centre frequencies $\hat{f}_{c_i}$ are computed as:

$$\hat{f}_{c_i} = \hat{f}_{c_1} + (i-1)\cdot\Delta\hat{f}, \quad \text{for}\ i = 2,\ldots,M-1. \qquad (26)$$

Next, through (14), the reverse transformation $\hat{f}_{c_i} \to f_{c_i}$ is performed, and through (20) the ERB_i for each f_{c_i} is computed. Finally, the low and high frequencies f_{low_i} and f_{high_i}, respectively, of the i-th filter are derived through:

$$f_{low_i} = -(700 + ERB_i) + \sqrt{(700 + ERB_i)^2 + f_{c_i}(f_{c_i} + 1400)} \quad (27)$$

$$f_{high_i} = f_{low_i} + 2\cdot ERB_i. \qquad (28)$$

With all parameters computed through $(21) \div (28)$, the design of the HFCC-E filter bank is completed.

Finally, as in the MFCC FB-20 of Davis and Mermelstein, the log-energy filter bank outputs are computed (11), and then (10) is applied to decorrelate the HFCC-E FB29 parameters.

3. Experiments and results

The MFCC implementations outlined in Section 2 were evaluated on the 2001 NIST SRE database by means of the PNN-based text-independent speaker verification system [10]. A common protocol was followed in all experiments according to the rules described in the 2001 NIST SRE Plan [9]. In brief, approximately 40 seconds of voiced speech were detected (in two-minute recordings) for training the target models. The common reference model was created by exploiting the male training speech available in the 2002 NIST SRE database. Approximately one hour and forty minutes of voiced speech was available for that purpose. After training, the user models were tested carrying out all male trials as defined in the complete one-speaker detection task. Each experiment comprised 850 target and 8500 impostor trials with a duration from 0 to 60 seconds of speech.

To accommodate to sampling rate of 8 kHz, we have excluded from all filter banks the filters which spread beyond the 4 kHz border. Thus, in the experiments with the MFCC FB-20 of Davis and Mermelstein we have used 19 filters – ten with linearly spaced centre frequencies and nine with logarithmically spaced ones. Following the instructions in [5], we used a filter bank of 20 filters for computing the HTK MFCC FB-24 features. In the experiment with the Slaney's MFCC FB-40, we kept the first 32 filters, which cover the frequency range [133, 3954] Hz. Finally, in the experiment with the HFCC-E FB-29 (using E-factor E=1) we tested various number of filters (19, 24, 29) to cover the frequency range of [0, 4000] Hz. In all experiments, the full number of cepstral coefficients, except the first one, was employed. Cepstral mean subtraction and dynamic range normalization were used for all speech features.

Table 1 presents the experimental results. As it was expected, there is no significant difference among the results for the MFCC FB-24 HTK, Slaney's MFCC FB-40, and the HFCC FB-29 (with 29 filters in the range [0, 4000] Hz). Next, the MFCC FB-20 of Davis and Mermelstein performed slightly worse, and finally, the HFCC-E FB-29 features with 24 and 19 filters provided the highest Equal Error Rate (EER). Assuming a filter bank of 24 filters for the frequency range [0, 4000] Hz, Skowronski and Harris [3] suggested 29 filters for the frequency range [0, 6250] Hz. However, the speaker verification results demonstrated that 29 filters (in the frequency range [0, 4000] Hz) provide lower EER than 24 or 19 filters. We deem the reason for this is (at least in part) in the irrelevant overlapping between the first few filters in the HFCC-E filter bank, especially when the number of filters is low. This results in a bad frequency resolution at low frequencies. In addition, examining the results for MFCC FB-40 and HFCC FB-29, it seems that more filters in the filter bank provide a better speaker differentiation. The only exception here is the result of HTK MFCC FB-24 – apparently, other factors influence the speaker verification performance as well.

To study the importance of the E-factor we experimented with various values. In experiments with a filter bank of 24 filters for the frequency range [0, 4000] Hz, it was observed that E=1 provides the lowest EER. Table 2 presents results for the best HFCC-E FB-29 – with 29 filters in the frequency range [0, 4000] Hz. For this filter bank it was found that E=0.5 provides the lowest EER. Deviating from E=0.5 in either direction increases the EER. We deem the reason is that for lower values of the E-factor, the filters with the lowest centre frequencies barely overlap, and thus the filter bank resolution for these frequencies is low – threshold phenomena were observed. For higher values of E, the filters are very broad and thus smooth some details in the spectrum which are important for speaker differentiation. In addition, in the HFCC-E scheme the filters with highest centre frequencies overlap widely. Each filter overlaps not only with its immediate neighbours but also with more distant ones. This was reported useful for speech recognition [3], but does not favour the speaker recognition task. The MFCC FB-40 and MFCC FB-24 HTK were found to provide the lowest decision cost.

Table 1. The Equal Error Rate (EER) and normalized optimal Decision Cost Function (DCFopt) for various MFCC implementations

Speech Features	# filters for [0, 4000] Hz	*DCFopt*	EER [%]
MFCC FB-20 D&M	19	0.554	14.00%
MFCC FB-24 HTK	20	0.538	13.76%
MFCC FB-40 Slaney	**32**	**0.541**	**13.65%**
HFCC-E FB-29 S&H	19	0.638	15.41%
HFCC-E FB-29 S&H	24	0.640	14.71%
HFCC-E FB-29 S&H	29	0.592	13.65%

Table 2. The Equal Error Rate (EER) and normalized optimal Decision Cost Function (*DCFopt*) for HFCC-E FB-29 with different E-factors

Speech Features	E factor	*DCFopt*	EER [%]
HFCC-E FB-29	E=0.25	0.604	14.00%
HFCC-E FB-29	E=0.35	0.589	13.77%
HFCC-E FB-29	**E=0.50**	**0.567**	**13.06%**
HFCC-E FB-29	E=1.00	0.592	13.65%
HFCC-E FB-29	E=1.50	0.585	14.12%
HFCC-E FB-29	E=2.00	0.609	14.45%

4. Conclusions

Comparative evaluation of various MFCC implementations was performed. As expected, the speaker verification performance did not vary vastly when different approximations of the non-linear pitch perception of human were used. However, some observations suggest that regardless of the specific filter bank design, a larger number of filters favours the speaker detection performance. Beside the number of filters in the filter bank, the overlapping among the neighbouring filters also proved a sensitive parameter. Increase or decrease of the overlapping beyond a given range increases the error rates.

5. References

[1] Zheng F., Zhang, G., Song, Z., "Comparison of different implementations of MFCC", *J. Computer Science & Technology*, 16(6):582-589, Sept. 2001.

[2] Shannon B.J., Paliwal K.K., "A comparative study of filter bank spacing for speech recognition", *Proc. of Microelectronic engineering research conference*, Brisbane, Australia, Nov. 2003.

[3] Skowronski, M.D., Harris, J.G., "Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition", *Journal of the Acoustical Society of America*, 116(3):1774–1780, Sept. 2004.

[4] Davis, S.B., Mermelstein, P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", *IEEE Trans. on Acoustic, Speech and Signal Processing*, 28(4):357–366, 1980.

[5] Young, S.J., Odell, J., Ollason, D., Valtchev, V., Woodland, P., "The HTK Book. Version 2.1", *Department of Engineering, Cambridge University*, UK, 1995.

[6] Slaney M. "Auditory Toolbox. Version 2", *Technical Report #1998-010*, Interval Research Corporation, 1998.

[7] Fant, G. Speech Sounds and Features. *The MIT Press*, Cambridge, MA, USA, 1973.

[8] Moore, B.C.J., Glasberg, B.R. "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", *Journal of the Acoustical Society of America,* 74(3):750–753, 1983.

[9] "The NIST Year 2001 Speaker Recognition Evaluation Plan", The NIST of USA, 2001. Available: http://www.nist.gov/speech/tests/spk/2001/doc/2001-spkrec-evalplan-v05.9.pdf.

[10] Ganchev, T., Fakotakis, N., and Kokkinakis, G., "Text-Independent Speaker verification Based on Probabilistic Neural Networks", *Proc. of Acoustics*, Patras, Greece, 2002. 159-166.