Performance Evaluation of Mel and Bark Scale based Features for Text-Independent Speaker Identification

S. B. Dhome, Amol A. Chaudhari, M. P. Gajare

Abstract: The performance of Mel scale and Bark scale is evaluated for text-independent speaker identification system. Mel scale and Bark scale are designed according to human auditory system. The filter bank structure is defined using Mel and Bark scales for speech and speaker recognition systems to extract speaker specific speech features. In this work, performance of Mel scale and Bark scale is evaluated for text-independent speaker identification system. It is found that Bark scale centre frequencies are more effective than Mel scale centre frequencies in case of Indian dialect speaker databases. Mel scale is defined as per interpretation of pitch by human ear and Bark scale is based on critical band selectivity at which loudness becomes significantly different. The recognition rate achieved using Bark scale filter bank is 96% for AISSMSIOIT database and 95% for Marathi database.

Keywords: Formants, MFCC, Text-independence, VQ

I. INTRODUCTION

A filter bank model and LPC model are widely used as signal processing front end tools for speech processing. The advantage of using filter banks is that they can be designed to cover an analysis band from 150 Hz to 7000 Hz. In a given frequency band, energy of speech signal is measured using filter bank analyser. Filter banks based on nonuniform spacing have been used in many practical systems [1]. Nonuniform filter banks are used to reduce overall computation and to characterize the speech spectrum in a manner considered more consistent with human perception [1]. The filters can be spaced along a logarithmic frequency scale which is often justified from a human auditory perception. Alternatively, the critical band scale can be used directly for designing of a nonuniform filter bank. The filters are placed according to perceptual studies for the intension of choosing bands that give equal contribution to speech articulation.

Mel scale is used to define Mel filter bank in the computation of MFCCs for speaker recognition system. MFCC feature extraction is widely used speaker recognition systems [2, 3]. In recent years, there have been efforts for the robust speaker identification system. Robustness of speaker identification system mainly depends upon the features extracted from speech signal. MFCC features are modified in some ways either by concatenating them with complementary information or modifying the pre-processing steps of speech signal. Multitaper MFCC features [4, 5], Inverted MFCC features [6], Combining MFCC with phase information [7] and even replacing the Mel filter bank by Radon based features [3] or wavelet based features [8 - 12] have been extensively studied.

However, the frequency warping scales such as Bark scale and ERB scale are less studied in speaker recognition systems. The critical band i.e. bandwidth at which subjective response such as loudness perceived becomes significant can be approximated by Bark scale [13]. The use of Bark scale for audio signal processing have been studied in [14] [15]. Also, performance of Bark frequency cepstral coefficients (BFCC) has been studied for speech recognition in [16].

In this paper, the frequency scales Mel and Bark are studied and their performance is evaluated in speaker identification system.

II. FREQUENCY WARPING SCALES

A. Mel scale

According to psychoacoustics studies, the content of frequency in pure tones or speech signal perceived by human ear follows a nonlinear scale. The understandings from human auditory system are used to define subjective pitch of pure tones or speech signal [13]. Mel scale is used to measure a subjective pitch. When stimulus frequency is increased linearly, the subjective pitch in mels increases less rapidly [13]. Mel-filter bank represents different perceptual effects at different frequency bands. Mel scale is linear below 1 kHz and logarithmic above 1 kHz [13]. As shown in figure 1, the edges of filters are placed such that they coincide with the centre frequencies in adjacent filters. The following equation represents the formula for conversion of Linear to Mel frequencies.

\[ F_{\text{mel}} = 2595 \times \log_{10} \left(1 + \frac{f_{\text{linear}}}{700}\right) \] (1)
B. Bark scale

Bark scale is psychoacoustical scale which is based upon loudness measurement. The scale has values ranging from 1 to 24. These 24 values relates to 24 critical bands of hearing. In Bark scale, the distance from the centre to left edge is different from that from the centre to the right edge. The band edges and band centres are given in [14]. The interpretation of the centre-frequencies and bandwidths are discussed in [14]. Following equation shows relationship between frequencies in Hz and Bark scale frequencies

$$\omega = 6 \ln \left( \frac{\omega}{1200 \pi} + \left[ \left( \frac{\omega}{1200 \pi} \right)^2 + 1 \right]^{0.5} \right)$$

(2)

where, $\Omega$ represents the angular frequency in Bark scale, and $\omega$ represents angular linear frequency = $2\pi f$

Apart from equation 2, many analytical formulae have been proposed for Bark scale [17] [18].

$$bark = 13 \tan^{-1} \left( \frac{0.76 F(\omega)}{1000} \right) + 3.5 \tan^{-1} \left( \frac{F(\omega)}{1200} \right)^2$$

(3)

$$Critical\ Bark\ Rate = \frac{\left[ \frac{26 \pi F}{1000} \right]}{\left[ \frac{1200 \pi f}{1000} \right]} = 0.53$$

(4)

$$F(bark) = 6 \sinh^{-1} \left( \frac{F(\omega)}{600} \right)$$

(5)

The important constants required to specify to define Bark filter bank are the number of filters, the minimum frequency, and maximum frequency [17]. The frequency range for these filters will be specified by minimum and maximum frequencies. As it can be seen from figure 1 and 2, Bark scale has high bandwidth as compared to mel scale particularly at higher frequency.
III. EXPERIMENTAL SET-UP

The experiments have been evaluated on our recorded database AISSMSIOIT database [19] of 75 speakers. Another databases used for speaker identification experiments are TIMIT [20] consisting of 630 speakers and subset of 120 speakers and Marathi database [21] consisting of 120 speakers (60 male and 60 female). A speaker model is trained using five SX and three SI sentences (almost twenty-four seconds of duration) from TIMIT database and remaining two SA sentences (each of almost three seconds of duration) are used for testing phase. In Marathi database, a speaker model is trained using ten sentences approximately three seconds of duration and for testing two sentences roughly each of duration three seconds are used. The pre-processing steps of speech signal are carried out as mentioned in [19]. In framing step 256 samples per frame in case of TIMIT database, 512 samples per frame for Marathi database and 1024 samples per frame for AISSMSIOIT database are carried out along with 50% overlap and hamming windowing. Twelve MFCCs are computed by changing number of MFCC filters and speaker model is created using LBG algorithm (VQ). The testing procedure is carried out as mentioned in [19].The same procedure is carried out to calculate Bark scale in which as increment in the filter number increases the critical band selection are primary for Bark scale. Mel scale is more suitable for pitch perception and phonetic approaches whereas, Bark scale is more suitable for mel scale. This is because Bark filters are effective than Mel filters in case of formant F1. For speaker recognition, the formant F1 is important as it brings the maximum energy along with it which reveals the language structure and speaking style [22]. It is observed that formants in voiced part has more closeness with bark centre frequencies than mel centre frequencies. This is observed specially in case of formant F1. Speaker recognition, the formant F1 is important as it brings the maximum energy along with it which reveals the language structure and speaking style [22]. It is observed that formants in voiced part has more closeness with bark centre frequencies than mel centre frequencies. This is observed specially in case of formant F1.
Table- V: Location of formants with respect to Mel centre frequency and Bark centre frequency for AISSMISOIT Database

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Speaker Number</th>
<th>Formant F1</th>
<th>Mel centre Frequency</th>
<th>Bark centre Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>558</td>
<td>77.88</td>
<td>77.52</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>521</td>
<td>164.43</td>
<td>152.74</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>479</td>
<td>260.60</td>
<td>226.76</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>575</td>
<td>367.47</td>
<td>300.31</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>407</td>
<td>486.23</td>
<td>376.49</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>463</td>
<td>618.21</td>
<td>457.99</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>501</td>
<td>764.88</td>
<td>545.38</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>523</td>
<td>927.85</td>
<td>639.32</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>544</td>
<td>1109</td>
<td>740.58</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>618</td>
<td>1310</td>
<td>850.04</td>
</tr>
<tr>
<td>11</td>
<td>21</td>
<td>439</td>
<td>1554</td>
<td>968.76</td>
</tr>
<tr>
<td>12</td>
<td>32</td>
<td>420</td>
<td>1782</td>
<td>1098</td>
</tr>
<tr>
<td>13</td>
<td>36</td>
<td>463</td>
<td>2059</td>
<td>1239</td>
</tr>
<tr>
<td>14</td>
<td>40</td>
<td>320</td>
<td>2366</td>
<td>1394</td>
</tr>
<tr>
<td>15</td>
<td>56</td>
<td>289</td>
<td>2707</td>
<td>1564</td>
</tr>
<tr>
<td>16</td>
<td>69</td>
<td>457</td>
<td>3086</td>
<td>1753</td>
</tr>
<tr>
<td>17</td>
<td>70</td>
<td>591</td>
<td>3507</td>
<td>1963</td>
</tr>
</tbody>
</table>

In this table, bold faces are for Bark centre frequencies and bold with underlined are for Mel centre frequencies.

V. CONCLUSION

The performance of Mel scale and Bark scale is evaluated. Mel scale and Bark scale are filter bank structure which is used in speaker recognition system. This filter banks are mainly based human auditory system. It is found that centre frequencies of Bark scale are closely matched with the formant F1, than Mel scale centre frequencies in case of Indian dialect speaker database. Formant F1 is mainly useful for inter-variability between speakers. Also, Bark scale is good approximation for measurement of critical band at which loudness becomes significantly different. Whereas, Mel scale is suitable for pitch perception and phonetical information. Therefore, Mel scale has shown better performance as compared to Bark scale on TIMIT database. The recognition rate achieved is 96% on AISSMISOIT database and average recognition rate achieved is 95% on Marathi database using 20 Bark filters.

REFERENCES

17. Dr. Shaila D. Apte, “Speech Processing Applications”, in Speech and Audio Processing, Section 1, Section 2 and Section 3, pp. 1-6, 67. 91,92, 105-107, 129-132, Wiley India Edition.
21. Department of Computer Engineering, Dr. Babasaheb Ambedkar Marathwada University.

AUTHORS PROFILE

Dr. S. B. Dhonde has completed his Ph.D. in Electronics Engineering from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. He has 19 Years of experience in teaching, research and industry. His area of interest includes speech signal processing, computer networking, wireless sensor network. Currently, He is working as an Associate Professor in A.I.S.S.M.S. Institute of Information Technology, Pune. He has published several papers in the area of speech signal processing, networking, embedded system at international conference and journals.

Mr. Amol A. Chaudhari has completed his M.E. in E & TC Engineering from Savitribai Phule Pune University. His area of interest includes speech signal processing and Embedded Systems. Currently, He is working as an Assistant Professor in A.I.S.S.M.S. IOIT, Pune. He has published several papers in the area of speech signal processing including speaker identification.

Mr. M. P. Gajare has completed M.E. and pursuing Ph.D. from Savitribai Phule Pune University. His area of interest includes signal processing and CMOS VLSI. Currently, He is working as an Assistant Professor in A.I.S.S.M.S. IOIT, Pune. He has published several papers in the area of CMOS.