Studying the Effect of Frame-Level Concatenation of GFCC and TS-MFCC Features on Zero-Shot Children's ASR

Ankita(✉), Shambhavi, and Syed Shahnawazuddin

National Institute of Technology Patna, Patna, Bihar, India
{ankita.ph21.ec,s.syed}@nitp.ac.in

Abstract. The work presented in this paper aims at enhancing the recognition performance of zero-shot children's speech recognition task through frame-level concatenation of two complementary front-end acoustic features. The acoustic features chosen are TANDEM-STRAIGHT-based Mel-frequency cepstral coefficients (TS-MFCC) and Gamma-tone frequency cepstral coefficients (GFCC). The GFCC model the cochlear response of the human auditory system. The MFCC features, on the other hand, model the human pitch perception. Therefore, the GFCC and TS-MFCC features capture the acoustic information differently and that too with very low correlation. Consequently, concatenation of TS-MFCC and GFCC feature vectors helps in modeling complementary and a wider range of relevant acoustic information. This, in turn, enhances the recognition performance significantly. The experimental evaluations presented in this paper show that a relative reduction of nearly 12% is achieved by feature concatenation.

Keywords: Zero-shot children's ASR · TS-MFCC · Feature concatenation · GFCC

1 Introduction

Automatic speech recognition (ASR) is the technology that aides in converting human speech into text. Cutting edge computational techniques such as highly efficient deep learning algorithms [5–7,17,20] have boosted the research work in this domain. As a result, ASR systems are employed in several applications such as voice-based digital assistance, voice-to-text conversion for hands-free computing, voice commands to smart home devices, virtual agents, reading tutors, interactive voice response (IVR) systems, live captioning, language learning tools, voice biometrics, automotives, entertainment and clinical note-taking.

To be effective and to generalize well for all kinds of users, ASR systems are supposed to be speaker-independent. For that purpose, a large amount of speech data is used for learning the statistical model parameters. Most of the ASR systems are designed for adult population and hence use data from adult speakers only. Therefore, such ASR systems have high recognition rates with

A. Karpov et al. (Eds.): SPECOM 2023, LNAI 14339, pp. 140–150, 2023.
https://doi.org/10.1007/978-3-031-48312-7_11

respect to adult's speech. However, their performance degrades substantially when they are subjected to the children's speech. Absence of speech data from the child domain in the training set leads to acoustic mismatch between the training and test conditions. This, in turn, results in severe degradation in recognition performance [4,11,19]. This task of recognising children's speech using statistical models trained on adults' speech is referred to as *zero-shot children's* ASR.

The acoustic mismatch between the training and test data can be alleviated by modifying children's speech test set prior to decoding by using techniques like prosody modification [26], formant scaling [9] and vocal-tract length normalization [10,22]. However, those approaches require two-pass decoding in order to optimally modify the test data which, in turn, results in increased computation time. Resorting to out-of-domain data augmentation [8,23,24] as well as developing robust front-end features specifically for children's speech can help overcome the issue of increased computation time. One such acoustic features, suitable for *zero-shot children's* ASR is referred as TS-MFCC, was proposed in [25]. The TS-MFCC feature extraction process employs pitch-synchronous spectrum estimation called TANDEM STRAIGHT (TS). This results in smoothed power spectra that suppresses the ill-effects of pitch harmonics. The Mel-frequency cepstral coefficients (MFCC) computed using the TANDEM-STRAIGHT power spectra are reported to be very effective for *zero-shot children's* ASR task.

In this study, we have revisited the TS-MFCC features and studied its effectiveness in combination with another front-end acoustic feature called Gammatone frequency cepstral coefficient (GFCC) [14]. The GFCC models the human auditory system's cochlear response whereas the MFCC models the human pitch perception. Consequently, the two kinds of features capture and model the acoustic information present in the speech signal differently and that too with a very low correlation. Therefore, it is expected that combining these two front-end acoustic feature vectors will capture a broader range of relevant acoustic information leading to improved recognition performance. Motivated by this fact, in our present work, we have studied the effect of frame-level concatenation of TS-MFCC and GFCC features for *zero-shot children's* ASR task. The ASR system trained on the concatenated feature vectors leads to significantly lower error rates as demonstrated by the experimental evaluations presented later in this paper.

The rest of the sections of this paper is organised as follows: In Sect. 2, the proposed approach is described and the experimental evaluations demonstrating the effectiveness of the proposed approach are presented in Sect. 3. Finally, the paper is concluded in Sect. 4.

2 Proposed Approach

In this work, we have studied the effect of concatenating TS-MFCC and GFCC features in order to enhance the recognition performance of *zero-shot children's* ASR task. The proposed feature concatenation approach is summarized in the block diagram shown in Fig. 1. It involves appending the coefficients of TS-MFCC and GFCC feature vectors at the frame-level. The resultant feature vectors are then used for the training purpose. In this section, we first describe the two kinds of features in detail. Next, we discuss the motivation behind concatenating those two feature vectors.

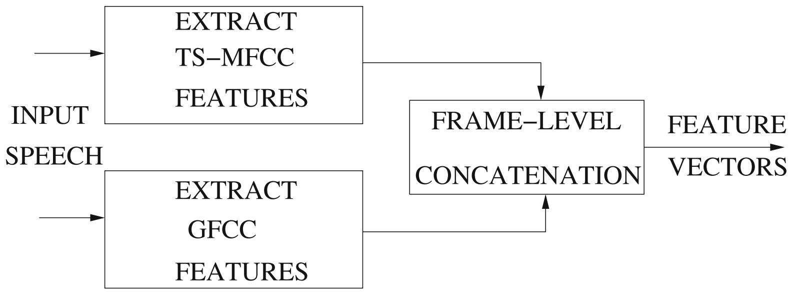

Fig. 1. Block diagram outlining the proposed approach for frame-level concatenation of TS-MFCC and GFCC features.

2.1 Overview of GFCC Features

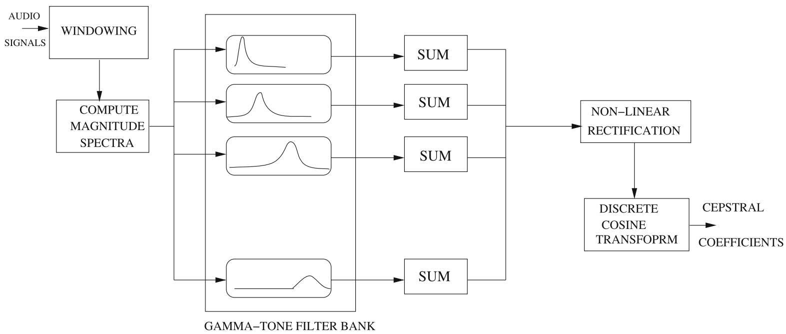

Fig. 2. Block diagram illustrating the process of exctracting GFCC features.

We have borrowed the idea of using Gamma-tone frequency cepstral coefficients [14,29] from other speech-related research fields where they have been successfully employed for speech recognition [2,21,27] and speaker identification [28]. However, it's application in children's speech recognition has not been explored yet. The computation of the GFCC features is similar to that of the MFCC extraction process. The speech signal is first analyzed into short-time frames. The non-stationary speech signal is known to show stationary behaviour in such short frames. This aides in the spectro-temporal signal analysis. Next, each of the frames is processed using a bank of Gamma-tone filters. The Gamma-tone filters are derived by observing the psychophysical and physiological behaviour of the auditory peripheral and hence serve as a standard model for Cochlear filtering. As a consequence, Gamma-tone filtering helps in effectively capturing acoustic information that is left out due to the use of Mel-filterbank.

The cochlea not only amplifies sound waves and converts them into neural signals, but also decomposes complex acoustic waveform into simpler elements. Thus, it acts as mechanical frequency analyzer where each position along the basilar membrane corresponds to a particular frequency. The Gamma-tone filters are designed as such to replicate this process. For this purpose, the magnitude or the power spectrum of the signal is passed through a Gamma-tone filterbank. We have used a bank of 40 filters spaced linearly on the equivalent rectangular bandwidth (ERB) scale whose central frequency varies between 50 Hz and 8000 Hz. The ERB is a psychoacoustic measure of the auditory filter width at each point along cochlea. The frequency conversion from Hz to the ERB scale is given by:

$$ERB = A \times \log_{10}(1 + 0.00437f) \tag{1}$$

where, f is in Hz and A is given by:

$$A = 1000\frac{\ln(10)}{24.7 \times 4.37} \tag{2}$$

Next, nonlinear cubic-root function is applied on the obtained time-frequency representation to model human loudness perception. To reduce dimensionality and de-correlate the resulting components, discrete cosine transform is applied. The overall GFCC feature extraction process is summarized in Fig. 2.

2.2 Review of TS-MFCC Features

A periodic signal $h(t)$ has a temporally stable power spectrum usually calculated over a sum of two power spectra. To serve this purpose, a pair of time windows are chosen such that they are separated for half of the fundamental period [13]. Let, $h(t)$ has a Fourier transform $H(\omega)$ and assuming that only two harmonic components of the fundamental frequency ($\omega_0 = \frac{2\pi}{T_0}$) occupy the main lobe of $H(\omega)$, then

$$h(t) = e^{jk\omega_0 t} + \alpha e^{j(k+1)\omega_0 t + \beta}. \tag{3}$$

where α and β represent real numbers. Taking Fourier transform of the above equation (assuming $k = 0$ for simplicity):

$$H(\omega) = \delta(\omega) + \alpha e^{j\beta}\delta(\omega - \omega_0). \tag{4}$$

The respective power spectra of the windowed test signal is then

$$P(\omega, t) = |H(\omega)|^2 + \alpha^2|H(\omega - \omega_0)|^2 + 2\alpha H(\omega)H(\omega - \omega_0)\cos(\omega_0 t + \beta). \tag{5}$$

The third term in the above equation is time-dependent and represents the temporal dependency in the spectrum estimation. It can be cancelled by taking an opposite polarity with a window at $t + T_0/2$. The spectrum without any temporal fluctuation i.e., the TANDEM spectrum $T(\omega, t)$ is now given as:

$$T(\omega, t) = \frac{1}{2}\left\{P(\omega, t) + P(\omega, t + T_0/2)\right\}. \tag{6}$$

The TANDEM spectrum $T(\omega, t)$ results in smoothed vocal-tract response. The suppression of pitch-harmonics through spectral smoothing due to TANDEM STRAIGHT analysis was demonstrated in [25]. MFCC features extracted after smoothing out the pitch-harmonics were noted to be effective in the context of *zero-shot children's* ASR task.

2.3 Motivation for Feature Concatenation

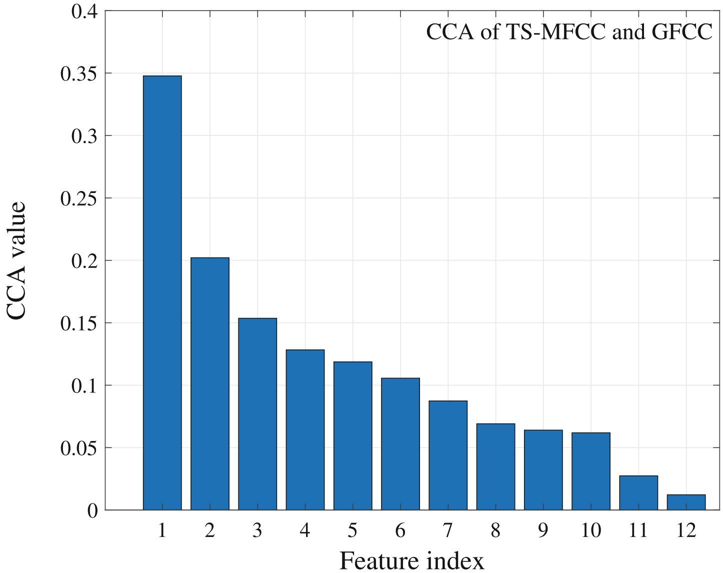

Fig. 3. Canonical correlation between TS-MFCC and GFCC features demonstrating that the two kinds of feature vectors are highly uncorrelated.

As mentioned earlier, the contribution of this work is to explore the effect of frame-level concatenation of two front-end features, i.e., TS-MFCC and GFCC on *zero-shot children's ASR* task. Due to inherent differences in the way the Mel- and Gamma-tone filterbanks are designed and act on a frame of speech, the two kinds of features capture and model complementary acoustic information. To demonstrate that the two kinds of features model the speech data differently and represent a wider range of acoustic attributes, canonical correlation analysis (CCA) was performed on these two features. As evident from Fig. 3, the CCA results in low values (≤ 0.15) for most of the indexes. This, in turn, implies that the TS-MFCC and GFCC features are highly uncorrelated. Hence, their frame-level concatenation is expected to represent a wider range of acoustic attributes as intended. Modeling those will, in turn, help in capturing the missing targeted attributes more robustly and hence improve the recognition performance.

3 Experimental Evaluations

3.1 Database and Experimental Specification

For experimental evaluations, we have used two different British English speech corpora, namely, WSJCAM0 [18] and PF_STAR [1]. The motivation behind using the said corpora is that the mismatches in the recording conditions and the accent of the speakers are minimal. Furthermore, both WSJCAM0 and the PF_STAR databases contain read speech. In our study, the training set was derived from WSJCAM0 and it consisted 15.5 h of speech data from 92 adult speakers (39 females). In order to deal with the unavailability of speech data from child domain, the acoustic attributes of adults' speech training set were modified to make them similar to that of children's speech. For that purpose, we up-scaled the pitch and formant frequencies as well as increased the duration of the adults' speech [8,23]. In addition to that, adults' speech was also subjected to voice-conversion using a generative adversarial network (GAN) to synthetically generate children's like speech [24]. The pitch of the adults' speech training set was increased by a factor of 1.35 while the duration was increased by a factor of 1.4 using the technique reported in [3]. The formant frequencies were up-scaled by a factor of 0.08. For formant modification, the approach proposed in [9] was used which employed scaling of the linear prediction coefficients [12]. These scaling factors were determined by performing experiments on a development set described later. The modified data-sets were then pooled into training. This out-of-domain augmentation approach helps in capturing the missing targeted attributes of children's speech. In addition to that, the overall duration of the training data is increased which, in turn, helps in more robust estimation of model parameters.

Children's speech test set was derived from the PF_STAR corpus and it comprised of 1.1 hours of speech data from 60 speakers (28 females). The age of the child speakers in this test set varied from 4 to 13 years. Furthermore, a development set of children's speech was also derived from the PF_STAR corpus. The development set consisted of 2.1 h of speech data from 63 speakers whose age varied between 6 and 14 years. This set was used for determining the optimal values for the tunable parameters. To gain better insight into the effect of feature concatenation, the test set was split into two, based on the age of the speakers. The first split consisted of nearly 0.6 h of data from children in the age group 4 to 8 years. The second split comprised of nearly 0.5 h of data from speakers belonging to the ages 9 to 13 years. Further to that, another split was done based on the gender of the speakers.

The Kaldi toolkit was used to perform all the experiments [16]. However, front-end speech parameterization was done using MATLAB. The TS-MFCC features reported in [25] were used for front-end speech parameterization in the case of baseline ASR system since those are observed to be more suitable than other existing features in the context of children's speech recognition task. Speech data was analyzed through short-time frames using overlapping Hamming windows of duration 25 ms with a frame-shift of 10 ms. A 40-channel log-Mel-filterbank was used to compute the 13-dimensional base TS-MFCC feature vectors. The base features were time-spliced with context size of ±4 frames and then projected to a 40-dimensional subspace and de-correlated using linear discriminant analysis (LDA) and maximum-likelihood linear transform (MLLT). For feature normalization, cepstral mean and variance normalization (CMVN) as well as feature-space maximum likelihood linear regression (fMLLR) were used. This helps in imparting robustness towards speaker variations. In the case of the GFCC features, frame-size and frame-overlap were chosen as 25 ms and 10 ms. The Gammatone-filterbank consisted of 40 channels. Cubic-root function was used for non-linear rectification prior to the application of DCT. The base features extracted in this case were also 13-dimensional. LDA, MLLT, CMVN and fMLLR were then applied in succession to obtain 40-dimensional feature vectors.

Hidden Markov models (HMM) were used for acoustic modeling. The observation probabilities for the HMM states were generated using Gaussian mixture models (GMM) as well as time-delay neural network (TDNN) [15,30]. Cross-word triphone models consisting of eight diagonal covariance components per state were used for the GMM-HMM-based ASR system. Furthermore, decision tree-based state tying was performed with the maximum number of senones being fixed at 2000. Speaker-adaptive training employing fMLLR transforms was used to optimize the final GMM-HMM system. The time-alignments generated using this GMM-HMM-based ASR system were used for initializing the TDNN-HMM. The lattice-free maximum mutual information (LF-MMI) criterion [17] was used for training TDNN-HMM-based ASR system. The TDNN consisted of 13 hidden layers with 1024 nodes per layer. The initial and final learning rates were set to 0.0005 and 0.00005, respectively. Prior to learning the TDNN parameters, 100-dimensional i-vectors were extracted and appended to the base acoustic feature vectors. The universal background model employed for extracting i-vectors consisted of 512 Gaussian components.

A domain-specific 1.5k bi-gram language model (LM) was used while decoding the children's speech test set. This LM was trained on the transcripts of the speech data from PF_STAR corpus after excluding the utterances from the test set. The employed LM had an out-of-vocabulary rate of 1.20% and a perplexity of 95.8 for the children's speech test set. The lexicon consisted of 1969 words including pronunciation variations. The metric used for performance evaluation are word error rate (WER) and character error rate (CER).

3.2 Results and Discussions

The WERs and CERs for the children's speech test set with respect to an ASR system trained on adults' speech and its' modified versions pooled into training

Table 1. WERs and CERs for the children's speech test set with respect to an ASR systems trained on augmented data. The recognition performances are given with respect to the explored front-end features as well as their fame-level concatenation.

Front-end features	Evaluation metrics	
	WER (%)	CER (%)
TS-MFCC	10.01	7.20
GFCC	10.07	7.20
TS-MFCC + GFCC	**8.86**	**6.29**

are given in Table 1. The baseline ASR system is trained using TS-MFCC features as already stated earlier. It is worth mentioning here that, a WER of 19.5% is achieved if only adults' speech is used for training. In other words, the WER gets nearly halved when data augmentation is employed. The WER and CER for GFCC features are almost the same as those obtained using TS-MFCC features. However, on concatenating the two kinds of features, an absolute reduction in WER by 1.15% over the baseline is obtained. Similarly, the absolute reduction in CER is 0.91%. The relative changes in WER and CER over the baseline are shown in Fig. 4. These, results statistically substantiate the efficacy of the proposed approach in the context of *zero-shot children's ASR* task.

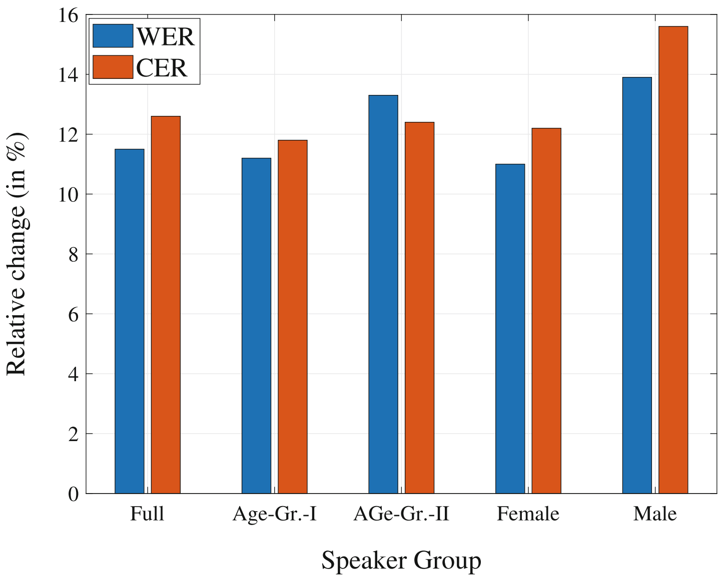

Fig. 4. Age-group and gender-wise relative change in WERs and CERs over the respective baselines obtained by the concatenation of TS-MFCC and GFCC features.

Table 2. Age-group as well as gender-specific WERs and CERs for children's speech with respect to an ASR systems trained on augmented data.

Front-end Features	Speaker Group	Evaluation metrics	
		WER (%)	CER (%)
TS-MFCC	Age-Gr.-I	15.08	11.34
	Age-Gr.-II	6.71	4.36
	Female	11.58	8.59
	Male	8.86	6.16
GFCC	Age-Gr.-I	15.93	12.03
	Age-Gr.-II	6.09	3.91
	Female	11.58	8.59
	Male	8.76	6.16
TS-MFCC + GFCC	Age-Gr.-I	**13.39**	**10.00**
	Age-Gr.-II	**5.82**	**3.82**
	Female	**10.32**	**7.54**
	Male	**7.63**	**5.20**

Next, we performed another study to determine the age-group-specific and gender-specific recognition performances. The age-group as well as gender-specific WERs and CERs are given in Table 2. As evident for the tabulated results, both TS-MFCC and GFCC give similar WER and CER values for each of the speaker groups. However, when the two kinds of feature vectors are concatenated, there are significant reductions in WERs as well as CERs in each of the case. The relative changes in WER and CER obtained over the respective baselines are shown in Fig. 4. In each of the cases, the relative reduction is more than 10%. These results show that the proposed approach is equally powerful not only for Age-Gr.-I kids where the pitch is relatively very high but also for the Age-Gr.-II children having relatively lower pitch values. Similarly, the gains are similar for both male as well as female speakers. Its worth mentioning here that, the pitch values for female speakers are somewhat higher than those for the male children. Thus it can be concluded that the proposed feature concatenation approach imparts pitch-robustness to the ASR system.

4 Conclusion

The work presented in this paper outlines our efforts towards enhancing the recognition performance of *zero-shot children's ASR* system. In this regard, we have implemented frame-level concatenation of two complementary features namely, TS-MFCC and GFCC. The TS-MFCC features employ Mel-filterbank for spectral warping while Gamma-tone filterbank is used in the case of GFCC. Consequently, the two kinds of features model speech data differently and with very low correlation. Hence, on concatenating those at the frame-level helps in capturing a wider range of acoustic attributes. This, in turn, enhances the recognition performance significantly. In our experimental setup, a relative reduction in WER by nearly 12% over the baseline is obtained.

References

1. Batliner, A., et al.: The PF_STAR children's speech corpus. In: Proceedings of Interspeech, pp. 2761–2764 (2005)
2. Cheng, O., Abdulla, W., Salcic, Z.: Performance evaluation of front-end algorithms for robust speech recognition. In: Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, 2005, vol. 2, pp. 711–714 (2005). https://doi.org/10.1109/ISSPA.2005.1581037
3. Damskägg, E.P., Välimäki, V.: Audio time stretching using fuzzy classification of spectral bins. Appl. Sci. **7**(12), 1293 (2017). https://doi.org/10.3390/app7121293
4. Gerosa, M., Giuliani, D., Brugnara, F.: Acoustic variability and automatic recognition of children's speech. Speech Commun. **49**(10–11), 847–860 (2007)
5. Graves, A., Jaitly, N., Mohamed, A.R.: Hybrid speech recognition with deep bidirectional LSTM. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 273–278. IEEE (2013)
6. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649. IEEE (2013)
7. Hinton, G.E.: Deep neural networks for acoustic modeling in speech recognition. Signal Process. Maga. **29**(6), 82–97 (2012)
8. Kumar, V., Kumar, A., Shahnawazuddin, S.: Creating robust children's ASR system in zero-resource condition through out-of-domain data augmentation. Circuits Syst. Signal Process. **41**(4), 2205–2220 (2022). https://doi.org/10.1007/s00034-021-01885-5
9. Kumar Kathania, H., Reddy Kadiri, S., Alku, P., Kurimo, M.: Study of formant modification for children ASR. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7429–7433 (2020). https://doi.org/10.1109/ICASSP40776.2020.9053334
10. Lee, L., Rose, R.: A frequency warping approach to speaker normalization. IEEE Trans. Speech Audio Process. **6**(1), 49–60 (1998)
11. Lee, S., Potamianos, A., Narayanan, S.: Acoustics of children's speech: developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am. **105**(3), 1455–1468 (1999)
12. Makhoul, J.: Linear prediction: a tutorial review. Proc. IEEE **63**(4), 561–580 (1975). https://doi.org/10.1109/PROC.1975.9792
13. Morise, M., Takahashi, T., Kawahara, H., Irino, T.: Power spectrum estimation method for periodic signals virtually irrespective to time window position. Trans. IEICE **90**(12), 3265–3267 (2007)
14. Patterson, R.: Auditory filters and excitation patterns as representations of frequency resolution. In: Frequency Selectivity in Hearing (1986)
15. Peddinti, V., Povey, D., Khudanpur, S.: A time delay neural network architecture for efficient modeling of long temporal contexts. In: Proceedings of Interspeech (2015)
16. Povey, D., et al.: The Kaldi Speech recognition toolkit. In: Proceedings of ASRU (2011)

17. Povey, D., et al.: Purely sequence-trained neural networks for ASR based on lattice-free MMI. In: Proceedings of Interspeech, pp. 2751–2755 (2016)
18. Robinson, T., Fransen, J., Pye, D., Foote, J., Renals, S.: WSJCAM0: a British English speech corpus for large vocabulary continuous speech recognition. In: Proceedings of ICASSP, vol. 1, pp. 81–84 (1995). https://doi.org/10.1109/ICASSP.1995.479278
19. Russell, M., D'Arcy, S.: Challenges for computer recognition of children's speech. In: Proceedings of Speech and Language Technologies in Education (SLaTE) (2007)
20. Sainath, T.N., Vinyals, O., Senior, A., Sak, H.: Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4580–4584 (2015). https://doi.org/10.1109/ICASSP.2015.7178838
21. Schluter, R., Bezrukov, I., Wagner, H., Ney, H.: Gammatone features and feature combination for large vocabulary speech recognition. In: Proceedings of ICASSP, vol. 4, pp. IV-649–IV-652 (2007). https://doi.org/10.1109/ICASSP.2007.366996
22. Serizel, R., Giuliani, D.: Vocal tract length normalisation approaches to DNN-based children's and adults' speech recognition. In: Proceedings of Spoken Language Technology Workshop (SLT), pp. 135–140 (2014)
23. Shahnawazuddin, S., Adiga, N., Kathania, H.K., Sai, B.T.: Creating speaker independent ASR system through prosody modification based data augmentation. Pattern Recogn. Lett. **131**, 213–218 (2020). https://doi.org/10.1016/j.patrec.2019.12.019
24. Shahnawazuddin, S., Adiga, N., Kumar, K., Poddar, A., Ahmad, W.: Voice conversion based data augmentation to improve children's speech recognition in limited data scenario. In: Proceedings of Interspeech, pp. 4382–4386 (2020). https://doi.org/10.21437/Interspeech.2020-1112
25. Shahnawazuddin, S., Adiga, N., Kathania, H.K., Pradhan, G., Sinha, R.: Studying the role of pitch-adaptive spectral estimation and speaking-rate normalization in automatic speech recognition. Digital Signal Process. **79**, 142–151 (2018)
26. Shahnawazuddin, S., Adiga, N., Kathania, H.K.: Effect of prosody modification on children's ASR. IEEE Signal Process. Lett. **24**(11), 1749–1753 (2017)
27. Shao, Y., Jin, Z., Wang, D., Srinivasan, S.: An auditory-based feature for robust speech recognition. In: Proceedings of ICASSP, pp. 4625–4628 (2009). https://doi.org/10.1109/ICASSP.2009.4960661
28. Shao, Y., Wang, D.: Robust speaker identification using auditory features and computational auditory scene analysis. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1589–1592 (2008). https://doi.org/10.1109/ICASSP.2008.4517928
29. Valero, X., Alias, F.: Gammatone cepstral coefficients: biologically inspired features for non-speech audio classification. IEEE Trans. Multimedia **14**(6), 1684–1689 (2012). https://doi.org/10.1109/TMM.2012.2199972
30. Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K.: Phoneme recognition using time-delay neural networks. IEEE Trans. Acoust. Speech Signal Process. **37**(3), 328–339 (1989). https://doi.org/10.1109/29.21701