Chapter 9
Speech Enhancement Using Neuro-Fuzzy Classifier

Judith Justin
School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India

Vanithamani R.
School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India

ABSTRACT

In this chapter, a speech enhancement technique is implemented using a neuro-fuzzy classifier. Noisy speech sentences from NOIZEUS and AURORA databases are taken for the study. Feature extraction is implemented through modifications in amplitude magnitude spectrograms. A four class neuro-fuzzy classifier splits the noisy speech samples into noise-only part, signal only part, more noise-less signal part, and more signal-less noise part of the time-frequency units. Appropriate weights are applied in the enhancement phase. The enhanced speech sentence is evaluated using objective measures. An analysis of the performance of the Neuro-Fuzzy 4 (NF 4) classifier is done. A comparison of the performance of the classifier with other conventional techniques is done for various noises at different noise levels. It is observed that the numerical values of the measures obtained are better when compared to the others. An overall comparison of the performance of the NF 4 classifier is done and it is inferred that NF4 outperforms the other techniques in speech enhancement.

INTRODUCTION

The autonomous, computer-driven translation of spoken language into readable text in real-time can be termed as Automatic Speech Recognition (ASR). In short, ASR is a technique that permits a computer to recognize the speech conversed into a microphone or telephone and translate it to written text (Stuckless, R., 1994). The composition of a characteristic continuous-speech recognizer comprises a feature examination block as front-end, followed by a statistical pattern classifier (Thomas Eisele et al. 1996). The interface connecting these two, should ideally hold all the relevant data of the speech signal appro-

DOI: 10.4018/978-1-7998-6659-6.ch009

priate to succeeding classification, be insensitive to irrelevant variations and simultaneously have a low dimensionality to reduce the computational burden of the classifier (Shigeru Katagiri & Chin-Hui Lee, 1993). The objective of speech recognition is to build up mechanisms to make improvements in the static representation of input speech. Nowadays, ASR has extensive applications that need a human-machine interface like automatic call processing (Reddy, D.R, 2009).

The field of ASR (Young.S, 1996 & Nurul Huda et al., 2010) mainly depends on statistical techniques, retreating from approaches that were primarily put forward such as template matching, dynamic time warping, and non-probabilistically motivated distortion measures. Consequently, methods for robust ASR have been developed to accomplish a near-perfect allocation of the acoustic mixture into contributions from constituent sources (Martin Cooke et al., 2001). The Hidden Markov Model (HMM) (Rojathai. S. & Venkatesulu. M., 2013) is one of the well-known algorithms and has proved to be an efficient technique of dealing with large units of speech amongst numerous categories of speech recognizers utilized in ASR products, implemented as well as proposed (Rabiner L.R. & B.-H. Juang, 1993, Bahl L.R, 1993). ASR has been employed for purposes like business dictation and particular requirements accessibility and the market presence for language learning has improved noticeably in recent years (In –Seok Kim, 2006). There are ASR systems that are based on stochastic replicas of speech acoustics that take out a set of major acoustic features from the speech signal and make use of statistical models to symbolize the distribution of these features for speech objects such as vocabulary, syllables or phonemes. Typical scalable signal processing methods are employed by a speech feature extractor to gather feature vectors from input audio forms (Jike Chong et al. 2010).

Speech Recognition is a pattern identification system (Ben Nasr, et al. 2013). Training and Testing are the two phases of pattern recognition (Santosh K. Gaikwad, et al. 2010). The modification caused by room reverberation is particularly unsafe for ASR systems (Nakamura.S & Shikano. K, 1997, Couvreur. L, et al. 2000, Y. Pan and A. Waibel, 2000). To minimize the inconsistency between the training conditions (close-talking / anechoic speech) and the operating conditions (distant-talking/reverberated speech) because of room reverberation, several approaches have been proposed (Laurent Couvreur & Christophe Couvreur, 2014). The illustration of the speech signal performs a significant role in the noisy speech recognition (Gong. Y, 1995). A person could speak slow or fast. It influences both the sequential and spectral uniqueness of the signal, upsetting the acoustic models. Specific models of frequency warping (which depends on vocal tract length differences) as well as more universal characteristic compensation and model alteration techniques, relying on Maximum Likelihood or Maximum a Posteriori criteria have been presented earlier. These model adaptation methods recommend a universal formalism for re-estimation based on reasonable amounts of speech data (Benzeghiba. M., 2007).

The proposed technique aims to enhance the speech quality in a noisy signal using a weighted mask and neuro-fuzzy classifier. The enhancement technique is divided into three phases. In the initial phase, the feature selected from the input noisy signal is the Modified Amplitude Magnitude Spectrogram (MAMS). Here, the noisy speech is split into 25 Time-Frequency (TF) units using a bandpass filter. The resultant channel is rectified and decimated by a factor of three. The envelope is segmented and windowed and then, Discrete Cosine Transform (DCT) is applied and multiplied with a triangular function to obtain the feature set. In the training phase, the estimated signal is computed and the ratio of the estimated to the original signal is obtained. Based on the values obtained, the respective training is implemented. In the third and the final phase, the noisy signal is filtered and windowed. Here, the weight of the mask is determined with the help of the probability function of feature vectors in the respective class. Weights

Related Content

Palmprint And Dorsal Hand Vein Multi-Modal Biometric Fusion Using Deep Learning
Norah Abdullah Al-johaniand Lamiaa A. Elrefaei (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 18-42).*
www.irma-international.org/article/palmprint-and-dorsal-hand-vein-multi-modal-biometric-fusion-using-deep-learning/257270

Churn Prediction in a Pay-TV Company via Data Classification
Ilayda Ulku, Fadime Uney Yuksektepe, Oznur Yilmaz, Merve Ulku Aktasand Nergiz Akbalik (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 39-53).*
www.irma-international.org/article/churn-prediction-in-a-pay-tv-company-via-data-classification/266495

Deep Learning for Feature Engineering-Based Improved Weather Prediction: A Predictive Modeling
Partha Sarathi Mishraand Debabrata Nandi (2021). *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science (pp. 195-217).*
www.irma-international.org/chapter/deep-learning-for-feature-engineering-based-improved-weather-prediction/268756

Industry Revolution 4.0 and Its Impact on Education
Riccardo Minasi (2023). *Encyclopedia of Data Science and Machine Learning (pp. 2211-2222).*
www.irma-international.org/chapter/industry-revolution-40-and-its-impact-on-education/317617

Product Offer and Pricing Personalization in Retail Banking
Wei Ke, Rogan Johannes Vlemingand Rohan Shah (2023). *Encyclopedia of Data Science and Machine Learning (pp. 1952-1963).*
www.irma-international.org/chapter/product-offer-and-pricing-personalization-in-retail-banking/317596