ADVERSE ENVIRONMENT AND ITS COMPENSATION APPROACHES IN SPEECH RECOGNITION

Saumya Sah[1]

[1]*Department of Electronics and Communication Engineering,Birla Institute of Technology, Mesra, Ranchi, India*

ABSTRACT: *The paper focuses on the causes of the low performances of the speech recognition system due to the stressful and noisy environment in real time applications like car, airplane cockpit, etc. This paper deals with the major problems like Adverse Environment and other issues that is noise, distortion of speech recognition system which causes the failure in real time applications. The adverse environment is simply the mismatched between the environments of training and testing of the speech signal. The paper briefly summarized different approaches to deals with the adverse environment.*

Key Words: Cause of speech variation, Noise, Source of Noise, Distortion, Adverse Environment.

1. INTRODUCTION

The voice of the human being is the primary mode of communication among humans. The researchers are working to improve the accuracy level so that the voice or speech could be use in real time applications [1]. Speech Recognition is the process of enabling a computer to identify and respond to the sounds produced by human in their speech. Speech recognition system often fails to provide an excellent accuracy level in real applications whereas it provides a high accuracy in experimental level**.** This is mainly occurring due to the speech recognition system which is trained in normal and clean environment but applied in stressful and noisy environment [2]. Most of the speech recognition system degrades rapidly in their performance due to the presence of noise and distortion in the speech signal. The speech recognition systems are designed by keeping in mind the prediction or assumptions of the ambient conditions, like background speech style, transducers, low noise, channels etc. [2, 3]. The different causes of the variation in speech due to which the distortion occurred are discussed in below section [4].

1.1 Causes of Speech Variation:

Few causes of speech variations are discussed below. Mainly there are three parameters over which the speech of human can be varied from one another and could produce noise in the speech.

These are:

a) Environment:

It is the acoustic properties of the environment including the impact on a speaker's voice [5]. There are two type of environment noise: Speech Correlated noise and uncorrelated noise.

b) Speaker

Speaker itself could be the cause of variation in speech. There may be two reasons for this noise to produce. These are: **Attributes of speakers** which includes dialect, gender, age and the other one is **Manner of speaking** which includes breath and lip noise, stress, Lombard effect, rate of delivery, level and dynamic range, pitch, cooperativeness.

c) Input Equipment

Input Equipments are used to deliver human voice to a place distant to the user are one of the cause of speech variation. These are due to microphone, distance to microphone, filter, recording equipment and transmission system which includes channel distortion, noise, and echo.

2. ADVERSE ENVIRONMENT

The use of the term "adverse environments" implies unknown, mismatched and often severe differences in variables and conditions between training and testing. A speech recognizer often encounters three main causes of adverse conditions: noise, distortion, and (human) articulation effects.

2.1 Noise

By nature, noise is usually considered additive. It can be wideband (assumed to be Gaussian distributed and white) or coloured, stationary or non-stationary. High levels of ambient noise are one of the primary concerns for speech recognizer.

2.2 Source of Noise

The source of additive noise can be found in several environments including the office, car and the aircraft cockpit.

The sound pressure level (SPL) in a normal personal office is around 45-50dBA. In a business office it is around 15-20dB higher than the personal office. Inside an automobile, the noise level due to engine, cooling fan, wind tire and road is usually considerably high, particularly when the automobile is moving at a high speed. In the cockpit of a modern jet fighter aircraft, SPLs

of 90dB or more across the speech frequency band have been reported.

There are many type of noises like electric noise and quantization noise, which are present inside of any electronic speech recognition system, are at a low level below the concern threshold. Noise due to transmissions and switching equipments in a telephone network is a factor which is affecting recognizer performance. At the noise level of equal to the noise level of cockpit, the speech signal is barely intelligible even to a human listener, not to mention an automatic speech recognition machine.

2.3 Distortion

The speech signal before recorded and processed for speech recognition, undergoes a series of spectral distortions. The microphone transducer can also distort the speech spectrum depending on its type and mounting positions. The transducer configuration used in testing is different from that used during training of the reference patterns, the mismatch in spectral distortion becomes one of the major problems.

It was reported that a large vocabulary speech recognition system with a performance of 85% word accuracy in a matched transducer condition could only achieve less than 19% word accuracy when a different microphone was used during testing.

Telephones channels are characterized by a high frequency pre emphasis with significantly decrease the dynamic range of the speech signal and increase the signal correlation. There is a wide range of variation in attenuation in the frequency spectrum and this can cause a spectral mismatch unless the telephone channel is measured and adapted to beforehand. Figure below compares the amplitude of frequency response for two different telephone channels. The distance between the microphone and the speaker's mouth can cause a poor SNR and the acoustics of the room or the microphone may be differ from the original set up in which the reference patterns were created.

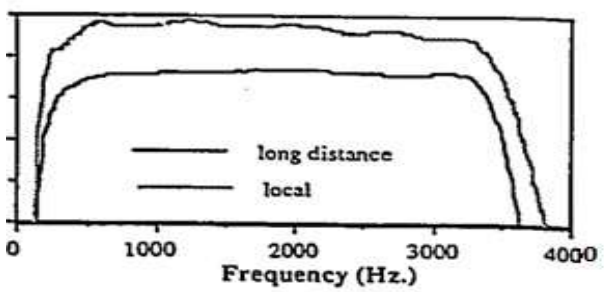

Fig-1: Frequency Response for two different telephone channels.

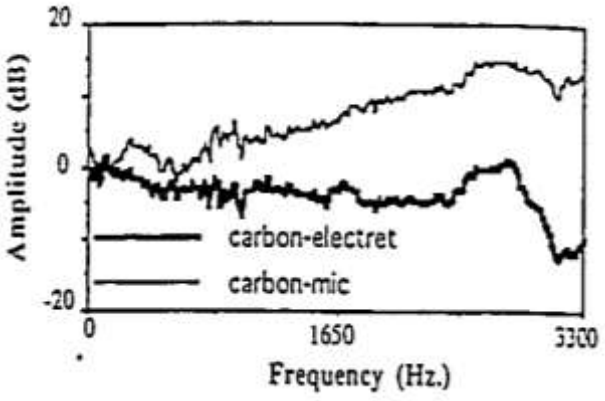

Fig-2: Frequency Response for two different microphone transducers.

2.4 Articulation Effects

There are many factors which affect the manner of speaking of individual talker. The main difficultly in dealing with the articulation effects is the lack of understanding that how to quantify them.

In terms of spectral characteristics, we can measure or model the effects to improve the recognizer design with noise or channel distortion which usually does not vary as rapidly as the speech itself. For example, a white noise is usually specified by its power level, where as a channel is often represented by its frequency response.

2.5 Lombard Effect

Lombard Effect is the characteristics changes in articulation due to environmental influence. The Lombard Effect is manifested in such changes as:

1. Louder speech (increases energy).
2. Energy migration in the frequency domain, F1 increases while F2 decreases.
3. Changes in spectral balance.

A speaker dependent isolated word recognizer that had accuracy of 91% when trained and tested in clean condition could only achieve accuracy of 62% when the test utterances contained the Lombard Effect, even though the test tokens were free from masking noise.

3. DEALING WITH ADVERSE EFFECTS

A speech recognizer that uses reference patterns trained from speech with the corrupting noise whose characteristics are approximately unknown performs more robustly than one that uses clean reference patterns. The key idea is to train the recognizer with a multi-style training schemes in which speech signals of talker of different talking styles are used as the training data.

In an isolated word recognition experiment using hidden Markov models had reported that the error rate of a recognizer, when tested with the speech of talkers having various talking styles was reduced by more than a factor of 2, from 17.5% using only normally spoken reference pattern to 6.9% with multi style training. Multi-style training seemed to be most effective with the speech which exhibits the Lombard effect and with speech produced under high emotional conditions.

Some of the methods and algorithm that have been proposed to combat the unknown environment in which speech recognizer must operate:

1. Signal enhancement pre-processing

When the adverse condition is alone due to the additive noise, one can use speech enhancement methods to suppress the noise before applying the recognition algorithms.

One of the most widely used signal enhancement method is adaptive noise cancellation using two signal sources.

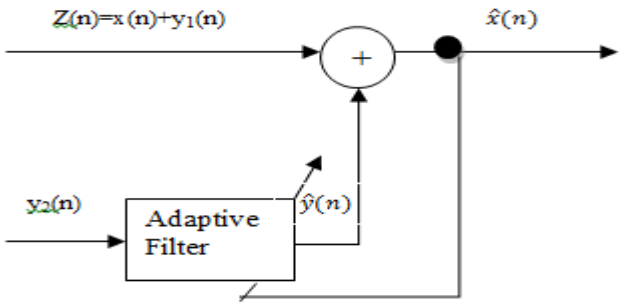

Fig-3: Schematic of the two input noise cancelling approach.

In the above figure x is the clean signal whereas y_1 and y_2 are two noise signals and z is the noisy signal. The technique adaptively adjusts the filter coefficient so that on subtracting the filtered (noise) reference input from the primary (speech plus noise) input leads to an output signal with minimum energy. This technique requires noise component in the corrupted signal and the reference noise have high coherence.

2. Special transducer arrangement

A noise cancelling microphone is a specially designed microphone in which both sides of the diaphragm are exposed to the sound field in such a way that sounds coming from a relatively large distance (far field) are cancelled because the sound pressure causes virtually no net force on the diaphragm. For sound sources close to the Microphone, the back of the diaphragm is effectively shaded from the sound field and the sound pressure is received only by the front of the diaphragm. It can be effective in suppressing low frequency noises and can be used in Car, automobile or aircraft cockpit environment.

3. Noise masking and adaptive models

In the presence of broadband noise, certain regions of the speech spectrum which are of lower level will be more affected by the noise. This makes the calculation of the spectral distortion difficult as more corrupted regions represent less reliable spectral measurements. By recognizing this difficulty, the use of noise masking in conjunction with a filter-bank analyser is suggested. The key idea is to choose first, for each channel of the filter-bank output, the masking noise level is the greater of the noise level in the reference signal and that in the testing signal, and then replace that channel output by the mask value if it is below the corresponding mask level. This helps prevent spurious distortion accumulation because those channels that are determined to have been seriously corrupted by noise will have the same spectral value in both the training and the testing tokens. The technique of noise compensation was also employed by adapting the spectral prototypes to the noise condition in the autocorrelation domain. The underlying assumption, unlike the above masking model, was that the power spectra of speech and noise are additive and so are the autocorrelations.

4. Stress compensation

The purpose of stress compensation is to provide offset for the spectral distortion caused by extraordinary speaking effort due to the talker's reaction to ambient conditions. Because of the difficulties in modeling such characteristic changes, no analytical result is available. Several proposed heuristic techniques are noteworthy because of their effectiveness. A traditional template-based system that incorporates both vector quantization and dynamic time warping. Another stress compensation technique is operating in the cepstral domain.

5. Robust distortion measures

The idea of robust distortion measures is to emphasize selectively and automatically the distortion pertaining to certain regions of the spectrum that are less corrupted by noise. A noise compensation scheme can be interpreted as implicitly defining a robust distortion.

Many distortion measures are there:

1. Cepstral distance

2. Likelihood ratio distortion

3. Weighted Likelihood Ratio distortion

4. Asymmetrically Weighted Likelihood Ratio distortion

6. Novel representation of speech

Apart from the above discussed schemes and algorithms, there are other attempts to compensate the noise problem by finding characteristic representations of speech that are invariant or resistant to noise corruption. In other words, the approach is to find a robust "front-end" that will reliably produce measurements of speech even in the presence of noise. There are two attempts; one is from a signal processing viewpoint and the other tries to duplicate the human auditory capability.

IV CONCLUSION

Different approaches and algorithms have been proposed to cope with ambient noise, distortions and other noise-induced problems in speech recognition system. These approaches can be categorized as: signal enhancement preprocessing, special transducer arrangements, noise masking, stress compensation, robust distortion measures and novel representations of speech. Proper adoption of these methods proves beneficial when dealing with noisy recognition problems.

V REFERENCES

[1]. Chavan, Karishma, and Ujwalla Gawande. "Speech recognition in noisy environment, issues and challenges: A review." In Soft-Computing and Networks Security (ICSNS), 2015 International Conference on, pp. 1-5. IEEE, 2015.

[2]. Rajasekaran, P., G. Doddington, and J. Picone. "Recognition of speech under stress and in noise." In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86., vol. 11, pp. 733-736. IEEE, 1986.

[3]. Starks, David Ross. Speech recognition in adverse environments: Improvements to IMELDA. University of Ottawa (Canada), 1995.

[4]. Rabiner, Lawrence R., and Biing-Hwang Juang. Fundamentals of speech recognition. Vol. 14. Englewood Cliffs: PTR Prentice Hall, 1993.

[5]. Shrawankar, Urmila, and Vilas M. Thakare. "Adverse conditions and ASR techniques for robust speech user interface." arXiv preprint arXiv:1303.5515 (2013)

[6]. Juang, B. H. "Speech recognition in adverse environments." Computer speech & language 5, no. 3 (1991): 275-294.

[7]. Rahim, Mazin G., Biing-Hwang Juang, Wu Chou, and Eric Buhrke. "Signal conditioning techniques for robust speech recognition." IEEE Signal Processing Letters 3, no. 4 (1996): 107-109.