Acoustic Driving Simulator Design for Evaluating an In-car Speech Recognizer

Lee, Seongjae[1] · Kang, Sunmee[2]

ABSTRACT

This paper is on designing an indoor driving simulator to evaluate the performance of in-car speech recognizer when influenced by the elements, which lower the success rate of speech recognition. The proposed simulator simulates vehicle noise which was pre-recorded in diverse driving environments and driver's speech. Additionally, the proposed Lombard effect conversion module in this simulator enables the speech recorded in a studio environment to convert into various possible driving scenarios. The relevant experimental results have confirmed that the proposed simulator is a feasible approach for realizing an effective method as it achieved similar speech recognition results to the real driving environment.

Keywords: Lombard effect, Acoustic driving simulator, Speech recognition, Performance evaluation

1. Introduction

Speech recognition technologies have been widely used for various areas such as robots and electronic home appliances. Especially the importance of speech recognition technologies under environments for vehicles which limits manipulations of user convenient devices such as navigation, audio and conditioning equipments during driving becomes more and more critical, and markets for telematics and infotainment also become more expanded. Conventional methods to evaluate in-car speech recognizer in its development phase for commercializing it considering driving noise which causes deterioration of speech recognition include: a method to record recognition result in a way that an experimenter gives speech recognition commands or gives speeches instead of human using HATS (Head and Torso Simulator) while he or she directly drives an vehicle. As this method is required for experimenter to drive vehicles by themselves or use highly expensive devices to evaluate all

processes for speech recognizer, problems related to evaluation time and cost occur.

This paper proposes an acoustic driving simulator to carry out virtual evaluation for speech recognizer in an indoor environment similar to actual driving environment by displaying recorded speech in a studio and driving noise without having to directly evaluate the speech recognizer improving difficulties mentioned earlier.

In order to design an acoustic driving simulator, two important factors should be considered. First is driving noise and it has different acoustic components depending on road surface and speed. Second is Lombard effect that is a body feedback of human to improve intelligibility of speech against surrounding noise and it has different acoustic features from ordinary speech and is a factor to affect success rate of speech recognition [2].

The proposed acoustic driving simulator enhances reality of virtual evaluation in a way that designs it based on acoustic features in driving environment above mentioned.

The organization of this paper is as follows. The next section presents overall structure of the acoustic driving simulator. section 2 describes conversion module for Lombard effect. In section 3, we illustrate the results of experiment for the proposed acoustic driving simulator. Finally, we conclude in

1) Korea University, sjlee@ispl.korea.ac.kr
2) Seokyeong University, smkang@skuniv.ac.kr, corresponding author

section 4.

1. Acoustic driving simulator

The acoustic driving simulator proposed in this paper is intended to implement acoustic components from actual vehicle driving environments similar to indoor environment and it consists of hardware module and software module. The former is for outputting speech file and inputting a speaker's speech into the speech recognizer. The latter is to convert speeches from hardware module so that they are well matched with each driving environments. Figure 1 indicates overall structure for the proposed simulator. Hardware and software module are harmoniously operated depending on driving conditions. Driving noise and speech transferred from software module are displayed via speaker in hardware module. Signal outputted through 2 speakers is inputted into microphone in speech recognizer. When the speech is outputted, whether or not the Lombard effect conversion module is operated is determined by driving conditions.
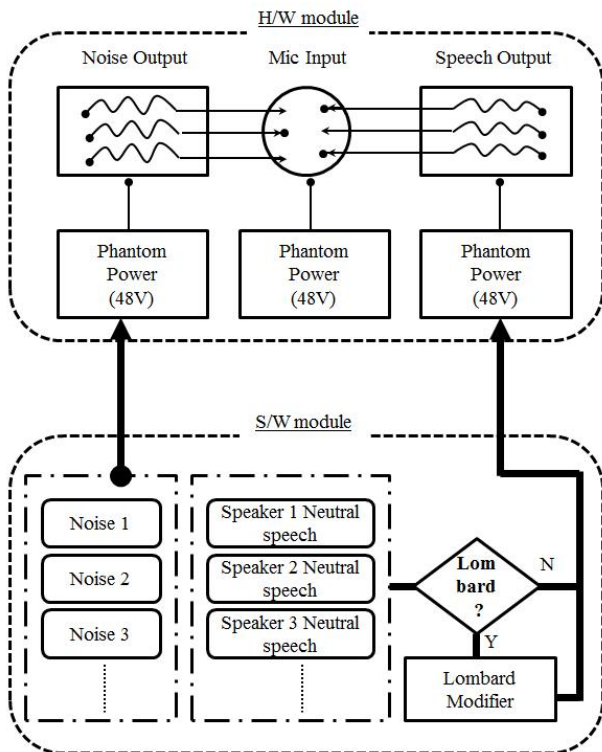

Figure 1. Overall structure of the proposed simulator

1.1 Hardware module

Hardware module to implement driving acoustic environment in a room consists of 2 speakers to output recorded driving noises and speaker's speech and a microphone to input sound from speakers into speech recognizer Among above equipments, speakers use monitor speakers that outputs sounds without distortion for all frequency ranges used by a recording studio in order to output recoded sounds and speeches similar to actual environment. Microphone uses directivity microphone that is usually using in in-car speech recognizer.

1.2 Software module

Software module to send speaker's speech and noise to hardware module brings about driving noise depending on driving conditions selected users and then converts recorded Neutral speech for each speaker using Lombard effect conversion module which will be described in the next section.

For instance, this module outputs idle noise and Neutral speech while a vehicle is idling. However when a vehicle is running at 60km/h and 120km/h, it outputs corresponding driving noise and converted Mimicked Lombard speech together.

2. Lombard effect conversion module

The proposed acoustic driving simulator in this paper is intended to input driving noise and speaker's speech into speech recognition simultaneously. As mentioned in the introduction, Lombard speech that is spoken when a vehicle is running has different acoustic feature from Neutral speech. The most ideal in-car speech recognizer evaluation is realized by inputting recorded speeches for various noise conditions into speech recognizer. However, it is very difficult in actual environments. In order to overcome these problems, therefore, the proposed Lombard effect module converts recorded speaker's Neutral speech in a studio into Lombard speech using speech signal processing techniques and then creates Mimicked Lombard speech data suitable for various driving conditions by a single Neutral speech data. For this conversion, conversion items are duration, intensity and fundamental frequency which significantly affect speech recognition rate among acoustic features from Lombard speech. More specific conversion values are set referring to existing researches [2]. Pellon and Hansen [4] analyzed characteristics of Lombard effect from SUSAS database (13 speakers, and 70 Utterances for each speaker) which include Lombard speech.

Although vowels have almost no variations in terms of speaking, semi-vowels are increased approximately by 3% for Lombard speech compared to the Neutral speech. Consonants

showed a little bit of reduction. Among this phenomenon, increased time of continuity from semi-vowels is a phenomenon coming from transition of formant, which becomes a factor to enhance intelligibility. Therefore, time of Mimicked Lombard speech which is converted through Lombard effect conversion module applies increment which is calculated by Formula 1 below.

$$T_{scale} = \frac{T_{Lomabrd} - T_{Neutral}}{T_{Neutral}} * 100 \qquad (1)$$

Where $T_{Lombard}$ and $T_{Neutral}$ indicate average values for speech time respectively which are analyzed from SUSAS database. Similarly in intensity for fundamental frequency, this creates Mimicked Lombard speech by applying increment calculated by Formula 1 to Lombard effect conversion module.
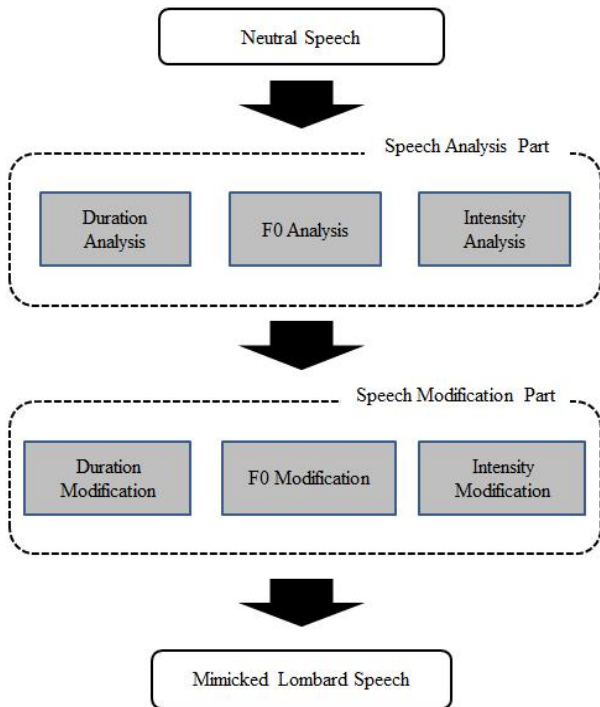
Figure 2. Flow diagram of the proposed
Lombard conversion module

Figure 2 shows a flow diagram for the proposed Lombard effect conversion module. First of all, it brings about speaker's Neutral speech that is recorded in a studio and extracts features for Duration, Fundamental frequency, and Intensity. Next, by applying the extracted features to each changed condition, it finally creates Mimicked Lombard speech.

Mimicked Lombard speech that converted Neutral speech using Lombard effect conversion module is acoustically most

similar to Lombard speech spoken by human. Figure 3 and Figure 4 show intensity by each time and fundamental Frequency for following cases: (1) Speech spoken under clean environment, (2) Speech spoken under noise environment and (3) Mimicked Lombard speech from using the proposed Lombard effect conversion module. Fundamental frequency, durations and intensity for both Lombard speech and Mimicked Lombard speech are all increased compared to Neutral speech and these 2 items are proved to be very similar.
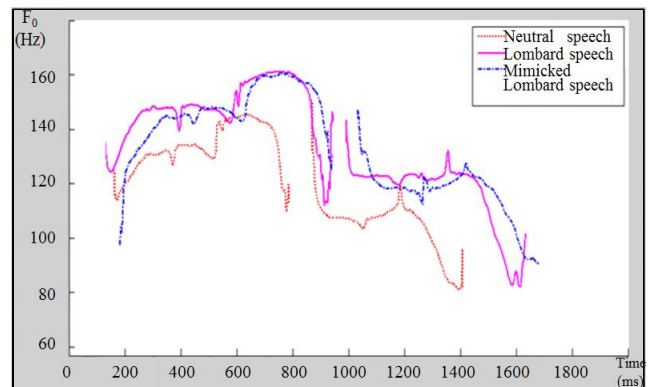
Figure 3. Fundamental frequency comparison between
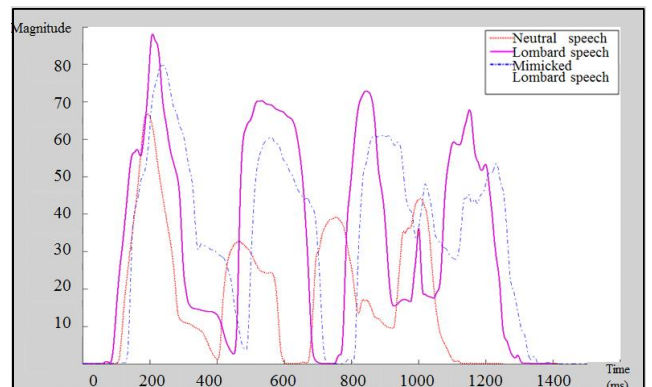Neutral speech, Lombard speech
and Mimicked Lombard speech

Figure 4. Intensity comparison between Neutral speech,
Lombard speech and Mimicked Lombard speech

3. Experiments and results

The purpose of the experiment in this paper is to verify whether or not speech recognition results performed by the proposed simulator are similar to that of actual vehicle. For this purpose, we recorded speaker's speech command data by PCM WAVE file in the driving vehicle and studio respectively. And a vehicle noise data affecting speech recognition results uses

recorded data from 'Gasoline fueled YF Sonata' of Hyundai Motors Corporation for each condition. Commands use for recognition used 50 words which consist of AV, BT, HELP, NAVI and ETC commands used for in-car speech recognition operation. Speech recognition module is adopted Sphinx engine which is developed by Carnegie Mellon University. Language model for training speech recognizer has used database suitably trained for above 50 words.

For the experiment, we first recorded driving noises for each condition driving vehicle and measured its intensity by noise measuring instrument and then performed speech recognition. After that, based on the results, we quantized them by 16 bit and then outputted commands of recorded driving noise and speaker using the proposed acoustic driving simulator. Directional microphone developed by AKG Corporation is used to input speech and driving noise in this experiment. The speech and noise are sent to PC using USB-type Audio Interface device. Studio monitor speaker made by BEHRITONE Corporation is used for outputting speech commands and driving noise. Finally, we calculated average value of recognized results by outputting 50 words 3 times for each driving condition.
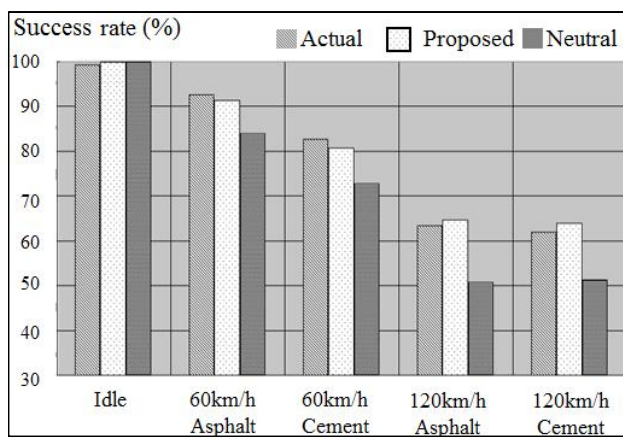

Figure 5. Results of the speech recognition rate
for each driving condition

Figure 5 shows results performed by this paper. As driving speed increases, the rate of recognition is proved to be more deteriorated in road of cement than that of asphalt. In addition, the result of speech recognition by outputting Neutral speech showed more different recognition results from actual driving conditions as driving speed increases.

On the contrary, the result of speech recognition using the proposed acoustic driving simulator indicated a very little difference of 1.7% showing approximately 3.3% for stopped

vehicle and 1.7% at 120km/h of driving, which proved that it is similar to actual driving conditions regardless of speed and surface of road.

Therefore, the proposed acoustic driving simulator showed good performance for virtual evaluation of speech recognizer in the vehicle, which in turn proved its effectiveness.

4. Conclusion

This paper proposed design of an acoustic driving simulator to enhance efficiency of time and cost by establishing most similar environments to actual driving environment to improve the performance of speech recognizer in the vehicle. This paper performed evaluation experiment for speech recognition under actual driving conditions and simulations to verify the proposed acoustic driving simulator environment. Command words used for speech recognizer and noise are outputted through each monitor speaker. The experiment result proved the effectiveness of the acoustic driving simulator showing similar speech recognition rate between actual driving conditions and acoustic driving simulator. The proposed acoustic driving simulator in this paper acquired similar speech recognition results to actual driving conditions simulating generation of mimicked Lombard speech and driving noise by improving conventional method which was required to check the recognition result by a method that human directed ride on a vehicle or by expensive device for speech recognition. Therefore, the proposed simulator is expected to result in reduction of time and cost.

The future work expects that more realistic evaluation is performed by adding various noise conditions such as road surface and weather.

Acknowledgement

References

[1] Lee, S. J. & Kang S. M. (2012). A study of automatic evaluation platform for speech recognition engine in the vehicle environment. *The Journal of Korea Information and Communications Society*, 37(7), 538-543.

[2] Huang, D. Y., Rahardja, S. & Ong, E. P. (2010). Lombard

effect Mimicking. *ISCA Workshop on Speech Synthesis*, 22-24.

[3] Wakao, A. & Takeda, K. & Itakura, F. (1996). Variability of Lombard effects under different noise conditions. *Spoken Language, ICSLP Proceedings,* 4, 2009-2012.

[4] Hansen, H. L. (2009). Analysis and compensation of Lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(2), 366-378.

[5] Bapineedu, G. (2004). *Analysis of Lombard effect speech and its application in speaker verification for imposter detection.* M.S. Thesis, Hyderabad University.

[6] Marvin, C. & McCallum & John, L. & Campbell (2004). Speech recognition and in-vehicle telematics devices: Potential reductions in driver distraction. *International Journal of Speech Technology,* 7, 25-33.

[7] Priscilla, L. (2008) The Lombard effect as a communicative phenomenon. *UC Berkeley Phonology Lab Annual Report.*

[8] Song, H. J. & Jeon, H. B. & Kim, H. S. (2009) Fast speaker adaptation based on Eigenspace-based MLLR using artificially distorted speech in car noise environment. *Journal of the Korean Society of Speech Sciences,* 4(1), 119-125.

• **Lee, Seongjae**
Department of Electrical Engineering
Korea University
5ka-1 Anam-dong, Seongbuk-gu,
Seoul, Korea
Tel: 02-3290-4995
Email: sjlee@ispl.korea.ac.kr

• **Kang, Sunmee** corresponding author
Department of Electronic Engineering
Seokyeong University
309 Bukakgwan Bldg. 124 Seokyeong-ro, Seongbuk-gu,
Seoul, Korea
Tel: 02-940-7737
Email: smkang@skuniv.ac.kr