INVESTIGATING THE ROLE OF THE LOMBARD REFLEX IN NON-AUDIBLE MURMUR (NAM) RECOGNITION

Panikos Heracleous, Tomomi Kaino, Hiroshi Saruwatari, Kiyohiro Shikano

Nara Institute of Science and Technology, Japan
e-mail:{panikos,tomomi-k,sawatari,shikano}@is.naist.jp

10.21437/Interspeech.2005-251

Abstract

In this paper, we report non-audible murmur (NAM) recognition results in noisy environments and investigate the effect of the Lombard reflex on non-audible murmur recognition. Non-Audible murmur is speech uttered very quietly and captured through body tissue by a special acoustic sensor (e.g., NAM microphone). A system based on non-audible murmur recognition can be applied in cases when privacy is preferable in human-machine communication. Moreover, due to direct body-transmission, the environmental noises do not affect the performance markedly. Previously, we reported non-audible murmur automatic recognition in a clean environment with very promising results. We also carried out experiments using clean models and simulated noisy data, showing that the performance did not change significantly. Using, however, real noisy test data, the performance decreased markedly. To investigate this problem, we studied the Lombard reflex and conducted non-audible murmur recognition experiments using Lombard data. Results show, that Lombard reflex affects non-audible murmur recognition.

1. Introduction

Non-Audible murmur (NAM) is very quietly uttered speech that cannot be heard by listeners near the talker. It is captured using a NAM microphone [1], which is a special acoustic sensor attached behind the talker's ear. A NAM microphone is a body-conductive acoustic transducer, in which speech is captured directly from the talker's body through tissue or bone. Thus, such a transducer shows high robustness against noise and can capture voices with a very low intensity. Similar studies have been proposed by Zheng et al. [2], Graciarena et al. [3], and Jou et al [4] for noise robust speech recognition or soft whisper speech recognition.

Similarly to whisper speech, non-audible murmur is unvoiced speech produced by vocal cords not vibrating and does not incorporate any fundamental (F0) frequency. Moreover, body tissue and loss of lip radiation acts as a low-pass filter and the high-frequency components are attenuated. However, the non-audible murmur spectral components still provide sufficient information to distinguish and recognize sounds accurately. To realize this, new hidden Markov models (HMMs) have to be trained using non-audible murmur data.

Previously, we reported HMM-based non-audible murmur automatic recognition with very promising results [5, 6]. More specifically, using a small amount of training data and adaptation techniques, we achieved a 93.9% word accuracy for a 20k dictation task in a clean environment. We also reported experiments for integrated non-audible murmur recognition and audible speech recognition using a NAM microphone [7].

In this paper, we investigate non-audible murmur recognition in noisy environments. However, because of the nature of non-audible murmur (e.g., privacy), it is of high importance to also deal with noisy conditions, such as background speech and office noise, in automatic non-audible recognition. We carried out experiments using simulated noisy test data and data recorded under noisy conditions. Although using simulated noisy data the performance did not decrease significantly compared with that of the clean case, using real noisy data the performance decreased markedly. To investigate this problem, we studied the role of the Lombard reflex [8, 9] in non-audible murmur recognition and conducted experiments using Lombard non-audible murmur data. Results showed, that the Lombard reflex seriously affects the performance of non-audible murmur.

2. Non-Audible murmur recognition in noisy environments

In this section, we report experimental results for non-audible murmur recognition in noisy environments. We carried out two types of experiment. In the first experiment, noise recorded using a silicon NAM microphone was superimposed on clean test data and recognition was performed using HMMs trained with clean non-audible murmur data. In the second experiment, various noises were played back at different levels and the test data were uttered by a speaker and recorded under those conditions.

The recognition engine used was the Julius 20k vocabulary Japanese dictation toolkit. The initial models

Table 1: System specifications

Sampling frequency	16 kHz
Frame length	25 ms
Frame period	10 ms
Pre-emphasis	$1 - 0.97z^{-1}$
Feature vectors	12-order MFCC, 12-order ΔMFCCs 1-order ΔE
HMM	PTM , 3000 states
Training data	JNAS/Non-audible murmur
Test data	Non-audible murmur

were speaker-independent, gender-independent, 3000-state phonetic tied mixture (PTM) HMMs, trained with the JNAS database and the feature vectors were of length 25 (12 MFCC, 12 ΔMFCC, ΔE). The non-audible murmur HMMs were trained using a combination of supervised 128-class regression tree MLLR [10] and MAP [11] adaptation methods. Table 1 shows the system specifications.

2.1. Speaker-dependent experiment using simulated noisy data

In this experiment, office noise was played back at different levels (dBA) and recorded using a NAM microphone attached to a female talker. We recorded noises at 50 dBA and 60 dBA levels. The recorded noises were then superimposed on 24 clean non-audible murmur utterances, uttered by the same female speaker, to create the simulated noisy data. The acoustic models were trained using 100 non-audible murmur utterances recorded in a clean environment.

The results showed that the performance remained almost equal to that of the clean case when noise was superimposed on clean test data and recognition was performed using clean HMMs. More specifically, we achieved 83.7%, 82.9% and 80.9% word accuracies for the clean case, the 50 dBA noise level, and the 60 dBA noise level, respectively.

2.2. Speaker-dependent experiments using real noisy data

In this section, we report experimental results for non-audible murmur recognition using real noisy database. The noisy test data were recorded in an environment, where different types of noise were playing back at 50 dBA and 60 dBA levels, while a speaker was uttering the test data. Four types of noise were used (office, car, poster, and crowd). For each noise and each level 24 utterances were recorded.

Figure 1 shows the obtained results when using office noise in comparison with the case when the same noise

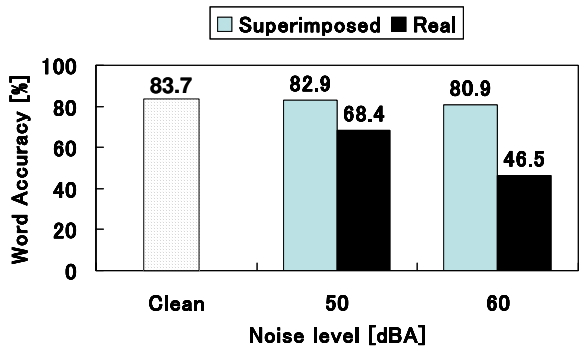

Figure 1: Non-Audible murmur recognition using noisy test data (office noise)

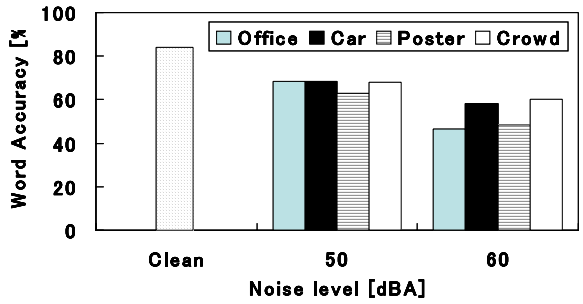

Figure 2: Non-Audible murmur recognition using various types of noise

was superimposed on the clean data. As can be seen, using real noisy test data, the performance decreases. Namely, at the 50 dBA noise level the obtained word accuracy was 68.4% and at the 60 dBA noise level 46.5%.

Figure 2 shows the word accuracies for the four types of noise. The results are similar to the previous ones. With increasing noise level, word accuracy decreases significantly. For the clean case we achieved an 83.7% word accuracy, for the 50 dBA noise level a 66.9% word accuracy on average, and for the 60 dBA noise level a 53.3% word accuracy on average. In the case of car and crowd noises, the difference between the 50 dBA and 60 dBA performances is not very large. In the case of poster and office noises, the difference is larger.

Although, the performance using real noisy data is not markedly low and non-audible recognition is still possible, further investigations are necessary. In several studies, a negative impact effect of the Lombard reflex on automatic recognizers for normal speech has been reported. It is possible, therefore, that the degradations in word accuracy for non-audible murmur recognition when using real noisy data, are also related to the Lombard reflex. To realize this, we also addressed the Lombard reflex problem.

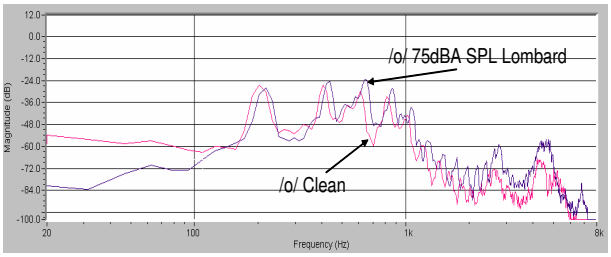

Figure 3: Power spectrum of clean vowel /O/ and Lombard vowel /O/

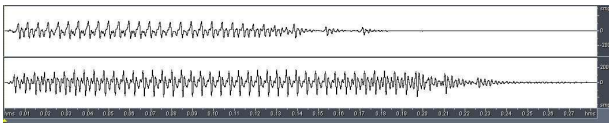

Figure 4: Waveform of clean vowel /O/ (upper) and Lombard vowel /O/

3. Role of Lombard reflex in non-audible murmur recognition

When speech is produced in noisy environments, speech production is modified leading to the Lombard reflex. Due to the reduced auditory feedback, the talker attempts to increase the intelligibility of his speech, and during this process several speech characteristics change. More specifically, speech intensity increases, fundamental frequency (F0) and formants shift, vowel durations increase and the spectral tilt changes. As a result of these modifications, the performance of a speech recognizer decreases due to the mismatch between the training and testing conditions.

To show the effect of the Lombard reflex, Lombard speech is usually used, which is a clean speech uttered while the speaker listens to noise through headphones or earphones. Even, though, Lombard speech does not contain noise components, modifications in speech characteristics can be realized.

Figure 3 shows the power spectrum of a normal-speech clean vowel /O/ and a Lombard vowel /O/ recorded while listening to office noise through headphones at 75 dBA noise level. The figure clearly shows the modifications leading to the Lombard reflex; power increased, formants shifted and spectral tilt changed. Figure 4 shows the waveforms of the clean and Lombard /O/ vowels. As can be seen, the duration and amplitude of the Lombard vowel also increased. These differences in the spectra cause feature distortions (e.g., Mel Frequency Cepstral Coefficients (MFCC) distortions), and acoustic models trained using clean speech might fail to correctly match speech affected by the Lombard reflex.

Figure 5 shows the waveform, spectrogram, and FO

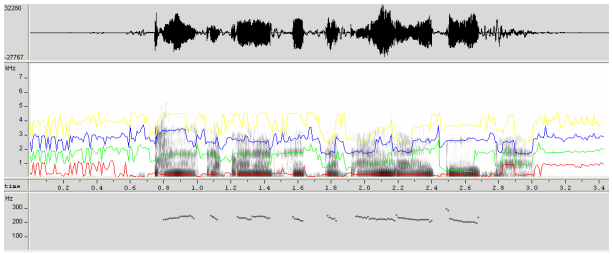

Figure 5: Lombard non-audible murmur recorded at 80 dBA

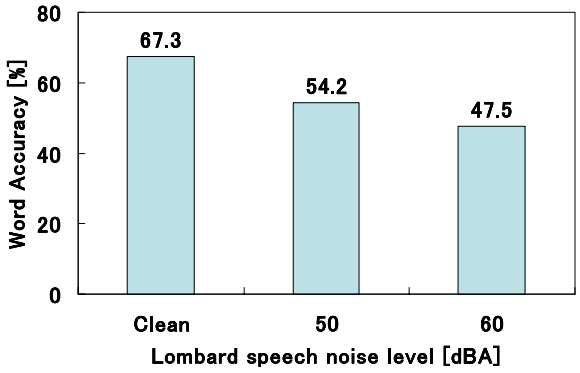

Figure 6: Non-Audible murmur recognition using Lombard data

contour of a Lombard non-audible utterance recorded at 80 dBA. As can be seen, this Lombard non-audible murmur speech has characteristics similar to those of normal speech. Therefore, when non-audible murmur recognition is performed in noisy environments, the produced non-audible murmur characteristics are different than those of the non-audible murmur used in the training. As a result, the performance is degraded, even though the NAM microphone can capture non-audible murmur without a high sensitivity to environmental noise.

To show the effect of the Lombard reflex on non-audible murmur recognition, we carried out an experiment using Lombard non-audible murmur test data. The data were recorded in an anechoic room, while the speaker was listening to office noise through headphones. Since we used high-quality headphones, we assumed that no noise from the headphones was added to the recorded data. We recorded 24 clean utterances, 24 utterances at 50 dBA and 24 utterances at 60 dBA noise levels. The acoustic models used were trained with clean non-audible murmur data using 50 utterances and MLLR adaptation.

Figure 6 shows the obtained results and the effect of the Lombard reflex on non-audible murmur recognition. Using clean test data, we achieved a 67.3% word accuracy, using 50 dBA Lombard data a 54.2% word accuracy,

and using 60 dBA Lombard data a 47.5% word accuracy. These results show an analogy between the experiments using real noisy data and the experiment using Lombard data. In both cases, the performances decreased almost equally.

In non-audible murmur phenomena, the Lombard reflex is also present when there is no masking noise. However, due to the very low intensity of non-audible murmur, speakers might not hear their own voice. To make their voice audible, they increase their vocal levels, and as a result, non-audible murmur changes to voicing.

4. Conclusion

In this paper, we presented non-audible murmur recognition in noisy environments using NAM microphones. A NAM microphone is a special acoustic device attached behind the talker's ear, which can capture very quietly uttered speech. Non-Audible murmur recognition can be used when privacy in human-machine communication is desired. Since non-audible murmur is captured directly from the body, it is less sensitive to environmental noises. To show this, we carried out experiments using simulated and real noisy data. Using simulated noisy data at 50 dBA and 60 dBA noise levels, the non-audible murmur recognition performance was almost equal to that of the clean case. Using, however, data recorded in noisy environments, the performance decreased. To investigate the possible reasons for this, we studied the role of the Lombard effect in non-audible murmur recognition and we carried out an experiment using Lombard data. The results showed that the Lombard reflex has a negative impact effect on non-audible murmur recognition. Due to the speech production modifications, the non-audible murmur characteristics under Lombard conditions are changed and show a high similarity to normal speech. Due to this fact, a mismatch appears between the training and testing conditions and the performance decreases. As future work, we plan to investigate methods of decreasing the effect of the Lombard reflex on non-audible murmur recognition. A possible solution might be the adaptation of clean acoustic models to several Lombard conditions.

5. Acknowledgment

This research is supported by the *NAIST 21st Century Center of Excellence (COE) Program*.

6. References

[1] Y. Nakajima, H. Kashioka, K. Shikano, N. Campbell, " Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin", *Proceedings of ICASSP*, pp. 708–711, 2003.

[2] Y. Zheng, Z. Liu, Z. Shang, M. Sinclair, J. Droppo, L. Deng, A. Acero, Z. Huang, " Air- and Bone-Conductive Integrated Microphones for Robust Speech Detection and Enhancement", *Proceedings of ASRU*, pp. 249–253, 2003.

[3] M. Graciarena, H. Franco, K. Sonmez, H. Bratt, "Combining Standard and Throat Microphones for Robust Speech Recognition", *IEEE Signal Processing Letters*, Vol. 10, No 3, pp.72–74, 2003.

[4] S. C. Jou, T. Schultz, Alex Waibel, "Adaptation for Soft Whisper Recognition Using a Throat Microphone", *Proceedings of ICSLP*, 2004.

[5] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano, " Accurate Hidden Markov Models for Non-Audible Murmur (NAM) Recognition Based on Iterative Supervised Adaptation", *Proceedings of ASRU*, pp. 73–76, 2003.

[6] P. Heracleous, Y. Nakajima, A. Lee, H. Saruwatari, K. Shikano, " Non-Audible Murmur (NAM) Recognition Using a Stethoscopic NAM microphone", *Proceedings of ICLP*, pp. 1469–1472, 2004.

[7] P. Heracleous, Y.Nakajima, A. Lee, H. Saruwatari, K. Shikano, "Audible (normal) speech and inaudible murmur recognition using NAM microphone", *Proceedings of EUSIPCO*, pp. 329–332, 2004.

[8] Junqua J-C, "The Lombard Reflex and its Role on Human Listeners and Automatic Speech Recognizers", *J. Acoust. Soc. Am.*, Vol. 1 pp. 510–524, 1993.

[9] A. Wakao, K. Takeda, F. Itakura, "Variability of Lombard Effects Under Different Noise Conditions", *Proceedings of ICSLP*, pp. 2009–2012, 1996.

[10] C. J. Leggetter, C. Woodland, "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models", *Computer Speech and Language*, Vol. 9, pp. 171–185, 1995.

[11] C.H. Lee, C.H. Lin, and B.H. Juang, "A study on speaker adaptation of the parameters of continuous density hidden Markov models", *IEEE transactions Signal Processing*, Vol. 39, pp. 806–814, 1991.

[12] P.C. Woodland, D. Pye, M.J.F. Gales, "Iterative Unsupervised Adaptation Using Maximum Likelihood Linear Regression", *Proceedings of ICSLP*, pp. 1133–1136, 1996.