Evaluation of Measures Applied to Monitor the Lombard Speech Signal in the Presence of Noise

Gražina Korvel[1], Povilas Treigys[1], and Bożena Kostek[2]

[1]Institute of Data Science and Digital Technologies, Vilnius University, Vilnius, Lithuania
[3]Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Gdansk, Poland

Introduction

The noisy environment decreases the quality and intelligibility of the speech signal. Contrarily, the so-called Lombard effect occurring in a person's talk while exposed to environmental noise enhances the speech signal. Even though the noise may contaminate Lombard speech, it remains more intelligible. Our long-term goal is to build a machine learning-based system for generating speech with the Lombard effect that can automatically adapt to noise inference. In this research, a method enabling to monitor of the Lombard speech signal in the presence of noise is to be proposed. For this purpose, the variation of frequency characteristics of Lombard speech at different noise distortions is investigated. The signal frequency domain is considered in the form of frequency tracks containing the location of fundamental frequency and formats. To quantify the effect of noise on the Lombard effect, an average formant track error is calculated.

Experimental setup

- ✓ The effect of noise is investigated at varying levels of SNR, from -10 dB to 40 dB.
- ✓ For environmental noise, real-life noise recordings are employed.
- ✓ Experiments are carried out on the recordings made in the studio with a room with acoustic treatment.
- ✓ To obtain the Lombard effect while speaking, closed headphones playing back the interfering noise are used.
- ✓ The recording scenario includes both reading sentences and a conversation between two people.

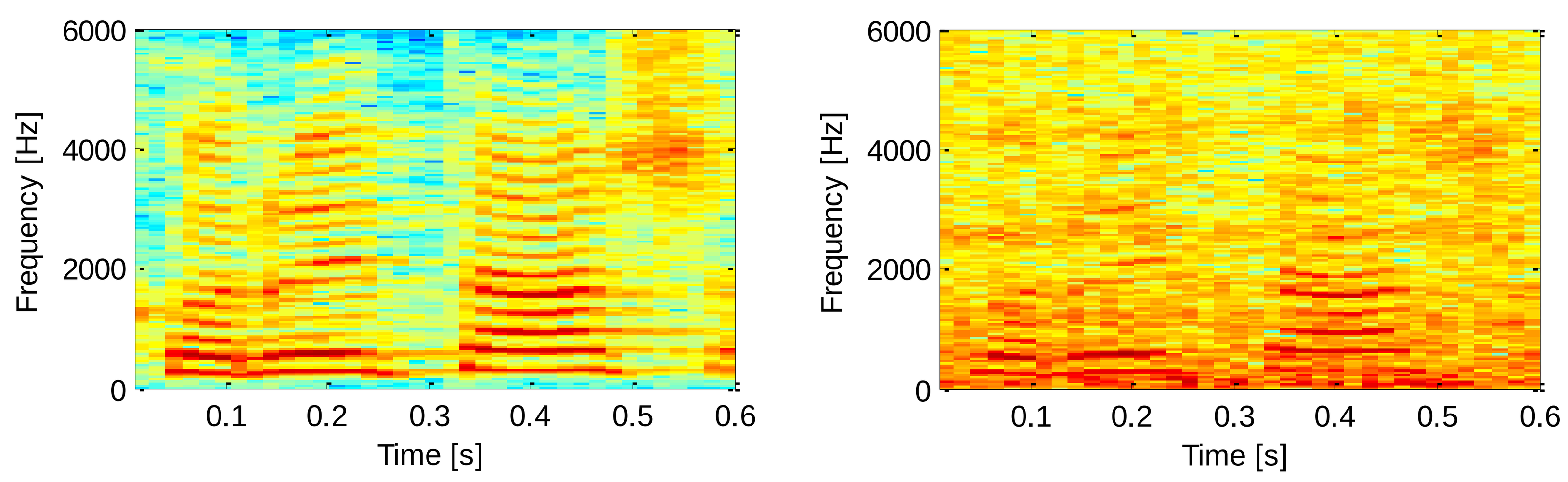

Fig. 1. The spectrogram of clean (the left side of the figure) and the noisy (the right side) Lombard speech signal (speech segment corrupted by nonstationary street noise at 0 dB SNR)

Investigation of the noise interference effect on Lombard speech:

- ✓ Rapidly changing areas of speech such as voiced/unvoiced transitions are investigated.
- ✓ To indicate such changes in spectral energy, frequency tracks are estimated.

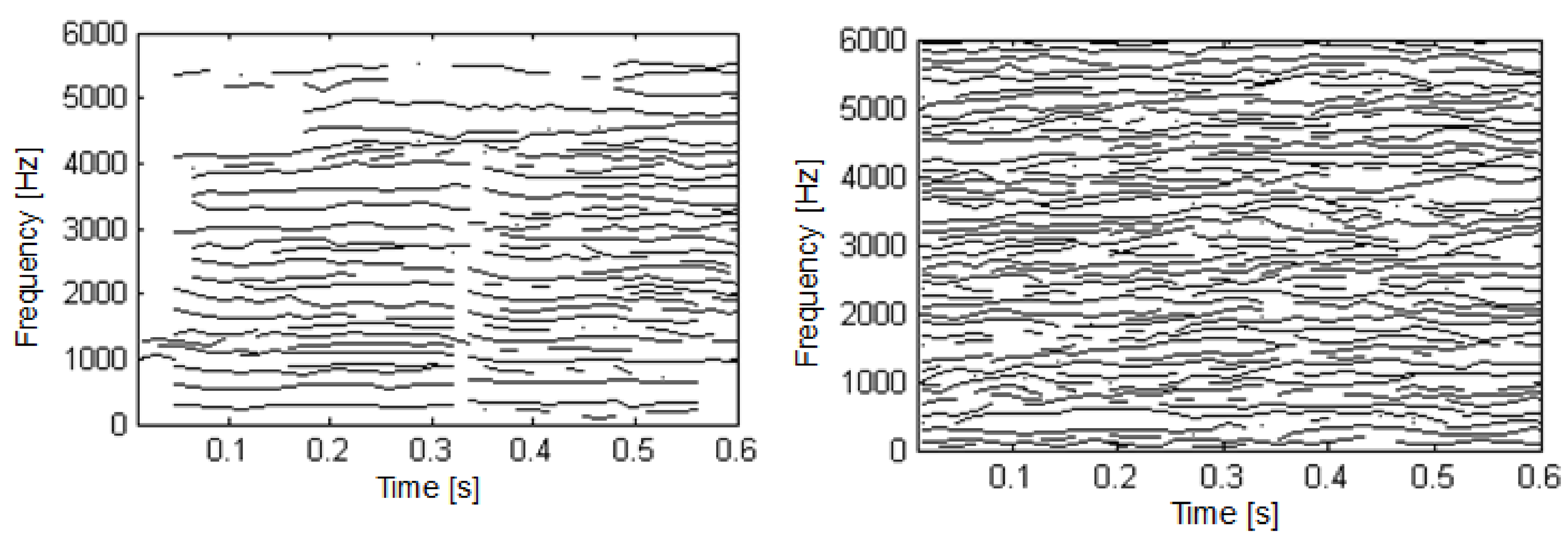

Fig. 2. The estimated frequency tracks of clean (the left side of the figure) and the noisy (the right side) Lombard speech signal (speech segment corrupted by nonstationary street noise at 0 dB SNR)

Measures applied

1. Structural Similarity Index Measure (SSIM)

$$S(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}$$

Where $l(x,y)$ is the luminance comparison function, $c(x,y)$ is the contrast comparison function, $s(x,y)$ is the structure comparison function, (in this research $\alpha = \beta = \gamma = 1$)

2. The spectrum-based parameters

- Spectral Entropy
- Spectral RollOff
- Spectral Brightness

3. Temporal characteristics

- Temporal centroid
- Zero crossing (ZC) rate
- Root mean square (RMS) energy

Experimental results

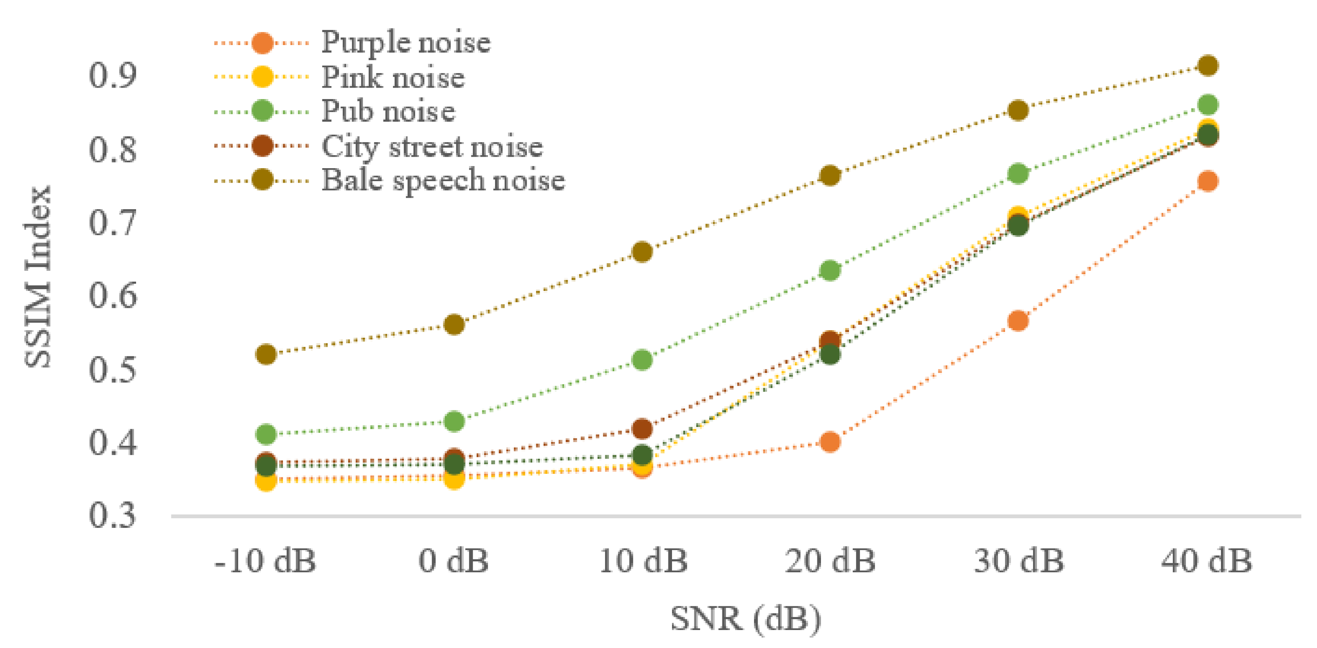

Fig. 3. The indexes of the SSIM for different recordings

Table 1. The normalized spectral characteristics of the noise signals

	Spectral Entropy	Spectral RollOff	Spectral Brightness
Purple noise	1	1	1
Pink noise	0.76	0.29	0.15
Pub noise	0.68	0.10	0.05
City street noise	0.75	0.16	0.13
Bale speech noise	0.63	0.09	0.03
Rain noise	0.78	0.19	0.15

Table 2. The normalized temporal characteristics of noise signals

	Temporal centroid	ZC rate	RMS energy
Purple noise	0.97	1.00	0.76
Pink noise	0.97	0.28	0.93
Pub noise	0.98	0.15	1.00
City street noise	0.99	0.15	0.42
Bale speech noise	1.00	0.09	0.89
Rain noise	0.97	0.25	0.51

The recommendations for machine learning-based system

To create an efficient method of noise profiling, it is critical to identify the sound characteristics specific to a given type of sound. It was demonstrated that stable and predictable noise profiling is possible using spectral noise characteristics. Spectral entropy, which provides a measure of spectrum irregularity, reflects the unpredictability of these signals. This may have led to lower values of the SSIM indexes for these noises. Also, the amount of high-frequency information, which Spectral Brightness and RollOff reflect, has a direct impact on the SSIM indexes.

Acknowledgment

This research is funded by the European Social Fund under the No 09.3.3-LMT-K-712 “Development of Competences of Scientists, other Researchers and Students through Practical Research Activities” measure.