

Title	ロンバード効果に着想を得た雑音中での音声了解度および自然性の向上
Author(s)	NGO, THUANVAN
Citation	
Issue Date	2020-09
Type	Thesis or Dissertation
Text version	ETD
URL	<a href="http://hdl.handle.net/10119/16995">http://hdl.handle.net/10119/16995</a>
Rights	
Description	Supervisor:赤木 正人, 先端科学技術研究科, 博士

氏名	NGO, Thuan Van		
学位の種類	博士(情報科学)		
学位記番号	博情第 435 号		
学位授与年月日	令和 2 年 9 月 24 日		
論文題目	Improvement of intelligibility and naturalness of speech in noise inspired by Lombard effect		
論文審査委員	主査	赤木 正人	北陸先端科学技術大学院大学 教授
		党 建武	同 教授
		鶴木 祐史	同 教授
		河原 英紀	和歌山大学 名誉教授
		坂本 修一	東北大学 教授

## 論文の内容の要旨

### Research content and purpose

In public announcements in train stations or airports, the presence of noise often smears presented speech, thus makes it hard for listeners to understand it. By reducing noise throughout the presentation, speech is still intelligible and natural to listeners. However, this seems impractical due to complex architectures of these places and costly installed devices. Besides, it is practical and less expensive to enhance speech itself before presenting to complement degradation in intelligibility and naturalness by the smears.

Lombard speech is intelligible speech produced by humans in noise due to the Lombard effect. Investigation of Lombard speech could reveal essential features to increase speech intelligibility and naturalness. Therefore, the purpose of this research was to improve the intelligibility and naturalness of speech in noise using conversion rules inspired by the Lombard effect. It came up with two sub-goals: (I) obtaining feature understanding and control which contributes to the intelligibility of Lombard speech under noise-level-varying and various noise and (II) identifying and applying the effective feature control methods for exceeding the intelligibility and naturalness of Lombard speech. From the previous research and the properties of Lombard speech that varies with noise levels and noise types, three problems arose to cover the search space (features, noise levels, feature variations, SNRs, and spectral-varied noise) for finding features and applying them.

(1) Contribution of acoustic features of Lombard speech has no consideration of their articulatory features at one noise level: For this problem, the modification of acoustic features of Lombard speech was often done without any considering the articulatory changes or challenging to be obtained in acoustical levels. The multiple contributions of the features to the intelligibility and naturalness were unclear.

(2) Control and contribution of acoustic features of Lombard speech in multiple noise levels of backgrounds: In this problem, acoustic features contributing to Lombard speech in various noise levels were difficult to be modeled and controlled by the conventional methods.

(3) Unclear effective features to the intelligibility and naturalness of speech varying noise levels and various types of noise: This problem followed the second problem. Acoustic features, when varying that contribute to the intelligibility of speech in noise, remained unclear. Recent studies differently reported effective features for the intelligibility and naturalness of speech in noise. Thus, the precise set of effective features was unidentified.

Thus, the investigation has three steps: (1) Mimicking Lombard speech by controlling articulatory and acoustic features, (2) Effective features for the intelligibility and naturalness of speech in noise, (3) Application to improve the intelligibility of speech under noisy reverberant conditions.

Following these steps, this study obtained the articulatory and acoustical controls in mimicking Lombard speech. The contributive features, including spectral tilts,  $f_o$ , and formants, were also explored in the first step. In the second step, the effective features for intelligibility and naturalness in all kinds of noise were identified. In the final step, the effective features were successfully applied to increase intelligibility and naturalness of speech under noisy reverberant conditions with a pair of effective time-frequency features.

Consequently, the originality of this study was firstly the leading investigation of the contribution of articulatory features to the intelligibility of speech in noise. Secondly, it was mimicking Lombard speech under various noise levels. Finally, this study presented a brutal-force method for extracting the effective acoustic features to vary to increase the intelligibility of speech in noise. Besides, the present study obtained novelty in two aspects. Firstly, it was the concept of applying rule-based methods and the Lombard effect model for the rule generation model to mimic Lombard speech concerning multiple noise levels. Secondly, it was the concept based on the modulation spectrum and modulation transfer function concepts in relationship with listening tests to identify the effective features to increase speech intelligibility and naturalness in noise.

As a scientific orientation, this research can enlighten the fields of speech enhancements, objective intelligibility measurements, voice conversion, and synthesis. Especially, it provides essential, necessary information for the areas of speech enhancement engineering.

Keywords: Lombard speech, mimicking, articulatory features, acoustic features, effective features, speech intelligibility and naturalness in noise.

### **Research accomplishment (publications)**

Journal (Peer reviewed)

[1] Thuanvan Ngo, Rieko Kubo, Daisuke Morikawa and Masato Akagi, "Acoustical Analyses of Tendencies of Intelligibility in Lombard Speech with Different Background Noise Levels," *Journal of Signal Processing*, vol. 21, no. 4, pp. 171–174, 2017.

[2] Thuanvan Ngo, Masato Akagi, and Peter Birkholz, "Effect of articulatory and acoustic features on the intelligibility of speech in noise: An articulatory synthesis study," *Speech Communication*, vol. 117, pp. 13–20, 2020.

[3] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Mimicking Lombard effect: An analysis and reconstruction,” IEICE Trans. Inf. & Syst., vol. E103.D, no. 5, pp. 1108–1117, 2020.

International conference (Peer reviewed)

[4] Thuanvan Ngo, Rieko Kubo, Daisuke Morikawa and Masato Akagi, “Acoustical analyses of Lombard speech by different background noise levels for tendencies of intelligibility,” NCSP’17, 2017, pp. 309–312.

[5] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Evaluation of the Lombard effect model on synthesizing Lombard speech in varying noise level environments with limited data,” APSIPA ASC, 2019, pp. 133–137.

Domestic conference (Non peer reviewed)

[6] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Intelligibility improving and naturalness preserving for evacuation speech in noisy environments,” IEICE Technical Report, Engineering Acoustics, 2019.

[7] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Improved quality and intelligibility of mimicking Lombard speech by source-filter and coarticulation model-based synthesis,” ASJ Spring Meeting, 2019.

[8] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Speaker-independent control model for mimicking Lombard speech uttered in background noises with various levels,” ASJ Spring Meeting, September, 2018, pp. 1371-1374.

[9] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Acoustical control method for increasing intelligibility based on Lombard speech uttered in background noises with various levels,” ASJ Fall Meeting, 2018, pp. 313-316.

[10] Thuanvan Ngo, Rieko Kubo, and Masato Akagi, “Acoustical rules for mimicking Lombard speech produced in a various noise level background,” IEICE Technical Report, Engineering Acoustics, vol. 117, no. 170, 2017.

## 論文審査の結果の要旨

本論文は、雑音中でも了解度が高いロンバード音声に着想を得た音声変換規則を使用することで、雑音中での音声アナウンスの了解度と自然性を改善することを目的とした研究報告である。

駅や空港などの公共空間での音声アナウンスは、雑音の存在によりしばしば音声の不鮮明になり、聞き手がそれを理解することが難しくなる。これを補うために、雑音で音声不明瞭になる過程をあらかじめ想定しておき、この過程を見越したうえで聴取点での音声了解度を最大化するように、音声を提示前に操作する手法が、近年脚光を浴びている。他方ヒトは、雑音中での発話において、自身が発した音声を自らの耳で聞き、その聞こえをフィードバック情報として利用することで、自身にとってより聞こえ易い音声となるように発話を適応的に制御することができる。この能力はロンバード効果と呼ばれる。

本研究では、このような有効な特性を持つヒトのロンバード効果に着想を得た音声変換規則を提案し、雑音中の聴取点での音声了解度と自然性を改善した。具体的には、ロンバード効果により生成されるロンバード音声に対して、次の課題を設定し研究を実施した。(1) 音声の調音および音響特徴を制御することによるロンバード音声の模倣、(2) 雑音レベルが変化するさまざまな雑音下でのロンバード音声の了解度に寄与する特徴の理解と制御方法の取得、(3) ヒトが発話したロンバード音声の了解度を超えるための効果的な特徴制御方法の特定。

研究の結果、課題(1)については、雑音のある環境での音声発話における調音機能の寄与を調査することで、ロンバード音声を模倣する際の調音および音響制御方法を得た。課題(2)については、複数の雑音レベルにおけるロンバード音声を模倣するために、スペクトル傾斜、 $f_0$ 、フォルマントなどの音響特徴に対する変形規則の生成を行い、ヒトが発話したロンバード音声と同等の了解度および自然性を持つ音声を生成した。さらに課題(3)については、より了解度向上に効果的である特徴として、変調スペクトルと変調伝達関数の概念にもとづいた特徴を提案し、聴取実験により音声了解度と自然性を向上させる効果的な機能を特定した。そして、これらの知見を統合して了解度向上手法を提案し、複数の言語音声に対して雑音中での了解度向上を実現させた。

以上のように、本研究は第一にヒトの優れた能力(ロンバード音声)を模倣し、第二にそれを超える手法を提案することで、雑音中での音声了解度を向上させる手法実現したものであり、学術的に貢献するところが大きい。よって博士(情報科学)の学位論文として十分価値あるものと認めた。