Analysis of Lombard Speech using Excitation Source Information

G. Bapineedu, B. Avinash, Suryakanth V. Gangashetty and B. Yegnanarayana

International Institute of Information Technology, Hyderabad, India

{bapineedu,avinashb}@research.iiit.ac.in, {svg,yegna}@iiit.ac.in

Abstract

This paper examines the Lombard effect on the excitation features in speech production. These features correspond mostly to the acoustic features at subsegmental ($<$ pitch period) level. The instantaneous fundamental frequency F_0 (i.e., pitch), the strength of excitation at the instants of significant excitation and a loudness measure reflecting the sharpness of the impulse-like excitation around epochs are used to represent the excitation features at the subsegmental level. The Lombard effect influences the pitch and the loudness. The extent of Lombard effect on speech depends on the nature and level (or intensity) of the external feedback that causes the Lombard effect.

Index Terms: Lombard effect, excitation source, loudness

1. Introduction

For effective communication a speaker relies on the auditory self feedback of the speech of his/her own voice. If the self feedback is hampered, then the articulatory movement of the speech production process and the corresponding acoustic signals are affected, thus resulting in speech which the listener perceives as not normal. The self feedback can be hampered by loss of hearing or by external environmental factors like noise or unwanted speech or music. The speaker tries to adjust the articulatory and acoustic parameter to produce speech as intelligible as possible to others. This psychological effect on speaker for producing speech in the presence of noise is termed as Lombard effect [1]. It is the changes in the articulatory movement that try to ensure better communication in noisy environment. The speech produced in such cases is different from the speech produced in silent conditions. Thus the Lombard effect speech not only affects the intelligibility in speech communication, but it also affects the performance of automatic speech and speaker recognition systems.

The Lombard effect on speech depends on the environment, speaker and the context of speech communication. Since the features extracted from Lombard effect speech will be different from normal speech, the extracted features have to be compensated, for using speech systems designed for normal speech. Likewise, to improve the intelligibility, the Lombard effect speech needs to be enhanced by modifying the parameters/features. These modifications at signal or parameter or feature levels have to be done by determining the level of compensation required. The first step in developing this process of modification is the analysis of features of Lombard effect speech.

Several studies have been reported on the analysis of Lombard effect speech [2][3][4]. The studies show that the durations of vowels generally increase, and the durations of unvoiced sounds generally decrease due to Lombard effect, in comparison with normal speech. The Lombard effect also produces louder speech, which results in decrease in the spectral tilt, with more energy in the high frequency region of the spectrum. It is also observed that the pitch or fundamental frequency (F_0) and the first formant in some vowels also increase due to Lombard effect. Some studies reported migration of energy from low and high frequency to middle range for vowels, and from low to high frequency for unvoiced stops and fricatives [3]. In some cases of Lombard effect speech, certain phonemes like /t/, /p/ and /f/ occurring at the end of a word are deleted, and aspiration after /m/ and /n/ increases [5].

Analysis of Lombard effect speech signal is based on time domain properties such as duration of voiced and unvoiced segments, and spectral domain properties such as spectral tilt and formants. The only source parameter that is used extensively is the variation of the fundamental frequency (F_0). On the other hand, perceptually several factors are noticed like loudness, stress and intensity. But very few attempts have been made in reporting the changes in the excitation source information due to Lombard effect. The objective of the present study is to analyze the Lombard effect speech in terms of features of excitation source in speech production, when the speech is produced under different types and levels of degradation. The noise signals are presented through headphones to the speaker, and the Lombard effect speech is recorded using a close speaking microphone. Hereafter, the noise which causes Lombard effect is termed as an external feedback, since normal speech can also be used as noise.

In Section 2 the features used for analysis of Lombard effect speech are described. In particular, the instantaneous fundamental frequency (F_0), the strength of excitation at epoch, a new measure of loudness, and durations of voiced and unvoiced regions are discussed. In Section 3 the changes in these features due to Lombard effect in relation to the features in normal speech are discussed. Section 4 discusses the analysis of Lombard effect speech for different types of noises at different levels. Section 5 describes the results of perceptual studies. Section 6 gives a summary of the studies reported in this paper.

2. Features of Lombard effect speech

In this section we propose a set of excitation source features along with duration of voiced/unvoiced segments for analysis of the Lombard effect speech. Here we use the excitation source features of instantaneous F_0 (pitch), strength of excitation at the epochs and a measure of loudness. We also describe the change in duration of the voiced and nonvoiced regions due to Lombard effect.

2.1. Fundamental frequency

The fundamental frequency (F_0) of speech varies from person to person, and also on the speech spoken by the person. Recently a method is proposed to extract the instantaneous F_0 from the zero-frequency filtered signal [6]. A zero-frequency

resonator is an all-pole system with two poles on the positive real axis in the z-plane [7]. Filtering the speech signal using a zero-frequency resonator emphasizes the characteristics of excitation, especially due to the abruptness of the closure of the glottis during vibration of the vocal folds. The system function for such a resonator is given by

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2}}, \quad (1)$$

where $a_1 = -2$ and $a_2 = 1$. This resonator removes the effects of the vocal tract system, as the resonances of the system are located at much higher frequencies. The speech signal is passed through the resonator twice so as to reduce the effect of all the resonances of the vocal tract system. Passing through the zero-frequency filter once is equal to successive integration twice, and passing the signal through the resonator twice equals 4 times successive integration. The output of the resonator is given by

$$x[n] = s[n] \star g[n], \quad (2)$$

where $s[n]$ is the speech signal, and $g[n]$ is the response of the cascade of two 0Hz resonators given by $G(z) = H(z)H(z)$. Due to integration the output $x[n]$ grows approximately as a polynomial function of time, but it contains the excitation features. To extract the excitation information the trend in $x[n]$ is removed by subtracting from $x[n]$ the average value of the output within a small window of about 10 msec. The resulting zero-frequency filtered signal $y[n]$ is given by

$$y[n] = x[n] - \frac{1}{2N+1} \sum_{k=-N}^{N} x[n+k], \quad (3)$$

where $2N + 1$ is the size of the window, and $y[n]$ is called the filtered signal. The zero-frequency filtered signal emphasizes the instants of significant excitation in the speech signal. The positive zero crossings correspond to the instants of significant excitation or epochs. The interval between the two adjacent epochs is the pitch period.

2.2. Strength of excitation

The strength of excitation at an epoch is measured by the slope of the zero-frequency filtered signal around the epoch. It gives an idea of the amplitude of the equivalent impulse-like excitation [8]. But the strength at an epoch may not give an indication of the sharpness of the impulse, as the sharpness of the impulse depends on the relative amplitudes of the signal samples around the impulse.

2.3. Loudness

A measure (η) of loudness is derived from the Hilbert envelope of the linear prediction (LP) residual [9]. The LP residual $e[n]$ is obtained using a 10^{th} order LP analysis on each 20 msec frame of speech signal with a frame shift of 5 msec. The Hilbert Envelope $r[n]$ of the LP residual $e[n]$ is given by

$$r[n] = \sqrt{e^2[n] + e_H^2[n]}, \quad (4)$$

where $e_H[n]$ denotes the Hilbert transform of $e[n]$. The Hilbert transform $e_H[n]$ is given by

$$e_H(n) = \text{IFT}(E_H(\omega)), \quad (5)$$

where IFT denotes the inverse Fourier transform, and $E_H(\omega)$ is given by

$$E_H(\omega) = \begin{cases} +jE(\omega), & \omega \leq 0 \\ -jE(\omega), & \omega > 0, \end{cases} \quad (6)$$

Here $E(\omega)$ denotes the Fourier transform of the signal $e[n]$. The sharpness of the peaks around the epochs in the Hilbert envelope $r[n]$ gives an indication of loudness [9]. It is measured as the ratio of the standard deviation (σ) and the mean (μ) of the samples of the Hilbert envelope in a 3 msec interval around each epoch. This measure (η) of loudness does not depend on the periodicity of the glottal vibration. It is observed that the loudness of different sound units in a speech signal are different. The loudness depends on the speaker also.

2.4. Duration

The durations of some vowels increase in the case of Lombard effect speech compared to normal speech. Some stressed vowels like /o/ in 'node' show increase in duration. The unstressed vowel like /o/ in 'Lombard' and the duration of silence regions in stops seem to decrease in Lombard effect speech. Overall, the durations of the voiced regions increase and those of the nonvoiced regions decrease in Lombard effect speech. The increase/decrease of the total duration of the sentence depends on the extent of voiced and nonvoiced regions in the utterance.

Table 1: Percentage duration of voiced region for normal speech and Lombard effect speech for 5 speakers.

	Normal speech	Lombard effect speech
	Voiced	Voiced
Speaker 1	84	89.67
Speaker 2	85.35	89
Speaker 3	73	86.5
Speaker 4	77.7	86
Speaker 5	75.77	77.9

Table 1 shows the percentage duration of voiced regions for 5 speakers for the sentence "I know Sir John will go, though he was sure it would rain cats and dogs" for normal speech and Lombard effect speech. We can see that there is increase in the percentage duration of the voiced region in the case of Lombard effect speech. Also this percentage increase depends on the speaker. The percentage decrease in the duration of the nonvoiced region is generally more than the percentage increase in the duration of the voiced region.

3. Lombard effect on excitation features

Here we study the Lombard effect on the features of excitation source. We examine the parameters of the excitation source features described in Section 2 for Lombard effect speech in comparison with normal speech.

The pitch frequency (F_0) is observed to be higher in the case of speech produced with external feedback, as shown by the pitch frequency contours in Figures 1(a) and 1(b) for normal speech and Lombard effect speech, respectively. The average pitch frequency for these normal and Lombard effect speech utterances are 157 Hz and 180 Hz, respectively. The strength of excitation (SoE) decreases for the Lombard effect speech compared to the normal speech as shown in Figures 2(a) and 2(b).

The Lombard effect speech is perceived to be louder than the normal speech. Figures 3(a) and 3(b) show segments of

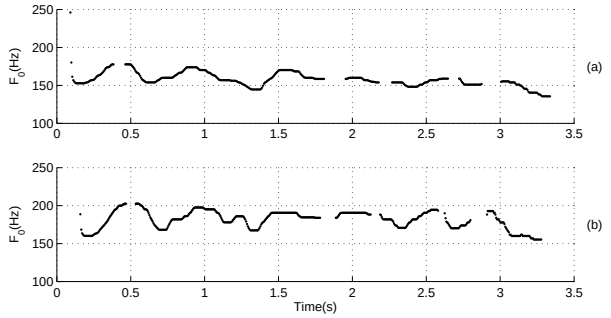

Figure 1: F_0 contours for (a) normal speech, and (b) Lombard effect speech for the utterances of the sentence "I know Sir John will go, though he was sure it would rain cats and dogs".

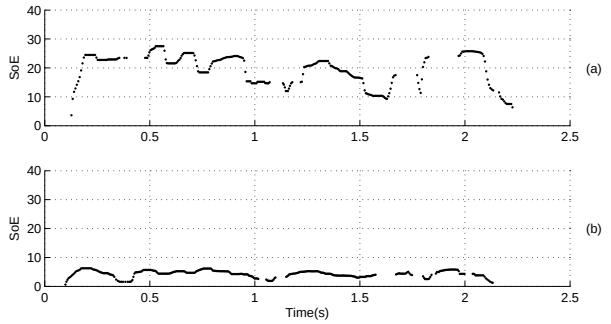

Figure 2: Strength of excitation contours for (a) normal speech, and (b) Lombard effect speech for the utterances of the sentence "I wish we may be able to tide over this difficulty".

the Hilbert Envelope around the peaks at the instants of significant excitation (epochs), which are superimposed for the cases of normal and Lombard effect speech, respectively [9]. Each segment has a duration of 3 msec, centered around epochs in the Hilbert Envelope. These plots show that the loudness in the Lombard effect speech increases compared to normal speech. The impulses around the instants of significant excitation are sharper in the case of Lombard effect speech, indicating higher loudness. Since the increased loudness cannot be present at all epochs (i.e., in all segments), only a few epoch locations show the sharpness, and hence it may be difficult to notice the effect of loudness prominently in these cluster plots.

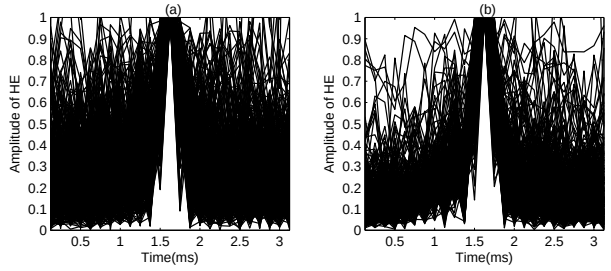

Figure 3: Superimposed segments of Hilbert envelope of the LP residual around the epochs for (a) Normal speech, and (b) Lombard effect speech [9].

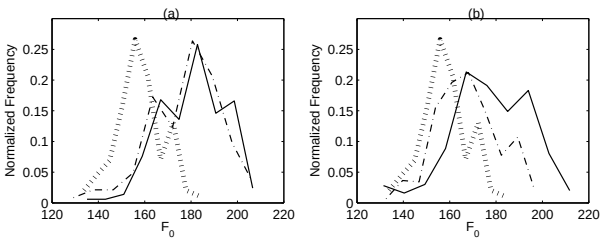

Figure 4: Distribution of F_0 for normal speech (dotted lines), Lombard effect speech with low intensity feedback (dash-dotted lines), Lombard effect speech with high intensity feedback (solid lines) for 2 cases of feedback: (a) Noise, and (b) Normal speech.

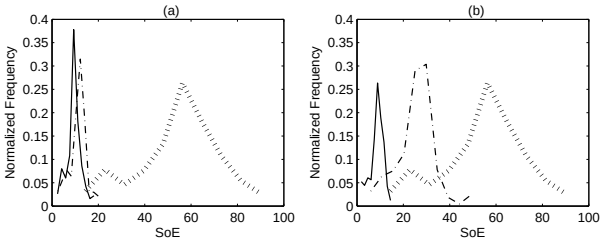

Figure 5: Distribution of SoE for normal speech (dotted lines), Lombard effect speech with low intensity feedback (dash-dotted lines), Lombard effect speech with high intensity feedback (solid lines) for 2 cases of feedback: (a) Noise, and (b) Normal speech.

4. Analysis of Lombard effect speech for different types and levels of external feedback

In this session we study the Lombard effect on the excitation features for two types of external feedback signals. The excitation source and the vocal tract system tend to change the excitation features to compensate for the loss of self feedback. This change depends on the extent to which the self feedback is lost, which in turn depends on the type and level of the external feedback signals.

Figures 4 and 5 show distributions of F_0 and SoE, respectively, for two types of feedback: (a) white noise and (b) normal speech, for 3 cases: (1) Speech under silent conditions. (2) Speech with low intensity of feedback. (3) Speech with high intensity of feedback. We find an increase in F_0 and a decrease in SoE with the increase in intensity of the external feedback signal. Another observation is that the distribution of the SoE (i.e., width of the spread) decreases with increase in the intensity level of the external feedback signal. The distribution of the SoE is also more for the case of normal speech as external feedback, when compared with the same intensity white noise as external feedback. Loudness increases with increase in intensity of the external feedback.

Loudness increases with increase in intensity of the external feedback. Loudness in the case of noise as external feedback is more in comparison with normal speech as feedback. But this increase is small due to the fact that all regions of the Lombard effect speech need not affect the loudness equally.

These studies show that the Lombard effect is different for different feedback conditions and for different levels of feed-

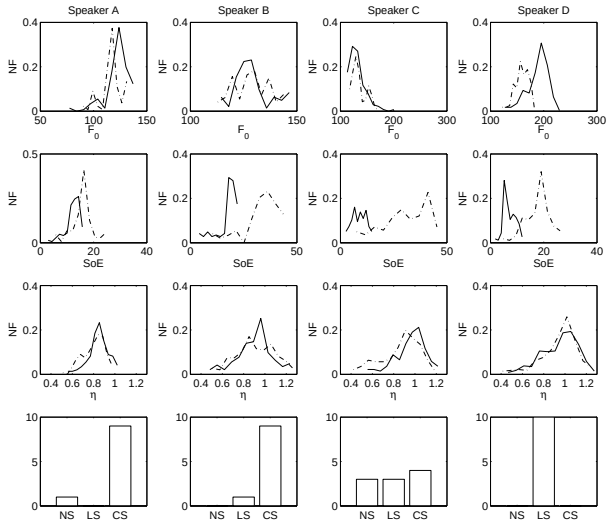

Figure 6: Distribution of F_0, SoE, η and perceptual evaluation results for normal speech (dash-dotted lines) and Lombard effect speech(solid lines) for 4 speakers. (NS=normal speech, LS=Lombard effect speech, CS=can't say, NF=Normalized frequency.)

back intensity. The excitation source features show significant differences for different conditions and levels of intensity of feedback.

5. Perception studies

Perceptual evaluation was carried out by conducting subjective test with 10 listeners in the age group of 21-23 years. Experiments were carried out in a laboratory environment. Two speech files, a normal speech and a Lombard effect speech of the same sentence spoken by the same person were played to the subjects through headphones. The listeners were asked to choose the louder of the two. If they perceive the pair as equally loud, they were asked to choose the option "can't say".

Figure 6 shows the distribution of perception evaluation results for four speakers. The results for these four speakers were selected here to illustrate the variety of responses that can be obtained in perceptual studies on Lombard effect speech. Row 4 in Figure 6 show the results of perceptual evaluation. We can see that the loudness of normal speech and Lombard effect speech was perceived to be similar(can't say (CS) response) in the cases of speaker A and speaker B. For speaker C, the perception results are varied, indicating that the Lombard effect may be small. For speaker D all the subjects perceived the Lombard effect speech as loud.

These observations can also be interpreted in terms of the distributions of the parameter, especially the F_0 parameter. The distribution of F_0 is distinctly different for Lombard effect speech compared to that for normal speech for speaker D, and it is also reflected in the perception results. The distribution of the loudness parameter η is not distinguishable for Lombard effect speech and normal speech. This may be due to the fact that only a few segments may be pronounced loudly due to Lombard effect, and the percentage of these segments may be small to get reflected in the distribution of η at all the epochs. The strength of excitation (SoE) at epochs is lower for Lombard effect speech compared to normal speech for all the speakers. SoE can differentiate between normal speech and Lombard effect speech even though the loudness measure (η) and F_0 are not distinctly different. This is reflected in case of speakers A and B. Note that SoE gives only the amplitude of the impulse-like excitation at each epoch, and it need not necessarily indicate loudness. The loudness is due to sharpness of the impulse-like behavior in excitation around the epochs, and it is better represented by the parameter η.

6. Summary and Conclusions

In this paper we have presented analysis of Lombard effect speech in terms of acoustic features representing the excitation source in speech production. We have used the instantaneous F_0, strength of excitation at epochs and a loudness measure describing the sharpness of the impulse-like excitation around epoch. The distributions of these parameters in Lombard effect speech show that the pitch frequency (F_0) increases due to Lombard effect on speech production. The Lombard effect also increases the loudness, in comparison with normal speech, as perceived by human listeners. The effect of external feedback signal that causes the Lombard effect was examined. The level of external feedback signal influences F_0 and also the loudness. The nature of the external feedback signal also influences the extent of Lombard effect.

Since the Lombard effect does not influence all segments of speech equally, it is important to study the segments or sound units that are most affected by the Lombard effect. It is interesting to study the effect of various acoustic features in producing Lombard effect speech, by synthesizing speech incorporating these features in normal speech. Such a study will help to process the Lombard effect speech appropriately for use in speech systems developed for normal speech.

7. References

[1] E .Lombard, "Le signe de l'elevation de la voix, annals maladiers oreille," *Larynx,Nez,Pharynx*, vol. 37, pp. 101–119, 1911.

[2] John h. L. Hansen and Vaishnevi Varadarajan, "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*, vol. 17, no. 2, pp. 366–378, Feb. 2009.

[3] B. J. Stanton, L. H. Jamieson and G. D. Allen, "Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions," in *ICASSP*, vol. 1, New York, Apr. 1988, pp. 331–334.

[4] Junqua J. and Anglade Y, "Acoustic and perceptual studies of Lombard speech: Application to isolated-words automatic speech recognition," in *ICASSP*, vol. 2, Apr. 1990, pp. 841–844.

[5] J. C. Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers," *Journal of the Acoustical Society of America*, vol. 93, pp. 510–524, Jan. 1993.

[6] B. Yegnanarayana and K. Sri Rama Murty, "Event-Based Instantaneous Fundamental Frequency Estimation From Speech Signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 4, pp. 614–624, May 2009.

[7] K. Sri Rama Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1602–1613, Nov. 2008.

[8] K. Sri Rama Murty and B. Yegnanarayana and Anand Joseph M., "Characterization of Glottal Activity from Speech Signals," *IEEE Signal Processing Letters*, vol. 16, no. 6, June 2009.

[9] Guruprasad Seshadri and B. Yegnanarayana, "Perceived loudness of speech based on the characteristics of excitation source," *Communicated to Journal of the Acoustical Society of America (under review).*