7

The Influence of Lombard Effect on Speech Recognition

Damjan Vlaj and Zdravko Kačič
University of Maribor, Faculty of Electrical Engineering and Computer Science
Slovenia

1. Introduction

The origin of Lombard effect dates back one hundred years. In 1911 Etienne Lombard discovered the psychological effect of speech produced in the presence of noise (Lombard, 1911). The Lombard effect is a phenomenon in which speakers increase their vocal levels in the presence of a loud background noise and make several vocal changes in order to improve intelligibility of the speech signal (Anglade & Junqua, 1990; Bond et al., 1989; Dreher & O'Neill, 1957; Egan, 1971; Junqua, 1996; Junqua & Anglade, 1990; Van Summers et al., 1988). In nowadays speech recognition applications appearance of Lombard effect can be expected in various domains, where spontaneous and conversational speech communication will take place in uncontrolled acoustic environments.

Two main interpretations of the Lombard effect have been proposed. The first argues that the effect is a physiological audio-phonatory reflex (Lombard, 1911), the second that Lombard changes are motivated by compensation on the part of the speaker for decreased intelligibility (Lane & Tranel, 1971). Some authors have also argued that both mechanisms may contribute to the changes made by the speaker in noisy environments (Junqua, 1993).

Detailed surveys of the literature on the Lombard effect phenomenon was made in (Lane & Tranel, 1971) and more recently in (Junqua, 1996). The conducted research showed that Lombard speech is different from normal speech in a number of ways. The main changes of characteristics of Lombard speech can be seen in increase in voice level, fundamental frequency and vowel duration, and a shift in formant center frequencies for F1 and F2 (Anglade & Junqua, 1990; Applebaum et al., 1996; Junqua, 1996; Junqua & Anglade, 1990). It was also reported in (Hanley & Steer, 1949) that speaking rate may be reduced when speech is produced in a noisy environment. A detailed acoustic and phonetic analysis of speech under different types of stress including the Lombard effect was carried out also in (Hansen, 1988). The studies showed that under the Lombard effect, duration of vowels increase while that of unvoiced stops and fricatives decrease. Also, spectral tilt decreases implying an increase in high-frequency components under the Lombard effect. An increase in pitch and first formant location also occurs in both cases. Also, energy migration from low and high frequency to the middle range for vowels, and movement from low to higher bands for unvoiced stops and fricatives was observed. In addition to the above, differences between male and female speakers was noted in (Junqua, 1993). Lombard changes are on the other hand greater in adults than in children and in spontaneous speech than in reading tasks (Amazi & Garber, 1982; Lane & Tranel, 1971).

It was concluded in (Bond et al., 1989) that the above mentioned changes of speech characteristics in Lombard speech are made to increase the vocal effort and to articulate in a more precise manner for better communication in a noisy condition.
Researchers (Pickett, 1956; Dreher & O'Neill, 1957; Ladgefoged, 1967) studied intelligibility of utterances under the Lombard effect. It was shown that the intelligibility of Lombard speech increases up to a certain level of noise, when presented at a constant speech-to-noise ratio, and sharply decreases when speech becomes shouted. It was also demonstrated that the presence of auditory feedback of speech is necessary to maintain the intelligibility of Lombard speech, as the primary purpose of Lombard effect is to increase speech intelligibility in communication with other speakers in noisy environments.
It was reported in (Junqua, 1996) and in (Van Summers et al., 1988) that acoustic changes that occur in speech in a noisy environment are different from person to person and are highly speaker-dependent (Junqua, 1996). This was confirmed also in (Van Summers et al., 1988), where the authors reported a significant increase in fundamental frequency for one male speaker, but not for the second, when they spoke in quiet and in different levels of noise. The characteristics of Lombard speech may also vary with the type of ambient noise, and with the language of the speaker (Junqua, 1996).
It was suggested in (Lane & Tranel, 1971) that the magnitude of the speakers' response to noise is likely to be governed by the desire to achieve intelligible communication. As an argument to support this idea they argue that in a noisy condition, speakers would not change their voice level when talking to themselves. In (Bond et al., 1989) the idea was confirmed as the authors observed that the magnitude of the Lombard effect is greater when speakers believe they are communicating with interlocutors. Encountering these Lombard reflex cannot be considered as an all-or-none response with some threshold level (Junqua, 1996; Lane & Tranel, 1971). According to (Junqua, 1996), the variability in Lombard speech appears to be distributed along a continuum. The acoustic differences that can be observed between Lombard speech and normal speech are believed to have an effect on intelligibility.
As reported in (Junqua, 1993; Van Summers et al., 1988; Dreher & O'Neill, 1957) the speech produced in noise is more intelligible than speech produced in quiet, when both types of speech are presented in noise at an equivalent signal-to-noise ratio. It was also shown in (Junqua, 1996) that the type of masking noise and the gender of the speakers used for the experiment are crucial to the difference in intelligibility of speech produced in noise-free and in noisy conditions. In (Junqua, 1993) it was also demonstrated that the babble noise degrades the intelligibility of English digit vocabulary more than white noise. He also showed that in such case the female Lombard speech is more intelligible than the male Lombard speech. It was further revealed that breathiness decreases the intelligibility of speech. In this sense it seems that female speakers tend to decrease the breathiness in their productions more than male speakers do (Junqua, 1993).
In this chapter we want to present the influence of Lombard effect on speech recognition, which presence can be expected in contemporary speech recognition application in numerous application domains. For this reason, we will use the Slovenian Lombard Speech Database, which was recorded in studio environment. Slovenian Lombard Speech Database will be presented in Section 2. The changes of Lombard speech characteristics will be presented in Section 3. With the experiments we want to confirm the influence of Lombard effect on speech recognition. In section 4, the experimental design for speech recognition will be presented. The results of experiments will be given in Section 5 and the conclusion will be drawn in Section 6.

2. Lombard speech database

For the analysis of the speech characteristics and speech recognition experiments, we used Lombard speech database recorded in Slovenian language. The Slovenian Lombard Speech Database[1] (Vlaj et al., 2010) was recorded in studio environment. In this section Slovenian Lombard Speech Database will be presented in more detail. Acquisition of raw audio material recorded in studio conditions is described in Subsection 2.1. Annotation of speech material and conversion of the audio material to the final format are presented in Subsection 2.2. The structure of Slovenian Lombard Speech Database is presented in Subsection 2.3.

2.1 Acquisition of raw audio material

The Slovenian Lombard Speech Database was recorded in studio environment. Each speaker pronounced a set of eight corpuses in two recording sessions with at least one week pause between recordings. Approximately 30 minutes of speech material per speaker and per session was recorded.

The recordings were performed using a hands-free microphone AKG C 3000 B, close talking microphone Shure Beta 53 and two channel electroglottograph EG2. Four channel recordings were performed:

- hands free microphone,
- close talking microphone,
- laryngograph and
- recordings of noise mixed with speaker's speech that was played on speaker's headphones during recordings.

The recording platform consisted of Audigy 4 PRO external audio card for 4 channel audio recording, Phonic MU244X mixer, and using 96 kHz sampling frequency, 24-bit linear quantization.

Two types of noises were used in recordings: babble and car noise. The noises were taken from the Aurora 2 database (Hirsch & Pearce, 2000) and were normalized. The noises were played to speaker's headphones AKG K271.

At the beginning of each recording the level of the reproduced background noise was adjusted according to the scheme proposed in (Bořil et al., 2006). The required noise level was adjusted by setting the corresponding effective voltage of the sound card open circuit VRMS OL. Noise levels of 80 dB SPL[2] and 95 dB SPL at a virtual distance of 1–3 meters were used for the Lombard speech recordings.

Three recordings of all corpuses were made within one recording session:

- without noise (reference recording),
- at 80 dB SPL and
- at 95 dB SPL.

A short pause was made between recordings of items of particular corpus (word, number, number string, and sentence) to allow speaker's recovery. After the complete corpus was recorded a longer pause was made to allow for speaker's recovery.

There was an interaction between the "Lombard" speaker and a listener. The listener heard the attenuated speech mixed with non attenuated noise, evaluated the intelligibility and reacted accordingly. The reaction of the listener was mediated to the speaker by means of

[1] The owner of the database is SVOX.

[2] SPL is abbreviation for Sound Pressure Level.

message displayed on the LCD display, where the speaker was notified that the pronunciation was intelligible or she/he was asked to repeat the pronunciation as it was not intelligible enough.

2.2 Annotation of speech material

The manual annotation of speech material is performed by the LombardSpeechLabel tool (Figure 1) developed at the University of Maribor. The program tool is written in the Tcl/Tk/Tix language, which is suitable for visual programming. It was developed on the Microsoft Windows platform and can be incorporated into other operating system platforms with small modifications.

The LombardSpeechLabel tool window is divided into three fields. The upper field contains four waveform views (hands free microphone, close talking microphone, laryngograph and recordings played on speaker's headphones) of the signal that have been captured during recording of the database. By clicking the buttons on the right hand side of the upper field, each signal can be played individually. The bottom of the tool window is divided into two parts. On the left hand side the information about the speaker and the recording is given. On the right hand side, the additional data of the recording and the orthographic transcription are presented.

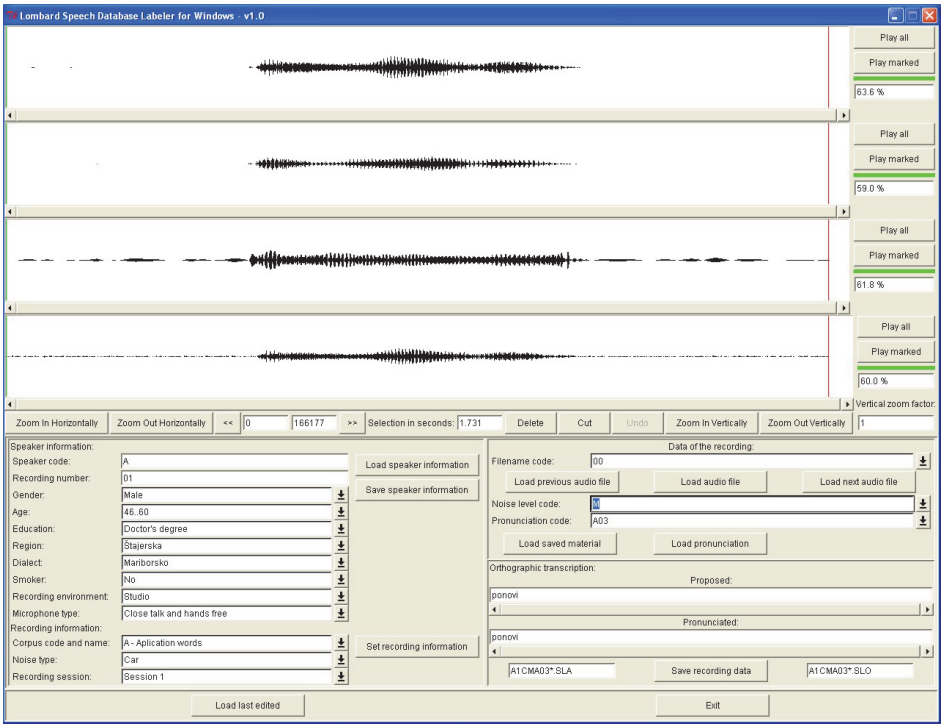

Fig. 1. LombardSpeechLabel tool for manual annotation of speech material.

The conversion of the audio material to the final format, which was set to 96 kHz sampling frequency, 16-bit linear quantization is also made with the LombardSpeechLabel tool.

2.3 The structure of the database

The Slovenian Lombard Speech Database consists of recordings of 10 Slovenian native speakers. Five males and five females were recorded. As we already mentioned, each speaker pronounced a set of eight corpuses in two recording sessions with at least one week pause between recordings. The corpus's structure is similar to SpeechDat II database (Kaiser & Kačič, 1997). In the following subsections more information about the database will be given.

2.3.1 Audio and label file format

Audio files are stored as sequences of 16-bit linear quantization at the sampling frequency of 96 kHz. They are saved in Intel format. Each prompted utterance is stored in a separate file. Each speech file has an accompanying SAM label file with UTF-8 symbols.

A	Speaker code (A-Z)
S	Session code (1-9) – used only 1 and 2
T	Code of the noise type: • R: without noise • C: Car noise • B: Babble noise
R	Code of the recording: • N: recording of the reference signal without presence of noise • L: recording of the signal without presence of noise • M: recording of the signal with presence of noise level of 80 dB SPL • H: recording of the signal with presence of noise level of 95 dB SPL
NNN	Code of the corpus (A00 – Z99): A – application words, B – connected digits, D – dates, I – isolated digits, N – natural numbers, S – phonetically rich sentences, T – times, W – phonetically rich words
C	Code of the recording channel: • 1: hands-free microphone • 2: close talk microphone • 3: signal captured by laryngograph • 4: signal in headphones that was heard by a speaker
LL	Two letter ISO 639 language code
F	File type code O=Orthographic label file, A=audio speech file

Table 1. Description of file nomenclature.

2.3.2 File nomenclature

File names follow the ISO 9660 file name conventions (8 plus 3 characters) according to the main CD ROM standard. Owing to the large amounts of audio material, the data were stored on a DVD-ROM media.

The following template for file nomenclature is used:

```
A S T R NNN C. LL F
```

The file nomenclature is described in Table 1.

2.3.3 Directory structure

The directory structure is set so that each speaker is located on his own DVD-ROM volume. Each speaker has two sessions. In each session the reference condition and two noise conditions are included. Each condition includes eight corpses. The following five levels directory structure is defined:

```
\<database>
   \<speaker>
      \<session>
         \<condition>
            \<corpus>
```

The Lombard speech database directory structure is presented in Table 2.

<database>	Defined as: <name><language code> i.e. LOMBSPSL Where: <name> is LOMBSP indicating Lombard Speech <LL> is the ISO 2-letters code SL for Slovenian
<speaker>	Defined as: SPK_<a> Where <a> is a progressive letter from A to Z. This letter is the same as the first letter used in file names (see subsection 2.3.2).
<session>	Defined as: SES_<s> Where <s> is a progressive number in the range 1 to 9. This number is the same as the second number used in file names (see subsection 2.3.2).
<condition>	Tree types of conditions are defined: • REF: recording of the reference signal without presence of noise, • CAR: recording of the signal with presence of car noise and • BABBLE: recording of the signal with presence of babble noise
<corpus>	Defined as: CORPUS_<c> Where <c> is a letter for one of corpus defined: A – application words, B – connected digits, D – dates, I – isolated digits, N – natural numbers, S – phonetically rich sentences, T – times, W – phonetically rich words

Table 2. Lombard speech database directory structure.

2.3.4 Corpus code definition

As it is useful for users to clearly identify the speech file contents by looking at the filename, we have specified the corpus code to support one letter corpus identifier and two numbers identifier. The corpus code definition is described in Table 3.

3. Changes of Lombard speech characteristics

In this section, we will present changes of three Lombard speech characteristics: mean value of pitch, phoneme duration and frequency envelope. To demonstrate changes of Lombard speech characteristics we used recordings of Slovenian Lombard Speech Database presented in Section 2.

In the analysis, the Lombard speech characteristics were measured for different voiced phonemes for the utterances of three words: "ustavi" (stop), "ponovi" (repeat) and

"predhodni" (previous). In this paper only the selected results of Lombard speech analysis will be presented.

Corpus identifier	Item identifier	Corpus contents
A	00-29	application words (30 words)
B	00-04	connected digits (10 digits sequence pronounced 5 times)
D	00-04	dates (5 dates)
I	00-11	isolated digits (12 digits)
N	00-04	natural numbers (5 numbers)
S	00-29	phonetically rich sentences (30 sentences)
T	00-06	times (7 times)
W	00-49	phonetically rich words (50 words)

Table 3. Corpus code definition.

3.1 Mean value of pitch

According to the literature, the value of pitch increases in Lombard speech compared to normal speech. In this section the results of mean pitch values of the first phoneme "O" of the word "ponovi" (Repeat) will be presented. Figures 2 and 3 show the mean pitch values of voiced speech (vowel "O") for five speakers, for two sessions and two noise types. Speakers 1 and 2 were male speakers, whereas speakers 3 to 5 were female speakers.

Significant increase of pitch in first vowel "O" of the word "ponovi" (repeat) compared to reference pronunciations can be seen on Figures 2 and 3 for Lombard speech recorded under 95dB noise level for all five speakers. The increase can be observed in both recording sessions and for both noise types, although the extent varies among speakers. The increase is almost the same for the first, second and the fifth speaker and varies most for the third speaker in case of babble background noise. In case of car background noise the difference is bigger for the first and the forth speaker. For utterances recorded under 80 dB noise level the increase of pitch is significant in case of babble noise (except for third speaker) but is less clear in case of car noise for most speakers

3.2 Phoneme duration

In this section the results of the duration of the vowel "A" of the word "ustavi" (stop) for all five speakers are presented. Figures 4 and 5 show the results of the analysis. It can be seen that the duration varies among speakers, but is more consistent per speaker regarding different recording sessions, background noise type and noise level. However, there is no clear distinction in phoneme duration concerning different recording sessions, background noise level or noise type. Figures 4 and 5 indicate that speakers tend to increase the phoneme duration at higher level of background noise, but this seems to be not as consistent as the increase of pitch.

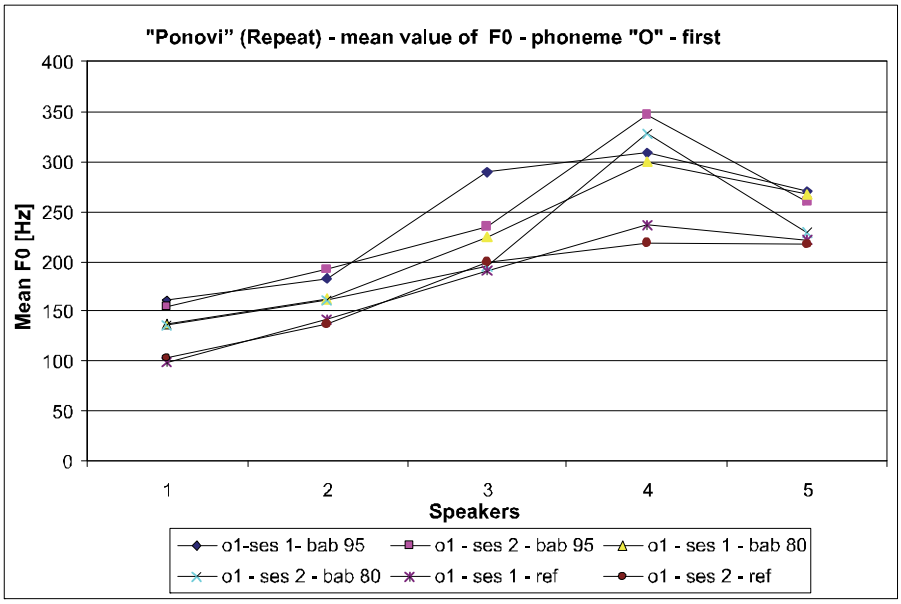

Fig. 2. Mean pitch values of the first phoneme "O" of the word "ponovi" (Repeat) recorded at different noise levels and at babble background noise.

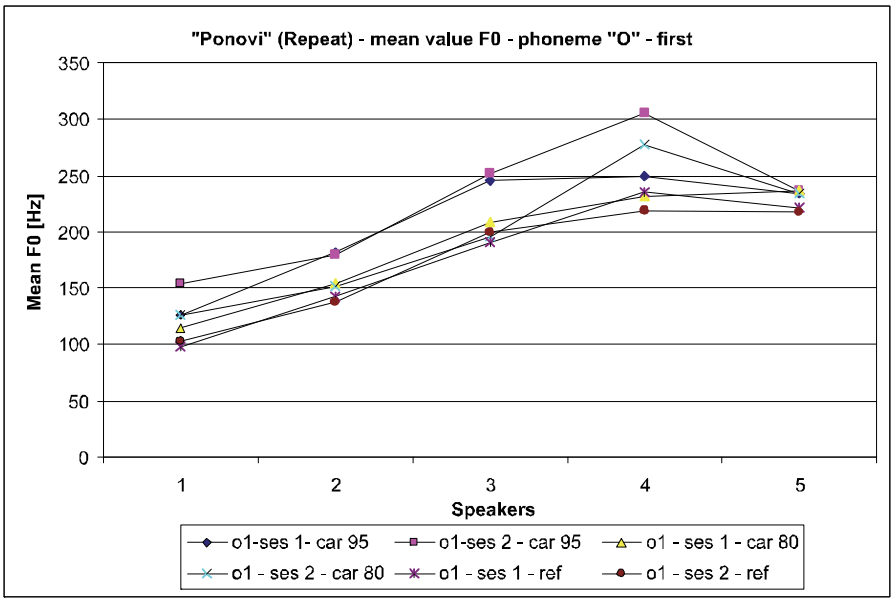

Fig. 3. Mean pitch values of the first phoneme "O" of the word "ponovi" (Repeat) recorded at different noise levels and at car background noise.

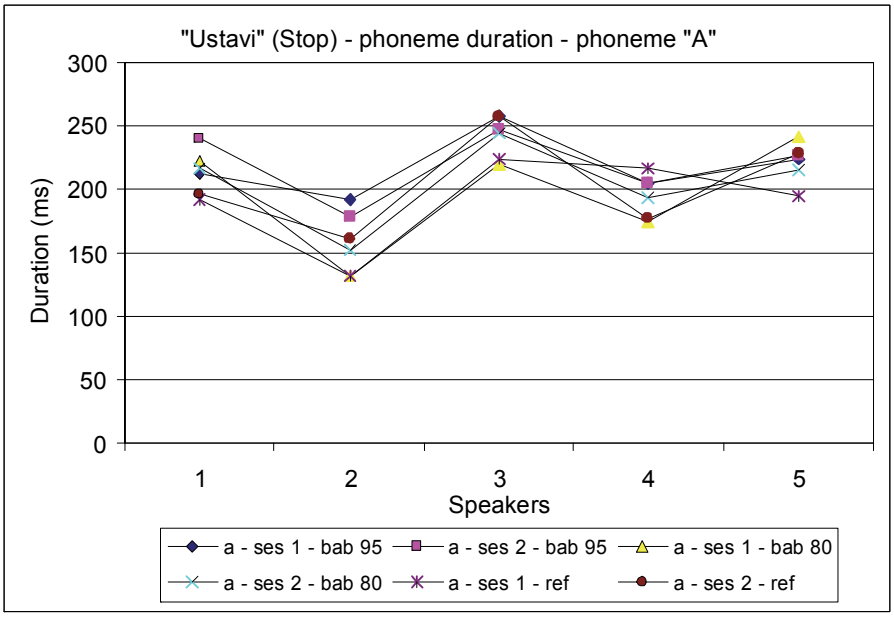

Fig. 4. Duration of the phoneme "A" of the word "ustavi" (Stop) recorded at babble background noise and at different noise levels.

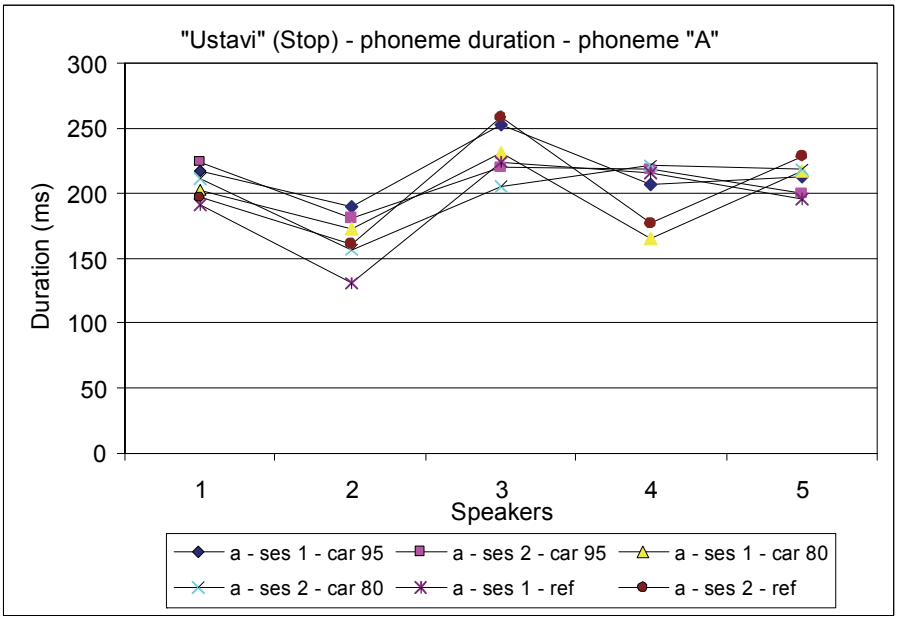

Fig. 5. Duration of the phoneme "A" of the word "ustavi" (Stop) recorded at car background noise and at different noise levels.

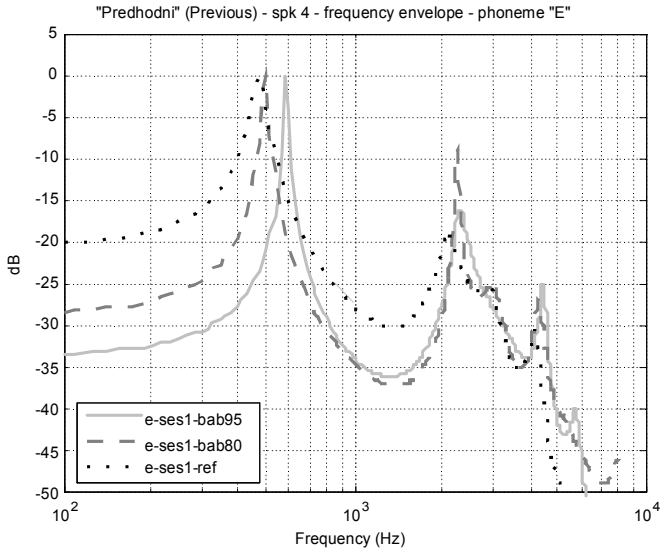

Fig. 6. Frequency envelope of phoneme "E" of the word "Predhodni" (Previous) recorded at babble background noise and at different noise levels for female speaker (speaker 4) and for the first recording session.

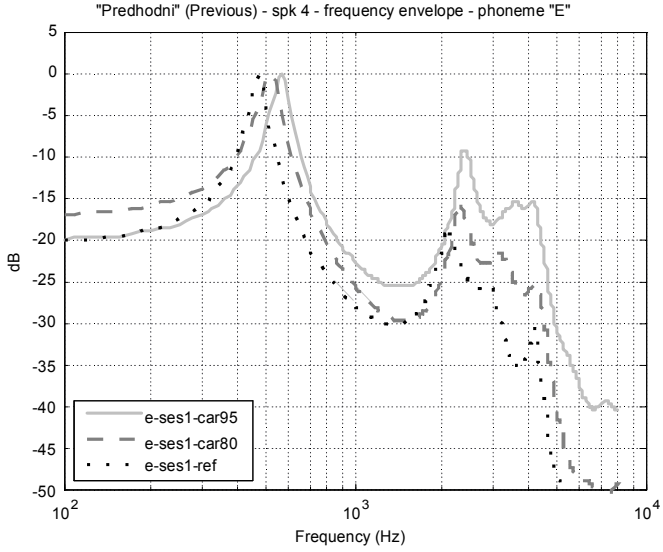

Fig. 7. Frequency envelope of phoneme "E" of the word "Predhodni" (Previous) recorded at car background noise and at different noise levels for female speaker (speaker 4) and for the first recording session.

3.3 Frequency envelope

In this section the results of frequency envelope of phoneme "E" of the word "Predhodni" (Previous) recorded at different background noises and at different noise levels for female speaker (speaker 4) are presented. Figures 6 and 7 show these results of the analysis. The increase of the first formant frequency is evident for both background noise types. Also an increase of energy in higher frequency range can be seen. Both features are known to occur in Lombard speech. The changes of these features are less obvious for utterance uttered at 80 dB background noise.

4. Experimental design

We created experimental design, which showed the influence of Lombard effect on speech recognition. It was carried out on the Slovenian Lombard Speech Database. The experimental design for acoustic modeling was based on continuous Gaussian density Hidden Markov Models. For hidden Markov modeling the HTK toolkit was used (Young et al., 2000). For training only recordings of the signal without presence of noise on speaker headphones (see code L of the recording in Table 1) were used. The training was done with monophone acoustical models. The reason why we decided to use monophone acoustical models and not triphone or word acoustical models lays in the content of the Slovenian Lombard Speech Database. For the training of triphone acoustical models the speech material of the Slovenian Lombard Speech Database is not big enough. Looking from the point of view of word acoustical models, the Slovenian Lombard Speech Database has too many various words to be trained well enough. The training procedure for monophone acoustical models is presented in Figure 8. The Gaussian mixtures were increased by power of 2 up to 32 mixtures per state. Monophone acoustical models were trained on all eight corpuses from the Slovenian Lombard Speech Database (see Table 3). For this reason 2880 recorded files with 9474 pronounced words were used. In the next paragraph we will shortly present the HTK tools, which were used in the training procedure.

The HTK tool *HCompV* scans a set of data files, computes the global mean and variance and sets all of the Gaussians in a given HMM to have the same mean and variance. The HTK tool *HERest* is used to perform a single re-estimation of the parameters of a set of HMMs using an embedded training version of the Baum-Welch algorithm. *HHEd* is a script driven editor for manipulating sets of HMM definitions. Its basic operation is to load in a set of HMMs, apply a sequence of edit operations and then output the transformed set. We used this program tool to add short pause model and for increasing the number of Gaussian mixture components for each state.

For the testing three types of the recordings were used:

- recordings of the signal without presence of noise on the speaker headphones,
- recordings of the signal with presence of noise level of 80 dB SPL on the speaker headphones and
- recordings of the signal with presence of noise level of 95 dB SPL on the speaker headphones.

The Slovenian Lombard Speech Database is recorded in two recording sessions with at least one week pause between recordings. For the training of monophone acoustical models the speech material of the first session was used and for the testing the speech material of the second session was used. We also made cross experiments, so that we trained monophone

acoustical models on the second session and tested them on the first session. The tests were made on four corpuses (application words, phonetically rich words, isolated digits and connected digits) from the Slovenian Lombard Speech Database. The test on application words contained 320 words and the test on phonetically rich words contained 500 words. The corpuses isolated digits and connected digits were combined in one test with 620 digits/words. Word loop was used in all tests, which simply puts all words of the vocabulary in a loop and therefore allows any word to be followed by any other word. The results will be presented in Section 5.

For the experimental design, we used Mel-cepstral coefficients and energy coefficient as features. We also used first and second derivative of the basic features. The features were created with the front-end using the basic distributed speech recognition standard from ETSI (ETSI ES 201 108, 2000).

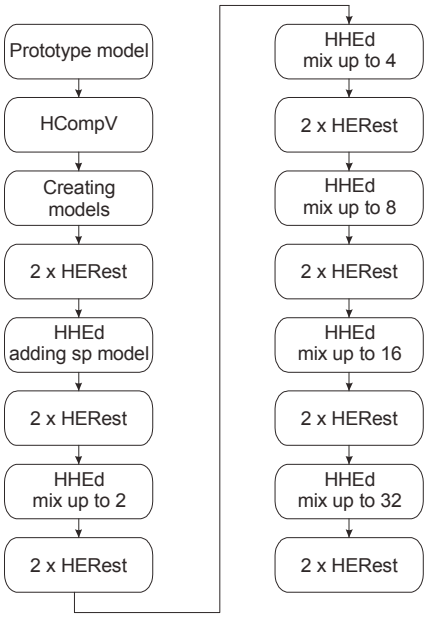

Fig. 8. The procedure for training of monophone acoustical models.

5. Results

The results obtained by the experiments will be presented in this section. Figures 9 to 14 present charts, which show the results on speech recognition accuracy. There are twelve groups with three speech recognition results presented on all charts. The first column in each group of results presents speech recognition accuracy when there was no noise played on the speaker headphones. The second column presents speech recognition accuracy when car or babble noise was played on the speaker headphones with the noise level of 80 dB SPL. The last third column presents speech recognition accuracy, when car or babble noise was played on the speaker headphones with the noise level of 95 dB SPL. At this point we must point out that recordings used for training of monophone acoustical models and testing

have no noise present. The noise mentioned was played on speaker headphones to encourage the speaker to speak louder. Speech recognition experiments were made on six different Gaussian mixtures per state. In the charts this is indicated by mix1 to mix 32. The speech recognition results are presented for both training scenarios. In the first scenario the monophone acoustical models were trained on the first session of the Slovenian Lombard Speech Database and then tested on the second one. In the second scenario the monophone acoustical models were trained on the second session and then tested on the first one. Bellow the title of the charts there is a row beginning with "Trained on" that indicates in which session monophone acoustical models were trained.
Figures 9 and 10 show speech recognition accuracy tested on corpus A (application words) with presence of car and babble noise on the speaker's headphones. Figures 11 and 12 show speech recognition accuracy tested on corpus W (phonetically rich words) with presence of car and babble noise on the speaker's headphones. And last two Figures 13 and 14 show speech recognition accuracy tested on corpuses B (connected digits) and I (isolated digits) with presence of car and babble noise on the speaker's headphones.
From the speech recognition results we can conclude that the Lombard effect is present in the recordings, which were recorded with noise present on the speaker's headphones. When the noise level on the speaker's headphones was increased from 80 dB SPL to 95 dB SPL, the speech recognition accuracy decreased.

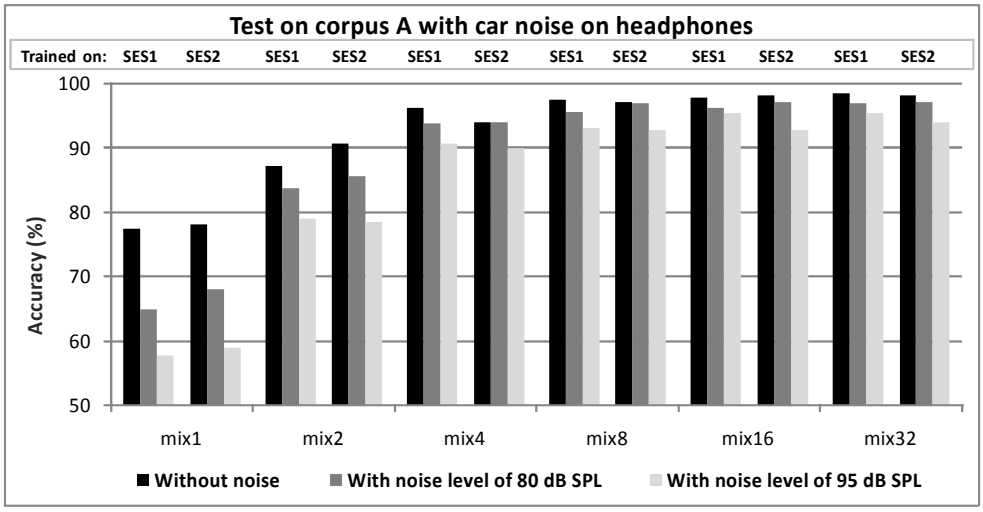

Fig. 9. Speech recognition accuracy tested on application words (corpus A) with presence of car noise on the speaker headphones.

The speech recognition accuracy was almost always better when the monophone acoustical models were trained on first sessions and tested on second session. The reason for this could lay in better trained monophone acoustical models on the first session or better acoustical environment in the second session of the Slovenian Lombard Speech Database. Should the second answer be correct, it could be concluded that speakers have adapted. Namely, when speakers recorded the second session, they had already known what to expect.
The best speech recognition results were achieved, when the tests were made on phonetically rich words (corpus W). The results were the worst, when the tests were made

on connected and isolated digits (corpuses B & I). If we analyze speech recognition results at only 32 Gaussian mixtures per state, we can see that the smallest differences between the tests when no noise was present on speaker's headphones and the tests when the noise level of 95 dB SPL was present on speaker's headphones were obtained on corpuses A (application words) and W (phonetically rich words).

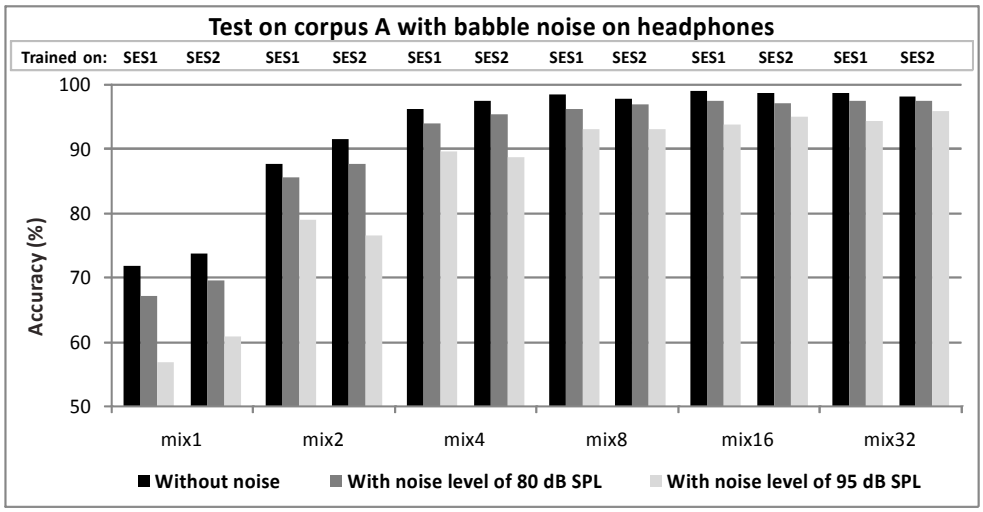

Fig. 10. Speech recognition accuracy tested on application words (corpus A) with presence of babble noise on the speaker headphones.

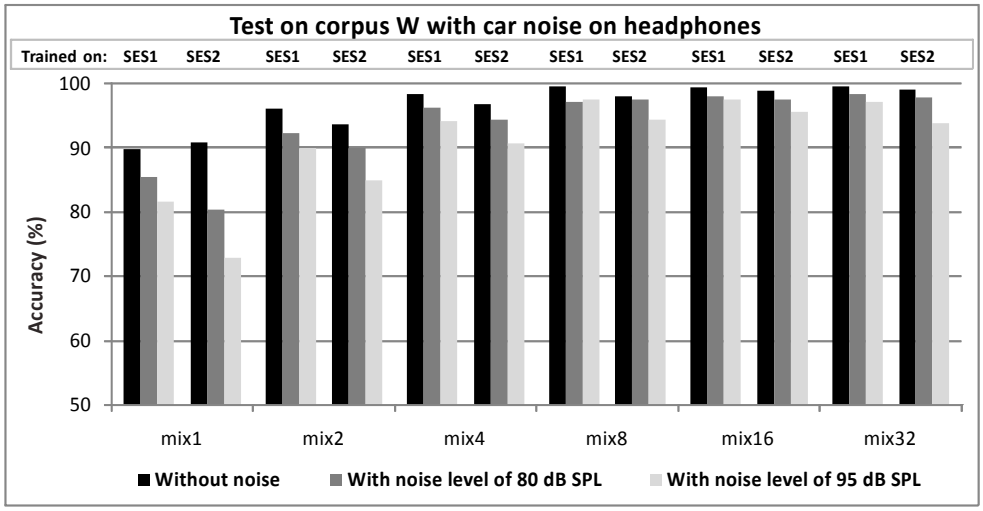

Fig. 11. Speech recognition accuracy tested on phonetically rich words (corpuses W) with presence of car noise on the speaker headphones.

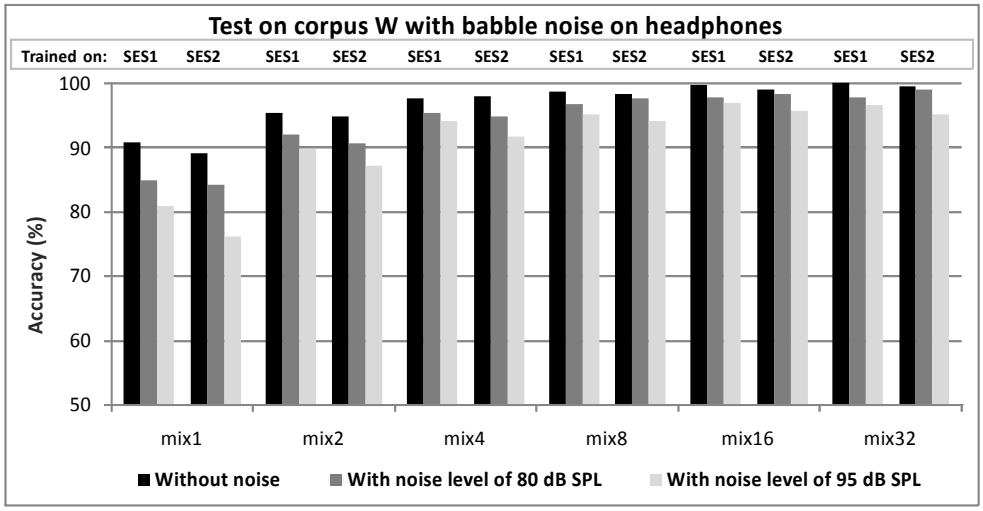

Fig. 12. Speech recognition accuracy tested on phonetically rich words (corpuses W) with presence of babble noise on the speaker headphones.

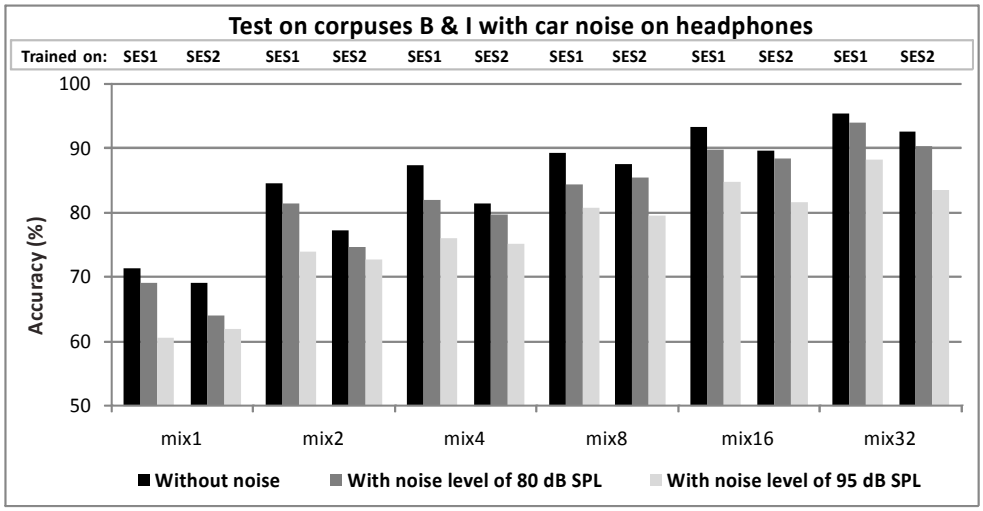

Fig. 13. Speech recognition accuracy tested on connected and isolated digits (corpuses B & I) with presence of car noise on the speaker headphones.

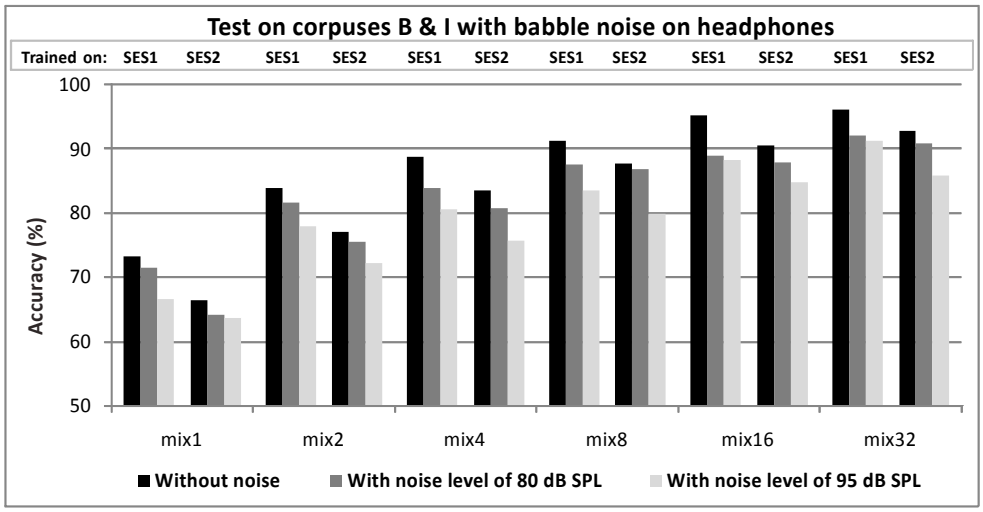

Fig. 14. Speech recognition accuracy tested on connected and isolated digits (corpuses B & I) with presence of babble noise on the speaker headphones.

By increasing the number of Gaussian mixtures per state, the speech recognition accuracy was also increased. If we look at the speech recognition accuracy presented in the charts, we can see that the influence of Lombard effect is smaller at 32 Gaussian mixtures per state than at lower number of Gaussian mixtures per state. This can be concluded from the fact that the difference of the speech recognition accuracy under different conditions (without noise, with noise level of 80 dB SPL and with noise level of 95 dB SPL) is the smallest at 32 Gaussian mixtures per state.

Considering the average of the obtained speech recognition results at 32 Gaussian mixtures per state, we can conclude that speech recognition accuracy was reduced by 1.59 %, when the noise level of 80 dB SPL was present on speaker's headphones and by 4.60 %, when the noise level of 95 dB SPL was present on speaker headphones.

6. Conclusion

In this chapter we made a short review of papers covering the topic of the Lombard effect, which were written by researchers in last hundred years. Nowadays, the presence of Lombard effect can be expected in contemporary speech recognition applications in numerous application domains. We made experiments to present the influence of Lombard effect on speech recognition. In order to do so, we used the Slovenian Lombard Speech Database, which was presented in Section 2. The Slovenian Lombard Speech Database was recorded in studio environment. In Section 3, we presented changes of three Lombard speech characteristics: mean value of pitch, phoneme duration and frequency envelope. In Section 4, the experimental design was presented. Results were presented in Section 5. With the experiments we confirmed the influence of Lombard effect on speech recognition accuracy.

7. References

Amazi, D. K. & Garber, S. R. (1982). The Lombard sign as a function of age and task. *The Journal of Speech and Hearing Research*, Vol. 25, No. 4, pp. 581–585.

Anglade, Y. & Junqua, J-C. (1990). Acoustic-phonetic study of Lombard speech in the case of isolated-words. *STL Research Reports*, Vol. 2, pp. 129-135.

Applebaum, T.; Hanson, B. & Morin, P. (1996). Recognition strategies for Lombard speech. *STL Research Reports*, Vol. 5, pp. 69-75.

Bond, Z.; Moore, T. & Gable, B. (1989). Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *Journal of the Acoustical Society of America*, Vol. 85, No. 2, pp. 907 -912.

Bořil H., Bořil T. & Pollák P. (2006). Methodology of Lombard speech database acquisition: Experiences with CLSD, *Proceedings of the fifth Conference on Language Resources and Evaluation – LREC'06*, Genoa, Italy, pp. 1644-1647.

Dreher, J. & O'Neill, J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *Journal of the Acoustical Society of America*, Vol. 29, No. 12, pp. 1320-1323.

Egan, J. (1971). The Lombard reflex: Historical perspective. *Archives of otolaryngology*, Vol. 94, pp. 310-312.

ETSI ES 201 108 v1.1.1 (2000*). Speech Processing, Transmission and Quality aspects (STQ), Distributed speech recognition, Front-end feature extraction algorithm, Compression algorithm*, ETSI standard document, Valbonne, France.

Hanley, T. & Steer, M. (1949). Effect of level of distracting noise upon speaking rate, duration and intensity. *Journal of Speech and Hearing Disorders*, Vol. 14, No. 4, pp. 363-368.

Hansen J. H. L. (1988). *Analysis and Compensation of Stressed and Noisy Speech with Application to Robust Automatic Recognition*, Ph.D. dissertation, School of Elect. Eng., Georgia Inst. of Technol., Atlanta.

Hirsch H. G. & Pearce D. (2000). The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions, *Proceedings of the ISCA ITRW ASR'00*, Paris, France.

Junqua, J-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, Vol. l, pp. 5 10-524.

Junqua, J-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, Vol. 20, No 1-2, pp. 13-22.

Junqua, J-C. & Anglade, Y. (1990). Acoustic and perceptual studies of Lombard speech: Application to isolated-words automatic speech recognition. *STL Research Reports*, Vol. 2, pp. 73-81.

Kaiser J. & Kačič Z. (1997). *SpeechDat Slovenian Database for the Fixed Telephone Network*, University of Maribor, Maribor, Slovenia.

Ladgefoged, P. (1967). *Three Areas of Experimental Phonetics*. Oxford Univ. Press., London, U.K.

Lane, H. & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, Vol. 14, pp. 677-709.

Lombard, E. (1911). Le signe de l'elevation de la voix, *Annals maladiers oreille, Larynx, Nez, Pharynx*, Vol. 37, pp. 101-119.

Pickett J. M. (1956). Effects of vocal force on the intelligibilty of speech sounds, *Journal of the Acoustical Society of America*, Vol. 28, No. 5, pp. 902–905.

Van Summers, W.; Pisoni, D.; Bernacki, R.; Pedlow, R. & Stokes M. (1988). Effects of noise on speech production: acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, Vol. 84, No. 3, pp. 917-928.

Vlaj, D.; Zögling Markuš, A.; Kos, M. & Kačič, Z. (2010). Acquisition and Annotation of Slovenian Lombard Speech Database, *Proceedings of the seventh Conference on International Language Resources and Evaluation – LREC'10*, Valletta, Malta, pp. 595-600.

Young, S.; Kershaw, D.; Odell, J.; Ollason, D.; Valtchev, V. & Woodland, P. (2000). *The HTK book*, Version 3.0, Microsoft Corporation, USA.