PAN ARCHIVES OF ACOUSTICS
Vol. 42, No. 4, pp. 561–569 (2017)

DOI: 10.1515/aoa-2017-0060

Research Papers

Lombard Effect in Polish Speech and its Comparison in English Speech

Piotr KLECZKOWSKI, Agnieszka ŻAK, Aleksandra KRÓL-NOWAK

Department of Mechanics and Vibroacoustics
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland; e-mail: kleczkow@agh.edu.pl

(*received April 26, 2017; accepted August 8, 2017*)

The first extensive investigation on the Lombard effect with Polish speech has been performed. Characteristic parameters of Lombard speech were measured: intensity, fundamental frequency, spectral tilt, duration of words, duration of pauses and duration of vowels. The effect was investigated in a task involving real communication – solving a Sudoku puzzle. The speakers produced speech in quiet and in three different backgrounds: competing speech, speech-shaped noise and speech-modulated noise. The experimental conditions were held as close as possible to those in the study by Cooke and Lu (2010) so that conclusions could be drawn whether differences between the Lombard effect in Polish speech and English speech existed. Most of the findings on the Lombard effect known from the literature have been confirmed with Polish speech. In three parameters, Polish speakers were more sensitive to modulated backgrounds while English speakers were more sensitive to a stationary background. In both languages, the modulated backgrounds induced speakers to extend pauses in the communication tasks.

Keywords: Lombard effect; speech production; Polish speech; speech level; fundamental frequency of speech; spectral tilt.
Abbreviations: RW – reference work, CS – competing speech, SSN – speech shaped noise, SMN – speech modulated noise, F0 – fundamental frequency of speech.

1. Introduction

The effect was first discovered by a French otolaryngologist Étienne Lombard (1911) who found that speakers involuntarily increased their vocal effort when speaking in intense noise, in order to enhance their audibility. Lombard considered that speakers being non aware of the effect was the most important finding. It was deemed that it could be used to unmask people simulating deafness (Zollinger, Blum, 2011b), or to make diagnosis on unilateral deafness (Lombard, 1911).

Some authors have referred to it as the Lombard reflex, e.g. (Egan, 1971) although further research showed that cognitive functions of the brain were also involved (Jürgens, 2009) and that humans were able to control it (Pick, 1989; Winkworth, Davis, 1997; Patel, Schell, 2008).

Further research showed that several other speech parameters were also affected during adverse listening conditions (Lane, Tranel, 1971; Junqua, 1993; Lau, 2008, Zollinger, Brumm, 2011a). Spectral energy was shifted from low towards middle frequencies (Garnier, 2008; Cooke, Lu, 2010). This is often referred to as flattening of spectral tilt and was found to increase the intelligibility of speech by enhancing frequencies in the range of ear's high sensitivity (Rostolland, 1982; Lu, Cooke, 2009). Fundamental frequency (F0) was increased (Cooke, Lu, 2010); the duration of words and vowels was increased (Garnier, 2010); F1 and F2 formant frequencies were shifted upwards (Cooke, Lu, 2010). Some non-acoustic effects were also observed, like larger facial movements (Vatikiotis-Bateson *et al.*, 2006) or larger lung volume used (Winkworth, Davis, 1997).

Lombard effect was also found in over 15 animal species, mainly birds and monkeys (Zollinger, Brumm, 2011a; Potash, 1972; Sinnott *et al.*, 1975; Brumm, Todt, 2002; Tressler, Smotherman, 2009). Birds raise loudness and frequency of their singing in the presence of noise (Nemeth, Brumm, 2010).

The magnitude of the Lombard effect in humans depends, among other factors, on the presence of another person or persons to whom we want to convey the message, the importance of the communicated mes-

sage, the kind of disturbance, or whether both the speaker and the listener or just one of them (and which one) are affected by noise (Lane, Tranel, 1971; Lau, 2008). One of the factors is gender, women tend to increase the level of their voice more than men (Egan, 1971), while in a bird from the finch family an opposite tendency was found (Kobayasi, Okanoya, 2003).

Several hundreds of works on the Lombard effect have been published. Most of the works were done in English, often based on closed speech corpora like numbers. Few works on the effect in other languages can be found: French, Czech and Slovenian (Garnier, 2010; Boril, Pollak, 2005; Vlaj, Kačič, 2011).

Recently, several works written in Polish by Mikulski and Jakubowska addressing the vocal effort appeared. In (Mikulski, Jakubowska, 2013a) the level of voice of teachers and the acoustic background in acoustically adapted and non-adapted classrooms was investigated. It was found that in adapted classrooms the vocal effort could be lower, as the background was lower. In the other paper (Mikulski, Jakubowska, 2013b) teachers were found to speak with different levels in rooms with the same acoustic properties.

There were two purposes for this work. First, to make an extensive examination of the Lombard effect in Polish. Second, to compare the effect in Polish with that found in English, on the basis of one of the published works. The published work that we chose for reference was by Cooke and Lu (2010), and will be further referred to as the reference work (RW). The work by Cooke and Lu was used as the reference since a wide scope of Lombard parameters was investigated there and it contained novel investigation of the effect of informational masking on the Lombard effect. Most of the previous work used stationary noise as masking noise, e.g. white noise (Junqua, 1993), or babble noise. A competing talker as a masker, as used in the RW, has rarely been investigated. As an outcome of interest in informational masking, the authors of the RW obtained for the first time the temporal effects on speech production in the face of modulated maskers. In an earlier work (Lu, Cooke, 2008) the authors of the RW have compared the Lombard effect induced by a competing talker and stationary noise, by using N-talker signals for six values of N, from one through infinity. They found that the effect increased with both the number of background talkers N and noise level.

Another reason for choosing the RW as the main reference was that the methods used in the RW could be fairly closely replicated. The method and conditions used by the authors were as close as possible to those in the RW. Rather than summarising the RW, the authors will report any differences that were unavoidable, or the lack of information on possible differences between their method and that used in the RW. Out of the speech parameters investigated in the RW, only those classified by Cooke and Lu as related to the Lombard effect were analysed in this work, plus two other, considered by the authors as important in speech communication.

The Lombard effect on the following speech parameters was investigated: intensity, changes in F0, the shift of energy towards higher frequencies, duration of words, duration of pauses and duration of vowels.

2. Speech recordings

Each participant spoke in four conditions: in quiet (Q), and in the presence of three different maskers: competing speech (CS), speech-shaped noise (SSN) and speech-modulated noise (SMN).

2.1. Masking signals

Each of the masking signals lasted 10 minutes and all were normalised to the same RMS value. The authors developed a specific procedure to minimise the effect of breaks in speech on RMS value of each signal.

2.1.1. Competing speech (CS)

The recordings to obtain the CS signal were held in an acoustically isolated recording studio of the AGH University. Speech was captured by the Studio Projects C4 $1/2''$ condenser microphone with omnidirectional capsules, amplified and digitised with the M – Audio Fast Track Pro interface and was recorded at 16 bit/44.1 kHz resolution. In the RW microphones with similar properties but a different recording setup were used, and the recorded speech was produced in quiet conditions as well. There were two female and two male speakers aged 20, they did not take part in the subsequent main experiment on Lombard speech.

In the RW, CS was obtained by recording participants solving a Sudoku puzzle individually in quiet. In this work the first attempt to follow this method failed, as the recording contained long periods of quiet or non-speech sounds to be removed. The participant commented that he found it very difficult to verbally express his way of thinking while concentrating on the solution. A variety of methods for eliciting quasi-spontaneous speech for the purpose of obtaining laboratory-quality recordings was used, many of them based on presenting pictures to talkers, like the Diapix task (Van Engen *et al.*, 2010; Baker, Hazan, 2011).

In order to obtain fluent and natural speech as CS signal a simple solution was chosen. Each of the talkers was given a set of picture cards from the Dixit association game and was asked to describe them freely (the game was not actually played). In order to encourage talking, they were presented three questions in printed form:

- What does a picture show?
- What is your association?
- What story is it likely to tell?

The final CS was collated from manually edited four individual speech recordings with the use of Praat software v 6.0.20. All breaks, breathing, hemming, coughing or unintelligible utterances were removed.

2.1.2. Speech-shaped noise (SSN)

The SSN has been generated by passing a pseudo-random white noise through a filter with a spectral characteristics equal to the long term average amplitude spectrum of a 5 s long fragment of the CS signal. The filter and the filtering operation were designed and implemented in MATLAB. The frequency characteristics of the SSN obtained is shown in Fig. 1.

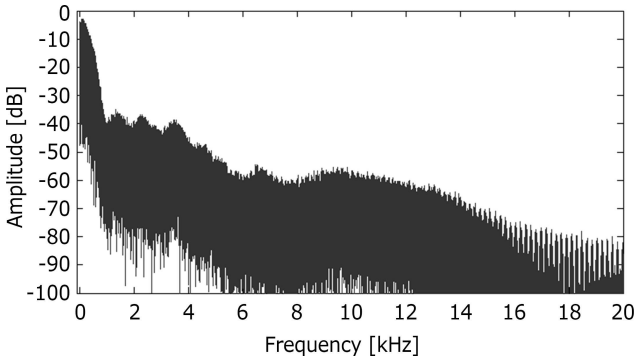

Fig. 1. The frequency characteristics of the SSN signal.

2.1.3. Speech-modulated noise (SMN)

The SMN signal was the effect of amplitude modulation of the SSN. The modulating signal was the time envelope of one of the four speech recordings made to produce the CS signal. The female recording evaluated as most fluent of all four was chosen. In order to obtain the envelope, first the analytic signal of the speech signal was computed with the use of the MATLAB *hilbert()* function, and then its absolute value was obtained.

2.2. Collection of Lombard speech

2.2.1. Participants

Six participants took part in the experiment, three female aged 24–28 and three men aged 23–26 (in the RW there were four women plus four men). Three participants were graduates of Acoustic Engineering studies at the AGH University, and three others were students or graduates of other studies, not related to acoustics or music. All participants were native Polish speakers. According to (Kleczkowski, Pluta, 2012) there is no effect of the listener's audiometric threshold (as long as impairments are mild) on his/her performance in listening tasks with test material well above threshold, but the audiograms were taken from all participants in 11 frequencies and were all found to be within 20 dB HL.

2.2.2. Task

In each of the four conditions of speaking (Q, CS, SSN, SMN), a speaker talked to a listener, so that all tasks were communication tasks, which are in general known to produce a more pronounced Lombard effect (Junqua, 1993; Lau, 2008; Garnier, 2010). The experiment was carried out in pairs, and members of each pair exchanged their roles. Members of pairs knew each other which encouraged natural conversation.

Following the method used in the RW, members of the pair collectively solved a Sudoku puzzle, which consists in filling the 9x9 matrix with digits, according to some rules. The advantages of using it are that speech is moderately natural and it elicits many repetitions of spoken digits. The last advantage facilitates acoustic analysis that follows. All participants declared they were familiar with Sudoku. Puzzle diagrams of intermediate difficulty were downloaded from *www.dailysudoku.com* website.

2.2.3. Experimental conditions

The recordings were held in the recording studio of the AGH University. A pair of participants sat at opposite sides of a table. The table was not covered with any sound absorbing material to reduce comb effects, as this was not mentioned in the RW. Between the participants there was a barrier formed with a wooden frame covered with an acoustically neutral cloth. Its purpose was to prevent participants from seeing Sudoku diagram of a partner and their possible communication through non-verbal means.

Each participant spoke to a microphone placed at the level of his or her mouth at the distance of 20 cm. When speakers spoke together, crosstalk between microphones should be minimised. In the RW a stiff cardboard was used for this purpose and in this work an additional Reflexion Filter from sE Electronics was used at the side of one of the speakers. Figure 2 presents a pair of participants during a recording session.

Fig. 2. Two participants at the Lombard speech recording facility.

Masking signals (CS, SNN, SMN) were reproduced diotically by closed headphones Beyerdynamic DT 770. This solution provides full control over masking signals, but brings the problem of attenuating the speaker's own speech. Following the RW the problem was solved by adding some amount of speaker's own voice to masking signals, before sending them to the headphones. The speaker's speech was routed from the audio interface M-Audio Fast Track Ultra 8R to the mixing headphone amplifier Presonus HP60, mixed with one of the maskers (CS, SNN, SMN) and passed to the headphones. The signal paths in the test setup are presented in Fig. 3. The devices in the setup were different than those used in the RW.

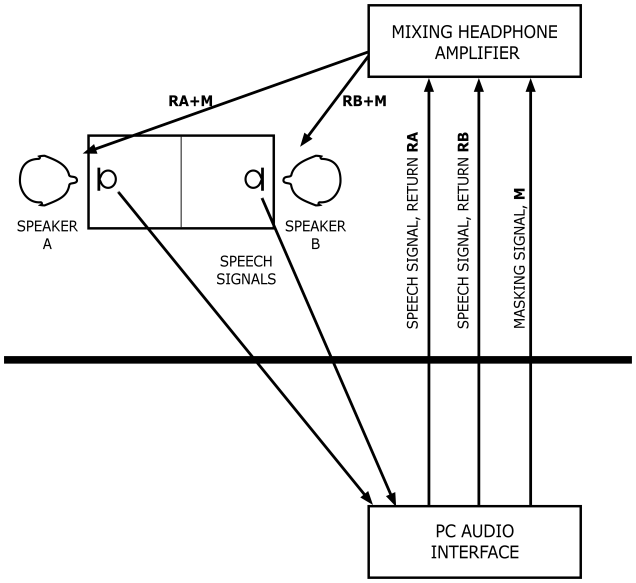

Fig. 3. Signal paths in the test setup.

Such a compensation procedure needs an appropriate setting of the levels of RA and RB signals. At the beginning of the experiment, each member of the pair put the headphones on and was asked to talk freely to her or his microphone. The level of the return speech RA or RB was adjusted manually until the speaker had the impression that the overall loudness level of his voice not wearing the headphones matched that when wearing them. Once the level of RA or RB had been set, the participants could not change it. Such a procedure may introduce an error, but it was used in the RW and the authors of the RW considered it appropriate (Cooke, 2016) and informed of another research using it successfully. The same test setup was also used during recording in the Q (quiet) condition.

Before each session the levels of the masking signals were calibrated in the headphones with the use of the Brüel & Kjær type 2209 sound level meter and an intermediate headphone adapter developed in the Department of Mechanics and Vibroacoustics of the AGH University. Since signals in all conditions had a normalised RMS level, the calibration was performed only once.

The sound level was kept at 82 dB. According to the RW, this level produces moderate Lombard effect but is adequate. Similar levels were used in other works. Summers *et al.* (1988) and Pittman and Wiley (2001) used 80 dB, while Garnier (2007) used 85 dB. Speech signals from the audio interface were recorded with 16 bit/44.1 kHz resolution.

2.2.4. Course of experiment

The members of each of the pairs were informed that the purpose was to solve a Sudoku puzzle together, and they were handed two copies of a Sudoku form.

The first condition for all three pairs was a quiet environment. No masking signal was presented over the headphones, beside some level of speaker's own voice, as explained above. The recording in each of the conditions, including the Q condition, lasted as long as each of the masking signals, i.e. 10 min. After each condition a break was held, the participants could have a rest and drink some water. Next tasks (conditions) were initiated only when the participants said they were ready. The recordings were held in two days. It took about two hours for a pair to complete all tasks.

2.3. Transcription

Transcription was performed with the use of Praat 6.0.20 software. All words were marked in the time domain window and transcribed in appropriate fields. All non-speech or unintelligible utterances were removed.

In the RW the digits "one" to "nine" were used as the speech corpus to analyse acoustical features, and many repetitions were available. This is more difficult in Polish because of a rich set of possible grammatical forms of numerals. In effect, basic forms of numerals occurred rarely. It was decided to include in the corpus all grammatical forms of numerals plus the words denoting position, like "up", "down", "left", right", with their declensions.

The microphone crosstalk mentioned in Subsec. 2.2.3 did not affect the location of boundaries of speech segments nor the decisions speech/non-speech segments. The average crosstalk was −10.62 dB, while it was −12 dB in the RW. The overlapped words were omitted from the analysis.

3. Analysis of acoustic features

3.1. Intensity

RMS values for entire words marked in Praat software were taken as their intensity. They were given by Praat.

3.2. Fundamental frequency

F0 estimates were also given by Praat, which uses autocorrelation-based method. 10 ms intervals were used. Because of the difference between average F0 values in women and men, the results were analysed for all listeners as well as for females and males separately.

3.3. Spectral tilt

The authors of the RW informed that spectral tilt estimates were obtained from a linear regression of the long-term average spectrum in the range 0–8 kHz, expressed in dB/octave, and implemented in MATLAB. As no further information was given, these authors developed their own similar procedure in MATLAB.

In the first step, seven octave-wide IIR Butterworth band-pass filters were designed with the Filter Design function in bands from 125 Hz through 8 kHz. In this analysis, only one word from the corpus was used: "siedem" (seven), as it had the highest occurrence. In the second step, each sample of this word was passed through the filters and the RMS in each band was calculated and then converted to dB values referenced to the full scale value. In the next step, average values for all words in each condition were calculated and in the last step the spectral tilt expressed in dB/octave in each condition was calculated from the linear regression.

3.4. Duration of words

In the RW the average word duration was computed as the average in a given condition from all words from "one" through "nine". Because of the multitude of forms in Polish, in this work it was impossible to obtain items from the corpus spoken by all participants in all conditions. Particular speakers tended to use one of the possible forms more often than others, and did not use some forms at all. It was decided to use only those words that had at least three occurrences in each condition. There were six such words. In view of these specific features of Polish language it seems that solving Sudoku is not really an appropriate communication task in this language and the battleship game could be a better option. This option was not attempted in this work in order to stay consistent with the RW.

The values of durations were obtained from Praat, and the medians of the duration for each word and each condition were calculated. Elongations or abridgements in conditions CS, SSN and SMN relative to Q were calculated. Both absolute and relative differences in percent were calculated.

3.5. Duration of pauses

Only pauses occurring amidst a longer utterance were taken, as there were numerous extraordinarily long pauses, for example when a participant thought over his next move in the puzzle. For each participant and for each condition 10 pauses between neighbouring words were chosen randomly. Thus, 60 values were obtained in each condition. These values were also obtained from Praat using the option of "silent/sounding" analysis. The analysis yielded by the software had to be carefully controlled manually, as its indications were often misleading when some low signals were present during periods of low intensity.

3.6. Duration of vowels

This experiment was not performed in the RW but was added to this work because of the importance of vowels in intelligibility (Kewley-Port *et al.*, 2007) and the scarcity of literature on that aspect of the Lombard effect (Junqua, 1993; Vlaj, Kačič, 2011).

A difficulty in vowel analysis in continuous speech is that its signal is usually a superposition of a vowel and consonants preceding and following it (Liberman *et al.*, 1967). Therefore, the analysis in this work was based on vowels occurring in the same words. The duration was evaluated in all four conditions for the following vowels: "y" in "trzy", "a" in "piątka" and "e" in "jeden" (second occurrence). Evaluations were based on auditioning, visual analysis of spectrograms, plots of energy versus time and formants versus time, all provided by Praat.

4. Results and discussion

One factor independent measures ANOVA (Analysis of Variance) was used to test hypotheses that conditions had statistically significant effects on acoustic features. Since there was just one group of listeners, the repeated measures ANOVA seemed more appropriate, but independent measures ANOVA was used in the RW. Besides, some results were based on different speakers hearing different words from the corpus, and that made independent measures ANOVA justified. Normality of distributions were verified by Kolmogorov-Smirnov test and the homogeneity of variance by Levene test. ANOVA was performed when both conditions were satisfied, and Welch's ANOVA was used when the data failed the test for homogeneity of variance. The latter was followed by the Brown-Forsythe post-hoc procedure. In cases where the assumption of normality was not met, the nonparametric Kruskal-Wallis test was used, followed by post-hoc pairwise comparisons by paired t-tests with the Bonferroni-adjustment.

The data set for each masker condition were all utterances of words from the corpus, spoken by all participants, unless specified otherwise. For all conditions the medians of the data sets were taken for comparisons.

4.1. Intensity

Figure 4 presents quartiles of RMS of speech for all conditions.

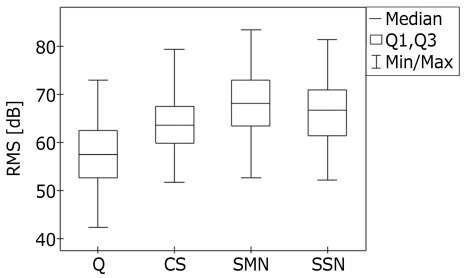

Fig. 4. Quartiles of RMS of speech in the entire panel of listeners for all conditions.

In all conditions the data satisfied the assumption of normality. The results in groups of different conditions showed significant differences. The type of masker affects the RMS values (FWelch (3; 603.48) = 157.86; $p < 0.001$, $\eta^2 = 0.31$). According to the scale proposed by Coohen (1988) this is an effect of high power. The post-hoc analysis confirmed the differences between all four groups of results, in three cases at the level of significance 0.001, and for the pair SSN-SMN at the level of significance 0.025. In this point there is a difference with the results in the RW, where no significant differences were found between CS and SMN conditions. In the RW the highest Lombard effect was noticed for SSN masker, then for CS and SMN. In this work the highest vocal effort can be seen in SMN condition, then follow SSN and CS conditions.

The spread of results in each condition ($X_{\max} - X_{\min}$) was fairly high, in SMN it was nearly 31 dB. Such a high spread could have four sources. The first is the difference between genders. Others could be the nature of the task and emotional context of an utterance. It was noticed that the participants spoke louder at moments when they stroke upon the idea of a solution than when they just considered a solution aloud. The last factor was changes in the distance from the mouth to the microphone. This was noticed and corrected by the test operator, but could not be eliminated altogether.

Further analysis in subgroups of males and females revealed that women produced significant differences between Q condition and other three conditions, at $p < 10^{-6}$ and insignificant other differences, while the tendencies among men were similar as in the entire panel of participants. The average increase of speech level in CS, SSN and SMN conditions over speech in quiet was 8.07 dB for women and 6.65 dB for men and was significant ($p < 0.05$) confirming the observations in (Egan, 1971). The gender differences were not evaluated in the RW.

4.2. Fundamental frequency

Table 1 presents the mean values of F0 for all participants and in subgroups, for all conditions, and Fig. 5 presents quartiles for all participants.

Table 1. Average values of F0.

	All	Women	Men	W-M
Q	158.1	202.5	115.6	86.9
CS	185.3	240.3	132.9	107.4
SMN	170.3	246.9	136.9	110.0
SSN	170.9	241.9	130.3	111.6

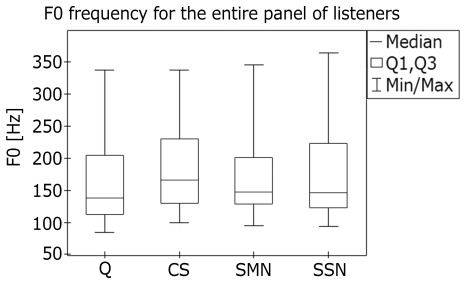

Fig. 5. Quartiles of F0 in the entire panel of listeners for all conditions.

The Kolmogorov-Smirnov test showed that none of the groups had normal distribution. Therefore the Kruskal-Wallis test was used and confirmed the significance of the differences between the medians ($p < 10^{-6}$). Pairwise comparisons with a post-hoc test revealed significant differences between all groups except for CS-SMN ($p = 0.0578$) and SMN-SSN ($p = 0.99$). Thus, it was demonstrated that all maskers raise F0, and CS was most effective. In the RW SSN was most effective and all differences were significant except for CS-SMN, which confirms our result.

4.3. Spectral tilt

Linear regression lines obtained according to Subsec. 3.3 are shown in Fig. 6. The values of spectral tilt per octave are given in Table 2.

Figure 6 and Table 2 demonstrate that spectral tilts get less steep when maskers are present, i.e. energy is

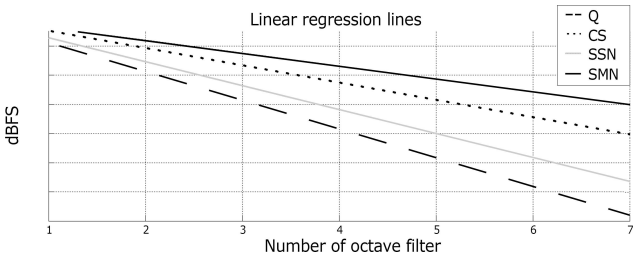

Fig. 6. Linear regression lines of spectral tilt in four conditions.

Table 2. Values of spectral tilt per octave in four conditions.

Condition	Q	CS	SSN	SMN
Tilt [dB/octave]	−1.96	−1.18	−1.63	−0.88

shifted up to higher frequencies. The data were found normally distributed and the differences for different conditions were found significant (FWelch (3; 50,57) = 5,68; $p < 0.002$). The post-hoc test indicated a significant difference only between Q and SMN conditions ($p = 0.003$). In the RW no significant differences were found between any pair of results.

The spectral tilt flattened for all maskers in Polish, with the biggest effect from SMN followed by CS. Thus, in Polish the shift of energy was evoked efficiently by modulated maskers. In English the opposite was found, SSN, i.e. stationary masker was most efficient, and both modulated maskers had lower and similar effect. This is the third parameter where Polish talkers were more sensitive to modulated maskers and English ones were more affected by the stationary masker.

4.4. Duration of words

Figure 7 shows relative change of word duration in conditions CS, SSN and SMN to the duration in quiet.

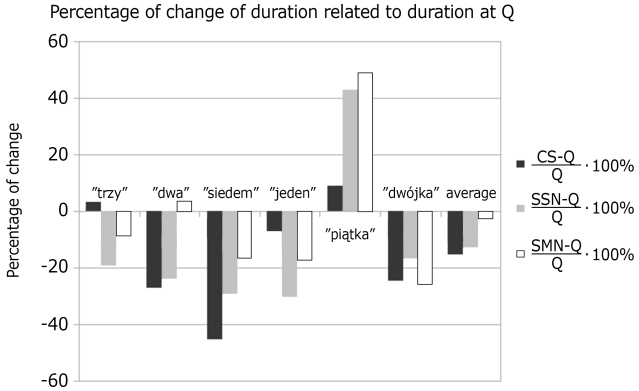

Fig. 7. Change of word duration in conditions CS, SSN and SMN expressed as percentage of duration in condition Q. Words in Polish are given above the zero value line.

The data distributions were normal and there were no differences between groups ($p = 0.85$). Therefore, no relation between the duration of words and the masker was found. This relation has not been found in the RW either.

A widely known phenomenon, e.g. (Baker, Bradlow, 2009), that there is a dependence of word duration on the context, was confirmed informally in this experiment. Words spoken in isolation lasted longer than those in continuous speech. A digit given firmly as a solution lasted shorter than one spoken with deliberation. This could be the reason why no effect was found neither in this work nor in the RW. The context effects could override any acoustic background effects, if present. The conclusion is that the use of Sudoku as a communication incentive is inappropriate in this case.

4.5. Duration of pauses

Figure 8 presents quartiles of pauses for all four conditions.

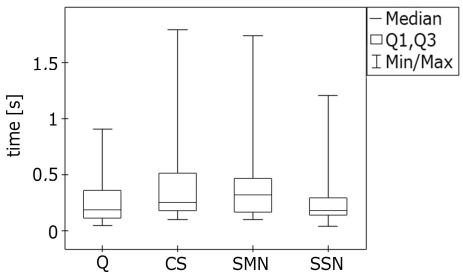

Fig. 8. Quartiles of pauses in the entire panel of listeners for all conditions.

Since the distributions were not normal, the same analysis as for F0 was used. The Kruskal-Wallis test showed significant differences between groups ($p = 0.0062$). The post-hoc test revealed significant differences between most of the pairs. In two cases there was no significant evidence to demonstrate that mean values were different: when Q was compared with SSN ($p = 0.98$) and when SMN was compared with CS ($p = 0.99$). In order to confirm that it was the type of a masker that had affected mean pause duration the post-hoc test was performed for two groups. One group consisted of Q and SSN conditions (stationary background), and the other of SMN and CS (modulated background). It confirmed a significant difference ($p = 0.005$). Thus, it was demonstrated that an amplitude modulated masker induces speakers to extend pauses. The same was shown in the RW, but only in tasks involving communication, like all tasks in this work. Authors of the RW speculated that the speakers could wait until an appropriate moment to make their contributions in the face of a modulated masker. This was confirmed during this work by an informal comment from one of the participants. Lau (2008) observed the Lombard effect in the condition where the speaker knew that the listener was in a noisy environment, while he himself was not. Another motivation to extend pauses can be speaker's own experience that he or she hears better during troughs in modulation.

4.6. Duration of vowels

The following figures present quartiles of the duration of vowels: "a" in Fig. 9, "e" in Fig. 10 and "y" in Fig. 11.

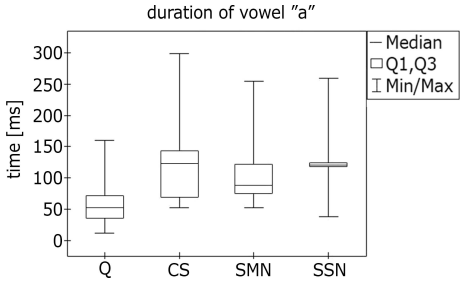

Fig. 9. Quartiles of the duration of vowel "a".

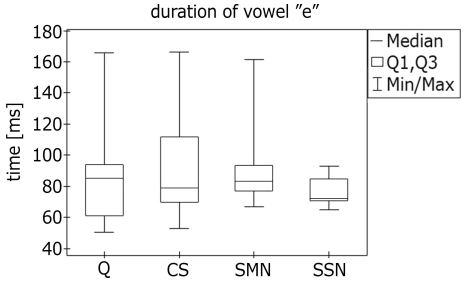

Fig. 10. Quartiles of the duration of vowel "e".

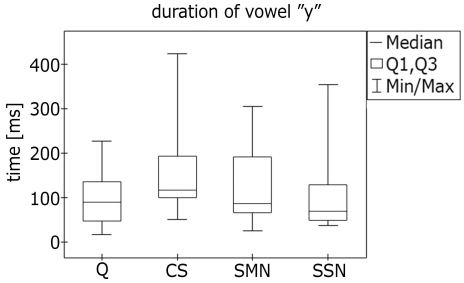

Fig. 11. Quartiles of the duration of vowel "y".

All the distributions were normal. The increase of duration of vowel "a" can clearly be seen in Fig. 9, but ANOVA did not confirm the significance of that observation. There are two likely reasons for this failure. In this analysis a relatively low number of words could be used (around 10 utterances of each word in one condition). It was also observed informally, that in this experiment an emotional context of speech considerably affected the results.

5. Conclusions

Three general findings on the Lombard effect known from the literature – increases of speech level, fundamental frequency and the shift of energy towards higher frequencies have been confirmed with Polish speech. Furthermore, the observation by Cooke and Lu (2010) of pause extension in the presence of modulated background was confirmed. Some increase in the duration of vowel "a" was also noticed.

Some similarities and some differences were found between the Lombard effect in Polish and English, on the basis of having performed the experiments in as close conditions as was possible.

Polish talkers were found to be more sensitive to modulated maskers and English ones were more affected by the stationary masker.

Separate analysis for males and females revealed that the latter increased their vocal effort more (by 8.07 dB on the average) than males (6.65 dB).

Acknowledgments

This research was partly supported by AGH University grant no. 11.11.130.995.

References

1. Baker R.E., Bradlow A.R. (2009), *Variability in Word Duration as a Function of Probability, Speech Style, and Prosody*, Language and Speech, **52**, 391–413.
2. Baker R., Hazan V. (2011), *DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs*, Behaviour Research Methods, **43**, 761–770.
3. Boril H., Pollák P. (2005), *Design and Collection of Czech Lombard Speech Database*, http://www.isca-speech.org/archive/interspeech_2005/i05_1577.html, accessed 2016 Aug. 25.
4. Brumm H., Todt D. (2002), *Noise-dependent song amplitude regulation in a territorial songbird*, Animal Behaviour, **63**, 891–897.
5. Coohen J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, New Jersey, Lawrence Erlbaum Associates, 283–287.
6. Cooke M., Lu Y. (2010), *Spectral and temporal changes to speech produced in the presence of energetic and informational maskers*, Journal of the Acoustical Society of America, **128**, 2059-2069.
7. Cooke M. (2016), private communication.
8. Egan J.J. (1971), *The Lombard Reflex. Historical Perspective*, Archives of Otolaryngology, **94**, 310–312.
9. Egan J.J. (1975), *Use of the Lombard response in cases of hysterical aphonia*, Archives of Otolaryngology, **101**, 557.
10. Garnier M. (2008), *May speech modifications in noise contribute to enhance audio-visible cues to segment perception?*, http://www.gipsa-lab.grenoble-inp.fr/∼maeva.garnier/Garnier-AVSP-08.pdf, accessed 2016 Aug. 25.
11. Garnier M. (2010), *Influence of Sound Immersion and Communicative Interaction on the Lombard Effect*, http://www.gipsa-lab.grenoble-inp.fr/∼maeva.garnier/Garnier-JSLHR-10.pdf, accessed 2016 Aug. 25.
12. Junqua J-C. (1993), *The Lombard reflex and its role on human listeners and automatic speech recognizers*, Journal of the Acoustical Society of America, **93**, 510–524.
13. Jürgens U. (2009), *The neural control of vocalization in mammals: a review*, Journal of Voice, **23**, 1–10.

14. Kewley-Port D., Burkle T.Z., Lee J.H. (2007), *Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners*, Journal of the Acoustical Society of America, **122**, 2365–2375.

15. Kleczkowski P., Pluta M. (2012), *Normally Hearing Subjects Have No Advantage of Better Audiograms in Listening Tasks*, Acta Phys. Pol. A, **121**, A120–A125.

16. Kobayasi K.I., Okanoya K. (2003), *Sex differences in amplitude regulation of distance calls in Bengalese finches, Lunchula striata var. domestica*, Animal Behaviour, **53**, 173–182.

17. Lane H., Tranel B. (1971), *The Lombard sign and the role of hearing in speech*, Journal of Speech, Language and Hearing Research, **14**, 677–709.

18. Lau P. (2008), *The Lombard Effect as a Communicative Phenomenon*, UC Berkeley Phonology Lab Annual Report 2008, 1–9.

19. Liberman A.M., Cooper F.S., Shankweiler D.P., Studdert-Kennedy M. (1967), *Perception of the speech code*, Psychological Review, **74**, 6, 431–461.

20. Lombard E. (1911), *The sign of the elevation of the voice* [in French: *Le signe de l'elevation de la voix*], Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx, **37**, 101–119, Eng. transl.: http://paul.sobriquet.net/wp-content/uploads/2007/02/lombard-1911-p-h-mason-2006.pdf

21. Lu Y., Cooke M. (2008), *Speech production modifications produced by competing talkers, babble, and stationary noise*, Journal of the Acoustical Society of America, **124**, 3261–3275.

22. Lu Y., Cooke M. (2009), *The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise*, Speech Communication, **51**, 12, 1253–1262.

23. Mikulski W., Jakubowska I. (2013a), *The results of the study on the reduction of the voice of teachers and the reduction of acoustic background noise in the rooms after the acoustic adaptation* [in Polish: *Wyniki badań zmniejszenia natężenia głosu nauczycieli oraz zmniejszenia hałasu tła akustycznego w salach po wykonaniu adaptacji akustycznej*], Bezpieczeństwo pracy – praktyka i nauka, **6**, 10–12.

24. Mikulski W., Jakubowska I. (2013b), *Level of lecturer's voice – results of lectures conducted* [in Polish: *Poziom natężenia głosu wykładowców – wyniki badań przeprowadzonych podczas wykładów*], Medycyna Pracy, **64**, 797–804.

25. Nemeth E., Brumm H. (2010), *Birds and anthropogenic noise: are urban songs adaptive?*, The American Naturalist, **176**, 465–475.

26. Patel R., Schell K.W. (2008), *The influence of linguistic content on the Lombard effect*, Journal of Speech, Language and Hearing Research, **51**, 209–220.

27. Potash L.M. (1972), *Noise-induced changes in calls of the Japanese quail*, Psychonomic Science, **26**, 252–254.

28. Rostolland D. (1982), *Acoustic features of shouted voice*, Acoustica, **50**, 118–125.

29. Sinnott J.M., Stebbins W.C., Moody D.B. (1975), *Regulation of voice amplitude by the monkey*, Journal of the Acoustical Society of America, **58**, 412–414.

30. Tressler J., Smotherman M.S. (2009), *Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats*, Journal of Comparative Physiology, **195**, 923–934.

31. Winkworth A.L., Davis P.J. (1997), *Speech breathing and the Lombard effect*, Journal of Speech, Language and Hearing Research, **40**, 159–169.

32. Van Engen K.J, Baese-Berk M., Baker R.E., Choim A., Kim M., Bradlow A.R. (2010), *The Wildcat Corpus of Native- and Foreign-Accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles*, Language Speech, **53**, 510–540.

33. Vatikiotis-Bateson E., Chung V., Lutz K., Mirante N., Otten J., Tan J. (2006), *Auditory, but perhaps not visual, processing of Lombard speech*, The Journal of the Acoustical Society of America, **119**, 3444.

34. Vlaj D., Kačič Z. (2011), *The Influence of Lombard Effect on Speech Recognition*, [in:] *Speech Technologies*, Ed. Ipsic I., http://www.utdallas.edu/∼hynek/citing_papers/Vlaj_The%20Influence%20of%20Lombard%20Effect%20on%20Speech%20Recognition.pdf, accessed 2016 Aug. 25.

35. Zollinger S.A., Brumm H. (2011a), *The evolution of the Lombard effect: 100 years of psychoacoustic research*, Behaviour, **148**, 1173–1198.

36. Zollinger S.A., Brumm H. (2011b), *The Lombard effect*, Current Biology, **21**, 16, 614–615.