9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Speaking style variation in laboratory speech: A perception study

Katarzyna Klessa & Maciej Karpiński

Adam Mickiewicz University in Poznań, Poland
{katarzyna.klessa|maciej.karpinski@amu.edu.pl}

Abstract

Laboratory speech is often implicitly assumed to be a stable and reliable source of data for phonetic studies and perception tests. At the same time, researchers are conscious that some of its parameters may significantly change due to minute changes in the internal states of the speaker or in recording conditions, including the environment and its acoustic properties. While we are able to detect tiny changes in the spectral and prosodic parameters of speech, it is usually more difficult to decide on their perception and communicative relevance. In the present study an attempt is made to find out whether speaking style changes occurring due to the change of experimental setting (type of microphone), related to the dialogue task stage (e.g. initial vs. final) as well as those resulting from explicit evaluation of the behavior of the participants, can be perceived in short samples of speech. Experimental stimuli are extracted from task-oriented dialogues and used in a perception test involving same-different and two-dimensional evaluation paradigms. The results show that the influence of the experimental setting factor can be detected, especially when combined with speaker factor, i.e. the specific realization of the differences does appear to be setting-dependent but also highly individual.

Index Terms: laboratory speech, variation, speaking styles, perception test, recording conditions

1. Introduction

The way people speak is not determined solely by their communicative intentions and goals but can be also altered depending on a number of factors related to the addressees, audience (witnessing the act of communication), discourse topic, and the environment in which the act of communication occurs. According to Communication Accommodation Theory (e.g., [1]), speech is "socially diagnostic". Different choice and sequencing of words as well as different phonetic realisations might be expected for utterances addressed to peers or to interlocutors representing higher or lower social (e.g., corporate) hierarchy, or a different age group. As Krauss & Chiu [2] claim, "language pervades social life". Remarkable findings have been reported in the field of honorifics or lexical choice mechanisms in high-context cultures [3, 4]. A similar perspective has been adopted in a range of studies of the acoustic aspects of speech, especially in the context of politeness (e.g., [5, 6, 7, 8]). Certain acoustic features of speech have been also reported to be shaped by the properties of the surrounding environment as another potentially significant factor. For example, automatic adjustment of speech volume and quality is a typical reaction to adverse acoustic conditions, specifically to noise that may reduce intelligibility. This phenomenon, known as the Lombard effect, has been found to occur not only in the domain of voice intensity [9, 10] but also in pitch [11] and formant frequencies [12] as well as in duration [13]. It was also shown that some acoustic properties of communication environment, like reverberation, may influence the acoustic properties of speech [14]. Taking these findings into account and considering a more general concept of how environment may influence human behaviour (e.g., according to Gibson's ecological psychology; [15]), it appears justified to assume that the features of physical environment as well as the way it is perceived by speakers may be considered as factors influencing acoustic parameters of spoken utterances. Similarly to what can be observed in everyday communication, laboratory speech remains under the influence of many factors, some of them similar to what is found outside the laboratory, and some laboratory-specific.

The analysis of various types of speech features as well as multilateral interactions among them can help to distinguish between speaking styles (e.g., [16]). It has been found that certain features may depend on the target audience (cf. adult directed speech vs. infant directed speech, e.g., [17, 18]). Schröder et al. [19] report on correlations between emotional characteristics and acoustic features related to *f0*, intensity, spectral slope or speech rate. Geumann [20] investigates differences between read and spontaneous speech using Long-Term Formant Distribution. Grawunder & Winter [8] are looking for acoustic correlates of politeness and verify the results of instrumental acoustic analyses by means of a perception test.

The need for perceptual verification of the findings obtained from acoustic or articulatory research brings in new challenges. One of them is related to rating scales or categorisations of perceived phenomena. The discussion on applying discrete or categorical [21] or continuous dimensions to linguistic – but especially paralinguistic – features uncovers a wide variety of views (e.g. [22]). For example, differing approaches are taken to attempts at applying time-sensitive continuous representations of emotions versus the criticisms of the dimension-based approach in judgement tasks by Clore, Ortony and Foss ([23], pp. 751–752).

The results reported in this paper belong to a series of experiments aimed at investigating the variation of speaking styles in laboratory speech and may be especially relevant for the purposes of preparation of highly controlled speech stimuli for detail-oriented perception studies, e.g., involving ERP (EEG) or ET (eye-tracking). Furthermore, our study may help to generalise findings based on laboratory speech and critically extrapolate them to everyday communication.

Section 2 provides a description of the speech corpus design and development (recording scenarios, tools, perception test procedures). In Section 3, we first briefly characterize the acoustic features of the test stimuli and then

10.21437/SpeechProsody.2018-105

report on the results of the perception study. The findings are summarized in Section 4.

2. Study design

2.1. Speech material

2.1.1. Recording scenarios

The present speech material was selected from a corpus designed and recorded by the authors specifically for the purpose of exploring both acoustic and auditory features of speaking styles. This study focuses primarily on the results of a perception test although certain acoustic properties of the speech signals are also discussed.

We designed and recorded a speech corpus in order to explore changes in speaking styles related to:

- the type of microphone (head-mounted vs. tripod-mounted) as an example of a change in the close physical environment;
- consciousness of being recorded as an example of an impact on the change in the mental state of the speaker;
- the influence of a (fake) negative judgement of the task as an example of a direct impact on the emotional state.

Each recording session followed the same scenario and was divided into four parts as shown in Figure 1.

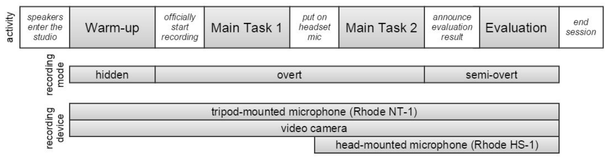

Figure 1: *Recording session design.*

Two speakers participated in each session: Instruction Follower (IF, our secret confederate) and Instruction Giver (IG). Before the recordings, IF had been instructed how to carry the conversation and to encourage the interlocutor to speak. IF and IG could hear each other only via a monitoring system and they could see each other through a soundproof window. No time limits were specified for the duration of any of the session parts.

At the beginning of the session, the speakers were asked to perform a preparatory task (henceforth, Warm-up) and were indirectly informed that it would not be recorded. The Warm-up and both parts of the Main Task were similar: IF's role was to sketch a blueprint of a room on the basis of information provided by IG. At the beginning of Main Task 1, participants were guided to sit in a specific distance from tripod-mounted microphones, and informed that the recording is about to begin. At a certain point of this part of the session (estimated as approximately half-time by the moderators), a technician entered the anechoic chamber and explained that from that moment on the recordings would be continued with a head-mounted microphone because of a technical issue. He assisted IG to wear and attach the head-mounted microphone, and informed the speaker that from that moment on s/he did not have to pay attention to the distance and position related to the table microphone anymore. This part of the conversation is referred to as Main Task 2.

After accomplishing Main Task 2, subjects were asked to stay in their seats and wait for the result of their work to be evaluated by moderators. In fact, the moderators never evaluated anything but just waited for some time and announced a negative score. The IG was handed a written note on the negative assessment, together with a bullet list of possible points to discuss, and was requested to inform the interlocutor about it. The discussion of the results took place during the final stage of the session named Evaluation.

2.1.2. Speakers

Twenty-one students (16 females, 5 males) were recorded, all aged between 19 and 24 years (Mean = 22, SD = 1.7) and none diagnosed with speech fluency problems nor voice quality issues. The speakers had not been familiarized with the particular aims of the study in advance; participants interested in obtaining more information were sent the description of the experimental setup after finalizing the recording sessions for all speakers. The data on their previous experiences as participants of similar studies were collected. The majority of the speakers claimed to be quite inexperienced in this respect: Eight of them had never visited any recording studio beforehand and had not participated in any recording sessions, eight speakers reported only a single visit a studio but no recording experiences, five speakers had more than one earlier experience (several recording sessions, e.g. recording vocals for music projects or teaching materials for language learning).

2.1.3. Tools and procedures

Speakers were recorded using condenser microphones and, in the case of IG (additionally), a head-mounted electret microphone, all connected to Focusrite Liquid Saffire PRO 56 Firewire Audio Interface (44.1 kHz, 24-bit).

The duration of the sessions ranged from 15’30” to 30’03”, including Warm-Ups and the microphone change. The recordings were annotated using multiple annotation layers with Annotation Pro [24]. First, recordings for each session were manually segmented into four stages corresponding to the four parts of the recording session described above (Section 2.1). An approximately 30-second long stretch of speech was selected from the middle part of each stage of the recording session for the needs of detailed analyses. The selected sections of recordings were manually segmented into intonational micro-phrases (MiP, e.g. [25, 26]) and transcribed orthographically on another annotation layer. On the basis of the orthographic transcript, speech signals were automatically segmented into phones and transcribed in broad phonetic transcription using Polphone [27] and Salian [28] integrated into Annotation Pro.

2.2. Perception test

2.2.1. Listeners

Altogether, 31 listeners participated in the test (14 males and 17 females). All of them were students of the first or second grade of the same university faculty (Modern Languages and Literature) aged between 18 and 25 years (Mean = 20.48, SD = 1.9).

2.2.2. Stimuli and experimental set-up

As the speech material for the perception test, 40 speech signals coming from ten speakers were used. The speakers were randomly selected from the corpus described above (only female voices in this case). The stimulus set comprised four signals from each of ten speakers, one per each recording session stage. The number of stimuli was intentionally reduced to avoid listeners' fatigue. The signals had been extracted from the annotated stretches of speech and consisted of one to three short MiPs.

The test was conducted using the perception test mode in Annotation Pro [24]. Participants listened to the stimuli individually, via headphones. The stimuli were presented in random order and each of them could be played no more than twice. Participants responded by clicking on two graphical representations (Figure 2) displaying:

- a horizontal strip representing the continuum between casual and formal speaking style;
- a circle representing valence-activation space for assessment of speaker's emotional state along negative-positive, and active-passive dimensions.

The dimensional approach [22, 29] to the description of paralinguistic features and affective states, potentially useful for dealing with "milder" affective states as opposed to prototypical emotions [30], was followed in the present study.

Listeners were informed that the middle area of the strip and the centre of the circle corresponded to neutral values, and that the values increased proportionally to the distance from the middle. They were instructed to mark their answers according to their own judgement and to their own understanding of the distinctions between the casual-formal, negative-positive/active-passive dimensions.

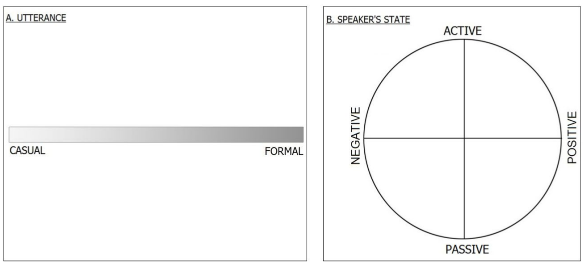

Figure 2: *Graphical representations of feature spaces used in Annotation Pro as perception test rating scales: formal-casual rating scale (a, left), valence-activation space (b, right).*

3. Results and discussion

3.1. Characteristics of the speech material

The recordings extracted from each of the session stages (Section 2.3) were analysed instrumentally to inspect potential prosodic and voice quality variability (often considered to form a group, cf. [31]) as correlates of affective states or speaking styles in varying recording conditions. Due to space limits, the measures will be discussed in detail in a separate publication [32]. Here, we report briefly on some of them as potential correlates of the perception based results.

When comparing the properties of the utterances across subsequent recording session stages, it was observed that the most significant differences occur between the three earlier stages and the Evaluation stage. For certain features, also an "entrance effect" was noticed, i.e. differences between the Warm-up stage and other stages.

The utterances produced in the Evaluation stage were characterized by a narrower *f0* range (relative pitch range calculated for each phrase as $(f_{0max} - f_{0min})/f_{0min}$) with Praat [33] autocorrelation-based algorithm) as well as shortening of MiPs. Pitch and duration values also tended to be more compressed around the maximum of the distribution. The differences in means between the four stages were statistically significant with regard to the *f0* relative range in MiPs (one-way ANOVA test results, $F(3,1794) = 8984$, $p < 0.001$), and to MiPs duration (one-way ANOVA, $F(3,1773) = 4.199$ $p < 0.006$). A simplistic explanation would be the growing fatigue of the speakers. Still, the recording procedure involved two breaks that were meant also to let the voice recover. Lower dispersion of the values in both cases for the last stage of the task can be understood as increasing stability of the way of speaking in terms of pitch frequency and phrase length.

Speakers tended to speak faster in the Evaluation stage than in any of the earlier session stages. The difference was statistically significant ($p < 0.0001$, $F(3, 945) = 12.19$). Considering this observation as well as the lower mean duration regression intercepts in interpausal time groups and slightly higher regression slopes [34, 35], we may conclude that more dynamics in timing patterns (not only faster speech on average but also more acceleration and deceleration) can be found in the utterances produced in the Evaluation stage.

The parameters for a spectral tilt model (a * frequency + b, where a stands for decay rate and b for shift) were calculated using pitch-corrected LTAS and the robust method of spectral tilt parameter extraction of the Report spectral tilt function in Praat. One-way ANOVA tests performed independently for the values of a and b showed that the hypothesis of the equality of their means among the stages should be rejected ($F(3,5981) = 10.542$, $p < 0.001$ for a and $F(3,5981) = 20.219$, $p < 0.001$ for b) The values of a and b changed in a similar way across the stages, with a significant drop in the Evaluation stage of the session.

3.2. Perception test results

Mean values of formality ratings normalised by perception test participants are shown in Figure 3. The overall mean of formality ratings was the lowest for Main Task 2 and highest for the Warm-up part. According to a factorial ANOVA, statistical significance was confirmed for the factors: "speaker's ID" ($p < 0.0001$, $F(9, 1185) = 32.48$), the "recording session stage" factor ($p < 0.022$, $F(3,1185) = 3.19$), and the interactive influence of the "speaker's ID x recording session stage" factor ($p < 0.0001$, $F(27, 1185) = 4.84$). According to a post-hoc Tukey's HSD multiple comparison test results, the significance of the „recording session stage" factor was due to the differences in the mean values of ratings between the Warm-up stage and Main Task 2 stages.

The above observations tentatively contribute to the claim that at the beginning of the recording session the speakers sounded slightly more formal than in the two following stages, especially in the stage Main Task 2. The higher perceived formality at the beginning of the recording session might have been expected due to the possible impact of the novelty of the location and situation context.

In the case of the activation dimension, a factorial ANOVA test confirmed statistical significance for the factor "speaker's ID" ($p < 0.0001$, $F(9, 1185) = 57$) and the interaction effect of the factors "speaker's ID x recording session stage" ($p < 0.0001$, $F(27,1185) = 6.08$). In the case of the valence dimension, the significance results were confirmed for the same factors as in the case of formality ratings (but with lower F value thus showing a smaller contribution to the explanation of valence rating variability): "speaker's ID" ($p < 0.0001$, $F(9, 1185) = 17.17$) and the interactive factors: "speaker's ID x recording session stage" ($p < 0.0001$, $F(27, 1185) = 4.81$). The valence-activation ratings depended most importantly on speaker-related features of the utterances while the influence of the recording session stage factor was only found as an interaction effect when combined with speaker information. The bag plots in Figure 4 illustrate the results of perception tests for three selected speakers (normalized by perception test participant).

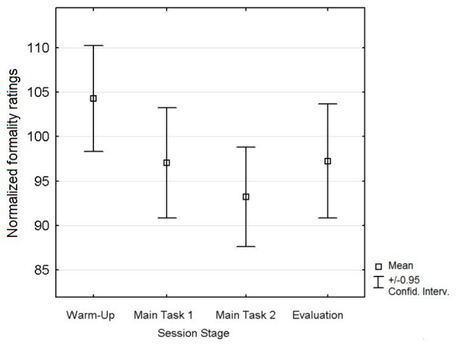

Figure 3: *Mean values of formality ratings normalised by perception test participants.*

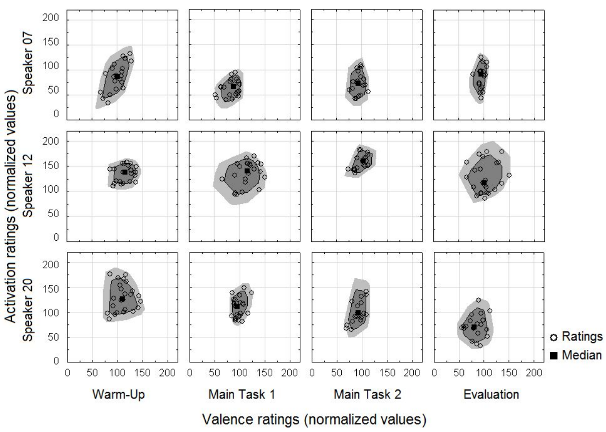

Figure 4: *Activation and valence ratings (normalized values) for three speakers in the four recording session stages*.

The differences between the recording session stages are distinguishable for each of the cases, however, the nature of these differences varies depending on particular speakers. In case of Speaker 20, the median was the highest for Warm-Up (in terms of valence and especially – activation) than in all the subsequent session stages (the lowest values were obtained for Evaluation). For Speaker 12, the median was the smallest both for valence and activation for Evaluation stage (similarly to Speaker 20) but in case of activation, the highest value was noted for Main Task 2. In case of Speaker 7, the differences in the median values are smaller, however when comparing the ratings dispersion and shapes of the bag, it can be observed that the results obtained for Warm-Up are more dispersed than in the three further session stages, especially with regard to valence judgements. For Speaker 20, the dispersion of the ratings was the largest for Warm-Up and the smallest for Main Task 1. For Speaker 12, the dispersion of ratings was visibly larger for Main Task 1 and Evaluation than for Warm-Up and Main Task 2.

4. Summary and conclusions

Participants of the perception test tended to evaluate utterances coming from the initial stage as more formal than those coming from the subsequent stages. More formal style occurring at the beginning of the session can be explained by an "entrance effect": the subjects only starting to adapt to the studio environment. This attitude appeared to be lost during the Main Task 1 and (most considerably) in Main Task 2. Afterwards, it tended to rise back again in the Evaluation stage, after the subjects had been consulted by the moderator again and given the (fake) negative evaluation of their efforts. The results of perception-based evaluation of utterances with the use of the valence-activation two-dimensional space showed slightly higher clustering of scores in the positive-active quadrant in the Warm-up stage as compared to the Evaluation stage. More important differences in ratings were related to the individual features of the speakers' voices; the effect of the recording session stage was statistically significant only when combined with the speaker's ID factor.

Speech produced in the Evaluation stage turned out to be significantly different in terms of many of the parameters under study. Apart from the factors such as growing fatigue or getting used to the environment, the change in the affective state (caused by the negative score) may be hypothesized as responsible for shifts in the analysed speech parameters and perception judgements. The effect of the affective factor seems to be more evident than those of microphone setup changes which is confirmed both by perception test results and instrumental analyses. An issue that becomes apparent here is individuality and idiosyncrasies of the speakers' voice productions and their assessment. In the case of dialogue, it is even more difficult as changes may occur depending on the quality of relationship between the subjects [36]. The reported subtlety of changes in voice quality and prosody might be just one explanation of listeners' difficulties to abstract from inter-speaker differences.

Obtaining results of higher significance and reliability would require much larger groups of subjects as well as more control over the environmental recording conditions. Although influences of many environmental factors were found to be weak or barely detectable, some turned out to be statistically significant. Further tests based on other sets of short and longer utterances as well as exploration of the links between acoustic features of speech and perceptually relevant cues to speaking styles are currently being carried out.

5. Acknowledgement

This study was conducted within the project: *NeuroPerKog: Development of the phonematic hearing and working memory in infants and children* funded by National Science Centre, Poland in the years 2013-2018.

6. References

[1] H. Giles & N. Coupland, *Language: Contexts and Consequences*. Keynes: Open University Press, 1991.

[2] R.M. Krauss & C.Y. Chiu, "Language and Social Behaviour," in D. Gilbert, S. Fiske & G. Lindzey (Eds.), *Handbook of social psychology, Vol. 2*, (4th ed.), Boston: McGraw-Hill, pp. 41-88, 1998.

[3] A. Agha, "Stereotypes and registers of honorific language," *Language in Society*, 27 (2), pp. 151-194, 1998.

[4] S. Strauss & J. Eun, "Indexicality and honorific speech level choice in Korean," *Linguistics*, 43, pp. 251-651, 2005.

[5] P. Brown & S. Levinson, *Politeness: Some Universals in Language Usage*. Cambridge: CUP, 1987.

[6] E. Ofuka, J. McKeown, M. Waterman, P. Roach, "Prosodic cue for rated politeness in Japanese speech," *Speech Communication, 32*, pp. 199–217, 2000.

[7] M. Ito, *The contribution of voice quality to the expression of politeness: an experimental study*. PhD Thesis, University of Edinburgh, 2005.

[8] S. Grawunder & B. Winter, "Acoustic correlates of politeness: Prosodic and voice quality measures in polite and informal speech of Korean and German speakers," in M. Hasegawa-Johnson, A. Bradlow, J. Cole, K. Liviescu, J. Pierrehumbert, & Ch. Shin (Eds.), *Proceedings of Speech Prosody Conference 2010*, Chicago, 2010.

[9] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, 37, 25, pp. 101-119, 1911.

[10] W.S. Brown, J.F. Brandt, F. John, "The effect of masking on vocal intensity during vocal and whispered speech," *Journal of Auditory Research,* 12, pp. 157-161, 1972.

[11] C. Rivers & M.P. Rastatter, "The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults," *The Journal of Auditory Research*, 25, pp. 37–45, 1985.

[12] W.V. Summers, D.B. Pisoni, R.H. Bernacki, R.I. Pedlow & M.A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analysis," *The Journal of the Acoustical Society of America, 84*, pp. 486–490, 1988.

[13] S. Kim, "Durational characteristics of Korean Lombard speech," *Proceedings of Eurospeech 2005, the 9th European Conference on Speech Communication and Technology*, pp. 2901–2904, 2005.

[14] F. J. Lundin, The influence of room reverberation on speech-an acoustical study of speech in a room. *STL-QPSR*, 2, pp. 24- 59, 1982.

[15] A. Chemero, "An outline of a theory of affordances," *Ecological psychology*, 15(2), pp. 181-195, 2003.

[16] J. Llisterri, "Speaking styles in speech research," *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language*, Dublin, Ireland, 15-17 July, 1992.

[17] M. Soderstrom, "Beyond babytalk: Reevaluating the nature and content of speech input to preverbal infants," *Developmental Review*, 27(4), pp. 501-532, 2007.

[18] O. Räsänen, T. Altosaar & U.K. Laine, "Comparison of prosodic features in Swedish and Finnish IDS/ADS speech," *Proc. of Nordic Prosody X*, Helsinki, 2008.

[19] M. Schröder, R. Cowie, E. Douglas-Cowie, M. Westerdijk & S. Gielen, "Acoustic correlates of emotion dimensions in view of speech synthesis," in Dalsgaard, P., Lindberg, B., Benner, H., Tan, Z-H. (Eds.) *Eurospeech 2001*, Vol. 1, Aalborg, pp. 87–90, 2001.

[20] A. Geumann, "Vocal intensity: acoustic and articulatory correlates," In B. Maassen, W. Hulstijn, R. Kent, H, Peters, & P. van Lieshout (Eds.), *Proceedings of the 4th International Speech Motor Conference 2001*, Nijmegen: Uitgeverij Vantilt, pp. 70–73, 2001.

[21] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, 6(3-4), pp. 169-200, 1992.

[22] R. Cowie & R. R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Communication, 40(1-2)*, pp. 5–32, 2003.

[23] G.L. Clore, A. Ortony & M.A. Foss, "The psychological foundations of the affective lexicon," *Journal of Personality and Social Psychology*, 53, pp. 751-766, 1987.

[24] K. Klessa, M. Karpiński & A. Wagner, "Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features," In D. Hirst & B. Bigi (Eds.) *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*, Aix en Provence, pp. 51-54, 2013.

[25] A. Wagner, *A comprehensive model of intonation for application in speech synthesis*. An unpublished PhD thesis, available at the University Library AMU in Poznań, 2008.

[26] M. Karpiński, "Acoustic properties and functions of phrase-final rises in Polish task-oriented dialogues. *Studies in honour of Wiktor Jassem. Speech and Language Technology*, Poznań: PTFon, pp. 147-156, 2012.

[27] G. Demenko, M. Wypych, & E. Baranowska, "Implementation of Grapheme-to-Phoneme Rules and Extended SAMPA Alphabet in Polish Text-to-Speech Synthesis," *Speech and Language Technology, Vol. 7*. PTFon, Poznań, pp. 79-97, 2003.

[28] M. Szymański & S. Grocholewski, "Transcription-based automatic segmentation of speech," *Proceedings of 2nd Language and Technology Conference,* Poznań, pp. 11–15, 2005.

[29] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, M. Schröde, "FEELTRACE: An instrument for recording perceived emotion in real time," *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, Newcastle, 2000.

[30] P. Laukka, *Vocal expression of emotion: discrete-emotions and dimensional accounts.* PhD Thesis. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Science, Department of Psychology, Uppsala University, 2004.

[31] N. Campbell & P. Mokhtari, "Voice quality: the 4th prosodic dimension," *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS),* pp. 2417-2420, 2003.

[32] M. Karpiński, M. & K. Klessa (in preparation). Acoustic correlates of speaking styles under different recording conditions.

[33] P. Boersma & P. Weenink, "Praat – doing phonetics with a computer," [Computer programme] Version 5.3.51, 2015, retrieved from: http://www.fon.hum.uva.nl/praat/

[34] D. Gibbon, "TGA: a web tool for Time Group Analysis," in D. Hirst & B. Bigi (Eds.) *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop*, Aix en Provence, pp. 66-69, 2013.

[35] K. Klessa & D. Gibbon (2014). Annotation Pro + TGA: automation of speech timing analysis. *Proceedings of Language Resources and Evaluation Conference (LREC) 2014*, Reykjavik, Iceland.

[36] S.D. Farley, S.M. Hughes, J.N. LaFayette, "People Will Know We Are in Love: Evidence of Difference Between Vocal Samples Directed Toward Lovers and Friends," *Journal of Nonverbal Behaviour*, 37, pp. 128-138, 2013.