Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Who wants to be a blabbermouth? Prosodic cues to correct answers in the WWTBAM quiz show scenario

Oliver Niebuhr

Department of Design and Communication, Innovation Research Cluster Alsion (IRCA),
University of Southern Denmark, Sonderborg, Denmark

olni@sdu.dk

Abstract

Starting from previous research on the prosodic patterns of emotion, psychological stress and deceptive speech, the paper investigates whether quizmasters convey telltale cues to correct answers in the popular four-alternatives (a/b/c/d) framework of "Who Wants to Be a Millionaire?" (WWTBAM). We simulated this game-show scenario in the lab, based on 20 naive German participants who took the roles of either quizmaster or contestant. Quizmasters were instructed to take care not to reveal correct answers to contestants. Despite this explicit instruction, our acoustic-prosodic analysis yielded clear telltale signs of correct answers. These telltale signs were consistent across all quizmasters, but complex insofar as they differed across question positions (a/b/c/d) could not be found in the introductory letters. Cues to correct answers involved timing and range of F0 and intensity patterns, speaking rate, and degree of final lengthening; pause durations between answers and introductory letters were irrelevant. The results are discussed with respect to their implications for real quiz shows and the elicitation of emotions and stress in the lab.

Index Terms: speaker state, prosody, emotion, question, stress

1. Introduction

Speech and social interaction are two sides of the same coin. One consequence of this irrefutable fact is that the speech signal not only includes the speaker's encoded message. It simultaneously mirrors the speaker's state that manifests itself most clearly, though not exclusively, in the prosodic domain of the conveyed signal. For example, there is a whole bunch of studies on the prosodic correlates of emotions in speech. Our understanding of this matter is constantly getting better in terms of level of detail, context-sensitivity, individual differences, and ecological validity [1,2,3]. Besides the *emotional* reflexes in speech, there are also more subtle and less well investigated signs of speaker state, such as *stress* factors. The studies of Hansen and colleagues (e.g., [4,5]) distinguish between psychological stress (mental/cognitive workload), perceptual stress (noise), physiological stress (fatigue, illness), and physical stress (physical activity). They found that all four types of stress leave an imprint on the speech signal.

Besides the contributions of Hansen and colleagues, the first studies that come to mind with regard to psychological stress are probably those on deceptive speech by Hirschberg, Benus, and colleagues (e.g., [6,7,8]). Although these are all recent studies, this line of research actually stretches back over at least 40 years, see [9,10,11,12,13,14].

Both the more obvious emotion-related and the more subtle stress-related reflexes of speaker state affect the same prosodic parameters of the speech signal. For example, fear (particularly cold fear) is often found to cause high flat F0 patterns, lower and less variable intensity patterns, a softer voice quality, and a faster speaking rate [15,16,17,18]. Similarly, deception as well as other types of increased mental/cognitive workload, for example, those due to multi-tasking, are typically characterized by an increase in all F0 and intensity parameters, including ranges, a tenser or creakier voice quality, and a lower speaking rate [8,9,12,13,14,19,20].

Quizmasters in game shows like "Who Wants to Be a Millionaire?" (WWTBAM) must be no blabbermouths in the sense that they constantly have to take care not to send out any telltale signs to their contestant. Most importantly, even if they know which of the four alternative answers to the quiz question is correct, they have to be extremely careful to not reveal this knowledge to their contestant. So, in terms of speaker states, this WWTBAM situation is – on the part of the quizmaster – clearly related to emotions like fear (fear of failure under social pressure) and psychological stress (increased mental/cognitive load) due to deceiving the contestant.

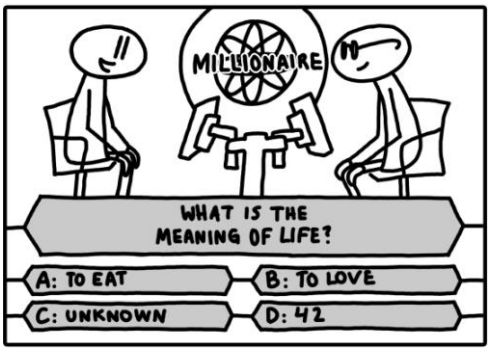

Figure 1: *The typical quizmaster-contestant and question-answer setups of WWTBAM, illustrated by Nathalie Schümchen.*

Given that we know *that* and *how* these emotion- and stress-related factors shape prosodic patterns, we wondered whether we can detect by means of multiparametric prosodic analyses which answer among the four alternative answers in the typical WWTBAM setup (Fig.1) is the correct one, even though the quizmaster tries hard to hide this information from the contestant. Admittedly, this research question has in part a practical or merely entertaining value outside the linguistic sciences. However, there is also a clear linguistic relevance to this question: First, if we find telltale cues to the correct answer, then we can further substantiate the empirical basis of the prosodic reflexes of speaker states. Moreover, as many of the prosodic characteristics of (cold) fear on the one hand and

 doi: 10.21437/SpeechProsody.2016-204

psychological stress/deception on the other deviate in opposite directions from a neutral speaking condition, our findings can shed some light on the origin of possible telltale signs, which, in turn, would provide the basis for targeted speaker/quizmaster training. In addition, some initial conclusions could be drawn on how emotion and stress factors interact, for example, in terms of coping strategies and/or prosodic correlates.

We addressed these issues by simulating the WWTBAM scenario that is illustrated in Figure 1 in a controlled laboratory production experiment, with students taking the roles of quizmaster and contestant.

2. Method

2.1. Participants

Twenty native speakers of Northern Standard German, 10 males and 10 females, participated in the production experiment. They were all undergraduate students at Kiel University and between 23 and 32 years old. All of them were naive with respect to the aim of the experiment, but equally well familiar with the WWTBAM show. The participants were grouped into pairs of speakers of the same sex, and care was taken to combine only speakers who knew each other only by sight. One speaker in a pair was to be the quizmaster, and the other speaker was to take the role of the contestant. Using same-sex speaker pairs was to minimize effects of prosodic convergence [21]. The fact that the speakers of a pair were both students but only vaguely familiar with each other was to create that kind of friendly but formal relationship that is characteristic of quizmaster and contestant in a game show. Speakers were rewarded for their participation with sweets that they received at the end of their recording session.

2.2. Procedure

Each recording session lasted about half an hour and took place in a sound-treated recording studio of the Dept. of General Linguistics at Kiel University. The individual sessions always followed the same scheme. First, the two speakers of a pair were asked to choose which role they would like to take, quizmaster or contestant. The choice was left to the speakers.

When the speaker roles were assigned, the contestant was asked to leave the room, and the quizmaster received 20 file cards on which the question-answer sequences were printed in the typical WWTBAM layout (Fig. 1). That is, the question was printed justified at the top of the file card. Underneath the question were four alternative answers, introduced by lower-case letters from <a> to <d>, followed by colons. the quizmaster was given some time to familiarize him-/herself with the question-answer sequences. Then, s/he was explicitly instructed to speak in a clear manner, like a real quizmaster, and only read what was written on the file card. That is, questions and their individual alternative answers were to be separated only by pauses. Linking words of filler phrases like "could it be...", "and", "or maybe", etc. were to be avoided (in this way, we prevented uncontrolled prosodic contexts from entering our analysis). However, the lower-case letters were to be produced. Most importantly, the quizmaster was urged not to convey any telltale signs of correct answer alternatives to the contestant. Afterwards, the contestant re-entered the room and sat down on a chair opposite the quiz-master. S/he also got 20 file cards that only differed in one respect from those of the quizmaster: On the quizmaster's cards, the correct answers were underlined in red.

Both quizmaster and contestant were equipped with headset microphones. Then, the experimenter left the room, and quizmaster and contestant were given some time to familiarize themselves with the recording environment. The recording started when they signaled to the experimenter that they felt comfortable and were ready to begin with the first question.

Finally, at the end of each recording session, a questionnaire was filled out that asked each participant to provide some personal information and judge the level of difficulty of each of the 10 target questions.

2.3. Materials

The question-answer sequences printed on the file cards were inspired by the online versions (German editions) of Trivial Pursuit and WWTBAM. We selected 20 question-answer sequences; 10 of them were our target sequences. There were selected to meet two requirements. First, 5 questions had to be easy, whereas the other 5 had to be so difficult that the quizmaster would consider it very unlikely that his/her contestant could know the right answer. Second, the four answer alternatives of each question were selected to be comparable in terms of the prosodic structure they provide. In particular, this meant that the number of syllables had to be constant across all four answer alternatives. The same applied to lexical-stress and nuclear pitch-accent positions (there were no other pre-nuclear accents). Moreover, syllables bearing the nuclear pitch accent were all sonorant and in non-final position.

The 10 question-answer sequences of the target condition were complemented by 10 filler sequences. They were to make the quiz show task more worthwhile for our participants, and, more importantly, distracted them from the analogous prosodic make-up of the answers in the target condition. Moreover, the first three question-answer sequences in each recording session were filler sequences. These initial dummies ensured that the acoustically analyzed target sequences were free from any training or familiarization artifacts.

Besides these constant initial dummy sequences, the target and filler sequences were given an overall randomized order. This order was different for each recording session. Correct answers were equally distributed across the four answer alternatives <a>-<d> both within and across the target and filler conditions (only <c> was correct twice).

2.4. Analysis

The quizmaster's realizations of the four alternative answers per question were acoustically analyzed with respect to a set of 7 prosodic parameters that are known from previous studies to be involved in psychological stress/deception and (cold) fear [6-14], perceptual prominence and accentuation [22,23], and prosodic phrasing [24]. That is, if the quizmaster did subconsciously send out subtle cues to correct answers, then we wanted to detect these telltale details, irrespective of whether they originated from stress and emotion or just occurred as changes in the degree to which the correct answer was singled out from its alternatives, for example, by means of a stronger/weaker accentuation or a stronger/weaker phrase boundary. The following prosodic parameters were measured:

- (1) F0 range in semitones,
- (2) Intensity range in dB,
- (3) Nuclear-accent alignment, i.e. time interval (in ms) between the accented-vowel onset and the F0 peak or valley of H* or L*,

- (4) Intensity-peak alignment, i.e. the time interval (in ms) between the accented-vowel onset and the intensity maximum in the accented vowel,
- (5) speaking rate in syllables per second (syll/s),
- (6) final lengthening, i.e. final-syllable duration relative to the duration of the prosodic phrase,
- (7) pause duration, i.e. the time interval (in ms) between the offset of the introductory (lower-case) letter and the onset of the corresponding answer.

Prosodic parameters (1)-(4) were measured for both the answers and their preceding lower-case letters. As 10 speakers produced 10 question-answer sequences of the target condition, there were 100x4, i.e. 400 answers and the same number of lower-case letters to be analyzed. All measurements were taken manually in PRAAT [25] using default settings. F0 errors were corrected by either changing the default settings or determining the corresponding periods in the waveform. Durations were measured by integrating information from waveform and broad-band spectrogram.

In order to statistically analyze the data, we used two analogous three-way MANOVAs for answers and lower-case letters, based on the fixed factors Question Type (easy, difficult), Answer Type (right, wrong) and Answer Position (<a,b, c,d>). In addition, we conducted separate linear discriminant analyses per Answers Position in order to see how well correct answers can be predicted from their own prosodic parameters or from those of the preceding lower-case letters.

3. Results

3.1. Realization of alternative answers

The MANOVA on the alternative answers yielded significant main effects of Question Type on nuclear-accent alignment ($F[1,385]=4.8$, $p<0.05$) and final lengthening ($F[1,385]=4.9$, $p<0.05$). Accents were aligned about 20 ms later and final lengthening was about 10 % more pronounced when the alternative answers followed a difficult question than when they followed an easy question. These differences between the difficult and easy question contexts proved to be fairly robust, as they were produced by our quizmaster sample independently of the other experimental variables. That is, Question Type interacted with neither Answer Type nor Answer Position.

The factor Answer Type was statistically a lot more productive and, moreover, closely intertwined with the factor Answer Position. There were significant main effects of Answer Type on the F0 and intensity ranges ($F[1,385]=12.9$, $p<0.001$; $F[1,385]=11.3$, $p<0.01$) as well as on speaking rate ($F[1,385]=4.0$, $p<0.05$) and the degree of final lengthening ($F[1,385]=10.6$, $p<0.001$). However, as is displayed in Figures 2(a)-(f), the F0 range was the only prosodic parameter whose difference between right and wrong answers was independent of Answer Position. There were significant interactions of Answer Type with Answer Position for intensity range ($F[3, 385]=14.1$, $p<0.001$), speaking rate ($F[3,385]=8.2$, $p<0.001$), and degree of final lengthening ($F[1,385]=12.7$, $p<0.001$). The interactions were disordinal in nature, i.e. the direction in which right answers differed from wrong answers depended on whether the right answer was in position <a>, <b>, <c>, or <d>. More specifically, Figures 2(a)-(f) illustrate that the prosodic profile of right answers in positions <c> and <d> resembled that of wrong answers in positions <a> or <a> and <b>. Thus, what characterizes right answers in positions <a> or <a> and <b> is diametrically opposed to what characterizes right answers in positions <c> and <d>.

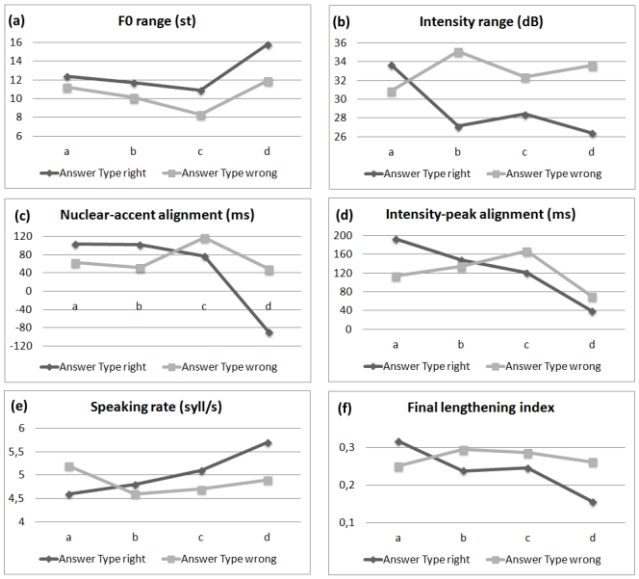

Figure 2: *Means of the prosodic parameters (1)-(6), measured for right (n=100) and wrong (n=300) alternative answers in positions <a>-<d>.*

It is due to the disordinal interactions that we found no separate main effects of Answer Type on the alignment characteristics of nuclear-accent and intensity peaks: Differences associated with Answer Type were leveled out across the four factor levels <a>-<d> of Answer Position. Instead, both nuclear-accent alignment and intensity-peak alignment yielded significant interactions of Answer Type and Answer Position ($F[3,385]=5.4$, $p<0.01$; $F[1,385]=7.2$, $p<0.01$). Post-hoc comparisons (t tests with Bonferroni correction for multiple testing) showed that the alignments of nuclear-accent and intensity peaks did differ highly significantly ($p<0.001$) between right and wrong answers in all answer positions, except for the intensity-peak alignment in position <b> (The three-way interaction was not significant).

Finally, in addition to the significant interactions of Answer Type and Answer Position, the MANOVA also resulted in significant main effects of Answer Position on all prosodic parameters ($2.9 \leq F[3,385] \leq 13.6$, $0.001 \leq p \leq 0.05$). Unlike for all other significant main effects and interactions, this also applied to the parameter pause duration. The main effects of Answer Position were overall similar: Duration characteristics like final lengthening, pause duration, and the alignments of nuclear accents and intensity peaks decreased across the alternative-answer positions, particularly after position <a> or at position <d>. The two range parameters showed a similar decrease, except that the answers' F0 ranges increased again at position <d>. Speaking rate is the only parameter whose mean values increased after <a>, see Figure 2(e).

3.2. Realization of lower-case letters

The MANOVA on the prosodic parameters in the lower-case letters yielded significant main effects of Answer Position on F0 range ($F[3,385]=6.3$, $p<0.001$), intensity range ($F[3,385]= 2.9$, $p<0.05$), nuclear-accent alignment ($F[3,385]=12.4$, $p<0.01$), and intensity-peak alignment ($F[3,385]=2.8$, $p<0.05$). The effects were qualitatively similar to those found for the

answers themselves, but more strongly pronounced. The MANOVA yielded no further significant main effects of Question Type and Answer Type, nor any significant interactions between the three fixed factors.

3.3. Linear discriminant analyses

For the alternative answers, we conducted 5 discriminant analyses in order to see how well right and wrong answers can be identified on the basis of the answers' prosodic profiles. One analysis was done across all answer positions <a>-<d>. The other four analyses dealt with each answer position separately. Results are summarized in Table 1. As can be seen from this Table, the canonical discriminant functions were all significant. That is, all 5 analyses were able to predict above chance level whether a produced alternative answer was right or wrong. However, in line with Figures 2(a)-(f) and the Answer Type * Answer Position interactions in the MANOVA, this prediction was far better when it was conducted for individual answer positions rather than across all answer positions: The discriminant analysis across all answer positions classified 62 % of all answers correctly as being right or wrong. In contrast, the best performance was achieved for the separate analysis of answer position <a>. It yielded 86.7 % correct right and 91.4 % correct wrong answers. Performance was worst for answers in position <b>. Here, only 60 % of right and 70 % of wrong answers were correctly identified by the discriminant analysis. This is, on the whole, only slightly better than in the analysis across all answer positions.

Table 1: Results summary of the 5 discriminant analyses on Answer Type, based on the prosodic parameters measured in the alternative answers (right, n=100; wrong, n=300).

	Eigenvalue	Chi-Squared[df]	p	Correct classification of right answers	Correct classification of wrong answers
All Answer Positions	0.64	$\chi^2[7] = 24.4$	<0.001	62.0 %	62.0 %
Position <a>	0.78	$\chi^2[7] = 53.7$	<0.001	86.7 %	91.4 %
Position <b>	0.19	$\chi^2[7] = 15.0$	<0.05	60.0 %	70.0 %
Position <c>	0.28	$\chi^2[7] = 23.3$	<0.001	70.0 %	76.3 %
Position <d>	0.72	$\chi^2[7] = 49.5$	<0.001	84.9 %	89.2 %

For the lower-case letters, none of the 5 discriminant analyses had a significant classification/prediction performance. That is, the analyses were not able to reliably distinguish between right and wrong answers. This outcome was expected, given the MANOVA's non-significant main effect of Answer Type on the prosodic parameters in the lower-case letters.

4. Conclusions

The results of the present study converge in the following main message: Yes, it is possible to predict from prosodic parameters which of the four alternatives in the WWTBAM scenario is the correct answer to the quiz question. Our quizmasters did send out subtle cues to correct answers, although they were explicitly instructed not to do so and tried hard to comply with this instruction, i.e. all quizmasters ticked a box in the de-briefing questionnaire stating that they were sure to not have revealed any correct answers to their contestants.

Cues to correct answers cover the whole range of prosodic parameters and relate to various perceptual domains like melody, prominence, and phrasing. Only the pause durations between the introductory lower-case letters (a/b/c/d) and their following answer alternatives were not a reliable cue to correct answers. In fact, the lower-case letters were in general not realized with any telltale cues. All telltale cues were contained in the answer alternatives themselves, and, in combination, they allowed statistical prediction performances of 70-87 %. Note that we have no indications that prediction performances or qualities of cues differed between genders.

Yet, it can be assumed that human listeners (contestants) would have a hard time identifying correct answers from the quizmasters' speech, as the links between correct answers and their prosodic cues are anything but straightforward. They strongly differ as a function of answer position. Accordingly, initial evidence from pilot perception experiments with and without de-lexicalized stimuli suggests that naive listeners are not able to identify correct answers in our pool of tokens. However, if listeners are trained based on our findings, then their performances increase significantly above chance level.

There also seems to be no simple explanation as to which type of speaker state created the involuntarily emitted prosodic cues to correct answers. The prosodic characteristics of psychological stress and fear that were summarized in the introduction point to a transition from psychological stress to fear (of failure) across answer positions <a>-<d>. It would indeed make sense that prosodic reflexes of psychological stress (mental workload/deception planning) are stronger at the beginning of the alternative-answer quartet. It also reasonable to assume that the fear to not make it through the task of avoiding any telltale cues to correct answers successively increases across the alternative-answer quartet.

We will deal with this question of the origin of telltale cues future studies that will also include analyses of biosignals like heart rate, blood pressure, and skin-conductance response. In any case, the range of involved prosodic parameters and the fact that differences between correct and incorrect answers were all located in the answers themselves and stronger for difficult than easy questions leave no doubt that, in general, the quizmasters' speech was shaped by stress- and/or emotion-related factors. This could mean that the quantity or quality of telltale cues to correct answers could vary with speaker age or social status. More self-confident speakers could convey fewer or less clear cues. Note in this context that the choice about who wanted to be quizmaster and contestant was deliberately left to the participants on the assumption that the more self-confident participant would take the role of the quizmaster. So, our results should already be conservative concerning the quality and quantity of effects.

The fact that we have effects that are obviously related to psychological stress and fear (of failure) stresses that quiz show scenarios are suitable for eliciting and measuring prosodic exponents of these speaker states, even in the laboratory.

Finally, in terms of specific practical advice, our study shows that contestants should focus on answers rather than their preceding introductory letters and train their ears before participating in quiz shows with alternative-answers scenarios, such as WWTBAM. Quizmasters, in turn, should hide correct answers preferably in middle positions like <b>, where prosodic cues to correct answers are weakest. Of course, this advice relies on the assumptions that (1) real and experienced quizmasters still send out the same telltale cues as our untrained student quizmasters, and (2) quizmasters in real quiz shows always know the correct answers to questions before they read the question-answer sequences to the contestant. Follow-up studies on these generalization issues are planned.

5. Acknowledgements

The author is greatly indebted to Sophie Beilfuß who developed, conducted, and analyzed the quiz show recordings.

6. References

[1] R. Jürgens, K. Hammerschmidt, and J. Fischer, "Authentic and play-acted vocal emotion expressions reveal acoustic differences," *Frontiers in Psychology 2*, 180, 2011.

[2] D.A. Sauter, F. Eisner, P. Ekman, and S.K. Scott, "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations," *Proc. Natl. Acad. Sci. U.S.A. 107*, pp. 2408–2412, 2010.

[3] K. Scherer, T. Johnstone, and G. Klasmeyer, "Vocal expression of emotion." In R. J. Davidson, K. R. Scherer, and H. Goldsmith (eds), *Handbook of the Affective Sciences*. New York/Oxford: Oxford University Press, 2010.

[4] H. Steeneken and J.H.L. Hansen, "Speech under stress conditions: overview of the effect on speech production and on system performance," *Proc. IEEE ICASSP'99, Phoenix, USA,* pp. 2079–2082, 1999.

[5] J.H.L. Hansen and S. Bou-Ghazale, "Getting started with SUSAS: A speech under simulated and actual stress database," *Proc. EUROSPEECH'97, Rhodes, Greece*, pp. 1743–1746, 1997.

[6] J. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis, B. Pellom, E. Shriberg, and A. Stolcke," Distinguishing Deceptive from Non-Deceptive Speech," *Proc. Eurospeech, Lisbon, Portugal*, pp. 1833-1836, 2005.

[7] S. Benus, F. Enos, J. Hirschberg, and E. Shriberg, "Pauses in Deceptive Speech," *Proc. 3rd International Conference of Speech Prosody, Dresden, Germany*, 2006.

[8] J. Hirschberg, "Deceptive Speech: Clues from Spoken Language." In F. Chen and K. Jokinen (eds), *Speech technology*. New York/London: Springer, pp. 79–88, 2010.

[9] L.A. Streeter, R.M. Krauss, V. Geller, C. Olson, and W. Apple, "Pitch changes during attempted deception," *Journal of Personality and Social Psychology 35*, pp. 345-350, 1977.

[10] M. Zuckerman, B.M. DePaulo, and R. Rosenthal, "Verbal and nonverbal communication of deception." *Advances in Experimental Social Psychology 14*, pp. 1- 59, 1981.

[11] B.M. DePaulo, R. Rosenthal, J. Rosenkrantz, and C.R. Green, "Actual and perceived cues to deception: A closer look at speech," *Basic and Applied Social Psychology 3*, pp. 291-312, 1982.

[12] P. Ekman, W.V. Friesen, and K. Scherer, "Body movement and voice pitch in deceptive interaction," *Semiotica 16*, pp. 23–77, 1976.

[13] M. Graciarena, E. Shriberg, A. Stolcke, J.H.F. Enos, and S. Kajarekar, "Combining prosodic, lexical and cepstral systems for deceptive speech detection," *Proc. IEEE ICASSP. Toulouse, France*, 2006.

[14] A. Mehrabian, "Nonverbal betrayal of feeling," *Journal of Experimental Research and Personality 5*, pp. 64–73, 1971.

[15] A. Paeschke, M. Kienast, and W.F. Sendlmeier, "F0-contours in emotional speech, " *Proc. 14th International Congress of Phonetic Sciences, San Francisco, USA*, pp. 929-933, 1999.

[16] H.R. Pfitzinger, H.R. and C. Kaernbach, "Amplitude and amplitude variation of emotional speech," *Proc. International Conference Interspeech 2008, Brisbane, Australia*, pp. 1036-1039, 2008.

[17] R. Banse and K. Scherer," Acoustic profiles in vocal emotion expression," *Journal of Personality and Social Psychology 70*, 614-636, 1996.

[18] T. Johnstone and K. Scherer, "The effects of emotions on voice quality," *Proc. 14th International Conference of Phonetic Sciences, San Francisco, USA*, pp. 2029–2032, 1999.

[19] S.E. Lively, D.B. Pisoni, W. Van Summers, and R.H. Bernacki, "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," *Journal of the Acoustical Society of America 923*, pp. 2962-2973, 1993.

[20] K. Huttunen, H. Keränen, E. Väyrynen, R. Pääkkönen, and T. Leino, "Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights," *Applied Ergonomics 42*, pp. 348-357, 2011.

[21] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," *Proc. 11th Interspeech Conference, Florence, Italy*, 2011.

[22] K. Kohler, "The perception of prominence patterns," *Phonetica 65*, pp. 257-269, 2008.

[23] P. Wagner, A. Origlia, C. Avesani, G. Christodoulides, F. Cutugno, M. D'Imperio, D. Escudero Mancebo, B. Gili Fivela, A. Lacheret, B. Ludusan, H. Moniz, A. Ní Chasaide, O. Niebuhr, L. Rousier-Vercruyssen, A.C. Simon, J. Simko, F. Tesser, and M. Vainio, "Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence," *Proc. 18th International Congress of Phonetic Sciences, Glasgow, UK*, pp. 1-5, 2015.

[24] C. Féry, R. Hörnig, and S. Pahaut, "Correlates of Phrasing in French and German from an Experiment with Semi-Spontaneous Speech." In C. Gabriel, C. Lleó (eds), *Intonational Phrasing in Romance and Germanic*, Amsterdam: John Benjamins, pp. 11–41, 2011.

[25] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International 5*, pp. 341-345, 2001.

[26] S.I. Levitan, A. Guozhen, M. Wang, G. Mendels, J. Hirschberg, M. Levine, and A. Rosenberg, "Cross-Cultural Production and Detection of Deception from Speech", *Proc. ACM Workshop on Multimodal Deception Detection, Seatle, USA*, pp. 1-8, 2015.