PHONATORY CHANGES DURING EMOTION-INDUCING GAME EVENTS: THE EFFECT OF DISCREPANCY FROM EXPECTATIONS AND GOAL CONDUCIVENESS

Márton Bartók

MTA–ELTE Lendület Lingual Articulation Research Group; ELTE Eötvös Loránd University, Budapest, Hungary, Doctoral School of Linguistics
bartokmarton@gmail.com

ABSTRACT

This paper[1] aims to describe how phonation changes during emotion-inducing stimuli. 34 Hungarian speakers (17 female, 17 male) were asked to compete in a simple computerised guessing game, using voice commands to proceed after faced with the result of their guess. We expected that acoustic measures taken on these voice commands differ based on two affective components: goal conduciveness (successful or unsuccessful result) and discrepancy from expectations (unexpected or expected result based on the uncertainty of the guess).

According to the results, only female subjects show phonatory variation as a result of varying emotional states: their fundamental frequency is higher at unexpected game events and their phonation is less breathy (lower H1-H2) at unexpected, unsuccessful events. Both of these changes can be caused by higher muscle tension in unexpected, unsuccessful situations and lower tension when experiencing goal conducive, expected events.

Keywords: emotion, phonation, appraisal, emotion regulation

1. INTRODUCTION

According to appraisal models of emotion (e.g. [16, 19, 25]), behavioral and physiological reactions to affective stimuli are a result of cognitive appraisal of the stimuli. Appraisal is described by the Component Process Modell (CPM, [25]) as a process consisting of several subsequent Stimulus Evaluation Checks (SECs). The result of these steps of evaluation determines the emotional state and physiological reactions of the organism. In this study, two such SECs, goal conduciveness and discrepancy from expectations were manipulated in a computer game, with the aim to describe their effect on vocal fold vibration. Speech was acquired during voice commands controlling the game, resulting in utterances that could capture the induced emotional effects right at the time they occur [1]. Hypotheses concerning these effects were formed based on the physiological changes predicted by the CPM for different results of these SECs and their possible effect on phonation, while also considering phonatory patterns observed in acted emotions, since such portrayals often build on representations of spontaneous emotional reactions. Phonatory variation was quantified in two ways: we determined phonation type manually, after which acoustic measurements were carried out on the modal parts of the analysed vowels. The two acoustic measures taken were fundamental frequency (f0) and the difference between the first two harmonics (H1-H2). H1-H2 is a measure well-suited to describe the degree of glottal constriction [12] as it correlates highly with the Open Quotient (OQ) [28], the proportion of a glottal cycle in which the glottis is open [9]. Higher values of H1-H2 would suggest a more breathy phonation, while low values indicate irregular phonation.

1.1. The effect of discrepancy from expectations

Events can be appraised as consistent to or discrepant from the organisms expectations, in other words, events can be expected or unexpected. Novel, unexpected stimuli can lead to a so called orienting response (OR) [29], which is characterized by increased muscle tension [17]. When the subject experiences stimuli that is discrepant from their expectations, the occurrence of such ORs could lead to changes in speech such as increased longitudinal tension of the vocal folds, resulting in higher f0 (as predicted in [11]). This would be in line with high f0 values in the portrayal of surprise (see [2] for Hungarian, [23] for an overview). Since both of the nonmodal phonation types we analysed can be characterized by low tension in some of the laryngeal muscle groups (for breathy phonation, weak longitudinal tension, for irregular phonation, weak medial compression), increased muscle tension when experiencing discrepant events would mean that these nonmodal phonation types are unlikely to occur, and

less nonmodal-like acoustic characteristics would be expected to be measured.

1.2. The effect of goal conduciveness

Organisms also evaluate how events could affect their chances to attain their current goals or needs. If an expected event is conducive to the goals of the organism, this could lead to a state of comfort and rest behavior, characterized by balanced muscle tone and relaxed respiration [24, 11]. Similar effects seem to appear when speakers are asked to portray contentment, an emotional state highly resembling these results of stimulus evaluation. This emotion is expressed with low f0 and there is a frequent occurrence of nonmodal phonation types: for females, phonation shifts towards irregular, while for males, breathy phonation becomes more frequent [2]. These effects could be both caused by relaxation of different muscle groups. For females, relaxation could result in weak longitudinal tension of the vocal folds, while for males, there could be a decrease in medial compression.

Goal obstructive game events can be further categorized based on how likely the organism is to cope with the obstructive events' consequences. In our case, this was the likelihood of the participant still getting a good final score in the game, which was possible, but harder to reach after every unsuccessful result. Therefore, we treated goal obstructive events as having moderate coping potential. Such obstructive events could lead to the activation of the sympathetic branch of the nervous system, to increased muscle tension [30], and thus phonation could be characterized by higher f0 and intensity (as predicted by [11] and observed in [10]). As in the case of unexpected stimuli, nonmodal phonation types would not be likely to occur, and less nonmodal-like acoustic characteristics would be expected to be measured.

1.3. Hypotheses

Hypotheses concerning phonation type frequency were formed with respect to frequency in the first two vowels of the voice commands, whereas hypotheses concerning acoustic measurements (f0, H1-H2) were formed with respect to the modal parts of those vowels.

- Goal conducive game events that are congruent with the subjects expectations lead to a more frequent occurrence of nonmodal phonation types, lower f0, lower H1-H2 for females and higher H1-H2 for males relative to the subjects' emotionally neutral speech.
- Goal obstructive game events and events discrepant from the subjects' expectations both lead to a less frequent occurrence of nonmodal phonation types, higher f0, and higher H1-H2 relative to the subjects' emotionally neutral speech.

Apart from assessing the effect of goal conduciveness, discrepancy from expectations and gender, the effect of syllable number was also analysed.

2. METHODS

A simple card game was developed in JavaScript, following the general experimental design described in [27], while we also made several methodological changes to the original game. In the game, participants had to guess whether an upcoming card would have a higher or lower value than the previous card. On the monitor in front of them, participants saw a row of six cards. At the beginning of each row, every card except the first was upside down, and the participants were asked to guess whether the next card's value will be higher. Guessing was executed using the voice commands *alacsonyabb* 'lower' and *magasabb* 'higher' (these voice commands will be analysed in a separate study). After guessing, the relevant card was turned, and if the participant guessed right, (s)he could go on to the next card in the row.

Knowing that the numbers ranged from 2 to 10, and cards that have already appeared in the row can not appear again, participants could easily estimate the probability of the upcoming card's value being higher or lower than the previous one. For example, if the first card's value was 8, the rational choice for the participant would be "lower", because the number of cards containing lower values (2, 3, 4, 5, 6, 7) than the previous card (8) is higher than the number of cards with values lower than that (9, 10).

To establish if the result of the guess was discrepant from the participant's expectations, we had to evaluate the potential of the participant to cope with a guess (coping potential, cp). For this purpose cp was calculated using the equation below. *cp* takes its maximum, 1, when the probability of the upcoming card's value being higher equals either 1 (the card's value will certainly be higher) or 0 (the value will certainly be lower). Therefore $cp = 1$ means that the subject can be absolutely certain while making that decision. If there is an equal chance of the value being higher or lower, meaning the participant faces a totally random decision, *cp* equals 0.

$$cp = 2 * |p_{higher} - 0.5|$$

A successful result after a decision with low *cp* or an unsuccessful result after a decision with high *cp* was considered to be discrepant from expectations

Table 1: Determining the discrepancy from expectations variable

	conducive	obstructive
low cp (cp<0.5)	discrepant	congruent
high cp (cp≥0.5)	congruent	discrepant

and vice versa (see Table 1). If the guess was correct, the participant then needed to make the same decision about the value of the next card and so on, until (s)he got to the end of the row of cards (this is referred to as one sequence). If the guess at any point was incorrect, the participant had to to move on to the next sequence. Before the new sequence started, the participant was required to utter *haladjunk* /hɒlɒɟːuŋk/ 'go on'. After the participants completed 5 sequences for familiarization, 32 sequences per participant were displayed in a random order, with a break after the 16th sequence. After the game was over, 5 repetitions of the voice command were read to acquire realizations in an emotionally neutral baseline condition.

Recordings were made in a sound-treated room using a Laryngograph Ltd. EGG-D200 electroglottograph, capturing acoustic and vocal fold contact area signals simultaneously. The recordings were automatically segmented using the BAS web services G2P [21] and MAUS [26], after which the segmentation was manually corrected. Both the segmentation and labeling of phonation types and acoustic measurements were carried out on the first two vowels of the target word /hɒlɒɟːuŋk/. Phonation type was determined based on audiovisual inspection of oscillographic, spectrographic and EGG signals, according to the acoustic descriptions and laryngograms in [15]. Mean f0 and H1-H2 were measured on the acoustic signal, on the modal part of the vowels in Praat [3]. For the estimation of f0, we used gender-specific f0-ranges suggested by [6] (100-300 Hz for females, 70-250 Hz for males). H1-H2 was calculated as the difference between peak amplitudes measured within a ± 10% frequency interval around the estimated harmonic frequency as in [5].

Altogether 36 Hungarian speakers were recorded. Two participants were excluded from the analysis due to possibly pathological phonation (nonmodal phonation in more than 50% of the duration of the analysed syllables). Therefore, speech of 34 subjects (17 female, 17 male, aged 18-35 years, mean age = 23 years) was analysed.

Statistical analyses were carried out in R [20]. The effect of discrepancy and conduciveness on the frequency of modal vs. nonmodal phonation types was assessed using Chi-Square tests. Separate linear mixed effect models were fitted to independently assess effects of discrepancy, conduciveness, gender, syllable number and their interactions on f0 and H1-H2, with random intercepts for subject. p-values were obtained with the lmerTest package [14], which uses the Satterthwaite's method of approximation for degrees of freedom. Since three different dependent variables (phonation type, f0, H1-H2) were used to test the same effects, we modified the alpha level to account for three tests using Bonferroni correction to $\alpha = 0.017$, as suggested by [22].

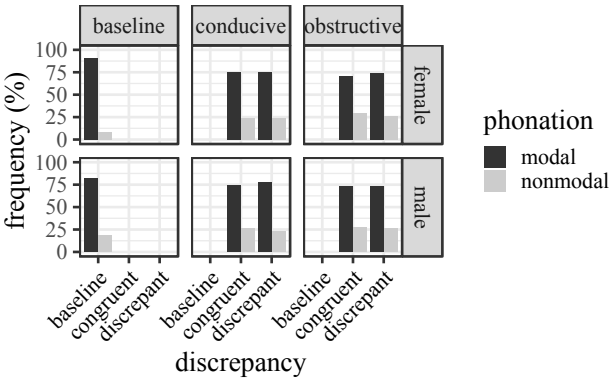
3. RESULTS

Altogether 2610 realizations of the vowel /ɒ/ were analysed in 1305 voice commands.

3.1. Frequency of phonation types

We found a significant effect of conduciveness (χ^2 = 55.88, $p < 0.001$) and discrepancy (χ^2 = 51.62, p < 0.001) on the frequency of phonation types. However, as it can be seen in Fig. 1, this effect can be mainly attributed to differences between the baseline condition and the experimental conditions (nonmodal phonation being generally more frequent during the experiment), while apparently there is no difference between the two experimental conditions.

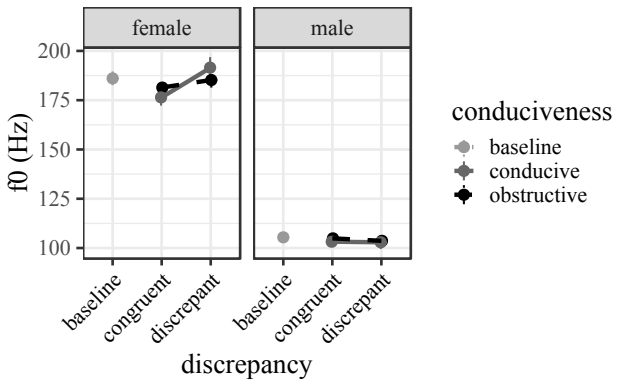
Figure 1: Frequency of modal vs. nonmodal phonation types as a function of discrepancy from expectations, goal conduciveness and gender.



3.2. Fundamental frequency

We found a singificant interaction effect of conduciveness, discrepancy and gender on f0 [F(1, 2082) = 5.88, $p < 0.017$]: as it can be seen in Fig. 2, while males showed no systematic variation as a function of conduciveness or discrepancy, females had higher f0 when experiencing discrepant, conducive stimuli than in the baseline condition, and events that were consistent with their expectations lead to f0-lowering. This effect was stronger in goal conducive situations than in goal obstructive ones.

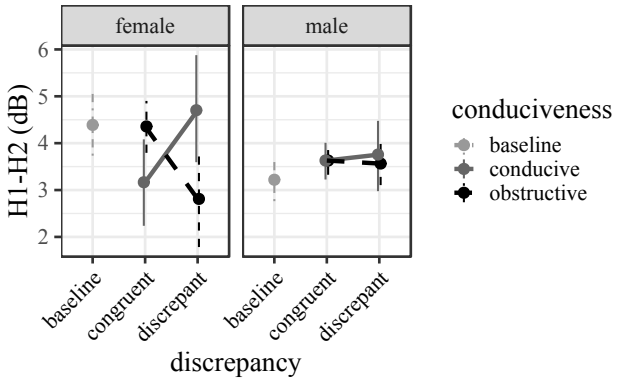
Figure 2: Mean f0 (Hz) with 95% confidence interval as a function of discrepancy from expectations, goal conduciveness and gender.



3.3. Difference of the first two harmonics

We found a singificant interaction effect of conduciveness, discrepancy and gender on H1-H2 [$F(1, 2085) = 9.743$, $p < 0.01$]: as it can be seen in Fig. 3, again, males showed little to no systematic variation, just a general raise in H1-H2 values in the manipulated (in-game) conditions relative to the baseline. For females, opposing effects of discrepancy were found for goal conducive and goal obstructive events: there was a lowering of H1-H2 values for discrepant, obstructive and congruent, conducive events, whilst in the other two conditions (discrepant, conducive and congruent, obstructive) values were similar to the baseline condition.

Figure 3: Mean H1-H2 (dB) with 95% confidence interval as a function of discrepancy from expectations, goal conduciveness and gender.



4. DISCUSSION

This paper aimed to describe how different results of stimulus evaluation could lead to emotional reactions observable in phonation. The effects of two Stimulus Evaluation Checks, discrepancy from expectations and goal conduciveness were explored using a computerized card game for emotion induction. We measured two acoustic parameters characterizing phonation and the frequency of modal vs. nonmodal phonation types on the voice commands that were used to control the game.

Although we found no difference in the frequency of phonation types between different results of the manipulated SECs, significant interaction effects of gender, discrepancy and conduciveness were found on the acoustic parameters measured on the modal parts of the voice commands. This could mean that emotions induced in this highly controlled, laboratory setting lead to subtle phonatory changes. The interaction effects of the two manipulated SECs and gender for both acoustic measures indicate that emotional reactions can only be captured in female speech. This could be explained by differences in the degree of emotional reactivity and emotion regulation between genders. Apart from the lay belief that females tend to be more emotional [7], several studies using physiological measures of emotional arousal and attention suggest that females are more reactive to emotional stimuli than males [4, 7, 13, 18].

We found that for females, f0 is higher when facing game events that are discrepant from expectations, while congruent events lead to a decrease in f0. This effect is likely to be caused by increased muscle tension when facing unexpected, discrepant stimuli and decreased tension in case of expected events [11]. The effect is stronger in case of relaxation at goal conducive events.

We also found a lowering of H1-H2 values for discrepant, obstructive and congruent, conducive events in the phonation of females. H1-H2 lowering indicates a shift from females' habitually breathy phonation [8] to a more modal one, as a result of the predicted increase in overall muscle tension [11]. In the case of conducive, congruent events, low H1-H2 together with the low f0 measured in this condition could mean that in this case, H1-H2 lowering does not simply indicate a more modal phonation, but rather a shift towards a more irregular phonation caused by relaxation, similarly to the frequent occurrence of irregular phonation when Hungarian females express contentment [2].

5. ACKNOWLEDGEMENTS

This study was supported by the Hungarian Ministry of Human Capacities, Hungarian Talent Program NTP-NFTÖ-17. I would like to thank my supervisors, Andrea Deme and Alexandra Markó for the help and support during this project.

6. REFERENCES

[1] Aharonson, V., Amir, N. 2006. Emotion Elicitation in a Computerized Gambling Game. *International Conference on Information Technology: Research and Education* Tel-Aviv, Israel. 179–183.

[2] Bartók, M. 2018. A gégeműködés variabilitása az érzelemkifejezés függvényében. *Beszédkutatás* 26(1), 30–62.

[3] Boersma, P., Weenink, D. 2018. PRAAT: Doing phonetics by computer.

[4] Bradley, M. M., Codispoti, M., Sabatinelli, D., Lang, P. J. 2001. Emotion and motivation II: sex differences in picture processing. *Emotion (Washington, D.C.)* 1(3), 300–319.

[5] DiCanio, C. 2007. https://www.acsu.buffalo.edu/~cdicanio/scripts/Get_spectral_tilt_2.praat.

[6] Eager, C. 2015. Automated voicing analysis in Praat: statistically equivalent to manual segmentation. *The Scottish Consortium for ICPhS* volume 18 551–585.

[7] Grossman, M., Wood, W. 1993. Sex Differences in Intensity of Emotional Experience: A Social Role Interpretation. *Journal of personality and social psychology* 65, 1010–1022.

[8] Hanson, H. M., Chuang, E. S. 1999. Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America* 106(2), 1064–1077.

[9] Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., Goldman, S. L. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research* 38(6), 1212–1223.

[10] Johnstone, T. 2001. *The effect of emotion on voice production and speech acoustics*. Thesis for the degree of doctor of philosophy. University of Western Australia.

[11] Johnstone, T., Van Reekum, C. M., Scherer, K. R. 2001. Vocal expression correlates of appraisal processes. In: *Appraisal processes in emotion: Theory, methods, research* Series in affective science. New York, NY, US: Oxford University Press 271–284.

[12] Keating, P., Esposito, C. 2007. Linguistic Voice Quality. *UCLA Working Papers in Phonetics* 105, 85–91.

[13] Kemp, A. H., Silberstein, R. B., Armstrong, S. M., Nathan, P. J. 2004. Gender differences in the cortical electrophysiological processing of visual emotional stimuli. *NeuroImage* 21(2), 632–646.

[14] Kuznetsova, A., Brockhoff, P., H. B. Christensen, R. 2015. LmerTest: Tests in linear mixed effects models. *R Package Version* 2.

[15] Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge University Press.

[16] Lazarus, R. S. 1968. Emotions and adaptation: Conceptual and empirical relations. *Nebraska Symposium on Motivation* 16, 175–266.

[17] Lynn, R. 1966. The orientation reaction. In: Lynn, R., (ed), *Attention, Arousal and the Orientation Reaction* International Series of Monographs in Experimental Psychology. Pergamon 1–13.

[18] Orozco, S., Ehlers, C. L. 1998. Gender differences in electrophysiological responses to facial stimuli. *Biological Psychiatry* 44(4), 281–289.

[19] Ortony, A., Clore, G. L., Collins, A. 1990. *The Cognitive Structure of Emotions*. Cambridge University Press.

[20] R, C. T. 2018. R: A language and environment for statistical computing.

[21] Reichel, U. 2012. PermA and Balloon: Tools for string alignment and text processing. *Proceedings of INTERSPEECH* paper no. 346.

[22] Roettger, T. B. 2019. Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1), 1.

[23] Scherer, K. R. 1986. Vocal affect expression: a review and a model for future research. *Psychological Bulletin* 99(2), 143–165.

[24] Scherer, K. R. 1987. Toward a dynamic theory of emotion: The component process model of affective states. In: *Geneva Studies in Emotion and Communication* volume 1. 1–98.

[25] Scherer, K. R. 2001. Appraisal considered as a process of multilevel sequential checking. In: *Appraisal processes in emotion: Theory, methods, research* Series in affective science. New York, NY, US: Oxford University Press 92–120.

[26] Schiel, F. 1999. Automatic phonetic transcription of nonprompted speech. *Proceedings of the International Congress of Phonetic Sciences* 607–610.

[27] Shahid, S., Krahmer, E., Swerts, M. 2007. Audiovisual emotional speech of game playing children: Effects of age and culture. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* volume 1 2681–2684.

[28] Shue, Y.-L., Chen, G., Alwan, A. 2010. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. *INTERSPEECH*.

[29] Sokolov, E. N., Cacioppo, J. T. 1997. Orienting and defense reflexes: Vector coding the cardiac response. In: *Attention and orienting: Sensory and motivational processes*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers 1–22.

[30] Svebak, S. 1983. The effect of information load, emotional load and motivational state upon tonic physiological activation. In: *Biological and psychological basis of psychosomatic disease* Advances in the biosciences, Vol. 42. Elmsford, NY, US: Pergamon Press 61–73.

[1] The research compendium for this paper is available at https://osf.io/v7u4x/?view_only=c7787320a75e4131856d1ad6f1ab79c5