Speak, Think, Act: A phonetic analysis of the combinatorial effects of respiratory mask, physical and cognitive stress on phonation and articulation

Carina Marquard[1], Christin Baasch[1], Michael Brodersen[1], Oliver Niebuhr[2], Gerhard Schmidt[1]
[1]Christian-Albrechts-Universität zu Kiel, Deutschland
[2]University of Southern Denmark

Abstract

Stress is perceptible in speech, but on which basis and in which way all the different stress types shine through and combine in the acoustic speech signal is only vaguely understood by now. This study addresses these questions and additionally examines how stress types in speech can be determined and quantified by means of acoustic-phonetic parameters. We used several stressors that form a taxonomy of stress factors according to Murray et al. [1]. Pink noise served as an external stressor. A further stress factor, cognitive load, was created by asking quiz questions. Physiological stressors were induced by training on an ergometer and the application of a respirator mask (full face mask). The speech signals produced by male German native speakers while answering the quiz and further reference questions were analyzed in terms of a number of prosodic F0 and duration measures. The results of our acoustic analysis allow drawing conclusions on if and how stress factors can be distinguished from each other, interfere with each other, and/or add up in the speech signal. Furthermore, we touch upon the issue whether measurable stress can increase ad infinitum or whether there is an upper limit for the manifestation of stress in speech.

Index Terms: prosody, Lombard, pitch accents, breathing, stress, emotion, taxonomy of stressors

1. Introduction

1.1 Speak, Think, Act

In everyday situations, it is very important to recognize the emotional state of a person. This applies to people like rescue paramedics, pilots, divers, and firefighterswho are in extreme situations. These situations inevitably involve emotional and physical components that stimulate aspects of language which are considered to be paralinguistic [1]. Prosodic parameters such as pitch, temporal structure, loudness, and voice quality are used for communicative purposes. A speaker can use them to indicate which information the listener should pay particular attention to. Through prosodic parameters, the human voice encodes a wealth of information about emotion, mood, and the mental state of the speaker. In addition, they reliably reflect the speaker's stress level. Estimating the stress of a speaker is useful for various applications, including emergency call sorting and voice communication monitoring of aircrafts. Here and in similar cases, stress can be defined as a condition that creates or increases variation in the production of speech relative to a of neutral reference condition [2].

1.2 Background and aims

Most evidence in stress research comes from question-specific individual case studies, conducted in various disciplines including engineering, medicine and psychology [3], [4], [5]. Being one of a few exceptions, the works of Meinerz [6] and Landgraf [7] took a different approach and analyzed stress from linguistic or phonetic points of view. Evidence converges in showing that stress is perceptible in speech. But, on which basis, and in which way all the different types of stress – i.e. cognitive, physical, and environmental stress – combine and shine through in the acoustic speech signal is only vaguely understood as yet. Our present work targets this empirical gap with acoustic analyses focusing on the special communicative characte-ristics of wearers of respiratory masks. The following three questions are addressed:

1. Can different stress factors be distinguished from each other in the speech signal? If so, how?
2. Do the stress factors interfere with each other and / or do they add up in the speech signal?
3. Can measurable stress increase ad infinitum or is there an upper limit for the manifestation of stress in the speech signal (in terms of additive effects of stress types)?

2. Method

2.1 Speakers

Five native speakers of Standard German took part in the production experiment, 4 males and 1 female. They were between 21 and 39 years old and voluntary firefighters from Kiel, Germany. One subject was wearing lenses which he also uses in real operations. None of the participants reported any hearing or speaking impairments.

2.2 Tasks

Stress is subdivided into three categories: (1) cognitive stress, (2) physical stress, and (3) mental stress. Cognitive stress in the form of quiz questions to be answered is a permanent stressor in our experiment, while physiological stress is temporarily induced by an ergometer. Physical stress is - also temporarily - created by pink noise. The participants went through individual recording sessions consisting of sequences of 8 stress conditions whose stress levels successively increased and decreased by adding or removing stressors. The full 8-step test sequence is summarized in Table 1.

Table 1: Test sequence over the individual conditions and their properties

1	vowel query	10 quiz questions	2 open questions	without effort	without mask
2	vowel query	10 quiz questions	2 open questions	with acoustical stress	without mask
3	vowel query	10 quiz questions	2 open questions	With physiological exercise	without mask
4	vowel query	10 quiz questions	2 open questions	with acoustical and physiological stress	without mask
5	vowel query	10 quiz questions	2 open questions	with acoustical and physiological stress	with mask
6	vowel query	10 quiz questions	2 open questions	With physiological exercise	with mask
7	vowel query	10 quiz questions	2 open questions	with acoustical stress	with mask
8	vowel query	10 quiz questions	2 open questions	without effort	with mask

2.3 Analysis

Our analysis were prepared and conducted on the basis of PRAAT [8] and supplemented by ProsodyPro [9], and FormantPro [10] sckripts. The following pitch and intensity characteristics were measured in the recorded speech signals using ProsodyPro:: (1) maxF0 (in Hz), (2) minF0 (in Hz), (3) meanF0 (in Hz), excursion size (i.e. F0 range in semitones), finalF0 (in Hz), average intensity (in dB), duration (in ms), maximum velocity of F0 change (in semitones per second), and final velocity of F0 change (in semitones per second). The FormantPro script, developed by [10], was used to analyze vowels that speakers produced at the segmental level. The average values of the first three formants (F1-F3) as well as their minima and maxima, durations (in ms) and the mean intensities (in dB) were measured. All ProsodyPro measurements were taken at three syntactic levels. The first level is the sentence-level, which in this work is not synonymous with a syntactic theorem, but just represents a single coherent answer of a participant to a given quiz question. At second level of analysis was the intonation phrase. The third syntactic level to which the ProsodyPro script was applied was the word-level. Here, we focused on the target word that every statement contained and that was cued to by the way the quiz question was formulated. The target words were selected such that they were primarily voiced and reflected a set of phonologically long and short German vowels including closed, open, and intermediate qualities.

3. Results

ANOVAs were performed based on the acoustic measurements. We excluded the female speaker in these ANOVAs in order to avoid the corresponding gender-related variability in our datasets. Due to the limit space of this paper and the number of tests performed, we omit all tests statistics in the following. Only significant results are reported, though. Results show for the vowel phoneme /a/ that F1 is significantly lower under physical stress, whereas F2 significantly increases. With the addition of pink noise, F1 increases and F2 decreases. This significant combinatorial effect remained also under respiratory mask conditions. However, without a respiratory mask, a significant fall of F1 is observed compared to the first condition. However, it has been categorized as "no effect", since no combinatorial effect shows up (see Table 3).

Table 2: Significant changes for F1 and F2 of the segmental examination of the vocal query

			F1	F2
[a]	Pink noise	*mask*		
		no mask	↘	↗
	Ergometer	*mask*	↘	↗
		no mask	↘	↗
[e]	Pink noise	*mask*	↗	
		no mask	↗	↘
	Ergometer	*mask*		
		no mask		↘
[o]	Pink noise	*mask*		
		no mask		↘
	Ergometer	*mask*	↗	↘
		no mask		↘
[i]	Pink noise	*mask*		↘
		no mask		↗
	Ergometer	*mask*		
		no mask		↗
[u]	Pink noise	*mask*	↘	↗
		no mask		
	Ergometer	*mask*	↘	
		no mask	↗	

The realization of the vowel phoneme /e/ shows that, under noise-induced physical stress that F1 increases significantly both without and with a worn mask. Without a mask, F2 drops below the values found under physical load (see Table 2). With a respirator mask, a counteracting effect is found for F1, resulting from a significant increase under pink noise and an almost constant physical load. For the conditions with physical and physiological stress combined, an additive effect shows up for F2 in the form of a significant increase (see Table 3). For the vowel phoneme /i/, the test persons in the conditions without a respirator mask articulated a significantly lower F2 under the influence of noise and also under physiological load (see Table 2). These two stressors produced an additive effect in combination and caused a further significant drop of F2.

For the combinatorial effect without mask, an additive effect is also observed for F1, but a significant increase is what was found here (see Table 3). For the respiratory mask, the effects of the ergometer manifest themselves in a significant increase in F1 and a significant drop in F2 (see Table 2). For F1, a relativizing effect with a significant increase in the combination of physical and physiological stress could be observed. This combination results in a significant, additive increase for F2 (see Table 3). For the anterior, closed vowel, only significant differences are found for F2 as the stressors are added up. That is, pink noise and also the ergometer caused a significant increase in F2 without a respiratory mask (see Table 2). The combination of both stressors caused a further significant increase. A significant reduction in the F2 could be seen with a mask when noise was added. This additive effect was not observed when our speakers wore the respiratory mask.

Table 3 Combinatory effects of the stressors on the vowels from the vowel query

		F1	F2
[a]	*mask*	**no effect***	**additional effect***
	mask	counteracting effect	counteracting effect
[e]	*mask*	counteracting effect	additional effect
	no mask	**counteracting effect ***	**additional effect***
[i]	*mask*	counteracting effect	**additional effect***
	no mask	counteracting effect	counteracting effect
[o]	*mask*	**counteracting effect ***	**additional effect***
	mask	**counteracting effect ***	**additional effect***
[u]	*mask*	no effect	no effect
	no mask	additional effect	counteracting effect

Table 4 Significant changes for each parameter in the examinations

			Max F0	Min F0	Mean F0	Mean intensity
Sentence-level	Pink noise	*mask*		↗	↗	↗
		no mask	↘			↗
	Ergometer	*mask*		↗		↗
		no mask	↘			↗
IP-Level	Pink noise	*mask*		↗		↗
		no mask				↗
	Ergometer	*mask*		↗		↗
		no mask				↗
Word-level	Pink noise	*mask*				↗
		no mask				↗
	Ergometer	*mask*				↗
		no mask				↗

For the condition without a respiratory mask, a significant increase in F1 emerged for the vowel-phoneme /u/ only under the physiological stressor of the ergometer. However, a significant reduction is observed with respiratory mask. Also, the addition of pink noise causes a significant drop of F1 with respiratory mask. On the other hand, a significant increase is found for F2 (see Table 2).

For all prosodic parameters measured at the sentence, intonation, and word levels, wearing the respiratory mask always led to higher values. The Lombard effect consistently manifests itself within intonation phrases, as well as at the word level in the form of significant increase in mean intensity as pink noise was added. This observation could be made for the conditions 1 to 4 without respiratory mask, as well as for the conditions 8 to 5 with mask. Furthermore, it could be observed that minimum pitch increases with pink noise and separate physiological stress at all prosodically analyzed levels. The combination of physical and physiological stress led to an additive effect in that it further significantly increased the mean intensity

Table 5 Summary of the combinatorial effects of the stressors at the prosodically analyzed levels

	Sentence		IP-Phrase		Word	
	no mask	*mask*	*no mask*	*mask*	*no mask*	*mask*
Max F0	**additional effect***	no effect	additional effect	counteracting effect	counteracting effect	**additional effect ***
Min F0	additional effect	**additional effect***	additional effect	**additional effect ***	additional effect	**additional effect ***
Mean F0	no effect	**additional effect***	no effect	**additional effect ***	no effect	additional effect
Final F0	counteracting effect	**additional effect***	counteracting effect	**additional effect ***	no effect	**additional effect***
Excursion size	**additional effect ***	additional effect	additional effect	counteracting effect	counteracting effect	no effect
Mean intensity	**additional effect ***	**additional effect ***	**additional effect ***	**additional effect ***	**additional effect ***	**additional effect ***
duration	no effect	no effect	no effect	counteracting effect	no effect	no effect

For the prosodic changes under physiological stress at the sentence-level, we found that the maximum pitch and mean intensity both increased significantly. The physical stress created by the ergometer, on the other hand, caused a significant drop in mean intensity, as well as a decrease in mean pitch. The combination results in a further significant increase, i.e. in an additive effect. Since the excursion size decreases under both noise and ergometer, the combination condition has an additive effect that then strongly reduced the excursion size. The significant decrease in maximum F0 under noise reported at the beginning, combined with physical stress, leads to an additive, significant, and strong fall in maximum F0. The influence of pink noise results, under the respiratory mask, in a significant increase in minimum F0 and mean F0, as well as in mean intensity. The ergometer had the same effect on minimum F0 and mean intensity (see Table 4). Significant combinatorial effects are also found in these two parameters (and additionally in mean F0, as well as final F0), causing a strong increase in minimum F0 and the mean intensity. Using a respiratory mask caused significantly higher values for all analyzed prosodic parameters related to the fundamental frequency (F0).

The prosodically significant changes under stress and without a respiratory mask include a significant increase in mean intensity under the influence of pink noise or physical stress (see Table 4). This parameter is besides minimum pitch the only one that is involved in a significant additive effect under combined stressors (see Table 5). Again, when using the respiratory mask, values are all significantly higher (see Table 5). Under the separate stressor of the pink noise and ergometer, the target word's mean intensity increases significantly. A significant increase was also observed at the word-level, in connection with the additive effect of the stressor combination (see Table 5). Even under the additional stressor of the respiratory mask, there are only significant increases in the mean intensity (see Table 4). This applies to the separate physical as well as the separate physiological stressor.. Furthermore, significant increases in the maximum, minimum and final F0 are observed in the combinatorial stressor condition (see Table 5). At the word level, significantly higher values have been obtained in the analyzed target words for the parameters of mean intensity as well as minimum F0 and maximum F0 when participants worse the respiratory mask.

4. Conclusions

On the basis of these results, the three questions raised at the outset can now be answered:

- (1) Can different stress factors be distinguished from each other in the speech signal, and if so, in which way? No. Based on this work, the individual stress factors cannot be reliably identified by means of the acoustic profile they create in the speech signal. This question needs further investigation with larger numbers of participants and further parameters (especially at the segmental level).
- (2) Is there an interaction/interference of stress-factor effects on the speech signal? Yes, stress factors interfere in most cases and show only significant, additive effects in the speech signal at sentence, intonation, and word level. At the segment levelthe analysis with regard to the combinatorial effect of the stressors on F1 and F2 shows that in the semi-closed vowels /e/ and /o/, there are relative effects in the range of F1. For F2, on the other hand, only additive effects have been found. The only exception with regard to these additive effects is the posterior, closed vowel /u/.
- (3) Can the measurable stress increase ad infinitum, or is there an upper limit for the amount of stress reflected in the speech signal (with regard to additive effects of stress types)?
  Since four different stressors have been used in this study (quiz paradigm, pink noise, ergometer and mask) whose effects added up in many cases, it can be said that there is no upper limit for the manifestation of stress in speech production. Thus, based on the present data, there is no clear indication that the additive effects of stress factors are limited in quality or quantitatively. Further research is needed to find such a limit.

Literature

[1] Murray, I. R., Baber, C., & South, A. 1996. *Towards a definition and working model of stress and its effects on speech*. In: *Speech Communication*, *20*(1-2), S. 3-12.

[2] Ververidis, D. und Kotropoulos, C. 2006. *Emotional speech recogni-tion: Resources, features, and methods.* In: Speech Communication 48. S. 1162-1181.

[3] Absil, Grammatica, Harmegenies, Legros, Poch und Ruiz. 1995. *Time-related variabilities in stressed speech under laboratory and real condi-tions: Proceedings of the ESCA-NATO Workshop on Speech under Stress. 14.-15. September 1995, Lisbon, Portugal, S. 53-56.*

[4] Hansen, J. H., Bou-Ghazale, S. E., Sarikaya, R., & Pellom, B. (1997): *Get-ting started with SUSAS: a speech under simulated and actual stress da-tabase*. In: *Proceedings of the Eurospeech* 1997. 22.-25 September, Rhodes, Griechenland (Vol. 97), No. 4, S. 1743-46.

[5] Steenken, H. J. M. und Hansen, J. H. L.. (1999) Speech under stress condi-tions: overview of the effect on speech production and on system perfor-mance. In: *Acoustics, Speech, and Signal Processing, 1999. Proceed-ings., 1999 IEEE International Conference on*. IEEE, 1999. S. 2079-2082.

[6] Meinerz, Christoph. 2010. *Effekte von Stress auf Stimme und Sprechen: Eine phonetische Untersuchung ausgewählter akustischer und sprechdynamischer Parameter unter Berücksichtigung verschiedener Stressoren.* Universität Trier. (Ph.D. dissertation)

[7] Landgraf, Rabea. 2015. Stimulating Complex Speech Production Environments. In R. Skarnitzl & O. Niebuhr (Eds.), *Tackling the complexity in speech* (pp. 97-112). Prague: Charlies University Press.

[8] Boersma, P. & Weenink, D. (2017): Praat: doing phonetics by computer [Computer program]. Version 6.0.26, retrieved 2 March 2017 from http://www.praat.org/. Stand: 21.03.2017

[9] Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France. 7-10. Stand: 21.03.2017

[10] Xu, Y. (2007-2015). FormantPro.praat. Available from: http://www.phon.ucl.ac.uk/home/yi/FormantPro/. Stand: 21.03.2017