Prosodic characteristics of read speech before and after treadmill running

Jürgen Trouvain[1], *Khiet P. Truong*[2,3]

[1]Saarland University, Computational Linguistics and Phonetics, Germany
[2]University of Twente, Human Media Interaction Group, The Netherlands
[3]VU University Amsterdam, Human Movement Sciences, The Netherlands
trouvain@coli.uni-saarland.de, k.p.truong@utwente.nl

10.21437/Interspeech.2015-734

Abstract

Physical activity leads to a respiratory behaviour that is very different to a resting state and that influences speech production. How speech parameters are exactly affected by physical activity remains largely unknown. Hence, we investigated how several prosodic parameters change under influence of physical activity and focused on temporal and breathing characteristics which have not been addressed in detail before. Speech from subjects reading aloud a text before and after a treadmill running exercise was analysed for prosodic differences between before and after running. The most important findings include a higher articulation rate, longer averaged pause and breath durations, a higher in-breath intensity, a higher out-breath rate, and a higher mean F0 for speech recorded immediately after vigorous treadmill running. These findings provide fundamental insights into how speech characteristics are affected by physical effort, and may help advance automatic classification of physical stress in speech.

Index Terms: speech under physical stress, prosodic analysis, treadmill running

1. Introduction

Often, talking is difficult after having performed vigorous physical activity. People for example feel out of breath and have trouble controlling their respiratory system and vocal apparatus. This trouble is mainly caused by competing needs between breathing for metabolic reasons and breathing for speech production. Due to this competition, one expects that physical load affects speaking and vice versa [1]. In the current study, we aim to investigate the former, namely how physical load affects speech production.

There have been several studies addressing the influence of physical load on speech production, but the results obtained are inconclusive. Fundamental frequency (F0), formants, syllable rate and pause placements are some of the speech parameters often investigated with respect to this topic [2, 3, 4]. However, surprisingly, one of the most salient changes in speech under physical load, namely breathing characteristics, has not or only marginally been addressed in the speech science community to the best of our knowledge. In more recent studies about speech and physical stress, the focus is on automatic classification of low and high physical stress [5, 6, 7, 8]. However, those studies are usually focused on algorithm and performance improvement without necessarily providing insights into exactly *how* speech characteristics change under influence of physical effort.

In this paper, we aim to provide more insights into how physical effort affects speech production and present an analysis of several prosodic parameters, focusing on temporal changes and breathing characteristics, in read speech that was recorded before and after vigorous physical activity. We explore characteristics of F0, speech and pauses, speaking rates, breathing and disfluencies, and test whether and how these characteristics are affected by physical load. The results will give us more insights into how breathing is regulated when used for different purposes simultaneously, and will help develop other speech features for automatic classification of speech under physical stress as well.

In the following sections we discuss related work and our hypotheses (Section 2), and we present the data used for our study (Section 3). The analysis and results are presented in Section 4 and 5 respectively. We present and discuss the results in Section 6.

2. Background

We describe previous work on speech and physical stress and subsequently present hypotheses based on previous literature about how speech is expected to change under influence of physical stress.

2.1. Related work

Previous studies focusing on unraveling how speech parameters are affected by physical load and what speech parameters can be used to predict levels of physical did not lead to conclusive results. In these studies, F0 (also referred to as its perceptual correlate 'pitch') is among the most frequently addressed one. Johannes et al. [2] studied changes in F0 with increasing exercise intensity in a bicycle task with participants who had to count from 1 to 10 during exercise. They found that F0 increases with increasing exercise intensity in a non-linear fashion. Within an individually well-tolerated range of physical load, F0 remains relatively unaffected by the level of physical effort. When approaching the point of physical exhaustion, F0 increases again. The increase in F0 with increasing physical effort was also found by Godin and Hansen [3]. The amount of voiced speech was found to decrease with increasing physical effort. The first two formants (F1, F2) and F0 variation did not seem to be affected by physical effort while utterance duration and several parameters of the glottal waveform do seem to change, albeit in a speaker-dependent manner. Finally, Sandage et al. [9] found in their study that voice use associated with physical activity requires additional laryngeal effort and closure forces.

Instead of focusing on acoustic changes in speech production caused by physical activity, Baker et al. [4] focused in their study on temporal changes in speech production, in particular, the placement of pauses. Participants performed exercise tasks (bicycle ergometer) at rest, 50%, and 75% of VO2 max

(maximal oxygen consumption). During these tasks, the participants were asked to read aloud the rainbow passage. The speech recorded during these tasks were analyzed for number of syllables between inspirations, articulation rate, and inappropriate/nonlinguistic placement of pauses. It was found that the values of all these speech parameters increased with increasing task intensity.

Our study mostly resembles the one carried out by Baker et al. [4] who also focused on temporal and pausing aspects in their study. However, our study complements Baker et al.'s [4] and other existing studies on speech under physical effort in that we additionally investigate breathing, disfluency and several temporal characteristics that have not been addressed before.

2.2. Hypotheses

Based on literature and our knowledge about our respiratory system and vocal apparatus, we can formulate several hypotheses about how physical effort affects breathing and speaking. Physical activity leads – among other physiological parameters – to a respiratory behaviour that is very different to a resting state. The breath cycle becomes shorter, the inspiratory phases become longer, the inspiration will be deeper and the sub-glottal pressure will be higher. It can be assumed that such a change in the respiratory setting will have consequences on various levels. A longer inspiratory phase will probably increase the duration of breath pauses, the deeper inspiration will lead to an increase in the frication noise of the inbreath, a higher flow level of the expiratory air probably effects in a breathier voice quality, a higher sub-glottal pressure normally leads to a higher F0, and more breath cycles should lead to more breath pauses.

The need for more pauses would require a re-arrangement of prosodic phrase structure: either more words are packed within one expiratory phase of a shortened breath cycle (leading to a higher articulation rate) or more prosodic phrases for the same number of words are produced (leading to fewer words per phrase). In either case a phonological re-structuring of prosodic phrasing is required, see also [10, 11] .

Due to the lack of control of the habitual respiratory cycles we can expect some planning difficulties not only for phrasing with possible pauses at unexpected ungrammatical locations, as was shown in [4]. The planning trouble can also affect segmental and syllabic articulation, leading to more disfluencies such as slips of the tongue, re-starts and fillers such as "uh" and "uhm".

If the above mentioned hypotheses are correct, then it would be interesting to observe the contradictory strategies with respect to speech tempo involved here, namely simultaneous slowing down (more pausing) with speeding up (higher articluation rate).

3. Data

For our analysis, we used data from the Talk & Run (TalkR) Speech database that is described in more detail in an accompanying paper accepted for this conference [8].

3.1. Participants

Dutch-speaking participants were recruited at the Radboud University in Nijmegen via the SONA participant pool system. We targeted healthy, young adults who had some experience with running or exercising in general but who were not professional runners. In total, 23 students participated of which two were excluded due to technical failures. Fifteen speakers were female, six male and the average age was 22.7 years (sd=3.4).

3.2. Procedure and task

Subjects were invited to the sportslab at Radboud University for two visits. In the first visit, a comfortable running speed was determined that could be maintained for about 20 minutes. In the second visit, subjects were asked to run at that speed on the treadmill until volitional exhaustion. Subjects had to read aloud the same two texts in Dutch before starting their treadmill exercise (pre-condition **pre**) and directly after stopping running (post-condition, **post**). Speech was recorded through a wireless Sennheiser EW 112 G2 system (connected to an USB audio interface) and a lapel ME2 microphone attached to the subject's shirt. Heart rate was measured continuously through a heart rate monitoring belt attached to the subject's chest and was 95 BPM (sd=12) and 163 BPM (sd=13) on average during **pre** and **post**, respectively. During the treadmill exercise, the subjects read aloud a text that was varied and that differed from the pre- and post-conditions. For the current study, we only used one of the texts that was read aloud during the pre- and post-conditions, see Table 1. The content of the text is designed as neutral as possible. Although the reading aloud task is not very realistic it is very useful here. Its high level of control allows a direct comparison of the pre- and post-recordings. Since the planning of upcoming speech in already formulated and scripted sentences is lower than in unscripted and unprepared speech we would expect a smaller control effort for prosodic phrase planning compared to online formulation (which can strongly influence duration and "content" of pauses such as inspiratory strength).

Annemieke vraagt Jos mee te gaan naar een feest en Roos vraagt Mark mee te gaan naar hetzelfde feest. Een van de oudere broers van Mark, die Roos niet heel goed kent, wil ook heel erg graag mee, omdat het een heel leuk feest schijnt te zijn. Ongeveer een half uur geleden zijn Annemieke en Jos vertrokken naar het feest.
Annemieke asks Jos to join a party and Roos asks Mark to join the same party. One of the older brothers of Mark, whom Rose does not know well, would also like to join, because it seems to be a very nice party. About half an hour ago Annemieke and Jos left for the party.

Table 1: *The text read aloud during **pre** and **post** in Dutch and English translation (60 words).*

4. Analysis

We describe the speech measures that were used in our study.

4.1. Speech measures

The acoustic signals from the 42 microphone recordings were manually segmented in *phrases*, i.e. articulatory phases, and *pauses* by using Praat [12]. A *pause* was defined as a non-speech phase in the acoustic signal that typically contains silence, breath noises and sometimes fillers. Please note that the perception of a pause does not necessarily need a pause in an acoustic sense. Prosodic breaks can also be marked without silence and/or breath noise just with intonation, intensity, segmental lengthening, see e.g. [13].

Based on our hypotheses formulated in Section 2.2, we calculated the following temporal parameters: **total speech duration** (sum of all phrases), **average phrase duration**.

Additional temporal measurement were **articulation rate**

prosodic parameter	expectation pre vs. post	level of signific.	mean (sd) pre	mean (sd) post	median pre	median post	expectation sustained	%
total speech duration (sec)	$>$	**	14.75 (2.36)	13.29 (1.42)	14.53	12.97	18	86
articulation rate (syll/sec)	$<$	*	5.43 (0.63)	5.87 (0.55)	5.49	6.00	19	90
speaking rate (syll/sec)	$<$		4.73 (0.69)	4.61 (0.57)	4.81	4.72	12	57
total pause duration (sec)	$<$	**	2.3 (1.52)	3.78 (1.54)	1.90	3.71	17	81
avg pause duration (sec)	$<$	***	0.39 (0.10)	0.74 (0.21)	0.38	0.73	21	100
pause rate (pau/min)	$>$		20.65 (6.48)	18.07 (5.14)	19.52	16.06	7	33
avg breath duration (sec)	$<$	***	0.28 (0.05)	0.40 (0.09)	0.28	0.39	21	100
in-breath intensity (dB)	$<$	***	27.48 (5.10)	36.27 (6.00)	26.23	36.16	21	100
avg in-breath duration (sec)	$<$	***	0.29 (0.05)	0.46 (0.09)	0.28	0.45	21	100
in-breath rate (tokens/min)	$<$		13.36 (3.91)	16.65 (5.53)	12.85	15.85	17	81
out-breath rate (tokens/min)	$<$	***	0.00 (0.00)	9.82 (7.86)	0.00	7.97	18	86
F0 mean (Hz) - female	$<$	**	193.89 (13.72)	225.03 (18.22)	191.05	230.48	15	100
F0 mean (Hz) - male	$<$		120.05 (12.86)	132.71 (13.76)	120.87	139.22	6	100
F0 range (Hz) - female	$<$		47.07 (29.23)	40.50 (16.67)	40.34	39.74	6	40
F0 range (Hz) - male	$<$		41.59 (45.18)	20.10 (10.01)	24.38	21.16	1	17
disfluency rate (tokens/min)	$<$		3.36 (3.71)	2.73 (3.31)	3.44	3.17	8	38

Table 2: *Descriptive statistics and statistical significance (*$p < .002$, **$p < .001$, ***$p < .000$) of differences in speech parameters measured in the **pre** and **post** conditions.*

(excluding pauses) and **speaking rate** (including the pauses) which were measured as syllabic rates (number of syllables per second). The text shown in Table 1 has 77 syllables – when syllables or entire words were repeated due to disfluent behavior, these syllables were added to the count. Possible omissions of syllables which is not unusual in fast speech were not considered.

Regarding the pauses in each recording we counted its number, determined its location in the text, calculated the **total pause duration**, the **average pause duration** and the **pause rate** (number of pauses normalized by duration).

Pauses could consist of silence and two forms of breathing: *in-breath* was considered as an audible breathing-in noise, and *out-breath* was considered present when the phase of expelling air at the end of a phrase clearly exceded the phonatory phase of the vowel or for plosives the usual duration of the aspiration. Instances of out-breath were only observed in the post-condition. In addition to the average pause duration, we also report the **average breath duration**, as well as the **average in-breath duration**. We also expected to see changes in the frication noise of in-breaths and the frequency of out-breaths, hence, we measured the **average in-breath intensity** and **out-breath rate** (number of out-breaths/sec).

For each phrase the mean of F0 was computed and averaged over all phrases (**F0 mean**) in a recording. The **F0 range** in Hertz was calculated based on the difference between the lowest and highest mean F0 found for a phrase in a recording.

Various types of disfluencies were distinguished for annotation such as *slips of the tongue*, *fillers*, *repetitions* and *false starts*. Since a closer analysis of the disfluency type goes beyond the current study, we decided just to report the number of disfluencies per minute, i.e. **disfluency rate**.

4.2. Magnitude of effect

In order to determine whether the differences in prosodic parameters found between **pre** and **post** are statistically significant, we carried out non-parametric repeated measures tests, i.e. Wilcoxon Signed Rank tests. Since the number of speakers considered in our study is not particularly large, we also report how many of the 21 speakers followed the specific pattern as expected for that prosodic parameter.

5. Results

The results are summarised in Table 2. We will present the results of each group of features in more detail.

5.1. Timing of phrases

As expected the articulation rate was generally higher in **post**. Although this is true for 90% of the speakers the level of significance was rather low. The speaking rate (including the pause time) was lower in **post**, though this effect was not significant. In **post** the total speech duration was shorter at a highly significant level.

5.2. Pausing

In accordance to our hypotheses the total pausing time increases in **post**. This increase is also very clearly visible for the mean pause durations. However, against the expectations most speakers reduced their pause rate in **post**, though not at a significant level.

If we assume that the usual way of where to place a pause directly corresponds to full stops and commas (as many text-to-speech synthesizers do), then this default case just happened 4 times in our data, i.e. in only 10% of all cases. There is no doubt that the proposed locations are the main choices for pauses, see Figure 1. However, it is extremely likely that pauses are placed at other locations as well which results in a huge diversity of possible pause structures for the same text. The assumption that expected planning difficulties in **post** would yield more pauses at unusual places could not be confirmed.

5.3. Breathing

As expected the duration as well as the intensity of inbreath segments increases from **pre** to **post**. This observation is valid for all speakers and shows a very high level of significance. Nearly all pauses in **post** are breath pauses, thus the number of

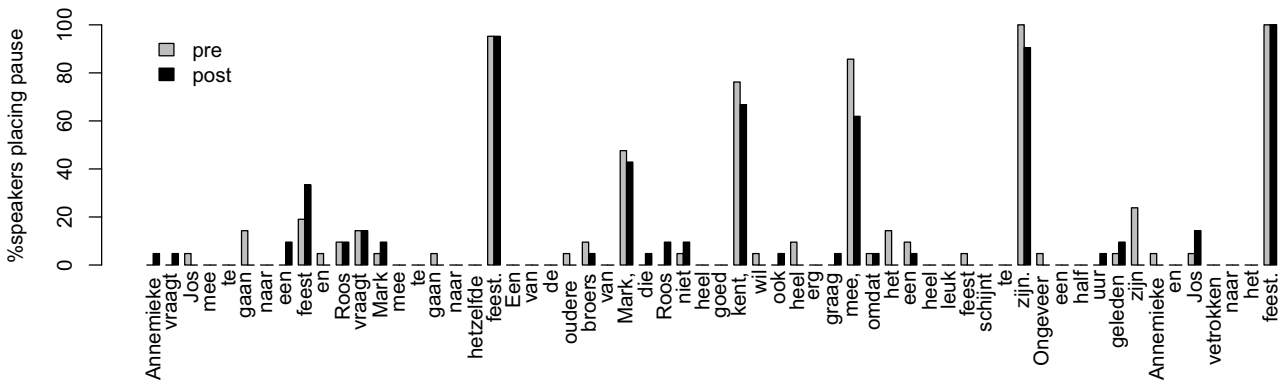

Figure 1: *Percentages of pause locations of subjects reading the text in during the **pre** and **post** conditions.*

in-breath tokens per second is larger **post** than in **pre**. However, this difference does not reach significance. Outbreath could not be observed for any speaker in **pre** but for 18 speakers in **post**.

5.4. Fundamental frequency

As expected the mean F0 is higher in **post** compared to **pre**. Higher F0 mean was found for all speakers with the result that his difference is highly significant. Against our expectations the F0 range is generally smaller (instead of larger) in **post**, though this difference does not reach statistical significance.

5.5. Disfluencies

In contrast to the hypotheses the number of disfluencies in **post** is not significantly higher. Only 8 out of 21 speakers followed the expected pattern. It even proved that the versions from **pre** contained more disfluencies than in **post**. Interestingly, there were only three speakers who were completely fluent, i.e. free of any disfluency.

6. Discussion

The results of this study confirm the findings from previous studies in that mean F0 is consistently higher for all subjects. Obviously, a higher subglottal pressure forces the vocal folds to vibrate more frequently. Our assumption that an increase in F0 mean also involves an increase in F0 range could not be confirmed. Speakers clearly reduced their pitch range which also reduce the possibilities for tonal movements to mark phrase boundaries and sentence accents with F0.

Articulation rate was faster for most speakers in **post** which may signal an increased level of arousal that can be assumed for physical stress. Interestingly, the mean phrase duration was kept constant which means that more words were packed into one phrase in **post**, a finding which is in contrast to Baker et al. [4]. It might be that putting more words between two inspirations helps to counterbalance the effect of having longer pauses.

Regarding breathing, our results confirm our hypotheses related to a different respiratory behavior in **post**. The intensity and the duration of in-breath segments are higher, probably as an effect of deeper inhalation and a lengthened inhalation phase under physical stress. Breath pauses are usually longer than pauses without breath segments which has been shown here again. Since virtually all pauses in **post** are breath pauses – in contrast to **pre** where several non-breath pauses occur – the total duration of pauses is longer in **post**. The observation of outbreath was not expected but happened regularly in **post**. This fact can be interpreted as difficulties in coordinating articulation and increased expiration on the one hand with phonation on the other.

Against our expectations and the evidence reported in [4] there were not more disfluencies and not more pauses at unusual and ungrammtical locations in **post**. An interpretation of this finding must take into account that the great majority of the subjects were already disfluent in **pre**. A further factor to be considered is that there might be an effect of familiarization with the text in **post**.

7. Conclusion

The prosodic characteristics of speech after treadmill running entails a higher articulation rate and longer duration of pauses (mean and total). The 'competition between metabolic requirements and linguistic phrasing' [4] leads to a contradictory strategy regarding speeding up and slowing down.

In addition, the increased F0 mean and the salient use of in- and out-breath segments further contribute to the acoustic-prosodic patterns of speech after physical exercise which can be seen as a type of arousal. Interestingly, live commentators of sports broadcasts show similar though not exactly the same patterns for an increased level of affective arousal. In horse race commentaries [14] strong inhalation noises accompany a highly increased F0 whereas in pre-goal scenes of football live commentaries [15] a higher articulation rate and an extreme increase of F0 provides the typical expressivity.

In contrast to broadcasted speech a reading task as used here does not aim at a high level of intelligibility. There are however more realistic speaking scenarios during physical exercise where speakers have an interest to speak to be understood such as exercise instructors [16] who are using their prosody also to gain attention and to maintain motivation. Thus, future studies should involve speaking situations beyond read speech as well as perception tests of speech under physical stress.

8. Acknowledgements

This research was funded by COMMIT/ and is part of the P3 project SenseI: Sensor-Based Engagement for Improved Health.

9. References

[1] Y. Meckel, A. Rotstein, and O. Inbar, "The effects of speech production on physiologic responses during sub-maximal exercise," *Medicine and Science in Sports and Exercise*, vol. 34, pp. 1337–1343, 2002.

[2] B. Johannes, P. Wittels, R. Enne, G. Eisinger, C. A. Castro, J. L. Thomas, A. B. Adler, and R. Gerzer, "Non-linear function model of voice pitch dependency on physical and mental load," *European Journal on Applied Physiology*, vol. 101, pp. 267–276, 2007.

[3] K. W. Godin and J. H. L. Hansen, "Analysis and perception of speech under physical task stress," in *Proceedings of Interspeech*, 2008, pp. 1674–1677.

[4] S. E. Baker, J. Hipp, and H. Alessio, "Ventilation and speech characteristics during submaximal aerobic exercise," *Journal of Speech, Language, and Hearing Research*, vol. 51, pp. 1203–1214, 2008.

[5] B. Schuller, F. Friedmann, and F. Eyben, "Automatic recognition of physiological parameters in the human voice: heart rate and skin conductance," in *Proceedings of ICASSP*, 2013, pp. 7219–7223.

[6] B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval, E. Marchi, and Y. Zhang, "The INTERSPEECH 2014 Computational Paralinguistics Challenge: Cognitive and Physical Load," in *Proceedings of Interspeech*, 2014, pp. 427–431.

[7] B. Schuller, F. Friedmann, and F. Eyben, "The Munich Biovoice Corpus: Effects of physical exercising, heart rate, and skin conductance on human speech production," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2014, pp. 1506–1510.

[8] K. P. Truong, A. Nieuwenhuys, P. Beek, and V. Evers, "A database for analysis of speech under physical stress: detection of exercise intensity while running and talking," in *Proceedings of Interspeech*, accepted.

[9] M. J. Sandage, N. P. Connor, and D. D. Pascoe, "Voice function differences following resting breathing versus submaximal exercise," *Journal of Voice*, vol. 27, no. 5, pp. 572–578, 2013.

[10] F. Grosjean and M. Collins, "Breathing, pausing and reading," *Phonetica*, vol. 36, no. 2, pp. 98–114, 1979.

[11] A. Rochet-Capellan and S. Fuchs, "The interplay of linguistic structure and breathing in German spontaneous speech," in *Interspeech*, 2013, pp. 2014–2018.

[12] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[13] A. Butcher, "Aspects of the Speech Pause: Phonetic Correlates and Communicative Function," Arbeitsberichte Institut für Phonetik, University of Kiel, 1981.

[14] J. Trouvain and W. J. Barry, "The prosody of excitement in horse race commentaries," in *Proceedings of ISCA-Workshop on Speech and Emotion, Newcastle (Northern Ireland)*, 2011, pp. 86–91.

[15] J. Trouvain, "Between excitement and triumph – live football commentaries in radio vs. TV," in *Proceedings of 17th International Congress of Phonetic Sciences (ICPhS)*, 2011, pp. 2022–2025.

[16] L. Skutella, L. Süsenbach, K. Pitsch, and P. Wagner, "The prosody of motivation. First results from an indoor cycling scenario." in *Elektronische Sprachsignalverarbeitung (ESSV)*, 2014, pp. 209–215.