Prosodic characteristics of read speech before and after treadmill running

Jürgen Trouvain¹, Khiet P. Truong²,³

¹Saarland University, Computational Linguistics and Phonetics, Germany
²University of Twente, Human Media Interaction Group, The Netherlands
³VU University Amsterdam, Human Movement Sciences, The Netherlands

trouvain@coli.uni-saarland.de, k.p.truong@utwente.nl

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
six male and the average age was 22.7 years (sd=3.4).
cluded due to technical failures. Fifteen speakers were female,
runners. In total, 23 students participated of which two were ex-
targeted healthy, young adults who had some experience with
Dutch-speaking participants were recruited at the Radboud Uni-
3.1. Participants
Based on literature and our knowledge about our respiratory
Physical activity leads – among other physiological parameters
A longer inspiratory phase will probably increase the duration
in the respiratory setting will have consequences on various levels.
more breath cycles should lead to more breath pauses.
The need for more pauses would require a re-arrangement
The text read aloud during
Table 1: The text read aloud during pre and post in Dutch and
English translation (60 words).
3.2. Procedure and task
Subjects were invited to the sports lab 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 in-
terface) and a lapel ME2 microphone attached to the subject’s
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 com-
parison 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 ex-
pect a smaller control effort for prosodic phrase planning com-
pared to online formulation (which can strongly influence dura-
and “content” of pauses such as inspiratory strength).

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 si-
ence, 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, seg-
mental lengthening, see e.g. [13].
Based on our hypotheses formulated in Section 2.2, we cal-
culated the following temporal parameters: total speech dura-
tion (sum of all phrases), average phrase duration.
Additional temporal measurement were articulation rate.
(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 exceeded 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 friction 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 increased in 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

<table>
<thead>
<tr>
<th>prosodic parameter</th>
<th>expectation</th>
<th>level of signific.</th>
<th>mean (sd) pre</th>
<th>median (sd) pre</th>
<th>median (sd) post</th>
<th>expectation sustained</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>total speech duration (sec)</td>
<td>&gt;</td>
<td>**</td>
<td>14.75 (2.36)</td>
<td>13.29 (1.42)</td>
<td>14.53 (1.27)</td>
<td>18.86</td>
<td></td>
</tr>
<tr>
<td>articulation rate (syll/sec)</td>
<td>&lt;</td>
<td>*</td>
<td>5.43 (0.63)</td>
<td>5.87 (0.55)</td>
<td>5.49 (0.60)</td>
<td>19.90</td>
<td></td>
</tr>
<tr>
<td>speaking rate (syll/sec)</td>
<td>&lt;</td>
<td>*</td>
<td>4.73 (0.69)</td>
<td>4.61 (0.57)</td>
<td>4.81 (0.72)</td>
<td>12.57</td>
<td></td>
</tr>
<tr>
<td>total pause duration (sec)</td>
<td>&lt;</td>
<td>**</td>
<td>2.74 (1.52)</td>
<td>3.78 (1.54)</td>
<td>1.90 (3.71)</td>
<td>17.81</td>
<td></td>
</tr>
<tr>
<td>avg pause duration (sec)</td>
<td>&lt;</td>
<td>***</td>
<td>0.39 (0.10)</td>
<td>0.74 (0.21)</td>
<td>0.38 (0.73)</td>
<td>21.100</td>
<td></td>
</tr>
<tr>
<td>pause rate ( pau/min)</td>
<td>&gt;</td>
<td></td>
<td>20.65 (6.48)</td>
<td>18.07 (5.14)</td>
<td>19.52 (16.06)</td>
<td>7.33</td>
<td></td>
</tr>
<tr>
<td>avg breath duration (sec)</td>
<td>&lt;</td>
<td>***</td>
<td>0.28 (0.05)</td>
<td>0.40 (0.09)</td>
<td>0.28 (0.39)</td>
<td>21.100</td>
<td></td>
</tr>
<tr>
<td>in-breath intensity (dB)</td>
<td>&lt;</td>
<td>***</td>
<td>27.48 (5.10)</td>
<td>36.27 (6.00)</td>
<td>26.23 (36.16)</td>
<td>21.100</td>
<td></td>
</tr>
<tr>
<td>avg in-breath duration (sec)</td>
<td>&lt;</td>
<td>***</td>
<td>0.29 (0.05)</td>
<td>0.46 (0.09)</td>
<td>0.28 (0.45)</td>
<td>21.100</td>
<td></td>
</tr>
<tr>
<td>in-breath rate (tokens/min)</td>
<td>&lt;</td>
<td>**</td>
<td>13.36 (3.91)</td>
<td>16.65 (5.53)</td>
<td>12.85 (15.85)</td>
<td>17.81</td>
<td></td>
</tr>
<tr>
<td>out-breath rate (tokens/min)</td>
<td>&lt;</td>
<td>***</td>
<td>0.00 (0.00)</td>
<td>9.82 (7.86)</td>
<td>0.00 (7.97)</td>
<td>18.86</td>
<td></td>
</tr>
<tr>
<td>F0 mean (Hz) - female</td>
<td>&lt;</td>
<td>**</td>
<td>193.89 (13.72)</td>
<td>223.03 (18.22)</td>
<td>191.05 (230.48)</td>
<td>15.100</td>
<td></td>
</tr>
<tr>
<td>F0 mean (Hz) - male</td>
<td>&lt;</td>
<td></td>
<td>120.05 (12.86)</td>
<td>132.71 (13.76)</td>
<td>120.87 (139.22)</td>
<td>6.100</td>
<td></td>
</tr>
<tr>
<td>F0 range (Hz) - female</td>
<td>&lt;</td>
<td></td>
<td>47.07 (29.23)</td>
<td>40.50 (16.67)</td>
<td>40.34 (39.74)</td>
<td>6.40</td>
<td></td>
</tr>
<tr>
<td>F0 range (Hz) - male</td>
<td>&lt;</td>
<td></td>
<td>41.59 (45.18)</td>
<td>20.10 (10.01)</td>
<td>24.38 (21.16)</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>disfluency rate (tokens/min)</td>
<td>&lt;</td>
<td></td>
<td>3.36 (3.71)</td>
<td>2.73 (3.31)</td>
<td>3.44 (3.17)</td>
<td>8.38</td>
<td></td>
</tr>
</tbody>
</table>

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.
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 observed. The intensity of any disfluency.

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


