

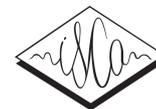
PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/150729>

Please be advised that this information was generated on 2021-03-07 and may be subject to change.



A database for analysis of speech under physical stress: detection of exercise intensity while running and talking

Khiet P. Truong^{1,2}, Arne Nieuwenhuys³, Peter Beek², and Vanessa Evers¹

¹University of Twente, Human Media Interaction Group, The Netherlands

²VU University Amsterdam, Human Movement Sciences, The Netherlands

³Radboud University, Behavioural Science Institute, The Netherlands

k.p.truong@utwente.nl, a.nieuwenhuys@bsi.ru.nl, p.j.beek@vu.nl, v.evers@utwente.nl

Abstract

One of the ways to gauge your own exercise intensity while running, is to assess your capability of talking while running: if you can still speak comfortably, you are running within the recommended intensity guidelines. This subjective way of estimating one's exercise intensity by talking (i.e. the Talk Test) motivated us to investigate how speech characteristics are affected during running and whether it is possible to develop a more objective way of estimating exercise intensity levels while running through voice analysis. To this end, we developed the Talk & Run Speech database that contains speech recorded from people before, during, and after running. We present our database and show that it is possible to detect exercise intensity below or above the anaerobic threshold in speech during running with a performance of 73.5% and 60.0% (unweighted average recall) for female and male speakers respectively.

Index Terms: speaker state analysis, physical stress, exercise intensity, heart rate, speech database, talk test

1. Introduction

Exercising regularly is good for your health. The American College of Sports Medicine (ACSM) recommends most adults to exercise at a moderate intensity level at least 5 days a week or to exercise at a vigorous intensity level 3–5 days a week. Frequency, Intensity, Time, and Type of exercise (FITT) are the 4 variables that make up an individual exercise prescription program and that can be specified according to an individual's specific needs and goals. Time (duration) and Intensity (exercise intensity) seem to be the most influential variables with respect to physical health improvement (e.g., [1]). Exercise intensity also plays an important role in exercise adherence. There is evidence that individuals with more exercise experience benefit more from higher intensity programs, while individuals who just started exercising might be better suited to moderate intensity programs [2]. Especially for beginning sports(wo)men, moderateness is important since exercising needs to remain enjoyable and the risk for overtraining and injuries low. Measuring one's own exercise intensity is thus crucial. Our question is whether we can measure exercise intensity unobtrusively through speech analysis.

As an alternative to commonly used methods that measure exercise intensity through heart rate monitoring equipment, the Talk Test [3, 4] is often used: "if you can still speak comfortably, the exercise intensity is about right". The Talk Test is a subjective verbal test that aims to measure exercise intensity and is popular with especially beginning sports(wo)men who do not have heart rate monitoring equipment (yet). Although the

ACSM does not recommend the Talk Test as a *primary* method of prescribing exercise intensity, the Talk Test has been validated against several physiologic markers, e.g., [5, 6, 7, 8]. In addition to these physiological markers, it is known that speech characteristics change under physical stress: the respiratory system that is affected by physical activity will also cause changes in speech production that makes use of the same respiratory system. However, less is known about how speech production changes in a way that is predictive of exercise intensity as the Talk Test suggests, and whether these changes in speech characteristics are sufficiently reliable and robust for a more objective and automatic assessment of exercise intensity. Although similar studies exist that aim to automatically classify low and high physical stress in speech, e.g. [9, 10], and in relation to heart rate or pulse [11], to the best of our knowledge, none of these studies consider classification of exercise intensity (expressed as individualized heart rate target zones) in speech *while running*.

We aim to investigate the feasibility of automatic and objective classification of exercise intensity in speech while running. This would enable runners to assess their level of exercise intensity by means of simple and available equipment such as a smartphone which would contribute to prevention of overtraining. To that end, we collected a speech database which is presented in this paper, as well as a first investigation into automatic detection of exercise intensity (expressed in heart rate target zones) in speech while running. In contrast to existing databases and studies involving research on speech under physical stress, the database presented contains speech with *varying lexical content* elicited from speakers *while running* under various exercise intensity levels.

The paper is structured as follows. Related work is discussed in Section 2. We present the database collected in Section 3. In Section 4, a classification experiment for exercise intensity is described and results are presented. Conclusions are given in Section 5.

2. Related work

Previous studies investigating speech under physical stress have found a number of speech characteristics that change under influence of physical stress. Works by Godin and colleagues [12, 13, 9] and Johannes et al. [14] found that speech parameters such as F0 and the percentage of voiced frames increase and decrease respectively under influence of physical stress (e.g. elicited by a stair stepper [12, 13, 9] and a bicycle stress test [14]), while variation in F0 and values of F1 and F2 remain rather stable. One of the few studies studying speech

change *during* exercise (i.e. graded test protocol on a bicycle ergometer) under various levels of physical stress (at 50%, 75% of VO₂ max) was performed by Baker et al. [15]. They found that the number of syllables between inspirations, articulation rate (number of syllables per second) and the number of inappropriate pause placements would increase with time while performing the exercise task at 50% or 75% of VO₂ max.

More recently, these speech characteristics have been used in automatic classification of low vs high physical stress [9, 11, 16] and even a challenge was organized around this theme at Interspeech 2014 [10]. In [16], using recordings of sustained vowel sounds and breathing periods made before and after a physical exercise were used as speech data, speaker-independent classification results (unweighted average recall) between 64%–70% for low vs high (> 90 BPM) physical stress were obtained. A similar classification task was addressed in [9] where an accuracy of 69%–74% was obtained using glottal and Mel-Frequency Cepstrum Coefficients (MFCCs).

However, rather than classifying low vs high physical stress measured before and after physical exercise, the task we tackle is more challenging: we analyse varying speech contents, and address levels of physical load *during* physical exercise. In order to carry out this analysis, speech needs to be collected from runners who are talking while running.

3. The Talk & Run (TalkR) Speech database

A speech database was recorded with the aim to collect speech samples from people while running under various levels of exercise intensity.

3.1. Participants

The target group consisted of healthy, young adults who are not professional runners. The participants were recruited via the Radboud SONA participant pool system - most of the participants were university students. Participants could sign up for the study, that consisted of 2 visits, between 5 August and 4 September 2014. In total, data from 21 out of 23 participants (15f, 6m) was used (data from 2 participants was discarded due to technical failures). As Table 1 shows, the participants were relatively physically active. At the end of the 2nd visit, each participant received a voucher worth of 15 euros.

participant characteristics	mean	std
age	22.7	3.4
average hours of running per week	3.4	1.0
strenuous exercise (times per week)	3.5	2.1
moderate exercise (times per week)	3.4	2.9
mild exercise (times per week)	1.1	1.5
engaging in regular activity long enough to work up a sweat (heart beats rapidly): 1=often, 2=sometimes, 3=never/rarely in a week	1.6	0.5

Table 1: Short summary of participant characteristics ($N = 21$, 15 female and 6 male). Questions are part of the Godin Leisure-Time Exercise Questionnaire [17].

3.2. Material and measurements

The participants were asked to read aloud, in Dutch, short paragraphs of texts. Text **A** consists of an excerpt of the well-known

(translated) text of *The North Wind and the Sun* (56 words) and text **B** consists of sentences with relatively neutral content developed by the authors:

A The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.

B 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 very 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.¹

Text **B** contains 60 words (roughly 15–20 seconds of duration), and it includes the following 5 syntactic breaks (to allow for investigation of pause placement): end of sentence, coordinated clause (conjunction, e.g., “and”, “or”), between adverbial phrase and preceding or following noun phrase or verbal phrase (e.g., “A long time ago...”), subordinate relative clause (e.g., “who”, “which”), and subordinate clause (other, e.g., “because”, “while”). The content of text **B** was varied in 8 different versions but followed the same criteria. All texts were presented to the participants on an LCD tv screen that was attached to the wall in front of the treadmill. Participants were asked to read aloud the texts before, during, and after running.

During the study, we measured and recorded the following:

- **Speech:** Speech was recorded through a lavalier microphone that was attached to the participant’s shirt. We used a wireless system (Sennheiser EW 112 G2 with an ME2 microphone) that was connected to a 4-channel usb audio interface (Alesis iO4) that allowed amplification of the input signal and simultaneous recording of separate audio channels. The recording was done through a laptop running Audacity.
- **Step frequency:** A second microphone was placed near the treadmill to capture the sounds of the steps from which step frequency could be derived.
- **Heart rate:** Heart rate was measured continuously by a Suunto t6d watch and heart rate chest belt that the participants wore while running.
- **Rate of Perceived Exertion (RPE):** The Borg scale [18] was used to measure the perceived exertion. On a scale from 6 (=no exertion at all) to 20 (=maximal exertion) participants were asked to verbally report a number on the Borg scale that represented their feeling of perceived exertion.
- **Feeling Scale (FS):** The Feeling Scale [19] was used to measure participant’s positive and negative feelings during running. On a scale from -5 (=very bad) to 0 (neutral) to +5 (=very good) participants were asked to verbally report the number that best represented their feelings: how would you say you are generally feeling now?
- **Talk Test (TT):** The Talk Test [3, 4] was used to measure the participant’s ability to speak comfortably. After reading a paragraph of text, participants were asked to rate on a scale from -5 (=very difficult) to +5 (=very easy) “how comfortable/how difficult or easy was it for you to read aloud these sentences?”.
- **Accelerometer data:** Accelerometer data was collected by letting the participants wear a smartphone on their arm.

¹In Dutch: 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.

• **Participants' general physical condition:** The participants were asked to answer the Stages of Change questions [20], to fill in the Godin Leisure-Time Exercise Questionnaire [17] and the International Physical Activity Questionnaire (IPAQ [21]). Some other general questions regarding their experience with running were also asked.

3.3. Procedure

The study was carried out in 2 separate visits. The purpose of visit 1 was to find a running speed that was comfortable for the participant to maintain for roughly 15–20 minutes. The actual measurements took place in visit 2 where the participant was asked to run until volitional exhaustion at a constant speed as determined in visit 1. The study took place in the sportlab of Radboud University in Nijmegen.

3.3.1. Visit 1

The participant first provided demographics and filled in questionnaires [20, 17, 21] about his/her general physical condition and his/her experience with running, and signed a consent form. Subsequently, (s)he was asked to run on the treadmill. Following a short warming-up, starting with a speed of 7 km/h, the belt speed was increased every 30s with speed increments dependent on the participant's estimates of how long (s)he think (s)he could maintain running at this speed (adopting a similar approach applied in Bood et al. [22]). The treadmill was stopped as soon as the participant said (s)he could run for no more than 20 minutes.

3.3.2. Visit 2

In the 2nd visit, where the actual measurements were carried out, the participant had to run until volitional exhaustion, but (s)he could stop at any moment during the exercise. Upon arrival, the participant signed a consent form and (s)he was explained the procedure. The participant wore a heart rate belt, a lavalier microphone attached to the shirt, and a smartphone on the arm. Subsequently, the participant took place on the treadmill and started with the running exercise that consisted of 4 phases, see Fig. 1.

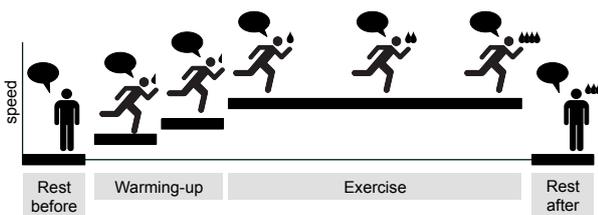


Figure 1: Running protocol in visit 2.

Rest before: The participant is in rest and is asked to read aloud both texts **A** and **B**.

Warming-up: The participant starts running at a speed of 65%, 75%, and 85% of the target speed (determined in Visit 1) that is increased every 1,5 min (i.e. at every stage). The measurements of the RPE, FS, and TT are performed at the end of the 75% and 85% stage, followed by the question whether (s)he can continue running. The content of text **B** used in the TT was varied in a randomized order.

Exercise: The setting is similar to the **Warming-up** phase but now the speed is increased to the target speed and kept constant (to keep the step frequency relatively stable) until the participant is exhausted and indicates that (s)he wants to stop the exercise. Also, each stage now lasts 2 min.

Rest after: Immediately when the participant indicates that (s)he would like to stop, the belt is stopped and the participant is asked to read aloud the same texts **A** and **B** as read aloud during **Rest before**: this allows for a direct and controlled comparison between read speech in low and high physical stress conditions.

Some descriptive statistics of the data obtained can be found in Table 2. To summarize, the main features of the database are that a) it contains speech, HR, RPE, FS, TT and movement execution (accelerometer data) measurements collected under various increasing levels of exercise intensity, ranging from rest to volitional exhaustion, b) the running speed was kept constant in order to keep the step frequency relatively stable, c) the texts recorded in **Rest before** and **Rest after** are the same (which allows for a controlled comparison) and the speech signals recorded in those phases are noise-free (no treadmill noise), and d) the texts recorded in **Warming-up** and **Exercise** have varying content which allows for better generalization of the results. With the data collected, we carried out a first study on the automatic assessment of exercise intensity in speech while running.

measure	min	max	mean	std
running speed (km/h)	8.5	15.8	10.9	1.8
HR _{first} (BPM)	109.0	166.0	144.4	14.3
HR _{last} (BPM)	172.0	198.0	183.9	6.9
TTE (seconds)	562.1	2393.4	1061.3	423.9

Table 2: Descriptive statistics of the data obtained (21 participants): HR_{first} indicates the heart rate measured during the first stage of the **Exercise** phase, HR_{last} indicates the heart rate measured during the last stage of **Exercise** before the participant stops voluntarily, TTE indicates time-until-exhaustion (sum of durations of **Warming-up** and **Exercise**).

4. Automatic detection of low and high exercise intensity in speech while running

We present results of our first study into the automatic assessment of exercise intensity in speech.

4.1. Task definition

Exercise prescription is often described in terms of target heart rate (THR) zones that are expressed in percentages of training intensity, and that can be associated with heart rate ranges. These heart rate ranges are individual and dependent on personal factors such as fitness level, age and gender. We used the Karvonen method [23] to calculate heart rate ranges at certain percentages of training intensity: $HR_{target} = ((HR_{max} - HR_{rest}) \times \%Intensity) + HR_{rest}$, where HR_{max} is the maximum HR (the HR measured during the last stage before exhaustion), $\%Intensity$ is the percentage intensity (THR), and HR_{rest} is the HR measured in rest. Since we do not have an accurate estimation of our participants' HR_{rest} , we used an average

HR_{rest} reported in [24] for males and females between 20 – 30 years old (70 and 74 BPM respectively).

In general, THR zones as described in Table 3 can be distinguished. A pivotal point occurs at the transition from aerobic to anaerobic zone where the exercise intensity turns into heavy load. Many exercise guidelines prescribe periods of moderate and vigorous exercise that correspond to training at below or above the anaerobic zone (e.g., [25, 26]) that in turn coincides with physiological thresholds such as the lactate and ventilatory thresholds at around a THR of 80%Intensity, e.g., [8, 7].

In our study, we calculated $HR_{target80}$ at the pivotal point of 80%Intensity for each participant and assigned all speech samples associated with a $HR < HR_{target80}$ to class **0** (low exercise intensity) and all other samples with $HR \geq HR_{target80}$ to class **1** (high exercise intensity).

Intensity	Description
90–100%	anaerobic: “you should only train in this zone if you are very fit”, lactic acid develops quickly as you are operating in oxygen debt to the muscles
80–90%	anaerobic: “training in this zone is hard, your muscles are tired, your breathing is heavy”, arrived at the point at which the body cannot remove lactic acid as quickly as it is produced (anaerobic threshold)
70–80%	aerobic: most effective for overall cardiovascular fitness
60–70%	aerobic: improves the ability of your heart to pump blood and improve the muscles’ ability to utilize oxygen

Table 3: Target heart rate zones.

4.2. Speech data

The speech data recorded during each Talk Test in the **Warming-up** and **Exercise** phase from 15 female and 6 male speakers were used for training and testing the classifier. In order to upsample the relatively small amount of data and to assess the feasibility of real-time processing, the speech data was segmented into 4s-long segments (with a shift of 2s). This resulted in a total of 1195 (**0**: 862, **1**: 333) and 510 (**0**: 321, **1**: 189) speech samples for female and male speakers respectively.

4.3. Speech features

openSMILE [27] was used for speech feature extraction. We modified an existing feature configuration file (from the Interspeech 2009 Challenge) and selected a subset of the features used in that file: Mel-Frequency Cepstrum Coefficients (MFCCs), energy, F0, and zero crossing rate were used, as well as their deltas (via regression). Finally, statistics, such as the range, mean, linear regression, standard deviation, skewness and kurtosis were extracted for each feature and delta feature, see Table 4. This configuration yielded a 224-dimensional speech feature vector.

LLD: MFCCs, energy, F0, zero crossing rate
Delta regression: MFCCs, energy, F0, zero crossing rate
Functionals: range, mean, standard deviation, skewness, kurtosis, slope and offset of a linear appr. of the contour

Table 4: Speech features used for automatic exercise intensity detection in speech

4.4. Classification experiment setup

We performed speaker-independent classification experiments in a leave-one-speaker-out fashion, separately for female and male speakers. Unweighted average recall (%UAR) was used as performance measure. A 2-class SVM with an RBF kernel

was trained for classification using [28]. The gamma g and cost c parameters of the SVM were optimized by performing a grid-search on pairs of exponentially growing g and c parameters (e.g. $g = [2^{-10}, 2^{-9}, \dots, 2^2, 2^3]$). Speaker-normalization was performed by scaling the speech features in training and testing linearly to $[-1, 1]$

4.5. Results

We report the unweighted average recalls (UARs) of the best performing classifiers that achieved the highest averaged UAR in the gridsearch. We found averaged UARs of 73.5% and 60.0% for female and male speakers respectively. Table 5 shows the confusion matrices and UARs of two of the best performing classifiers.

		Female			Male		
		high	low	recall	high	low	recall
true	high	238	95	71%	139	50	74%
	low	204	658	76%	174	147	46%

Table 5: Results of classifiers.

5. Conclusion and discussion

In contrast to previous studies [11, 9], we aim to address automatic detection of exercise intensity levels that correspond to physiological thresholds (anaerobic vs aerobic) in speech with varying lexical content elicited while running. For this purpose, we collected the Talk & Run (TalkR) Speech Database that was presented in this paper. A 2-class classification experiment (above or below the anaerobic threshold) was carried out with this database which yielded UARs of 73.5% and 60.0% for female and male speakers respectively. We suspect that the relatively low performance for the male speakers may be caused by the relatively small amount of training data available (6 male speakers). The results imply that a simplified Talk Test (2 classes) can indeed be automatized with reasonable reliability: an automatic and objective classification of exercise intensity would be possible by means of simple and available equipment such as a smartphone. The performances are in line with previous studies [11, 9] that reported similar values, however, the tasks in those studies involved speech with a fixed content, recorded before and after running.

For future work, we intend to improve classification performance by comparing more in depth different classes of features and classification algorithms. We also plan to investigate the robustness of the classifier for outdoor use. Further, we would like to study correlations between the subjective self-reported FS (emotions), RPE (perceived exertion) and the TT (Talk Test), and speech characteristics and human movement characteristics (accelerometer data). We are also currently carrying out a more detailed analysis of phonetic features that are affected before and after running [29]. Finally, we plan to release the database for research purposes in the near future.

6. Acknowledgements

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

7. References

- [1] H. A. Wenger and G. J. Bell, "The interactions of intensity, frequency and duration of exercise training in altering cardiorespiratory fitness," *Sports Medicine*, vol. 3, no. 5, pp. 346–356, 1986.
- [2] S. D. Anton, M. G. Perri, J. Riley III, W. F. Kanasky jr., J. R. Rodrigue, S. F. Sears, and A. D. Martin, "Differential predictors of adherence in exercise programs with moderate versus higher levels of intensity and frequency," *Journal of Sport & Exercise Psychology*, vol. 27, no. 2, pp. 171–187, 2005.
- [3] R. C. Goode, R. Mertens, S. Shaiman, and J. Mertens, "Voice, breathing, and the control of exercise intensity," in *Advances in Modeling and Control of Ventilation*, R. L. Hughson, D. A. Cunningham, and J. Duffin, Eds. New York: Plenum Press, 1998, pp. 223–229.
- [4] R. C. Goode, "A personal insight into the origin of the Talk Test," *Health and Fitness Journal of Canada*, vol. 1, no. 1, pp. 5–8, 2008.
- [5] M. Dehart-Beverley, C. Foster, J. P. Porcari, D. C. W. Fater, and R. P. Mikat, "Relationship between the talk test and ventilatory threshold," *Clinical Exercise Physiology*, vol. 2, pp. 34–38, 2000.
- [6] C. Foster, J. P. Porcari, J. Anderson, M. Paulson, D. Smaczny, H. Webber, S. T. Doberstein, and B. Udermann, "The Talk Test as a marker of exercise training intensity," *Journal of Cardiopulmonary Rehabilitation and Prevention*, vol. 28, pp. 24–30, 2008.
- [7] R. Persinger, C. Foster, M. Gibson, D. C. W. Fater, and J. P. Porcari, "Consistency of the talk test for exercise prescription," *Medicine and Science in Sports and Exercise*, vol. 36, no. 9, pp. 1632–1636, 2004.
- [8] T. J. Quinn and B. A. Coons, "The talk test and its relationship with the ventilatory and lactate thresholds," *Journal of Sports Sciences*, vol. 29, no. 11, pp. 1175–1182, 2011.
- [9] K. W. Godin, T. Hasan, and J. H. L. Hansen, "Glottal waveform analysis of physical task stress speech," in *Proceedings of Interspeech*, 2012.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] —, "Analysis of the effects of physical task stress on the speech signal," *Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 3992–3998, 2011.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] G. Godin and R. J. Shephard, "A simple method to assess exercise behavior in the community," *Canadian Journal of Applied Sport Sciences*, vol. 10, no. 3, pp. 141–146, 1985.
- [18] G. A. V. Borg, "Psychophysical bases of perceived exertion," *Medicine and Science in Sports and Exercise*, vol. 14, no. 5, pp. 377–381, 1982.
- [19] C. J. Hardy and W. J. Rejeski, "Not what, but how one feels: The measurement of affect during exercise," *Journal of Sport & Exercise Psychology*, vol. 11, pp. 304–317, 1989.
- [20] J. O. Prochaska and C. C. DiClemente, "Stages and processes of self-change of smoking: toward an integrative model of change," *Journal of consulting and clinical psychology*, vol. 51, no. 3, p. 390, 1983.
- [21] C. L. Craig, A. L. Marshall, M. Sjostrom, A. E. Bauman, M. L. Booth, B. E. Ainsworth, M. Pratt, U. Ekelund, A. Yngve, J. F. Sallis, and P. Oja, "International physical activity questionnaire: 12-country reliability and validity," *Medicine & Science In Sports & Exercise*, vol. 35, pp. 1381–1395, 2003.
- [22] R. J. Bood, M. Nijsen, J. van der Kamp, and M. Roerdink, "The power of auditory-motor synchronization in sports: enhancing running performance by coupling cadence with the right beats," *PLOS One*, vol. 8, no. 8, p. e70758, 2013.
- [23] J. Karvonen and T. Vuorimaa, "Heart rate and exercise intensity during sports activities," *Sports Medicine*, vol. 5, no. 5, pp. 303–311, 1988.
- [24] Y. Ostchega, K. S. Porter, J. Hughes, C. F. Dillon, and T. Nwankwo, "Resting pulse rate reference data for children, adolescents, and adults: United States, 1999–2008," *National Health Statistics Reports (U.S. Department of Health and Human Services: Centers for Disease Control and Prevention)*, vol. 41, pp. 1–17, 2011.
- [25] G. F. Fletcher, P. Ades, P. A. Kligfield, R. Arena, G. J. Balady, V. A. Bittner, L. A. Coke, J. L. Fleg, D. E. Forman, T. C. Gerber, M. Gulati, K. Madan, J. Rhodes, P. D. Thompson, and M. A. Williams, "Exercise standards for testing and training: a scientific statement from the American Heart Association," *Circulation*, vol. 128, pp. 873–934, 2013.
- [26] D. P. Swain, "Moderate- or vigorous-intensity exercise: What should we prescribe?" *ACSM's Health & Fitness Journal*, vol. 10, no. 5, pp. 7–11, 2006.
- [27] F. Eyben, M. Wollmer, and B. Schuller, "openSMILE - The Munich versatile and fast open-source audio feature extractor," in *Proceedings of ACM Multimedia*, 2010, pp. 1459–1462.
- [28] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [29] J. Trouvain and K. P. Truong, "Prosodic characteristics of read speech before and after treadmill running," in *Proceedings of Interspeech*, 2015, to appear.