Acoustical Analysis of Speech Under Physical Stress in Relation to Physical Activities and Physical Literacy

Si-Ioi Ng[1], Rui-Si Ma[2], Tan Lee[1], Raymond Kim-Wai Sum[2]

[1]Department of Electronic Engineering, [2]Department of Sports Science and Physical Education
The Chinese University of Hong Kong
siioing@link.cuhk.edu.hk, tanlee@ee.cuhk.edu.hk

Abstract

Human speech production encompasses physiological processes that naturally react to physic stress. Stress caused by physical activity (PA), e.g., running, may lead to significant changes in a person's speech. The major changes are related to the aspects of fundamental frequency level, speaking rate, pause pattern, and breathiness. The extent of changes depends presumably on the physical fitness and well-being of the person, as well as the intensity of PA. The general wellness of a person is further related to his/her physical literacy (PL), which refers to a holistic description of engagement in PA. This paper presents the development of a Cantonese speech database that contains audio recordings of speech before and after physical exercises of different intensity levels. The corpus design and data collection process are described. Preliminary results of acoustical analysis are presented to illustrate the impact of PA on F0 level, F0 range, speaking and articulation rate, and time duration of pauses. It is also noted that the effect of PA is correlated to some of the PA and PL measures.

Index Terms: acoustical analysis of speech, physical activity, physical literacy, speech corpus

1. Introduction

Speech, as one of the most natural and common media of human communication, carries a rich array of information that extends far beyond the verbal message being expressed. From an utterance, the speaker's gender, age, dialectal background, emotional status and personality could be identified by human listeners. Part of the paralinguistic information in speech carries physiological and healthy condition of a speaker. It is feasible to capture the natural speech sound in the form of acoustic signals, and extract the speaker's information via signal processing techniques and statistical modeling [1, 2]. Acoustic and linguistic analysis of speech signals are effective means of detecting and quantifying disorders, diseases, and other changes in human body.

Speech production is an effortful activity that involves the brain and the respiratory system. It is closely coupled with breathing. While talking, inhalations are shortened and accelerated, while exhalations are prolonged and slowed down [3]. The timing of inhalations depends on the control of habitual respiratory cycles [4], and is mostly determined in accordance with the linguistic structure of speech [5]. If the functioning of a person's respiratory system does not function properly, evident changes in terms of perceptual and acoustic properties of the produced speech can be observed. In our daily lives, the pattern of breathing is often altered by physical activity (PA). PA refers to any body movement that increases the energy expenditure from resting state and reaches exercise state. It may take place during running, swimming or weight lifting, etc. [6].

In exercise states, PA of different intensities pose different levels of physical stress to human body. Respiratory rate increases accordingly with the stress to provide adequate oxygen supply to human body. The effect of PA on breathing is expected to be manifested in the produced speech. In [7], it was shown that perceivable prosodic characteristic of speech, such as pause pattern, shifting of fundamental frequency (F0) and formants might be consistent indicators of physical stress. Detection of exercise intensity in PA was successfully performed using paralinguisic features of speech [8].

The well-being of a person is related to participation in PA. For older adults, regular physical exercises was found to improve subjective well-being measures, i.e. happiness, satisfaction with life [9]. Adolescents who reported greater engagement in PA tend to experience less stress and depression [10], and have higher self-perceived physical competence [11], which is one's ability to move with competence in a wide variety of activities [12]. Recently, lights are shed on the concept of physical literacy (PL), which extends beyond the participation of PA [13]. A physically literate individual should understand the value and responsibility, and is confident, motivated, skilful in participating in PA throughout the lifetime. Numerous instruments for assessing PL were proposed and validated [14]. Positive correlation between perceived PL and PA was reported in a large population study [15].

PL is correlated with PA, while PA determines the well-being and competence of a person, and affects the characteristics of breathing and speaking. Intuitively, a physically literate person with strong competence tends to have good respiratory control. This hypothetically implies less noticeable effect of physical stress, in terms of the difference in speech characteristics between resting and exercise states. The present study starts with the collection of speech data in PA of different intensities. The speech database is intended to support investigation on the relationship between PA/PL and speech; facilitate statistical analysis of speech features and physiological measures; provide training data for classification of PA intensity experienced by the subjects. The design of PA follows the Canadian Assessment of Physical Literacy-2 (CAPL-2) test protocol [16]. For general population, speech production may become laborious and difficult during and after PA. This results in breathy, intermittent and unstable voice as compared to that produced in resting state. A fundamental hypothesis of this research is that the influence of PA on speech production can be revealed by characteristic changes of acoustic speech signals. F0 and temporal characteristics are extracted from the speech signals to reflect the difference of speech between resting and exercise states. The correlation between the difference, and the measures of PA and PL, is investigated.

2. Design of Speech Corpus

2.1. About Cantonese

Cantonese is a major Chinese dialect widely spoken in Hong Kong, Macau, Guangdong and Guangxi Provinces of Mainland China as well as overseas Chinese communities. Each Chinese character is pronounced as a single syllable carrying a lexical tone. The lexical tone (T) is determined by the temporal pattern of fundamental frequency (F0) across the voiced portion of a syllable. Cantonese has six tones, including T1 (high-level), T2 (high-rising), T3 (mid-level), T4 (low-falling), T5 (low-rising) and T6 (low-level) [17]. As shown in Figure 1, a legitimate Cantonese syllable can be divided into an onset and a rime. Onset is a consonant, while rime can contain either a nucleus or a nucleus followed by a coda. Nucleus can be a vowel or a diphthong, and coda can be a final consonant. Both onset and coda are optional in a Cantonese syllable. In the inventory of Cantonese phonemes, there are a total of 19 initial consonants, 11 vowels, 11 diphthongs and 6 final consonants [17, 18].

A Cantonese Syllable			
Onset	**Rime**		**Tone**
*Initial consonant	Nucleus (Vowel / Diphthong)	*Coda (Final consonant)	1-6

Figure 1: *Structure of Cantonese syllable. '*' denotes optional.*

2.2. Design of speech material

The speech materials consist of three parts. The first part contains two text passages. Passage A is the story of 'the North Wind and the Sun'. It comprises 142 syllables. The story content was originated from [19]. It was used in the studies of Cantonese perceptual evaluation of voice [20] and speech rhythm [21]. Passage B is composed by following the corpus designed in [8] to elicit designated types of syntactic breaks and investigate how placing of pauses interacts with breathing. The second part of speech materials contains three sustained vowels, namely the low-central vowel /aː/, high-front vowel /iː/ and mid-back vowel /ɔː/. Sustained vowels are most useful to reflect voice characteristics in stable articulation states. The third part contains digit sequences "3947", "3948", and "0526". The digits cover all of the 6 Cantonese tones. The digit sequences are intended for reliably measuring F0 level and F0 range of a Cantonese speaker.

2.3. Recording

Speech recording took place in an outdoor sports ground of basketball court size. Under the instruction of qualified fitness trainers, each subject was required to go through three different parts of PA from CAPL-2. The first part was the Canadian Agility and Movement Skill Assessment (CAMSA) [22]. It measured the skill execution and completion time for a series of dynamic movement to assess the subject's movement capabilities. The second exercise was plank, a reliable measure of isometric core muscular endurance [23]. The third part was the Progressive Aerobic Cardiovascular Endurance Run (PACER). It assessed aerobic endurance and determined the maximal oxygen consumption (VO$_2$ max) of the subjects [24]. The physical intensity was presumably increasing from CAMSA to PACER. Each subject was required to wear heart rate monitors. The av-erage heart rate during speaking was regarded as an objective indicator of the PA intensity [25].

Speech recordings were made before and in-between the physical exercises. Since the recordings shared the same linguistic content, they were most suitable for contrastive investigation on the influence of PA on paralinguistic factors of speech. A digital recorder (TASCOM DR-44WL) was located at 20-50 centimeters in front of the participant's mouth. The gain of the recorder was adjusted to maintain the background noise level below -30 dB (relative to the maximum input level). A two-channel stereo recording was made for each subject at a sampling rate of 44.1 kHz. Each recording was later down-sampled to 16 kHz with mono-channel, and was further divided into three sections corresponding to the three parts of text materials.

2.4. Subjects and measurement of PA and PL

30 native Cantonese speakers (23 males, 7 females), aged between 18 to 21, were recruited by convenience sampling. All speakers were university students majored in sports science. The Physical Activity Readiness Questionnaires (PAR-Q) were distributed to all subjects [26]. All of them indicated "NO" to all questions, and became eligible to participate in the physical exercises. Informed consents were obtained prior to the recording session. The subjects were required to complete the questionnaire of the 9-item Perceived Physical Literacy Instrument (PPLI) [27]. The responses were rated on 5-point Likert scale ranging from "strongly disagree" to "strongly agree". The 9 items, denoted as PL1-9, considered the "knowledge and understanding", "self-expression and communication with others", and "sense of self and self-confidence" of a person. The subjects were also asked to report their daily routine on PA in sedentary, light, moderate and vigorous intensities. Metabolic equivalent of task (MET), which represents the amount of PA in daily life, was measured based on the reported routine [28]. One MET is defined as an oxygen uptake of 3.5 ml per kilogram per minute while a person is sitting at rest. MET minutes per week was calculated for each level of PA intensity. The subject's performance in terms of time and skill scores in CAMSA, total elapsed time in plank, and number of round finished in PACER, was measured.

3. Speech Feature Extraction

3.1. Extraction of F0

F0 in a person's speech carries linguistic, paralinguistic and physiological information. F0 was extracted using the auto-correlation (AC) function via Praat [29, 30]. Reliability of F0 estimation is sensitive to the parameter settings, specifically, the floor ($F0_{floor}$) and ceiling values ($F0_{floor}$). Extracting the F0 for different speakers within the same floor and ceiling values may not be optimal. Male voice typically has lower F0 than female. F0 range is also related to language and emotional status of the speaker [31]. We expect that physical stress would cause large alteration of F0. A two-stage strategy of F0 extraction was adapted to improve the accuracy of F0 estimation [32]. In the first-pass extraction, the floor and ceiling values were initialized to 75 Hz and 600 Hz. Subsequently the 35-th quantile (q_{35}) and the 65-th quantile (q_{65}) of the F0 values were obtained. Before the second-pass processing, the floor and ceiling values were updated as $q_{35} * 0.72 - 10$ and $q_{65} * 1.9 + 10$. These equations were manually tuned to better correlate between automatic and manual estimation of F0 [33]. The F0 contour was re-estimated

Table 1: *Mean (standard deviation) of speech features.*

Feature	Measure	Rest	CAMSA	Plank	PACER
F0 (utterance)	Hz	143 (38)	147 (41)	144 (40)	159 (41)
	DQ (%)	15.58 (2.90)	16.12 (3.10)	15.95 (3.34)	16.48 (2.65)
F0 (Tone 1)	Hz	162 (45)	166 (48)	163 (46)	178 (46)
	DQ (%)	14.29 (3.05)	15.01 (3.15)	15.22 (3.35)	15.63 (3.06)
F0 (Tone 3)	Hz	142 (39)	144 (41)	143 (40)	157 (40)
	DQ (%)	11.45 (2.73)	12.41 (2.90)	11.79 (2.70)	12.69 (2.3)
F0 (Tone 6)	Hz	133 (37)	137 (40)	135 (39)	147 (39)
	DQ (%)	10.95 (2.55)	11.30 (2.31)	12.18 (3.28)	12.60 (2.47)
Speaking rate	spm	246 (34)	267 (38)	267 (34)	252 (37)
Articulation rate		306 (36)	341 (41)	340 (42)	366 (36)
Speech to non-speech ratio	Ratio	4.33 (1.25)	3.90 (1.25)	4.02 (1.35)	2.35 (0.73)

Table 2: *Difference of speech features between resting and exercise states.*

Feature	Measure of Difference	Rest-CAMSA	Rest-Plank	Rest-PACER
F0 (utterance)	Cents	39 (84)	11 (81)	181 (141)
	DQ ratio (%)	104.30 (13.40)	102 (12.37)	107.15 (14.98)
F0 (Tone 1)	Cents	37 (99)	0 (103)	168 (153)
	DQ ratio (%)	105.96 (13.35)	107.10 (13.55)	110.99 (17.44)
F0 (Tone 3)	Cents	29 (95)	10 (86)	181 (159)
	DQ ratio (%)	112.55 (30.89)	105.63 (21.49)	115.12 (26.88)
F0 (Tone 6)	Cents	33 (80)	16 (77)	172 (128)
	DQ ratio (%)	105.52 (20.13)	112.46 (21.77)	119.23 (29.09)
Speaking rate	spm difference	21 (29)	21 (33)	6 (37)
Articulation rate		35 (27)	34 (33)	61 (34)
Speech to non-speech ratio	Ratio (%)	91.76 (24.12)	94.59 (24.40)	56.79 (19.49)

as the final output. The utterance-level F0 was obtained by averaging the F0 values of the vowel segments. Average F0 values of tone 1 (T1), tone 3 (T3) and tone 6 (T6) in each utterance were calculated. They represented the high, medium and low level of F0 in the utterance.

We are interested in how the F0 level changes resting state to exercise states. The change of F0 was measured in terms of cents (¢), defined as,

$$\text{¢} = 1200 * \log_2 \frac{F0_e}{F0_r}$$

The variation of the F0 contour within an utterance is represented by dynamism quotient (DQ), which is given by dividing the standard deviation by mean [31, 34]. Higher DQ indicates larger variation of F0. The change of DQ from resting state to exercise states was measured as DQ_e/DQ_r.

3.2. Timing characteristics

Increased respiratory rate is a typical response to physical stress. To understand how speech properties change with intensity level of physical stress, temporal characteristics of speech including speaking rate, articulation rate and speech-to-pause time ratio were examined. Speaking rate is defined as the number of elicited syllables per minute (spm). Articulation rate has the same definition as speaking rate, except that the time intervals occupied by non-speech segments are excluded. Speech-to-pause time ratio is the total time of speech divided by the total duration of silence. Beside syntactic pauses, breathing pauses are frequently found in speech under physical stress. The speech-to-pause ratio is expected to carry the information of breathing strategy in speech under physical stress. The change of speech-to-pause time ratio (tr) from resting state to exercise states is expressed by tr_e/tr_r. For speaking rate (sr) and articulation rate (ar), the difference between the two states are calculated as by $sr_e - sr_r$, and $ar_e - ar_r$, respectively.

To automatically locate the time boundaries of phones and syllables, forced alignment was applied to the utterances with the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) tri-phone acoustic model. We were focused on the passage of "The North Wind and The Sun". The acoustic model was trained on a large-vocabulary database of adult speech called CUSENT. The corpus contained about 20 hours

of speech recorded from 76 adult speakers [18]. The acoustic features for GMM-HMM training consisted of 13-dimensional Mel-frequency cepstral coefficients (MFCC) and their first- and second-order derivatives extracted at every 0.010 second. Feature space Maximum Likelihood Linear Regression (fMLLR) was applied to the acoustic features[35]. Forced alignment and acoustic model training were implemented by the Kaldi speech recognition toolkit [36].

4. Results and Discussion

4.1. Analysis of speech features

Table 1 reports the mean and the standard deviation of F0, speaking rate, articulation rate and speech-to-pause time ratio in resting and exercise states. The difference of feature values between resting and exercise states are shown in Table 2. Given limited amount of speech recordings, the gender effect is not considered in the analysis. ANOVA and the two-tailed t-test are applied. A significant threshold of 0.05 is assumed.

The results of ANOVA indicate the changes of F0 levels in different exercises are significantly different ($p < 0.001$). An increase of over 150 cents is observed in PACER. The DQs in CAMSA, plank and PACER do not differ significantly. For each type of exercise, significant difference is found in the DQs of T1, T3, and T6 ($p < 0.001$). Physical stress exhibits different extents of influence to different tones.

The two-tailed t-test shows that the speaking rate, articulation rate and speech-to-pause time ratio of exercise states are significantly different from that in resting state ($p < 0.01$). ANOVA indicates the increase of speaking rate and articulation rate, and the decrease of speech-to-pause time ratio, are significantly different among CAMSA, plank and PACER ($p < 0.001$). The temporal characteristics of speech in PACER are further compared to resting state via the t-test. A notable increase in articulation rate is observed ($p < 0.001$), but not in speaking rate. Subjects generally requires more frequent and deeper breathing in high-intensity exercise. The speech-to-pause time ratio is lower in the exercise states ($p < 0.001$). These observations suggest that the subjects tend to talk faster and use longer time to breath in PACER. The above results can help with the design of acoustic feature sets that distinguish be-

Table 3: *Correlation between the acoustic changes in exercise states and the PA/PL measures.*

Feature	Measure	Rest-CAMSA		Rest-Plank		Rest-PACER	
		PA/PL Measures	PCC	PA/PL Measures	PCC	PA/PL Measures	PCC
F0 (utterance)	Cents	-	-	-	-	-	-
	DQ ratio	PL1 Pacer round CAMSA skill score	-0.456 -0.399 -0.501	PL3 PL9 MET (total)	-0.377 -0.377 0.379	MET (total)	0.420
Speaking rate	spm difference	PL1	-0.410	PL1	-0.560	PL1 CAMSA time score	-0.540 -0.376
Articulation rate		PL1	-0.403	PL1	-0.466	PL1	-0.465
Speech-to-pause time ratio	Ratio	PL2 Pacer round	-0.388 0.374	PL1 PL2	-0.513 -0.448	PPLI CAMSA time score	-0.376 -0.370

tween PA of different intensities.

4.2. Correlation with PA and PL measures

We attempt to correlate the measures of PA/PL with the changes of acoustical/temporal characteristics of speech under physical stress. The daily exercise routine of the subjects and their performance in different exercises are provided as the PA/PL measures. The correlation is measured by the Pearson's correlation coefficient (PCC). PA/PL measures that are significantly correlated to the changes of speech features are reported in Table 3.

The changes of speaking rate and articulation rate in exercise states are correlated mostly to PL1 and PL2, which correspond to 'I am physically fit in accordance with my age' and 'I have a positive attitude and interest in sports' respectively. As subjects are confident with their physical fitness and enthusiastic in sports, the temporal characteristics of speech are less altered during exercises. The self-perceived physical fitness in youngsters is a reliable predictor of actual fitness [37]. Subjects who rated themselves physically fit could be in fact in good shape, and better control their breathing during exercises. This leads to similar speaking and articulation rates between resting and exercise states. For speech-to-pause time ratio, subjects who perform well in CAMSA, i.e. high CAMSA time score, have less changes in the time ratio in PACER. Subjects who endured more rounds in PACER enlarges the speech-to-pause time ratio more in CAMSA. A negative correlation is found in between the PPLI score and the speech-to-time ratio in PACER. The overall descriptor of a person's PL is connected to characteristic changes of speech.

PA/PL measures do not significantly correlate to the shift of F0 in exercise states ($p > 0.05$). Changes of DQ have negative correlation with PL1, PL3 and PL9. The descriptions of PL3 and PL 9 are 'I appreciate myself or others doing sports' and 'I am aware of the benefits of sports related to health'. Both PL3 and PL9 concern the knowledge and understanding of participating in PA. A better understanding towards PA, and better performance in PA, i.e. higher CAMSA skill score and more PACER rounds, lead to less deviation of F0 contour during exercises. The frequency of engagement in PA also plays a significant role in the changes of acoustic features in exercise states. Higher total MET score corresponds to larger DQs.

5. Conclusion

We present the collection of speech data for investigating the relation between physical stress and speech. Statistical analysis has shown that there are different extent of changes in fundamental frequency, speaking and articulation rate, and speech-to-pause time ratio in different physical activities. Variations of F0 in different Cantonese tones are different in the resting and exercise states. Correlation analysis indicates that part of the measures of physical activities and physical literacy, and daily habits of participating in physical activities, are associated with the changes in acoustic/ temporal characteristics of speech. To extend the present study, more speech data will be collected from subjects of diverse background. In-depth analysis of the 6 Cantonese tones under physical stress, investigation on the interaction between pauses and breathing, classification of physical stress, and automatic assessment of physical literacy will be among the future works.

6. Acknowledgements

7. References

[1] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2010.

[2] G. Mohammadi, A. Vinciarelli, and M. Mortillaro, "The voice of personality: Mapping nonverbal vocal behavior into trait attributions," in *Proc. of SSPW*, 2010, pp. 17–20.

[3] B. Conrad and P. Schoenle, "Speech and respiration," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 226, no. 4, pp. 251–268, 1979.

[4] J. Trouvain and K. P. Truong, "Prosodic characteristics of read speech before and after treadmill running," in *Proc. of Interspeech*, 2015.

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

[6] B. Kumar, R. Robinson, and S. Till, "Physical activity and health in adolescence," *Clinical Medicine*, vol. 15, no. 3, p. 267, 2015.

[7] K. W. Godin and J. H. Hansen, "Physical task stress and speaker variability in voice quality," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, pp. 1–13, 2015.

[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 *Proc. of Interspeech*, 2015, pp. 3705–3709.

[9] E. McAuley, B. Blissmer, D. X. Marquez, G. J. Jerome, A. F. Kramer, and J. Katula, "Social relations, physical activity, and well-being in older adults," *Preventive medicine*, vol. 31, no. 5, pp. 608–617, 2000.

[10] R. Norris, D. Carroll, and R. Cochrane, "The effects of physical activity and exercise training on psychological stress and well-being in an adolescent population," *Journal of psychosomatic research*, vol. 36, no. 1, pp. 55–65, 1992.

[11] J. Timo, Y.-P. Sami, W. Anthony, and L. Jarmo, "Perceived physical competence towards physical activity, and motivation and enjoyment in physical education as longitudinal predictors of adolescents' self-reported physical activity," *Journal of science and medicine in sport*, vol. 19, no. 9, pp. 750–754, 2016.

[12] L. C. Edwards, A. S. Bryant, R. J. Keegan, K. Morgan, and A. M. Jones, "Definitions, foundations and associations of physical literacy: a systematic review," *Sports medicine*, vol. 47, no. 1, pp. 113–126, 2017.

[13] M. Whitehead, *Physical literacy: Throughout the lifecourse*. Routledge, 2010.

[14] M. Tremblay and M. Lloyd, "Physical literacy measurement: The missing piece," *Physical and health education journal*, vol. 76, no. 1, pp. 26–30, 2010.

[15] S. M. Choi, R. K. W. Sum, E. F. L. Leung, and R. S. K. Ng, "Relationship between perceived physical literacy and physical activity levels among hong kong adolescents," *PLoS One*, vol. 13, no. 8, p. e0203105, 2018.

[16] P. E. Longmuir, K. E. Gunnell, J. D. Barnes, K. Belanger, G. Leduc, S. J. Woodruff, and M. S. Tremblay, "Canadian assessment of physical literacy second edition: a streamlined assessment of the capacity for physical activity among children 8 to 12 years of age," *BMC public health*, vol. 18, no. 2, pp. 1–12, 2018.

[17] R. S. Bauer and P. K. Benedict, *Modern cantonese phonology*. Walter de Gruyter, 2011, vol. 102.

[18] T. Lee, W. K. Lo, P. Ching, and H. Meng, "Spoken language resources for cantonese speech processing," *Speech Communication*, vol. 36, no. 3-4, pp. 327–342, 2002.

[19] E. Zee, "Chinese (hong kong cantonese)," *Journal of the International Phonetic Association*, vol. 21, no. 1, pp. 46–48, 1991.

[20] T. Law, K. Lee, J. Lam, A. C. van Hasselt, and M. C.-F. Tong, "The construction of the cantonese perceptual evaluation of voice (canpev): the content validation process," in *Proc. of VOICE*, 2010, p. 159.

[21] P. P. K. Mok and V. Dellwo, "Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English," in *Proc. Speech Prosody 2008*, 2008, pp. 423–426.

[22] P. E. Longmuir, C. Boyer, M. Lloyd, M. M. Borghese, E. Knight, T. J. Saunders, E. Boiarskaia, W. Zhu, and M. S. Tremblay, "Canadian agility and movement skill assessment (camsa): Validity, objectivity, and reliability evidence for children 8–12 years of age," *Journal of sport and health science*, vol. 6, no. 2, pp. 231–240, 2017.

[23] K. L. Schellenberg, J. M. Lang, K. M. Chan, and R. S. Burnham, "A clinical tool for office assessment of lumbar spine stabilization endurance: prone and supine bridge maneuvers," *American journal of physical medicine & rehabilitation*, vol. 86, no. 5, pp. 380–386, 2007.

[24] S. N. Scott, D. L. Thompson, and D. P. Coe, "The ability of the pacer to elicit peak exercise response in youth [corrected]." *Medicine and science in sports and exercise*, vol. 45, no. 6, pp. 1139–1143, 2013.

[25] J. Karvonen and T. Vuorimaa, "Heart rate and exercise intensity during sports activities," *Sports medicine*, vol. 5, no. 5, pp. 303–311, 1988.

[26] R. J. Shephard, "Par-q, canadian home fitness test and exercise screening alternatives," *Sports medicine*, vol. 5, no. 3, pp. 185–195, 1988.

[27] R. K. Sum, C.-F. Cheng, T. Wallhead, C.-C. Kuo, F.-J. Wang, and S.-M. Choi, "Perceived physical literacy instrument for adolescents: A further validation of ppli," *Journal of Exercise Science & Fitness*, vol. 16, no. 1, pp. 26–31, 2018.

[28] M. Fan, J. Lyu, and P. He, "Chinese guidelines for data processing and analysis concerning the international physical activity questionnaire," *Chinese journal of Epidemiology*, vol. 35, no. 8, pp. 961–964, 2014.

[29] Y. Jadoul, B. Thompson, and B. De Boer, "Introducing parselmouth: A python interface to praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.

[30] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the institute of phonetic sciences*, vol. 17, no. 1193, 1993, pp. 97–110.

[31] H. Traunmüller and A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults," *Manuscript, Department of Linguistics, University of Stockholm*, 1995.

[32] K. Evanini and C. Lai, "The importance of optimal parameter setting for pitch extraction." *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 2291–2291, 2010.

[33] C. De Looze and S. Rauzy, "Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration," in *Proc. of Interspeech*, 2009.

[34] R. Hincks, "Processing the prosody of oral presentations," in *InSTIL/ICALL Symposium 2004*, 2004.

[35] M. J. Gales, "Maximum likelihood linear transformations for hmm-based speech recognition," *Computer speech & language*, vol. 12, no. 2, pp. 75–98, 1998.

[36] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," in *Proc. of ASRU*, no. EPFL-CONF-192584, 2011.

[37] J. L. Germain and H. A. Hausenblas, "The relationship between perceived and actual physical fitness: a meta-analysis," *Journal of Applied Sport Psychology*, vol. 18, no. 4, pp. 283–296, 2006.