

主題：身體素養與健康茁壯成長 (Summer 03/22)

探討不同運動狀態下語音特徵的變化規律

馬睿思先生 香港中文大學體育運動科學系博士生
沈劍威教授 中國香港體適能總會主席
香港中文大學教育學院副院長(本科課程)
香港中文大學體育運動科學系副教授兼課程主任

摘要

本研究旨在探索不同運動強度下語音特徵的變化。所有參與者均來自香港中文大學，採用方便抽樣的橫斷面研究設計。本研究共招募了 31 名參與者，參與者被要求佩戴心率檢測器完成三個不同的練習，然後在每個練習結束時大聲朗讀一段話。從語音數據中提取基頻，採用單因素重複方差分析來研究每種運動狀態下的語音特徵差異。基頻及其範圍與其他狀態有明顯的不同，整個段落的語音持續時間也隨著運動強度的增加而增加，停頓的次數和持續時間在不同的運動狀態下更具有辨別力，與只有在劇烈運動下才有差異的基頻相比，停頓的特徵也可以區分中等強度和靜止狀態。綜上，本研究探索了不同運動狀態下語音特徵的變化模式，為進一步探討語音特徵與運動狀態之間的聯繫提供了可靠的參數選擇。

前言

語音是人類交流的最自然、最有效的方式之一^[1]。人類的聲音不僅包含有意義的資訊，而且還包含可識別的特徵，如性別、年齡、口音和情緒狀態。一些副語言資訊帶有說話者的心理和身體健康情況。這表明，以聲學信號的形式捕捉自然語音，並通過信號處理技術和統計建模提取有關說話人的各種資訊是可行的^[2, 3]。因此，語音技術的研究人員正在努力開發能夠理解語言資訊並能與人交流的系統^[4-14]。近年來，自然語言處理領域的學者已經開始利用語言學特徵對高低運動強度進行自動分類^[8-10, 15]。本研究採用加拿大身體素養評估工具第二版 (CAPL2) 作為設定運動強度標準的方案^[16-19]。這種戶外運動提供的聲學數據更接近於現實生活中的運動場景。因此，本研究利用這個資料庫來探索不同運動強度之間的語音特徵的變化。

方法

研究設計

所有運動都是在中大的戶外運動場進行的。在運動測試前，我們收集了 31 名參與者的人口統計資料和其他問卷。參與者被要求佩戴心率測試儀。然後，參與者開始大聲地朗讀閱讀材料。同時，訓練有素的幫助者記錄下講話和心率。訓練有素的幫助者指導參與者閱讀記錄講話。在休息時大聲朗讀一次閱讀材料並記錄心率。然後，受試者開始進行 CAPL2 身體能力測試，在每次運動後，都要大聲朗讀相同的閱讀材料。同時還記錄了說話和心率數據。說話時的平均心率被當作運動強度的客觀指標^[20]。

測量和閱讀材料

作為最常用的語音特徵，基頻 (F0) 承載著一個人的語言學、副語言學和生理學資訊。在這項研究中，所有參與者的 F0 都是用 openSMILE3.0 提取的^[21]。我們還在錄音中標記了說話者停頓的次數和時間，以直觀地反映運動壓力對說話的影響。閱讀材料為著名的語音研究文本《北風與太陽》(粵語版)^[22]。該故事內容被廣泛用於廣東人對語音和語速的感知評價研究中^[23, 24]。

(閱讀材料: 北風和太陽)

有一日，北風同太陽喺到嗰交究竟邊個犀利啲。
呢個時候，咁啱有個著住件大褸嘅人經過。
於是，佢哋決定邊個可以令到嗰個人剝咗佢件褸嘅話，邊個就贏。
北風出盡力咁吹，但係越吹得大力，路人反而將件褸包得越緊。
但係太陽出馬曬咗一陣，嗰個人就流晒大汗，即刻將件褸剝咗落嚟。
北風唯有承認太陽比佢更加犀利。

研究結果

本研究採用 IBM SPSS 26 進行單因素重複方差分析 (RM-ANOVA) 來研究各運動狀態下的語音特徵差異。共有 31 名母語人士 (24 名男性和 7 名女性) 參加了研究。學生的年齡在 18 至 21 歲之間 (總年齡: $M=18.97, SD=.91$; 男性年: $M=19.04, SD=0.96$; 女性年齡: $M=18.71, SD=.76$)。RM-ANOVA 在語音持續時間的平均值之間有統計學上的顯著差異, $F(3, 90) = 7.89, p=.001, \eta^2=.21$ 。經 Bonferroni 調整的多重檢驗的配對比較進一步顯示, 休息狀態下的語言持續時間 ($M=19.82, SD=3.21$) 明顯長於加速器狀態 ($M=17.35, SD=2.32$), $p=.003, 95\%CI [.68, 4.26], d=.70$ 。折返跑的講話時間也明顯短於 CAMSA 狀態 ($M=18.56, SD=3.13$), $p=.014, 95\%CI [-2.26, -.18], d=.60$ 。和平板支撐狀態 ($M=18.23, SD=2.74$), $p=.049, 95\%CI [-1.76, -.01], d=.51$ 。除平板支撐外, 其他三個運動狀態的言語持續時間沒有明顯差異 (圖 2)。同時, 每個狀態下的 F0 平均值, F0 的 25% 至 75% 區間, 停頓時長與停頓次數也用 RM-方差分析進行了檢驗 (如圖 2)。

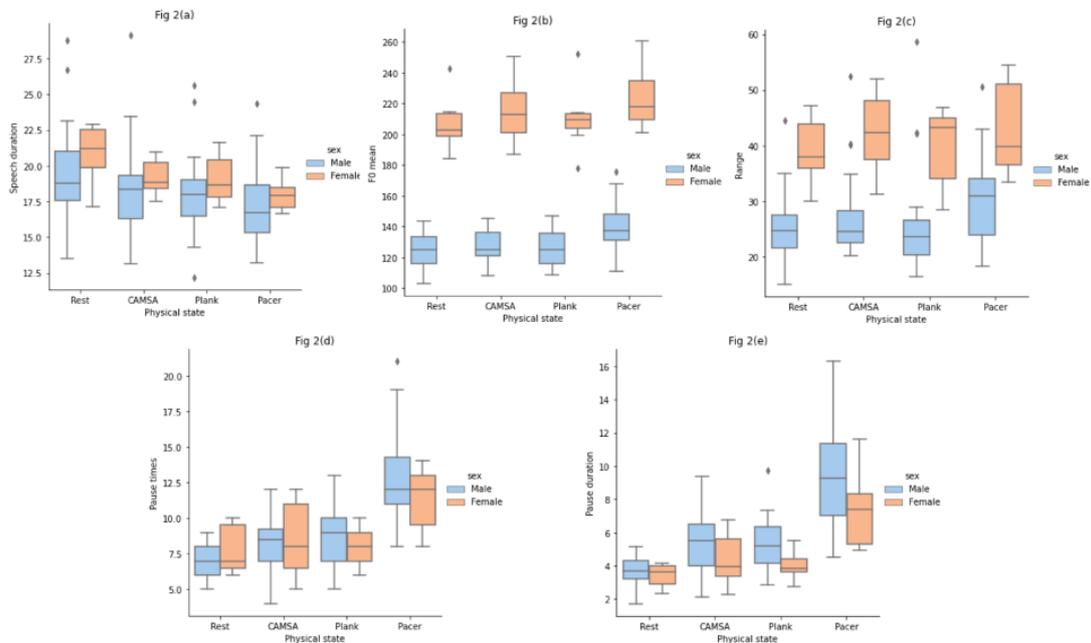


圖 2 · 不同運動強度下的語音特徵比較

討論

本研究的結果表明，特定的語音特徵在不同的運動狀態下有一定的模式變化。目前的研究設計包括兩個中等強度的運動和一個高強度的運動。結果顯示，這兩種運動之間以及與靜止狀態下的語音特徵存在著明顯的差異。此外，它們之間存在著內在的線性聯繫。劇烈的強度鍛煉會對說話的持續時間造成更明顯的影響。在逐漸達到身體極限的過程中，人們需要吸入更多的氧氣來保持身體的運轉。然而，由於說話和呼吸有著共同的系統，在急需呼吸的情況下，身體會本能地縮短說話的時間，以便順利地進行呼吸。不同運動強度下說話的停頓時間也有明顯差異。如前所述，在身體壓力下說話，由於呼吸頻率高，會增加停頓的次數。不僅如此，呼吸的時間也增加了。這也意味著，隨著運動強度的不斷增加，鍛煉者在說話時需要更多的時間來呼吸。換句話說，需要更多的氧氣攝入來繼續維持發音。停頓時間可以補充停頓次數，作為未來研究的一個參數選擇。因為停頓的時間因人而異，無論停頓選擇在句子的中間還是結尾，總的停頓長度都沒有太大變化。然而，與停頓次數相同，講話中的停頓時間雖然能很好地區分不同強度的練習狀態，但仍不能區分相同強度的不同練習。

參考文獻

1. Docio-Fernandez L, García Mateo C. Speech Production. In: Li SZ, Jain AK, eds. Encyclopedia of Biometrics. Boston, MA: Springer US; 2015:1493-1498. doi:10.1007/978-1-4899-7488-4_199
2. Dehak N, Kenny PJ, Dehak R, Dumouchel P, Ouellet P. Front-end factor analysis for speaker verification. IEEE Trans Audio, Speech Lang Process. 2011;19(4):788-798. doi:10.1109/TASL.2010.2064307
3. Mohammadi G, Vinciarelli A, Mortillaro M. Mapping nonverbal vocal behavior into trait attributions. Proc 2nd Int Work Soc signal Process. 2010:17-20.
4. B C, Schoenle P. Speech and respiration. Arch Psychiatr Nervenkr. 1979;226(4):251-268.
5. Amélie R-C, Fuchs S. The interplay of linguistic structure and breathing in German spontaneous speech. Interspeech 2013-14th Annu Conf Int Speech Commun Assoc. 2013:1228.
6. Jürgen T, Truong KP. Prosodic characteristics of read speech before and after treadmill running. Sixt Annu Conf Int Speech Commun Assoc. 2015.
7. World Health Organization. Physical activity.
8. Truong KP, Nieuwenhuys A, Beek P, Evers V. A database for analysis of speech under physical stress: Detection of exercise intensity while running and talking. Proc Annu Conf Int Speech Commun Assoc INTERSPEECH. 2015;2015-Janua:3705-3709.
9. Godin KW, Hasan T, Hansen JHL. Glottal waveform analysis of physical task stress speech. Heal Fit J Canada. 2008;1(1):5-8.
10. Schuller B, Friedmann F, Eyben F. Automatic recognition of physiological parameters in the human voice: Heart rate and skin conductance. ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc. 2013:7219-7223. doi:10.1109/ICASSP.2013.6639064
11. Godin KW, Hansen JHL. Analysis and perception of speech under physical task stress. Proc Annu Conf Int Speech Commun Assoc INTERSPEECH. 2008:1674-1677.
12. Godin KW, Hansen JHL. Analysis of the effects of physical task stress on the speech signal. J Acoust Soc Am. 2011;130(6):3992-3998. doi:10.1121/1.3647301
13. Johannes B, Wittels P, Enne R, et al. Non-linear function model of voice pitch dependency on physical and mental load. Eur J Appl Physiol. 2007;101(3):267-276. doi:10.1007/s00421-007-0496-6
14. Baker SE. Ventilation and Speech Characteristics During Submaximal Aerobic Exercise. 2008;51(October):1203-1215.
15. Schuller B, Friedmann F, Eyben F. The munich biovoice corpus: Effects of physical exercising, heart rate, and skin conductance on human speech production. Proc 9th Int Conf Lang Resour Eval Lr 2014. 2014:1506-1510.
16. Longmuir PE, Gunnell KE, Barnes JD, et al. Canadian Assessment of Physical Literacy Second Edition: A streamlined assessment of the capacity for physical activity among children 8 to 12 years of age 11 Medical and Health Sciences 1117 Public Health and Health Services. BMC Public Health. 2018;18(Suppl 2). doi:10.1186/s12889-018-5902-y
17. Longmuir PE, Boyer C, Lloyd M, et al. Canadian Agility and Movement Skill Assessment (CAMSA): Validity, objectivity, and reliability evidence for children 8–12 years of age. J Sport Heal Sci. 2017;6(2):231-240. doi:10.1016/j.jshs.2015.11.004
18. Schellenberg KL, Lang JM, Chan KM, Burnham RS. A clinical tool for office assessment of lumbar spine stabilization endurance: Prone and supine bridge maneuvers. Am J Phys Med Rehabil. 2007;86(5):380-386. doi:10.1097/PHM.0b013e318032156a
19. Scott SN, Thompson DL, Coe DP. The ability of the PACER to elicit peak exercise response in the youth. Med Sci Sports Exerc. 2013;45(6):1139-1143. doi:10.1249/MSS.0b013e318281e4a8

20. Karvonen J, Vuorimaa T. Heart rate and exercise intensity during sports activities. *Sport Med.* 1988;5(5):303-311.
21. Eyben F, Wöllmer M, Schuller B. OpenSMILE - The Munich versatile and fast open-source audio feature extractor. *MM' 10 - Proc ACM Multimed 2010 Int Conf.* 2010:1459-1462. doi:10.1145/1873951.1874246
22. E Z. Chinese (Hong Kong Cantonese). *J Int Phon Assoc.* 1991;21(1):46-48.
23. Law T, Lee K, Lam JH, van Hasselt AC, Tong MCF. The construction of the Cantonese perceptual evaluation of voice (CanPEV): the content validation process. in *Proc of VOICE.* 2010:159.
24. Mok P, Dellwo V. Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proc Speech Prosody 2008.* 2008.
25. Rockwood NJ, Hayes AF, Pj M, et al. MLmed: An SPSS Macro for Multilevel Mediation and Conditional Process Analysis. 2017:19. www.afhayes.com.
26. Li SZ, Jain A, eds. Fundamental Frequency, Pitch, F0. In: *Encyclopedia of Biometrics.* Boston, MA: Springer US; 2009:592. doi:10.1007/978-0-387-73003-5_775