

躍度による調音運動の流暢性評価の試み

大上 舞歌 北村 達也

甲南大学知能情報学部 〒 658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: t-kitamu@konan-u.ac.jp

あらまし 躍度は加速度の時間変化であり, 運動の円滑性を評価する指標である. 躍度は対象の運動が円滑であるほど小さくなる. 本研究では, 発話の流暢性の評価に躍度を利用することを目的として, 発話時の開口高さの躍度を計測した. 実験参加者にとって初見の文章を読み上げた場合と数回の練習後に読み上げた場合の発話の様子を録画し, 開口高さを求め, その躍度 (jerk cost) を計算した. 練習後に躍度が減少することを予想したが, jerk cost に有意な変化は見られなかった.

キーワード 流暢性, 躍度, Jerk cost, 口唇運動

Preliminary Study on Evaluation of Fluency of Articulatory Motion by Jerk

Maika Ohue and Tatsuya Kitamura

Faculty of Intelligence and Informatics, Konan University 8-9-1, Okamoto, Higashinada, Kobe, Hyogo,
658-8501, Japan

E-mail: t-kitamu@konan-u.ac.jp

Abstract Jerk is the temporal differentiation of the acceleration and an evaluation index for fluidity of motion. The more fluid motion, the less jerk. This study measured the jerk of the height of the mouth opening during speech production to utilize it as a fluidity index of articulatory motion. We took movies of the face of speakers while they read text aloud at sight and they read it after practice and calculated the jerk cost. The jerk costs showed no significant difference between before and after the practice contrary to our expectation.

Keywords Fluency, Jerk, Jerk cost, Lip motion

1 はじめに

発話 (音声) の評価尺度のひとつに「流暢性」があり, 一般的には流暢な (流暢性の高い) 発話が望ましいとされている. 音声コミュニケーションにおける非流暢性には意味や役割があるものの [1, 2], 流暢性を定量的に評価できれば発話訓練などに利用できると考えられる. しかし, 外国語習得, 吃音や失語症などの障害, 言語発達における流暢性に関する研究は多いのに対して, 健常な成人の母語の流暢性への注目度はそれほど高くない [2].

運動の円滑性を評価する指標として躍度 (jerk) がある. これは加速度の変化に対応する. 躍度が小さい運動は, 加速度の変化が小さく滑らかな動きであるといえる. 従って, 躍度を最小化 (jerk minimization) する

ことによって滑らかで効率的な運動が実現できる ([6] など).

流暢な発話の調音運動は円滑であると考えられるので, 躍度によって発話の流暢性を評価できる可能性がある. そこで, 本研究では, 参加者に初見の文章を読み上げさせた時と, その文章を数回練習した後に読み上げさせた時の口唇運動の躍度を比較する. 後者の方が滑らかに読み上げられると考えられるため, 練習後の躍度は減少すると予想される.

2 躍度

躍度は加速度の変化を表し, 加加速度とも呼ばれる. 躍度の次元は m/s^3 である. 躍度が小さいほど加速度の変化が小さいため, より円滑な運動であるといえる.

身近な例では、躍度は自動車 [3] や列車 [4] などの乗り心地に影響する。未熟な運転手が不用意に加減速を繰り返すと、躍度が増大し、乗り心地が悪くなる。

運動の変位を $x(t)$ とすると、その躍度 $j(t)$ は以下のように定義される。

$$j(t) = \frac{d^3x(t)}{dt^3} \quad (1)$$

さらに、ひとまとまりの運動を達成するための躍度を評価する指標として、躍度の 2 乗和である jerk cost が提案されている [7, 8]。運動の時間長を T とすると jerk cost は以下の式で得られる。

$$\text{jerk cost} = \sum_{t=0}^T \left(\frac{d^3x(t)}{dt^3} \right)^2 \quad (2)$$

本研究では、単位時間あたりの躍度を比較するため、jerk cost を時間長で正規化した以下の値も用いる。

$$\bar{j}_c = \frac{1}{T} \sum_{t=0}^T \left(\frac{d^3x(t)}{dt^3} \right)^2 \quad (3)$$

3 方法

3.1 実験参加者

発話に問題のない 20 歳から 23 歳までの大学生 18 名が実験に参加した。

3.2 方法

実験参加者は、27 インチ液晶ディスプレイ (BenQ, GW2270-T) とその上に固定された Web カメラ (Logi-cool, BRIO) に正対した位置に座った。そして、Web カメラと顔の距離が 45 cm になる位置に固定された顎台 (SR Research, SR-HDR) の水平方向のバーに前額をあて、頭部を固定した。顎台の顎置きは利用せず、下顎は自由に動くようにした。実験の様子を図 1 に示す。

実験参加者は、最初にディスプレイに表示された初見の文章を読み上げ、続いて 2, 3 回読み上げの練習を行い、その後改めてその文章を読み上げた。実験参加者に読み上げさせた文章を以下に示す。これは発声練習によく用いられる『^{ういろ}外郎売り』の一節である。

ふな きんかん しいたけ さだ ごだん そぼき そうめん
 鯛・金柑・椎茸・定めて後段な、蕎麦切り素麺、
 うどん ぐどん こしんばち こだな こした こおけ
 饅頭か愚鈍な子新発知。小棚の小下の小桶
 こみ そ こあ こしゃくし こも こすく
 に、小味噌が小有るぞ、小杓子小持って小掬
 こよ がってん こころえたんぼ
 って小寄せ。おっと合点だ、心得田圃の
 かわさき かながわ ほどが や とつか はし い
 川崎・神奈川・保土ヶ谷・戸塚は走って行



図 1: 実験の様子

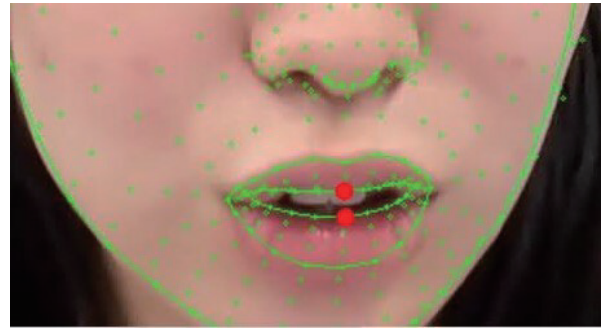


図 2: MediaPipe Face Mesh により抽出した顔ランドマーク (緑色の点) および正中面の上下口唇内側の点 (赤丸)

けば、^{やいと}灸を^す擦り^む剥く。三^{さんり}里ばかりか、^{ふじさわ}藤沢、
^{ひらつか}平塚、^{おいそ}大磯が^{こいそ}しや、^{やど}小磯の^{なな}宿を^お七つ^お起きして、
^{そうてん}早天^{そうそう}早々、^{そうしゅう}相州^{おだわら}小田原、^{とんちんこう}透頂香、^{かく}隠れ^{ごさ}御座
 らぬ。

読み上げ時の顔画像と音声は上記の Web カメラを用いて収録した。収録に利用したソフトウェアは Logicool Capture ver. 2.06.12 である。画像サイズは 1,080×1,080 ピクセル、フレームレートは 60 fps、保存形式は MP4 である。得られた動画から発話区間を切り出し、以降の処理を行った。

得られた動画から機械学習ライブラリ MediaPipe の Face Mesh [5] を用いて各フレームにおける顔ランドマークの 3 次元座標を計測した。顔ランドマークの抽出例を図 2 に示す。Face Mesh では画像の幅、高さを 1 で正規化した座標を出力する。すなわち、画像の左上は (0, 0)、中央は (0.5, 0.5)、右下は (1, 1) である。得られた顔ランドマークから正中面における上下口唇内側の点 (図 2 中の赤丸) を選択し、その $x-y$ 平面上

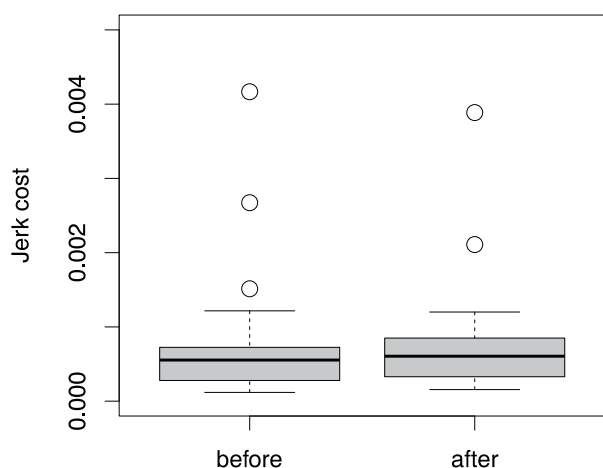


図 3: 練習前後の開口高さの jerk cost

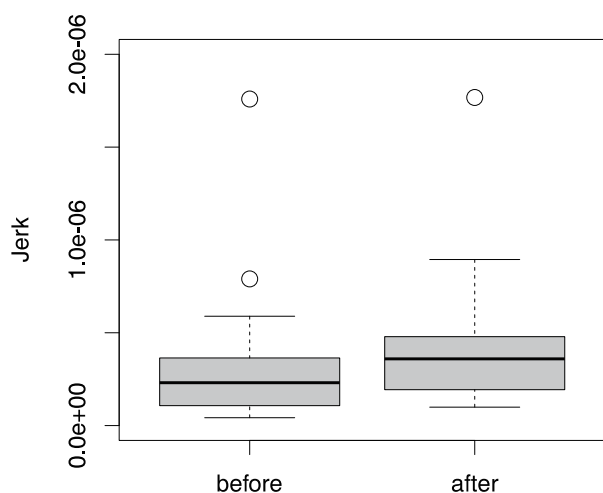


図 4: 練習前後の開口高さの $\bar{j}_c$

の距離を求め、開口高さとした。そして、式 (2) と式 (3) に基づいてそれぞれ jerk cost と $\bar{j}_c$ を計算した。

3.3 結果

文章の読み上げに要した時間長は練習前、練習後それぞれ 41 秒、30 秒であった。練習によって流暢に読めるようになった結果、読み上げに要する時間長が短縮された。実験者 (第 1 著者) が確認したところ、1, 2 名を除き、練習後の音声の流暢性が向上していた。

練習前後の jerk cost および $\bar{j}_c$ をそれぞれ図 3, 図 4 に示す。jerk cost の中央値は練習前が $5.53\text{E-}4$ 、練習後が $6.06\text{E-}4$ であり練習後の方がやや大きくなった¹。練習前後の jerk cost の平均値について t 検定を行ったところ、 $p = 0.200$ となり有意差があるとはいえない

¹MediaPipe Face Mesh が出力した座標 (画像の幅と高さを正規化した座標) に基づいて計算するため、本研究の躍度は非常に小さい値になる。

かった。練習によって調音運動が円滑になり、躍度が減少することを予想していたが、文章読み上げに要する調音運動全体の躍度は練習前後で有意な差異がないという結果になった。

一方、単位時間当たりの jerk cost である $\bar{j}_c$ の中央値は練習前が $2.31\text{E-}7$ 、練習後が $3.59\text{E-}7$ となり練習後に増大した。練習前後の $\bar{j}_c$ の平均値について t 検定を行ったところ、 $p = 0.0004$ となり、有意差があることが示された。

4 考察

読み上げ練習により発話の流暢性が高くなり、その結果、練習後は練習前よりも躍度が減少すると予想したが、予想に反する結果が得られた。調音運動における躍度 (の 2 乗) の総量である jerk cost は練習前後で有意な差が見られなかった。一方、単位時間当たりの jerk cost である $\bar{j}_c$ は練習後に増加した。

練習前後の読み上げに要する時間長の差異 (練習前: 41 秒、練習後: 30 秒) が示唆するように、練習前の音声にはフィラーや言いよどみなど発話の滞留区間が含まれている。これらの区間では開口高さの変化が小さい。一方、練習後の音声ではこれらの区間が減少しているため、結果的に $\bar{j}_c$ が増大したと考えられる。このように考えると、本研究の結果に関しては読み上げに要する時間長が練習前後で差異があり、口唇運動には顕著な差異がなかったと見なすことができる。

5 おわりに

本研究では、加速度の時間微分である躍度に着目し、調音運動の流暢性の定量的評価を試みた。実験の結果、予想に反して、発話の流暢性が上がっても調音運動 (開口高さの変化) の jerk cost に有意な変化は見られなかった。なお、本研究では、発話の流暢性の定義が曖昧で、流暢性の程度も評価しなかった。また、躍度の計算における高次微分に伴うノイズについても未検討である。さらに、本研究では上下口唇の動きに焦点を当てたが、他の顔ランドマークの躍度も利用できる可能性がある。今後はこれらの点も含めて再検討し、引き続き発話の評価における躍度の有用性を調査する。

謝辞

本研究は JSPS 科研費基盤 (A)(20H00291) の支援を受けて行われた。

参考文献

- [1] 定延利之, ささやく恋人, りきむレポーター: 口の中の文化, 岩波書店 (2005).
- [2] 伝康晴, 渡辺美知子, 音声コミュニケーションにおける非流暢性の機能, 音声研究, 13(1), 53–64 (2009).
- [3] 王鋒, 佐川貢一, 猪岡光, 自動車の加減速と乗り心地の関係に関する研究, 人間工学, 36(4), 191–200 (2000).
- [4] 小美濃幸司, 白戸宏明, 遠藤広晴, 清野寛, 乗り心地に配慮したブレーキパターン, 人間工学, 42(3), 164–171 (2005).
- [5] MediaPipe Face Mesh, https://google.github.io/mediapipe/solutions/face_mesh.html (2022 年 1 月 30 日参照)
- [6] A. Rajpal and H. A. Patil, Jerk minimization for acoustic-to-articulatory inversion, *Proc. 9th ISCA Speech Synthesis Workshop*, 82–87 (2016).
- [7] N. Hogans, An organizing principle for a class of voluntary movements, *J. Neuroscience*, 4(11), 2745–2754 (1984).
- [8] T. Flash and N. Hogans, The coordination of arm movements: An experimentally confirmed mathematical model, *J. Neuroscience*, 5(7), 1688–1703 (1985).