一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SP2022-16(2022-06)

［ポスター講演］コロナ禍における児童の音声および口唇運動の収録

北村　達也†　　白勢　彩子††

† 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
†† 東京学芸大学人文社会科学系　〒184–0015 東京都小金井市貫井北町 4–1–1
E-mail: †t-kitamu@konan-u.ac.jp, ††shirose@u-gakugei.ac.jp

あらまし　児童の長・短母音生成時の口唇運動を撮影するため，ヘルメットにアクションカメラを固定したシステムを製作した．撮影中に児童の頭部を安定させるため，カメラアームの重量やその固定位置に配慮した．収録した映像と音声より各フレームの開口高さと音素区間を求め，アノテーションツール ELAN 上で口唇運動を分析できるようにした．

キーワード　口唇運動, アクションカメラ, バンジーヘルメット, 音素セグメンテーション，ELAN

[Poster presentation] Recording of children's speech and lip movements in the Corona disaster

Tatsuya KITAMURA† and Ayako SHIROSE††

† Faculty of Intelligence and Informatics, Konan University　8–9–1 Okamoto, Higashinada-ku, Kobe, Hyogo, 658–8501 Japan
†† Humanities and Social Sciences Division, Tokyo Gakugei University　4–1–1 Nukuikita-machi, Koganei-shi, Tokyo, 184–0015 Japan
E-mail: †t-kitamu@konan-u.ac.jp, ††shirose@u-gakugei.ac.jp

Abstract　Children's lip movements and speech were recorded by a head-mounted, downward facing action camera and voice recorder during the production of Japanese long and short vowels. The weight of the camera arm and its fixed position on a helmet were considered to stabilize the child's head during the recording. The height of the mouth opening and phoneme segments were obtained from the recorded video and audio and displayed together with the media on the annotation tool ELAN to analyze the lip movements.

Key words　lip motion, action camera, bungee helmet, phoneme segmentation, ELAN

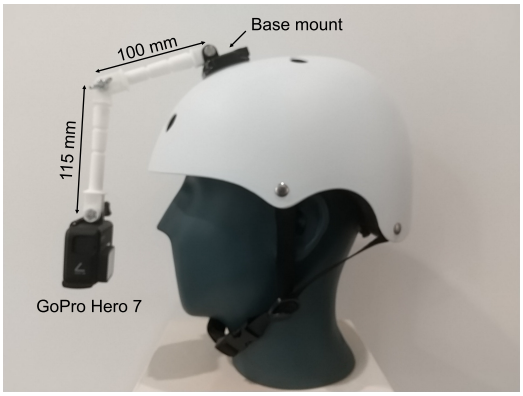
1. はじめに

著者らは，日本語の長母音，短母音のモーラリズムの生成機構およびその発達的過程を明らかにするために，成人および児童を対象としてこれらの母音を生成する際の調音運動を観測している．この研究では，成人を対象とした場合には磁気センサシステム (Electromagnetic Articulography, EMA) を用いて調音運動を計測しているが [1]，EMA では舌や口唇などの調音器官に小型のセンサを接着する必要があるため [2]，児童には負担が大きい．そこで，児童についてはカメラにて顔の正面方向より口唇運動を録画し分析している [3], [4].

児童の場合，収録中に頭を動かさないように指示してもそれを維持するのが難しく，カメラと口唇の位置関係が大きく変化してしまうことが少なくない．そこで，我々はアクションカメラをアームで固定したヘルメット (バンジーカメラ，バンジーヘルメットなどの呼び方がある) を自作し，実験参加者の顔とカメラの位置関係を固定して計測を行った．しかし，その後 COVID-19 の感染拡大を受け，大学等にて音声を収録することが難しくなった [5]．このような状況でもデータを収集するため，著者らは児童の自宅に機材一式を送り，保護者に収録の操作を依頼するという方法を採っている．本稿では，我々が用いたデータ収録用機材および収録手続き，その後のデータ処理について報告する．

2. 収録用ヘルメット

著者らが調べた限り，バンジーヘルメットは市販されていなかったため，市販のヘルメットにカメラ用アームを固定して自作した．製作した収録用ヘルメットを図 1 に示す．以下では各



側面

正面

図 1　収録用ヘルメット．(上) 側面，(下) 正面．

パーツについて説明する．

2.1　ヘルメット

子供サイズのスポーツ用ヘルメットを使用した．このタイプのヘルメットには，後部にサイズ調整用ダイヤルが付いており，児童個々の頭のサイズに合わせることができる．

2.2　カメラアーム

バンジーヘルメットのカメラアームの配置には，頭頂部から下方向にアームを伸ばすタイプと，側頭部あるいは後頭部から前方向にアームを伸ばすタイプがある．後者の方がカメラの重量を感じにくく，疲労が少ないと考えられるが，適当な市販のパーツがなく，自作も困難であった．そこで，まず市販のグースネックタイプの GoPro 用カメラマウントを試みたが，それ自体が重く児童の負担になると考えられた．そこで，カメラアームは以下のようにして自作した．

インターネット上には 3D プリント用 CAD データの共有サイトが存在する (例えば，`https://grabcad.com/library/`)．このようなサイトから GoPro 対応のカメラアームを入手し，必要に応じて CAD ソフトにて長さの変更などを加えれば，設計の労力を大幅に削減することができる．このようにして長さの異なる 2 本 (100 mm, 115 mm) のカメラアームの CAD データを作成した．それを 3D プリントサービスを提供する DMM.make に送付し，軽量なナイロン素材で造形した．

2.3　ビデオカメラ

実験当初は大学の実験室に児童を招いて収録を行っていた．その際にはカメラアームの先に Web カメラ (Logicool BRIO) を固定し，録画の操作はそれを USB ケーブルにて接続した PC にて行っていた．しかし，COVID-19 の影響により大学内にて収録を行うことができなくなったため，児童の自宅に機材一式を送付し保護者に収録の操作を行ってもらう方法に変更した (3 節参照)．その際，収録に PC が必要な Web カメラではなく，単独で収録できる液晶ディスプレイ付きのアクションカメラ GoPro Hero 7 を採用した．このカメラの重量は 114 g である．

2.4　カメラの固定

GoPro 付属のベースマウントをヘルメットの図 1 (上) に示す位置に固定した．この位置をあまり前にするとバランスが悪くなり，実験参加者が頭部を安定させづらくなる．また，GoPro のレンズの位置に合わせて，ベースマウントの位置を正面から見て左側方にずらしてある．このベースマウントにカメラアームを取り付け，その先に GoPro を固定した．できる限り参加者の個人を特定する情報を少なくするため，撮影範囲を口唇周辺に限った配置にした．

カメラアームは脱着可能であり，大人サイズのヘルメットにベースマウントを付ければ，大人を対象にした実験にも使用できる．また，三脚ネジ (1/4 インチ) と GoPro の固定部との変換アダプタを用いれば GoPro 以外のカメラも固定することができる．

3.　リモート収録の手続き

前述のように，COVID-19 の影響を受け，対面による実験の実施が困難となったことから，機材類を実験参加者の児童宅に送付して収録する手続によって実験を進めた．以下に，実施内容について示す．なお，リモートによる収録の実施について，東京学芸大学倫理委員会の審査を受け，承認されている．

3.1　収録機材の送付

送付物は，GoPro を取り付けたヘルメット，録音機 (SONY, ICD-UX533)，マイク (Audio-Technica, AT9912) など収録機材と，それらの簡易マニュアル，発音リストと，実験参加の同意書，話者情報の記録紙，返送用の宅配ラベルである．COVID-19 の影響を考慮し，機材を含め，全ての物品は，消毒作業を行ってから送付している．

送付物は 4 セット準備し，複数の家庭に並行して送付できるよう設定した．

3.2　収録手順の指示

実験参加者が児童であることから，実験は保護者等，監督者が実施することになる．収録の実行者が音声機材の扱いには不慣れであることが考えられ，簡易的かつ明瞭なマニュアル等を整備することが肝要と考えられた．加えて，模擬的に収録を行って動画に記録し，これらを視聴することにより文字情報だけでは伝えきれない内容を補えるよう工夫した．動画は YouTube にアップロードした．

動画の収録にあたり特に注意が必要な点として，口唇が適切に撮影されているかどうかがある．記録された口唇運動は，4 節に示すように，広角レンズの歪み補正の後，機械学習による自動検出を行うことから，正確に正面を向いて動画が撮影さ

図 2　口唇撮影の指示例

れていることが望ましい．この点を的確に指示するため，図 2 のように，具体的な事例を挙げて説明を示した．これは紙のマニュアルによる指示例であるが，前述のように動画によっても解説した．

他，収録機器類の取り扱いについても，紙媒体と YouTube 動画との双方により，説明を示した．

3.3　発音リスト

発音リストは，比較のため，成人を対象とした我々の研究 [1] で用いたものとほぼ同一のものとなっている．/kabu/（下部）–/kaRbu/（カーブ）（/R/は長音拍）のような母音の短・長の対である．実在語の複数対と，それらを模した無意味音列（例：/mama/–/maRma/）の複数対とした．実在語の一覧を表 1 に示す．これらに加え，「ア」，「イ」，「ウ」の孤立発声も含めた．

表 1　発音リスト

	a	i	u	e	o
short	kabu	siru	tsuru	beru	koto
	下部	汁	鶴	ベル	琴
long	kaRbu	siRru	tsuRru	beRru	koRto
	カーブ	シール	ツール	ベール	コート

3.4　実験参加者

小学校 1〜6 年，10 名の実験参加者があった．これらのうち 2 名は，撮影角度が適切でなく，口唇が正面から撮影できていなかったことからデータ処理の対象外とした．

4.　データ処理

4.1　動画データの処理

GoPro にて録画した動画には広角レンズによる歪みが含まれる．開口高さを正確に計測するために，動画処理ソフトウェア VideoProc Converter を用いてレンズ収差を補正した[1]．その際，GoPro にて方眼用紙を撮影した画像を用いて補正用のパラーメータを決定し，その値を用いて全ての動画を補正した．

次に，機械学習ライブラリ dlib [6] を用いて開口高さを自動

(注1)：Adobe Premire は GoPro のレンズ収差補正の機能を持っているが，著者らが用いた Hero 7 に対応していなかった．

補正前

補正後

図 3　広角レンズの影響の除去．(上) 補正前，(下) 補正後．

計測した．まず，各動画の 50 フレームを対象にして手作業で口唇の 4 点，すなわち，上下赤唇縁の内側の正中面を通る点の座標と左右の口角の座標を計測し学習データとした．この学習データを用いて dlib に実装されている Kazemi and Sullivan のアルゴリズム [7] にて学習し，得られたモデルを用いて全フレームにおける上記の 4 点を抽出した．それらの座標から各フレームの開口高さを算出し，CSV ファイルに出力した (現段階では開口幅のデータは利用していない)．

4.2　音声データの処理

IC レコーダにて収録した音声を書き起こし，それに基づいて音素ラベリングを行った．音素ラベリングには julius 音素セグメンテーションキット [8] を利用した．得られた音素ラベルデータを ELAN にて読み込むことができる Praat の TextGrid ファイル形式に変換した．

4.3　ELAN での表示

動画データ，開口高さデータ，音声データ，音素ラベルデータをアノテーションツール ELAN [9] [10] に読み込み，「メディアの同期化画面」にてこれらを同期させた．ELAN を用いることによって一覧性の高い状態でデータを分析することができる．

5.　おわりに

本稿では，アクションカメラを用いて児童の口唇運動を収録するシステム，それを用いたデータ収集，得られたデータの処理について報告した．コロナ禍の中で各家庭にて監督者の機器操作により収録を実施し，著者らの要求を満たしたデータを確実に得るために事前の準備を行ったが，改善の余地はある．原稿執筆時点 (2022 年 5 月) では，いつ大学にて児童を対象とし

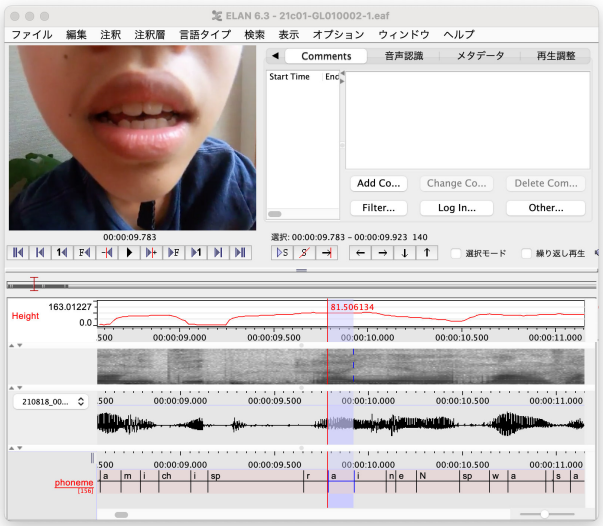


図 4　ELAN によるデータ表示．動画，開口高さ，音声，スペクトログラム，音素ラベルを同期させて確認できる．

た収録を再開できるかは不透明な状況であるため，多くの方の助言を得つつ，より良い収録手法を模索していきたい．

謝　　辞

本研究は JSPS 科研費 (17K02769) の支援を受けて行われた．実験の実施およびデータ処理にご協力いただきました小杉圭子さん，高濱明日香さん，狩野晃弘さん (東京学芸大学)，宮川裕士朗さん (東京学芸大学) に感謝します．

文　　献

[1] 白勢彩子, 北村達也, 能田由紀子, 長母音の発音における時間的な伸張と構音運動の相関性, 音講論 (秋), 801–802, 2021.
[2] 北村達也, 磁気センサシステムによる調音運動のリアルタイム観測, 音響誌, 71(10), 526–531, 2015.
[3] 白勢彩子, 北村達也, 北澤佐知子, 児童を対象にした日本語長短母音の構音動作の観察, 音講論 (春), 821–822, 2020.
[4] 白勢彩子, 北村達也, 児童の日本語母音発声時の口唇運動に関する動画記録による検討, 音講論 (秋), 675–676, 2020.
[5] 北村達也, 高野佐代子, 石本祐一, コロナ禍における音声収録の実態調査, 音響誌, 78(4), 187–188, 2022.
[6] D. E. King, Dlib-ml: A Machine Learning Toolkit, *Journal of Machine Learning Research*, 10, 1755–1758, 2009.
[7] V. Kazemi and J. Sullivan, One millisecond face alignment with an ensemble of regression trees, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
[8] A. Lee and T. Kawahara: Julius v4.5, `https://doi.org/10.5281/zenodo.2530395`, 2019.
[9] 宮澤幸希, コミュニケーション研究における ELAN の活用: 音声・映像データへのアノテーション, 音響誌, 75(6), 344–350, 2019.
[10] 細馬宏通, 菊地浩平編, ELAN 入門: 言語学・行動学からメディア研究まで, ひつじ書房, 2019.