

調音・音声データベースの構築

研究代表者

若 宮 幸 平

九州大学大学院芸術工学研究院 助教

1 はじめに

発話中の舌・唇などの調音器官の動き（調音運動）を音声信号と同時に測定することにより、人間が音声生成する際に必要な情報や知見を得ることができる。また、測定データに基づいて調音運動を音声信号に、あるいは音声信号を調音運動に変換することが可能になれば、調音データを用いた自然な音声の合成[1]～[4]やサイレントスピーチインタフェース[5]の構築等が可能となる。

上記のような調音運動と音声信号との対応関係や調音運動と言語情報との対応関係を学習データから機械学習により獲得することができれば、未知の音声信号もしくはテキストに対応する調音運動を自動推定する技術[6], [7]や、調音運動と音声信号との対応関係を同時確率分布により表現し、統計的声質変換及び音声合成システムを調音運動情報に基づき制御可能にする技術[2], [8]などにも応用可能である。

調音運動を直接測定する方法としては、リアルタイム磁気共鳴画像法（Real-time MRI）[9]や磁気センサシステム（EMA）[10]、電気的口蓋図（EPG）[11]、X線撮影[12]、超音波[13]の利用などが考えられるが、本研究では EMA を用いる。このシステムは小型のセンサの位置を磁気を用いて測定するものである。測定点がセンサ貼り付け位置の数点に限定されるものの、時間分解能が高く、測定時の騒音がないという特徴があるため、音声との同時測定に適している。

EMA を用いた調音運動データベースとしては、2名のイギリス人英語話者による TIMIT 文（460 文）の読み上げ時の調音運動に関するデータベースである MOCHA TIMIT[14]、1名のイギリス人英語話者による 1,300 文の読み上げに関する mngu0[15]、7名の健常英語話者および 8名の構音障害者（小児性脳性麻痺 7名、ALS 1名）の調音運動からなる、計 23 時間の TORGO データベース[16]、USC-TIMIT[17]等が知られている。現状で最も利用されている調音運動データベースは mngu0 である。このデータベースは EMA の他にも MRI を含む多様なデータを含んでいるものの、前述のように話者は 1 名である。そのため、同じ音素が別の調音方法で発音されうる場合は、それらのうちの 1 つに関するデータしか得られていないものと考えられる。

また、現在までのところ、英語の調音運動データベースは公開されているが、日本語のデータベースは公開されていない。日本語の調音・音声データベースを作成し、公開することができれば、日本語音声に固有の特徴を含む調音運動をモデル化できるなど、日本語の音声研究分野での活用が期待できる。さらに、音声波形自体は、時々刻々と不規則に変動するのに対し、調音運動はヒトの身体部位が実際に変位したその運動を捉えたものであるため、音声波形のように短時間のうちに大きく不規則的に変動しないと考えられることから、このような調音運動の変化パターンを音声圧縮に利用できれば、音声伝送時に必要な通信容量を減らせる可能性もあり、そうなれば、より効率的な情報通信へも寄与できる。そこで、本研究では、調音・音声データベースの構築と公開を目指し、3次元磁気センサシステム（3D-EMA）を構築し、3D-EMA を用いた調音・音声パラレルデータを収集した。

2 3次元磁気センサシステム

本研究で用いる 3次元磁気センサシステム（3D-EMA）は、8個の送信コイルにより生成した磁界中に位置マーカーとしての受信コイルを置くことにより、受信コイルに誘導起電力が生じ、その信号強度から受信コイルの 3次元位置を推定するシステムである。

この受信コイルを舌、顎、下唇などの調音器官に貼付すれば、発話中の調音運動が、コイルを貼付された調音器官の部位の運動という形で測定できる。また、3次元磁気センサシステムには、大きな雑音の発生源がないため、音声と調音運動の同時収録が可能である。

3次元磁気センサシステムを用いた位置推定法の概要は下記のとおりである。図 1 に示す位置に配置された 8 個の送信コイル（ $S_1, S_2, \dots, S_8$ ）は、互いに異なる周波数の交流磁界を生成する。生成された交流磁界中に置かれた受信コイルには受信信号が誘導される。この時の受信コイルの状態

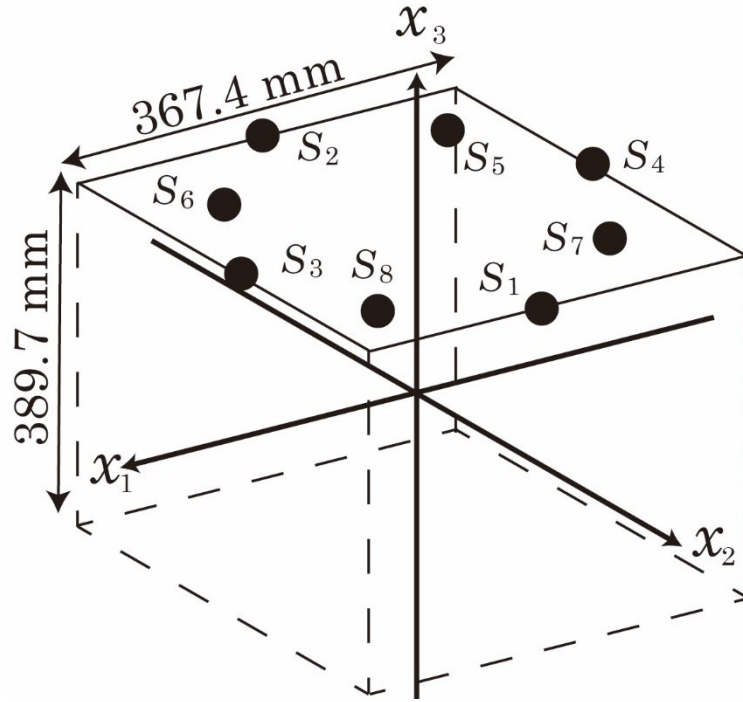


図 1 8 個の送信コイルを持つ 3 次元磁気センサシステムの送信コイルの配置

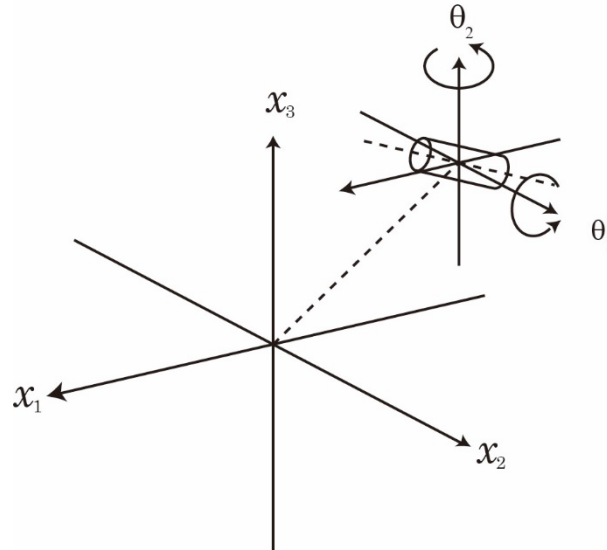


図 2 受信コイルの状態

$\mathbf{s} = (x_1, x_2, x_3, \theta_1, \theta_2)$ (位置及び傾き) を, 図 2 のように定める.

本研究では, 送信コイルを磁気双極子とみなし, 送信コイルの生成する磁界をクーロン則に基づく磁界モデルにより表現する. 本磁界モデルを用いることにより, 任意の位置 $\mathbf{x} = (x_1, x_2, x_3)$ における磁界 $\mathbf{y}(\mathbf{x})$ が求められる. 状態 $\mathbf{s}$ における受信信号予測値はこのとき, 下記の式で表現される.

$$\hat{z}(\mathbf{s}) = g\mathbf{y}(\mathbf{x}) \cdot \mathbf{e}(\theta_1, \theta_2) \quad (1)$$

ここで, g は受信コイルの校正時に決定されるゲインパラメータであり, $\mathbf{e}(\theta_1, \theta_2)$ は受信コイルの向きを表す単位ベクトルである. 受信コイルの状態 $\mathbf{s}$ は, 測定された受信信号 $z_l (l=1, 2, \dots, 8)$ とその予測値

$\hat{z}_l(\mathbf{s}) (l=1, 2, \dots, 8)$ の間の下記の自乗誤差 E が最小となるように定められる.

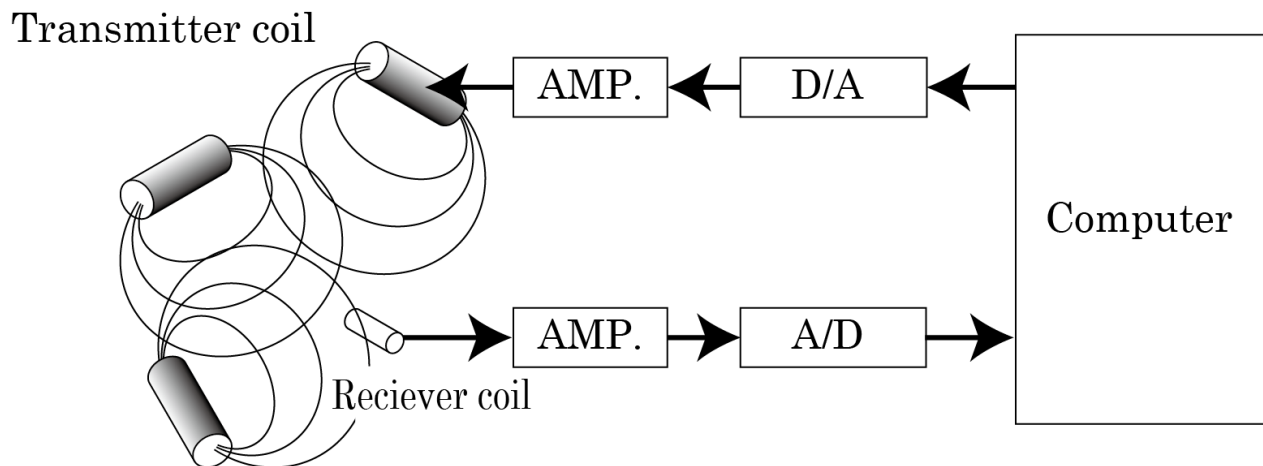


図 3 3次元磁気センサシステムのハードウェアシステム構成

表 1 送信コイルの送信周波数と向き

/	1	2	3	4	5	6	7	8
送信周波数(Hz)	12,375	8,500	13,375	11,375	9,375	10,375	11,875	10,875
$\theta_1(^{\circ})$	-90	-80	-80	-45	-70	-65	-15	-85
$\theta_2(^{\circ})$	0	345	95	240	350	135	320	355

$$E = \sum_{l=1}^8 (z_l - \hat{z}_l(s))^2 \quad (2)$$

ここで、 l は送信コイルのインデックスである。この問題は非線形最小自乗問題となるため、本研究では繰り返し最適化手法の一種である Gauss-Newton 法を用いて最適解を求めることで、受信コイルの状態 s を推定する。

3次元磁気センサシステムのハードウェアシステム構成を図3に示す。本システムは、8個の送信コイルを搭載した送信アンプ、12チャンネルの受信アンプ、サンプリングレート100kHzで8チャンネル同時DA可能なDA変換器（DASmin-E2000 model 1608-100k-DA、コーメックス電子）、サンプリングレート50kHzで16チャンネル同時AD可能なAD変換器（DASmin-E2600 model 1616-100k-AD、ケイテクノス）およびLinux PCで構成されている。PCは送信コイルに関する正弦波駆動信号の生成と受信コイルから得られる受信信号の取得に使用される。送信コイルは、図1に示すように測定領域の上部に半径18.37cmの円の円周上に等間隔に配置する。また、各送信コイルの向きを、先行研究[18]の方法を用いて求めた非決定性の問題が生じにくくなると考えられる向き[19]に設定する。各送信コイルの生成する交流磁界の周波数および各送信コイルの向きは、表1のとおりである。

2-1 3次元磁気センサシステムの推定精度

本研究では、3次元磁気センサシステムを用いて調音運動を測定する前に、本システムが調音運動の測定に使用可能かどうか確認するため、構築した3次元磁気センサシステムの位置推定精度について検討を行った。舌の調音運動を正確に認識するために必要とされる精度は、最低でも誤差が1mm以下であることが求められる。理想的な精度は、誤差が0.5mm以下であることである。

3次元磁気センサシステムの位置推定精度を求めるために、校正用ロボットとロボットにコイルを固定する受信コイルホルダを使用した。校正用ロボットは指定した座標にコイルホルダを動かし、その位置でのデータを取得する。このデータは、受信コイルの校正を行ったり、位置推定を行ったりするために使用される。

本研究で用いた手順は下記のとおりである。まず、校正用ロボットを使用して、一辺100mmと80mm

表 2 受信コイルの位置推定精度(単位:mm)

	ch1	ch3	ch4	ch5	ch6	ch7	ch8	ch9	ch10	ch11	ch12	平均
-30°	1.069	1.669	1.554	0.970	1.138	1.016	0.874	0.822	0.520	0.856	0.792	1.025
-20°	0.612	1.148	1.010	0.569	0.708	0.662	0.686	0.566	0.427	0.628	0.606	0.693
-10°	0.285	0.653	0.535	0.266	0.392	0.364	0.448	0.413	0.348	0.452	0.399	0.414
0°	0.252	0.380	0.354	0.227	0.282	0.295	0.372	0.400	0.238	0.435	0.355	0.327
10°	0.452	0.367	0.506	0.516	0.494	0.533	0.558	0.484	0.463	0.467	0.404	0.477
20°	0.633	0.551	0.815	0.695	0.680	0.740	0.875	0.560	0.524	0.382	0.532	0.635
30°	0.965	0.793	0.981	1.089	0.964	0.964	1.126	0.833	0.708	0.383	0.668	0.861
平均	0.610	0.794	0.822	0.619	0.665	0.653	0.706	0.583	0.461	0.515	0.536	0.633

の立方体内で等間隔に 27 点のデータを取得する。なお、一辺 100 mm のデータを取得する際には、コイルの角度を $\theta_1 = 0^\circ$, $\theta_2 = 0^\circ$ とした。このようにして得られた一辺 100 mm のデータを使い受信コイルのゲインの校正を行なった。次に、一辺 80 mm で取得された受信信号と校正されたゲインに基づいて受信コイルの位置推定を行い、理想的な位置 (± 40 mm, 0 mm) との二乗誤差を求め、その平均の平方根を位置推定精度として求めた。なお、位置推定を行う一辺 80 mm のデータを取得する際には、 $\theta_1 = 0^\circ$ に固定し、コイルホルダの角度 θ_2 を $0^\circ, \pm 10^\circ, \pm 20^\circ, \pm 30^\circ$ に調整し位置推定を行った。受信コイルホルダには、11 個の受信コイルを搭載し、それぞれの受信コイルについて、位置推定を行い、受信アンプのチャンネル番号を基に、チャンネルごとに位置推定精度を求めた。また、各角度における位置推定の平均値についても求めた。

結果を、表 2 に示す。表 2 から、誤差が 1 mm を超えるものは少なく、角度が $0^\circ, \pm 10^\circ$ の小さいものでは、角度ごとの平均を見ると誤差は 0.50 mm 以下となり、非常に優れた値を示していることがわかる。最小の誤差は、ch10 の 0° で 0.238 mm であり、調音運動の観測に申し分のない精度が得られていることがわかる。以上の結果から、本研究で使用する三次元磁気センサシステムは、角度が $\pm 10^\circ$ 以下の範囲であれば、誤差が 0.50 mm 以下という優れた精度を持つことが示された。ただし、チャンネルによっては、過度に傾けないように使用することが必要である。ここで得られた精度は、舌等の調音器官の調音運動を観測するために十分であると考えられる。したがって、本研究で構築した三次元磁気センサシステムは、調音観測に問題なく利用可能であると考えられる。

3 調音・音声同期データの収集

3-1 調音・音声同期データの収録条件

ここでは、調音・音声同期データの収録条件として、発話者、発話内容の提示方法、提示文、発話時間について述べる。

発話者は、成人の日本語母語話者 4 名（男性 1 名（M1）、女性 3 名（W1, W2, W3））で、うち 2 名は、正確な日本語発話ができるよう日本語の発話トレーニングを受けたアナウンサー（M1, W3）である。発話内容は、発話者の正面に設置したディスプレイに提示する形式を採用した。発話者には、収録開始の合図が聞こえたら、提示された文を読み上げるよう、あらかじめ指示を行った。

調音・音声データベース作成に向けた調音・音声同期データの収録においては、多くの音素パターンに対応するデータを収録することが望まれる。そのため、様々な音素パターンが出現するように構築された音素バランス文を提示し、発話者がそれらの文を発話している際の観測を行うことが望ましい。本研究では日本

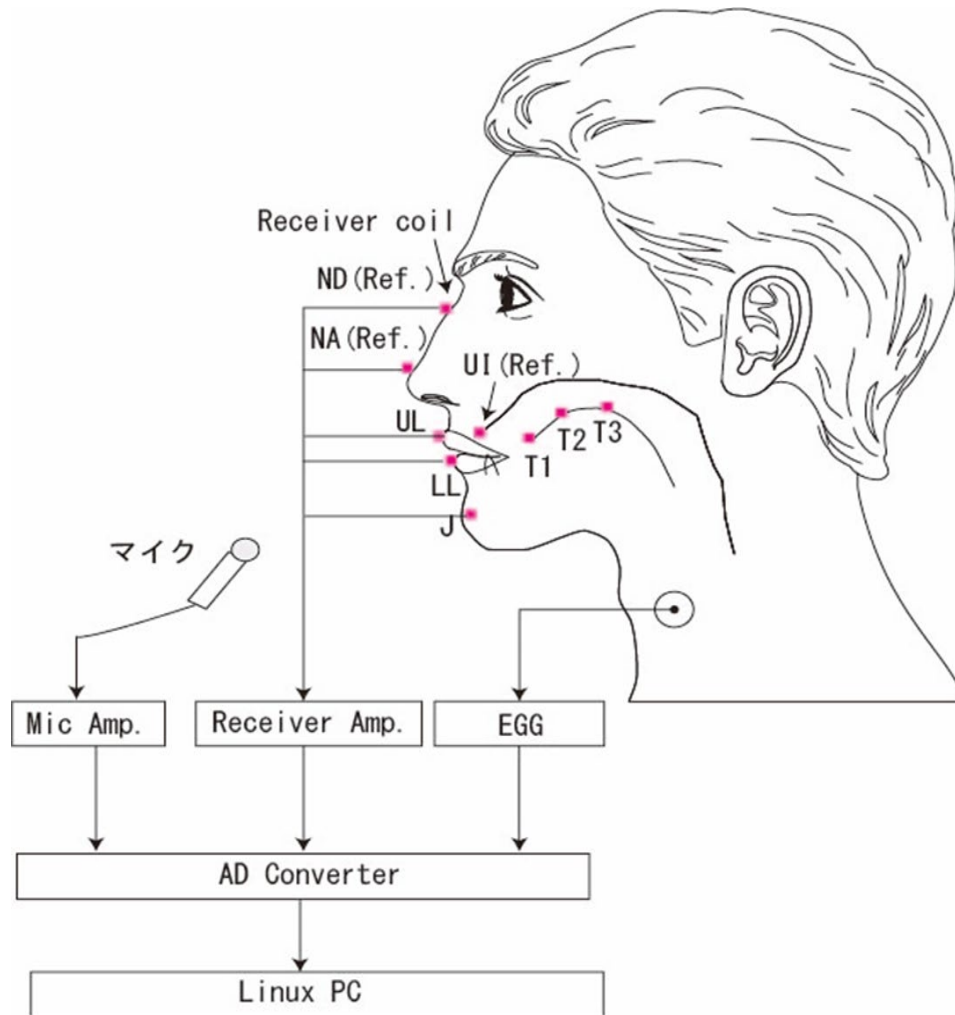


図 4 調音・音声同期データ収集システムの構成並びに 3D-EMA による観測点

語を対象とした音素バランス文として、ATR503 文を採用した。この ATR503 文は、503 文全体でみても音素バランスが取れているのに加え、A ～ J までのサブセットに分かれており、A ～ I セットが各 50 文、J セットが 53 文からなっていて、それぞれのセット内でも音素バランスがある程度取れたものとなっている。この文セットは、ATR (株式会社 国際電気通信基礎技術研究所) によって 1988 年に開発されたもので、各文の長さが 10～73 モーラの範囲であり、二音素 (diphone) 連鎖パターンは 503 種類、三音素 (triphone) 連鎖パターンは 2,834 種類が含まれており、そのうち、441 種類と 2,142 種類については、少なくとも 2 回以上提示文に含まれていた。

調音運動観測中に、3 次元磁気センサシステムのマーカである受信コイルが外れた場合は、受信コイルを貼りなおさなければならなくなる。その際、できるだけ以前の貼付位置と同一の場所に貼付するよう心掛けたが、実際にはずれが生じている可能性が高い。そのため、受信コイルが調音器官から外れた場合は、受信コイルを貼り直した後で、外れた際の発話中の文が含まれているサブセット内の発話についての観測を再度行うことにより、少なくともサブセット内でのマーカの調音器官上の観測点が同一に保たれるようにした。

データの総量でみると、本研究で収集された調音・音声同期データの発話時間は、一人当たり、平均で約 34 分、4 名分で総計約 137 分となった。これは、mngu0 のデータベースの総発話時間より多く、各話者の発話時間が 15～20 分程度の Wisconsin XRMB コーパスや MOCHA と比べても大規模なものとなっている。

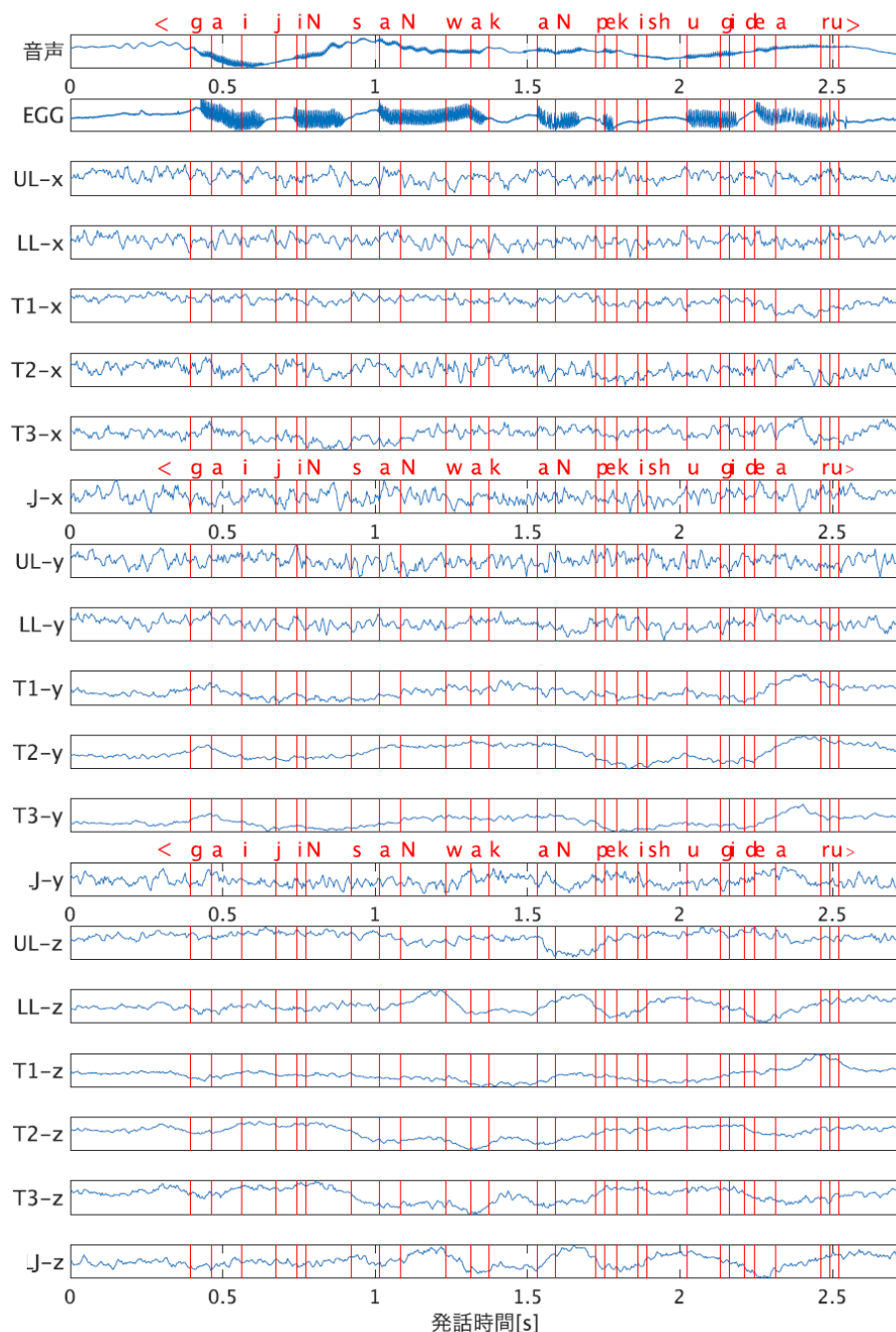


図 5 「外人さんは完璧主義である」の発話時の音声波形，EGG 波形と対応する調音運動

3-2 調音・音声同期データの収集システムの構成並びにデータ処理

図 4 には，本研究で使用する調音・音声同期データ収集時のシステム構成ならびに 3 次元磁気センサシステムによる観測点を示されている．本システムでは，3 次元磁気センサシステムによる調音運動，マイクを用いて収集される音声に加えて，発話中の声帯の接触の程度を EGG [20]により測定した．EGG 信号データは，音声の基本周波数の測定や，有声音・無声音の識別などに利用できると考えられるため，調音・音声同期データを使用した機械学習を行う際の補助データ等としての使用が期待できる．

3 次元磁気センサによる測定点は，上唇（UL），下唇（LL），下顎（J），舌上の 3 点（T1，T2，T3）とした．また，発話中の頭部の動きを補正するための参照点として，鼻背（ND），鼻尖（NA），上門歯（UI）の 3 点の位置を測定した．なお，全ての測定点は，頭部のほぼ正中面上に取った．発話中の各受信コイルの受

表 3 収集した調音・音声同期データの概要

発話者	成人の日本語母語話者 (男性 1 名, 女性 3 名)
一人当たりの発話文数	503 文
一人当たり発話時間	平均 約 34 分
一人当たり総発話モーラ数	16,970
各文のモーラ数	10 ~ 73
Diphon の種類数	503
(うち 2 回以上提示)	(441)
Triphon の種類数	2,834
(うち 2 回以上提示)	(2,142)
音声のサンプリング周波数	50 kHz
EGG のサンプリング周波数	50 kHz
EMA による調音運動測定点数	6
調音位置データのサンプリング周波数	250 Hz

信信号と音声信号ならびに EGG 信号を AD 変換器により 50 kHz で標本化し、PC に取り込んだ。

各受信コイルの受信信号から調音運動データを得るために、下記の処理を行った。まず、受信信号を送信コイルの送信周波数に対応した周波数成分に分離した。その際、窓長 8 ms (400 サンプル) の矩形窓を使用し、シフト幅を 4 ms (200 サンプル) とすることにより、サンプリングレート 250 Hz で各送信コイルにより生成される交流磁界の周波数に対応する受信信号を取得した。次に、この受信信号をもとに受信コイルの状態を推定することで受信コイルの位置情報を得た。最後に、得られた受信コイルの位置情報に対し、頭部参照点の位置情報を用いて発話中の頭部の動きの影響が出ないように補正を行なった。得られた位置情報データは、3 次元位置情報となっており、 x 座標値は頭部の左右方向の変位を、 y 座標値は前後方法の変位を、 z 座標値は上下方向の変位を示している。また、各軸の正方向を、 x 軸方向では発話者から見て左に、 y 軸方向では後方に、 z 軸方向では上方に取った。

音声信号データの収録には、Brüel & Kjær 社製 Type 4191 外部偏極型自由音場マイクロホン (測定範囲 3.15 Hz ~ 40 kHz)、Type 2669-L プリアンプ、NEXUS Type 2690-A-0S2 マイクコンディショニングアンプを使用した。音声収録に当たっては、マイクスタンドを用いて、発話者の右前方約 60 cm の距離にマイクを設置した。音声信号データの収録には、Brüel & Kjær 社製 Type 4191 外部偏極型自由音場マイクロホン (測定範囲 3.15 Hz ~ 40 kHz)、Type 2669-L プリアンプ、NEXUS Type 2690-A-0S2 マイクコンディショニングアンプを使用した。音声収録に当たっては、マイクスタンドを用いて、発話者の右前方約 60 cm の距離にマイクを設置した。

EGG 信号データの取得には、glottal enterprises 社製の EG-2 を使用した。

また、収録した音声波形を MATLAB 2023b を用いて wav 形式への変換、16 kHz へのダウンサンプリング、振幅の正規化を行ったのち、この音声と発話の書き下し文から Julius 音素セグメンテーションキット[21]を用いて音素セグメントデータを得た。

3-3 測定データの例

本研究の調音・音声同期データにおける発話時の調音運動測定データ例を図 5 に示す。この時の発話内容は、「外人さんは完璧主義である」であった。図の横軸は時間、縦軸は順に音声波形、EGG 波形、上唇(UL)、下唇(LL)、舌 (T1, T2, T3)、顎 (J) の測定点の変位である。ここで、○- x は各測定点の x 軸方向の変位を、○- y 、○- z も同様にそれぞれの軸方向の変位を示す。なお、○には調音測定位置の略語 (UL, LL, etc.) が入る。この図において、各グラフの縦軸の値はそれぞれ異なるため、相対的な変化であることに注意が必要である。また、音声波形、J- x 、UL- z の上部に発話内容をローマ字で、発話の開始を < で、発話の終了を >

両唇音：/p/ /b/

「女の子は、以前は、パンより御飯が好きで、喜んでお代わりをした。」
「プレー中絶えず笑いを浮かべる彼女の強さばかりが目立った。」

発話順序	a	p or b	a
マーク順序	○	○	○

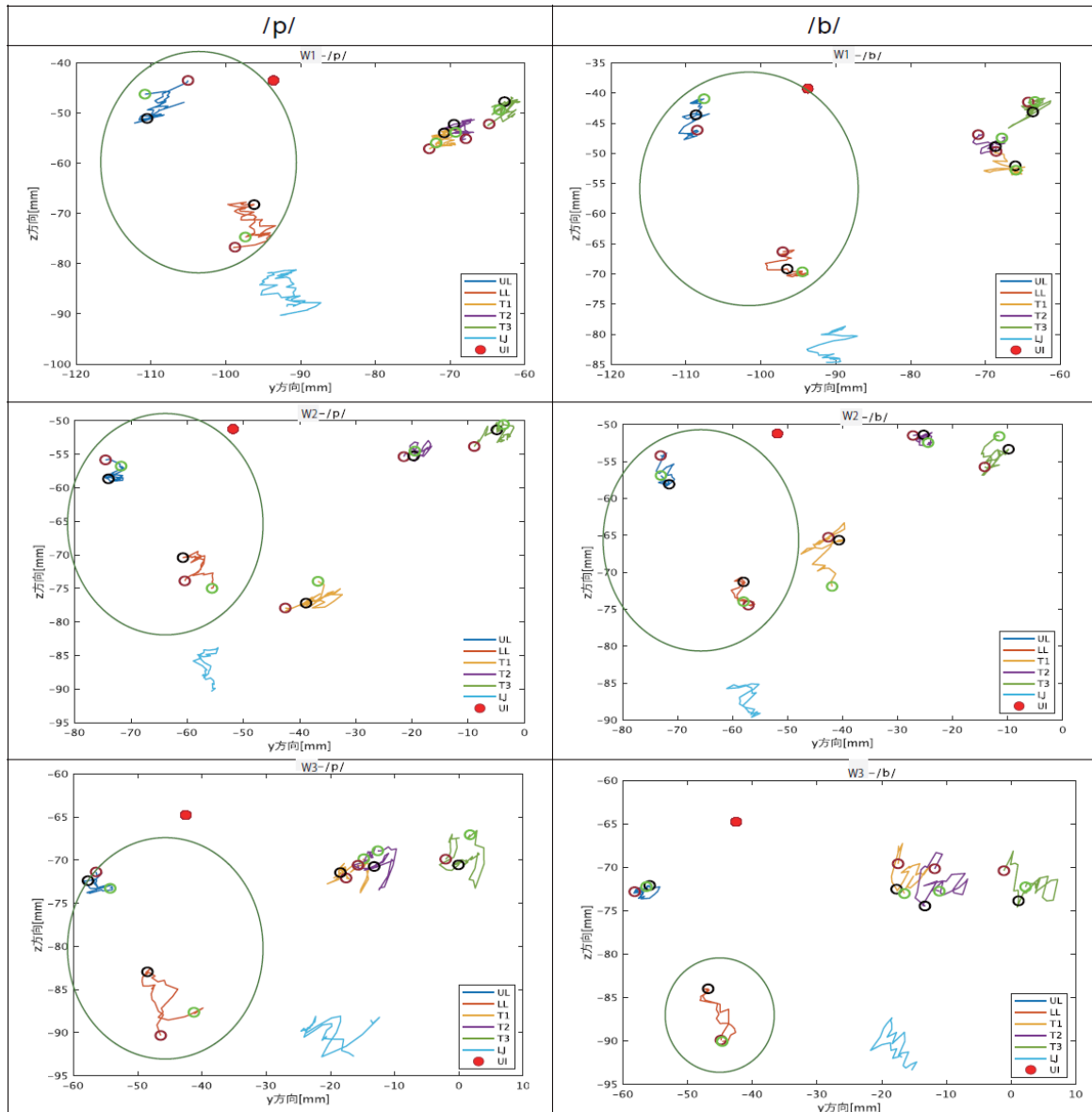


図 6 両唇閉鎖音 /p/ or /b/ を含む発話の調音運動例

で記載している。図中の縦の点線は、各音素セグメントの区切りを示している。

得られた結果から、この例では、録音開始の約 0.4 秒後より発話が始動され、約 2.5 秒後に発話が終了していることがわかる。

たとえば、録音開始後約 0.5 秒過ぎあたりの、母音 /a/ から /i/ への調音運動を確認すると、T1-y, T2-y, T3-y が徐々に減少し、T1-z, T2-z, T3-z が徐々に増加しているのがわかる。これは、後舌母音 /a/ では、舌の位置がやや後方にあり、口が大きく開いている状況であったものが、前舌母音 /i/ では、舌が前方に移動し、口が狭くなった状態へ移行するさいの調音運動を捉えられていることが見て取れる。録音開始後約 1.2 秒

あたりの、わたり音 /w/ から /a/ への変化について見ると、顎の上下の動き J-z をみれば、/w/ を発音する際には、顎がそれほど大きく開いていない状態（J-z が大きい値）であったものが、/a/ を発音する際には、顎が大きく開いている状態（J-z が小さい値）へと変化していることがわかる。さらに、録音開始後 1.6 秒あたりの撥音 /N/ から両唇閉鎖音 /p/ への移行部では、下唇が /p/ 発音の直前で高くなり、/p/ 発音の直後に下がっていること、前後方向でみると、/p/ 発音の直前で下唇が前に突き出され、/p/ 発音の直後に後ろに下がっていることが LL-z, LL-y それぞれより確認されたとともに、EGG 信号においても、/N/ の後半部から、/p/ の終了時点までは振動波形が観測されず、/e/ になると振動波形が観測されることがわかる。これらの調音運動の様子は、先行研究により調査されてきた各音素の一般的な調音的特徴に一致するものであり、本実験で収集された調音・音声同期データは、ヒトの自然な発話における声帯振動、調音運動と音声の対応データとなっているものと考えられる。

また、3 名の話者による無声両唇閉鎖音 /p/ および有声両唇閉鎖音 /b/ を含む発話時の各調音器官の調音運動の y-z 平面データを図 6 に示す。図では、無声両唇閉鎖音は、「女の子は以前は、パンよりご飯が好きで、喜んでお代わりをした。」、有声両唇閉鎖音は、「プレー中絶えず笑いを浮かべる彼女の強さばかりが目立った。」発話中の下線部の遷移 /apa/ または、/aba/ の際の調音運動を示しており、/a/ → /p/ → /a/、/a/ → /b/ → /a/ の各音素の特徴的な音が発音されている際の調音位置をそれぞれ、赤丸→黒丸→緑丸で示している。この例では、全ての被験者において、両唇音が発話されているタイミング（図の黒丸）において、上唇（UL）と下唇（LL）が最も接近していることが確認できる。

4 まとめ

ヒトの発話時の声帯震度状態、唇、舌等の発話器官の運動（調音運動）と音声との関係を明らかにするための基礎的データを取得するため、3 次元磁気センサシステムを構築し、構築したシステムを用いて、複数話者による日本語調音・音声データベース構築に向けた調音・音声同期データを取得した。本同期データにおける調音データは、3 次元磁気センサシステムにより測定した調音器官の 3 次元位置データとなっている。発話文としては、音素バランス文である ATR503 文を採用することで、様々な音素環境下でのデータを取得できるようにした。本研究で得られた調音・音声同期データを発展させて、データベースという形で公開することができれば、ヒトの発話機構の詳細の解明に資するとともに、観測された調音運動の変化パターンと音声の関係を音声圧縮に利用できれば、音声伝送時に必要な通信容量を減らし、より効率的な情報通信へも貢献できるものと考えている。

【参考文献】

- [1] 鎬木時彦, 譽田雅彰, 津村尚志, “音素ラベル付き調音・音声対コードブックの検索に基づく調音運動からの音声合成法の検討,” “日本音響学会誌, 54, 3, 207-214(1998).
- [2] T. Toda, A. W. Black, K. Tokuda, “Statistical mapping between articulatory movements and acoustic spectrum using a Gaussian mixture model,” “Speech Commun., 50, 215-227(2008).
- [3] P. L. Tobing, T. Toda, G. Neubing, S. Saki, S. Nakamura and A. Purwarianti, “Articulatory controllable speech modification based on statistical feature mapping with Gaussian mixture models,” 15th Annual Conf., Int. Speech commun. Assoc.(2014).
- [4] Z. C. Liu, Z. H. Ling, and L. R. Dai, “Articulatory-to-acoustic conversion using BLSTM-RNNs with augmented input representation,” “Speech Commun., 99, 161-172(2018).
- [5] B. Denby, T. Schult, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brunberg, “Silent Speech Interfaces,” “Speech Commun., 52, 4, 270-287(2010).
- [6] C. S. Blackburn and S. Young, “A self-learning predictive model of articulatory movements during speech production,” “J. Acoust. Soc. Am., 107, 1659-1670(2000).

- [7] T. Okadome and M. Honda, "Generation of articulatory movements by using a kinematic triphone model," J. Acoust. Soc. Am., 110, 353-463(2001).
- [8] I. Y. Ozbek, M. Hasegawa-Johnson and M. Demirekler, "Estimation of articulatory trajectories based on Gaussian Mixture model (GMM) with audio-visual information fusion and dynamic Kalman smoothing," IEEE Trans. Audio Speech Lang. Process., 19, 1180-1195(2011).
- [9] T. Baer, J. C. Gore, S. Boyce and P. W. Nye, "Application of MRI to the analysis of speech production," Magn. Reson. Imaging, 5, 1-7(1987).
- [10] P. W. Schönle, K. Gräbe, P. Wenig, J. Höhne, J. Schrader and B. Conrad, "Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract," Brain Lang., 31, 26-35(1987).
- [11] Y. I. Kuzmin, "Mobile palatography as a tool for acoustic study of speech sounds," Proc. 4th Int. Cong., on Acoust., G35(1962).
- [12] S. Kiritani, "X-ray microbeam method for the measurement of articulatory dynamics: Technique and results," Speech Commun., 45, 119-140(1986).
- [13] Y. Akgul, C. Kambhamettu and M. Stone, "Extraction and tracking of the tongue surface from ultrasound image sequences," IEEE Comput. Vision Pattern Recog., 124, 298-303(1998).
- [14] A. Wrench, "The MOCHA-TIMIT articulatory database," <http://www.sctr.ed.ac.uk/artic/mocha.html>(1999).
- [15] K. Richmond, P. Hoole and S. King, "Announcing the electromagnetic articulography(day 1) subset of the mngu0 articulatory corpus," Proc. Interspeech 2011, 1505-1508(2011).
- [16] F. Rudzicz, A. K. Namasivayam and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," Lang. Resour. Eval., 46, 523-541(2012).
- [17] S. Narayanan, A. Toutios, V. Ramanarayanan, A. Lammert, J. Kim, S. Lee, K. Nayak, Y. C. Kim, Y. A. Katsamanis and M. Proctor, "Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC)," J. Acoust. Soc. Am., 136, 1307-1311(2014).
- [18] H. Uchida, K. Wakamiya and T. Kaburagi, "Improvement of measurement accuracy for the three-dimensional electromagnetic articulograph by optimizing the alignment of the transmitter coils," Acoust. Sci. Tech., 37, 3, 106-114(2016).
- [19] K. Wakamiya, H. Uchida and T. Kaburagi, "Alignment of the Transmitter Coils in the Three-dimensional Electromagnetic Articulography having Eight Transmission channels," Proc. Youngnam-Kyushu Joint Conf. on Acoust. 2017, J15-1-4(Jan. 2017).
- [20] P. Fabre, "Un procede electrique percetance d'inscription d l'accolement glottique au coours de la phonation – glottographie de haute frequence-," Bull Acad Natl Med, 144, 66-69(1957).
- [21] <https://github.com/julius-speech/segmentation-kit>

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
日本語の調音-音声データベースの拡張	日本音響学会 2024 年度秋季研究 発表会	2024 年 9 月(予定)