

# 発話観測システム NDI Wave のセンサの改良

北村 達也<sup>1,a)</sup> 能田 由紀子<sup>2,b)</sup> 波多野 博顕<sup>3</sup> 吐師 道子<sup>4</sup> 西谷 実<sup>5</sup>

**概要:** Northern Digital Inc. の Wave speech research system は調音器官に貼り付けた小型のセンサの位置を磁気を利用して追跡する発話観測システムである。観測点は限られるものの空間分解能および時間分解能がともに高いという特長をもつ。本研究では、このシステムのセンサに接続されたワイヤを細く柔軟性のあるものに変更し、本体装置とはコネクタで接続できるようにした。オリジナルのセンサとの比較を行い、これらの改良による位置計測精度への影響はないことを確認した。今回の改良によって、センサを調音器官に貼り付けたときに被験者が感じる違和感は激減し、調音運動への影響も減少した。また、センサと本体装置との接続に要する手間も大幅に軽減された。

## 1. はじめに

調音運動のほとんどは体の外から直接観察することができないため、様々な調音運動観測法が考案されてきた。その中の1つに磁気センサシステムがある。この技術は、下顎、舌、口唇などの観測点に貼り付けた複数のセンサの動きを磁気を利用して追跡するものである。この方法は、観測点は限定されるものの、高い時間・空間分解能を得ることができ、さらに観測後のデータ処理が容易であるという特徴を有する [1]。また、磁気共鳴画像法 (MRI) と異なり観測時に騒音が生じることがない点も音声研究にとって有利である。

Northern Digital Inc. の Wave speech research system (以下 Wave と略) は、調音運動観測用磁気センサシステムの1種である。このシステムは、フィールドジェネレータと呼ばれる装置によりその近傍に磁場を発生させ、その磁場内でのセンサの動きを観測する。センサのサイズは

3 × 3 × 2 mm で、このセンサの動きを標準化周波数 400 Hz (最大) で追跡することが可能である。このシステムには、可搬性がある、被験者の姿勢を選ばない、センサのキャリブレーションが不要であるといった他の磁気センサシステムにはない特長がある。このシステムの精度については Berry が詳細に報告している [2]。日本では 2012 年に Wave のレンタルおよび技術支援サービスが始まり、音声生成の研究のみならず、音声合成技術における調音運動情報の活用、医療や外国語教育などの分野での利用が期待されている。

しかし、このシステムにはまだいくつか課題もある。第1に、上述のセンサに接続されているツイストペアワイヤ (1本のワイヤの直径は 0.4 mm) が柔軟性に欠けるため、発話を阻害する可能性がある。第2に、センサのワイヤを本体装置に接続する部分がターミナルブロックになっているため、接続に非常に手間がかかる上、実験中にワイヤが外れてしまうことがある。第3に、付属のソフトウェアの使い勝手が悪く、かつ結果の分析には機能が不十分と言わざるを得ない。ただし、第3の問題に関しては、Ouni らにより開発されている調音データ表示システム VisArtico [3] が Wave のデータ形式に対応したことによって、状況はいくぶん改善されている。そこで、本研究では第1、第2の問題に対して改善を試みた。

本稿では、我々が行ったセンサ改良について説明した後、改良後のセンサの動作確認および改良の効果に関して行った実験について報告する。

## 2. センサの改良

Wave による調音運動観測では、空間座標の基準となる

<sup>1</sup> 甲南大学知能情報学部知能情報学科  
Faculty of Intelligence and Informatics, Konan University,  
Kobe, Hyogo 658-8501, Japan

<sup>2</sup> ATR-Promotions 脳活動イメージングセンタ  
ATR-Promotions, Brain Activity Imaging Center, “Keihanna Science City”, Kyoto 619-0288 Japan

<sup>3</sup> 神戸大学大学院国際文化学研究所メディア文化研究センター  
Research Center for Media and Culture, Graduate School of Intercultural Studies, Kobe University, Hyogo 657-8501, Japan

<sup>4</sup> 県立広島大学保健福祉学部コミュニケーション障害学科  
Faculty of Health and Welfare, Prefectural University of Hiroshima, Mihara, Hiroshima 723-0053, Japan

<sup>5</sup> アドバンスシステムズ株式会社  
Advanced Systems Co., Ltd. Tachikawa, Tokyo 190-0022 Japan

a) t-kitamu@konan-u.ac.jp

b) ynota@atr.jp

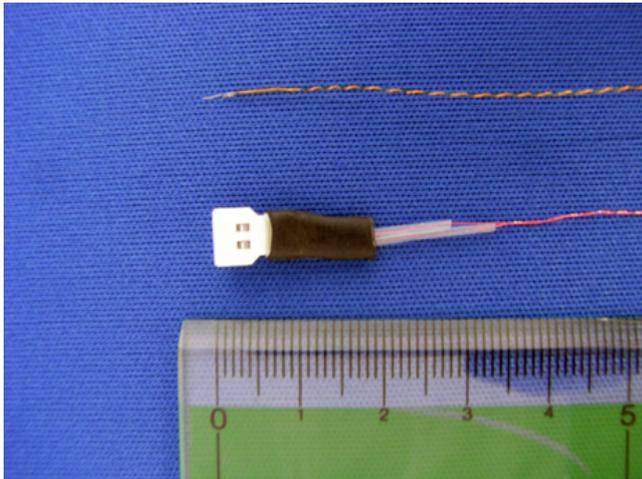


図 1 オリジナル (上) と改良後 (下) のセンサのワイヤ。改良後のセンサの末端にはコネクタを取り付けてある。

6 自由度 (6D) センサ 1 個と調音器官に貼り付ける複数の 5 自由度 (5D) センサを使用する。このうち、本研究では 5D センサのワイヤを細く柔軟性のあるものに交換し、さらに本体装置との接続部をコネクタに変更した。

センサの部分はもともと防水のためシリコンで被覆されている\*1。このシリコンを取り除き、ワイヤを極細 (直径 0.1020 mm) かつ柔軟なエナメル線と交換した。ノイズ対策のため、ワイヤは撚り線にした。ワイヤの交換後、センサの部分に歯科用シリコン印象材を塗布することによって改めて防水処理を施した。さらに、センサのワイヤと本体装置はモレックス社のコネクタで接続するよう改良した。改良前後のセンサの写真を図 1 に示す。

### 3. 実験 1 センサの動作確認

センサのワイヤ交換が Wave のセンサ位置計測精度に影響するか否かを静的状態と動的状態で調査した。あわせて、フィールドジェネレータからの距離による影響も調べた。

#### 3.1 実験条件

プラスチック板にオリジナルのセンサと改良後のセンサを 2 個ずつ約 23 mm 間隔に固定した。そして、静的状態の実験においては、フィールドジェネレータから各センサまでの距離が約 60 mm, 150 mm, 250 mm となるようプラスチック板を固定した。動的状態の実験においては、フィールドジェネレータから約 60 mm, 150 mm, 250 mm の位置にてプラスチック板を手で動かした。いずれの実験においても約 5 秒間のセンサ位置を標準化周波数 400 Hz で記録した。計測データから 2 つのセンサ間距離を求め、その平均値と標準偏差を求めた。

\*1 センサは舌に貼り付けることがあるため防水の必要がある。

表 1 静的状態での実験において計測データより求めたセンサ間距離の平均値と標準偏差 [mm]。フィールドジェネレータ (FG) からの距離による差異も示している。

	Distance from FG		
	60 mm	150 mm	250 mm
Original	23.56 (0.00)	23.40 (0.03)	24.11 (0.17)
Improved	22.28 (0.00)	22.18 (0.02)	22.20 (0.19)

表 2 動的状態での実験において計測データより求めたセンサ間距離の平均値と標準偏差 [mm]。フィールドジェネレータ (FG) からの距離による差異も示している。

	Distance from FG		
	60 mm	150 mm	250 mm
Original	23.23 (0.11)	23.29 (0.09)	23.93 (0.46)
Improved	22.14 (0.05)	22.14 (0.09)	22.05 (0.39)

### 3.2 実験結果

静的状態および動的状態での実験における、オリジナルのセンサ間の距離および改良後のセンサ間の距離をそれぞれ表 1, 表 2 に示す。2 種のセンサで得られたセンサ間距離が異なるが、これはセンサの性能に由来するものではなく、センサをプラスチック板に固定した際の位置の精度が原因であると考えられる。

オリジナルと改良後のセンサ間距離の標準偏差はほとんど変わらない。ワイヤを変更することによってノイズを受けやすくなることが懸念されたが、この結果はその影響がないことを示している。これはワイヤを撚り線にしたことが功を奏している。試作の当初はワイヤを撚り線にしていなかったため、ノイズの影響が極めて大きかった。

フィールドジェネレータからの距離によりセンサ間距離の平均値は変化し、標準偏差は増加した。これは、Berry [2] の報告と同様である。調音運動の測定の際には、可能な限り被験者の頭部に近い位置にフィールドジェネレータを配置するのが望ましいといえる。

### 4. 実験 2 センサ改良の効果

#### 4.1 実験条件

##### 4.1.1 被験者

著者のうち 2 名を被験者とした。以後、被験者 A, B と呼ぶ。

##### 4.1.2 センサの装着

上述のように Wave ではリファレンス用の 6D と一般用の 5D のセンサを使用する。6D センサはナジオン (nasion, 鼻根点) に、5D センサは図 2 に示す正中面上の 7 カ所に貼り付けた。舌尖を用いる子音の調音に配慮し、T1 は舌尖から約 10 mm 後方に貼り付けた。

舌 (T1, T2, T3) と歯茎 (UI と LI) へのセンサの接着には医療用接着剤アロンアルファ A (東亜合成) を用いた。この接着剤を使うと、10 秒ほどでセンサを舌に接着することができる。我々の経験では、センサは 1 時間程度で自

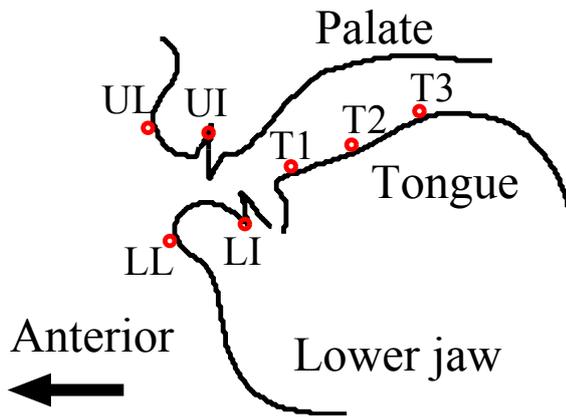


図 2 センサの装着位置 (UL: upper lip, UI: upper incisor, LL: lower lip, LI: lower incisor, Tn: tongue no. n)

然に外れ、皮膚を痛めるようなことはない\*2。ナジオンへの 6D センサの固定、口唇 (UL と LL) への 5D センサの固定は、つけまつげ用接着剤およびサージカルテープにて行った。

#### 4.1.3 データ収集

被験者 A のデータ収集はオリジナル、改良後のセンサの順序で行い、被験者 B のデータ収集はこの逆の順序で行うことによって、順序効果を相殺した。いずれの場合も、調音運動の計測の合間に口蓋用センサで口蓋形状を計測した。

発話資料は昔話「桃太郎」の冒頭部分である。センサ貼り付け完了の直後 (1 回目)、10 分後 (2 回目)、20 分後 (3 回目) にデータを収集した。調音運動の標本化周波数は 400 Hz である。Berry [2] による Wave の精度評価の結果および実験 1 の結果を参考に、フィールドジェネレータから正中面までの距離を 20 cm 以内に保つようにした。被験者にはデータ収集の合間に実験者と会話させて、可能な限りセンサを装着した状態での発話に慣れてもらうようにした。

調音運動と同時にマイクロフォン (Sony ECM-77B) と USB オーディオ (M-Audio Fast Track) を用いて音声も収録した。この他、センサ装着せずに発話した音声も録音した。録音の際の標本化周波数は 22.05 kHz、量子化ビット数は 16 bit である。

#### 4.1.4 音声分析

音声データを標本化周波数 16 kHz にダウンサンプリングした後、フレーム長 32 ms、フレーム周期 8 ms、窓関数として Hanning 窓を用いて対数スペクトルを求めた。そして、周波数軸上で平均した。

### 4.2 実験結果

#### 4.2.1 センサ装着作業への影響

柔軟性のあるワイヤに変更したことによって、ワイヤの取り回しが容易になった。オリジナルのワイヤはこすれ合うとひっかかるような抵抗があるため、ワイヤの配置に神

\*2 無理に外そうとすると皮膚を痛める可能性がある。

経を使う必要があったが、改良後のワイヤではそのようなことがない。その一方で、改良後のワイヤは強い力がかかると断線することがあることがわかった。

また、オリジナルのセンサを用いた場合には、ワイヤをターミナルブロックに接続するのに非常に手間がかかるのみならず、実験中にワイヤがターミナルブロックから抜けることがあった。しかし、センサのワイヤと本体装置との接続部をコネクタにした結果、接続が一瞬で完了し、実験中に抜けることも皆無となった。

#### 4.2.2 聴感上の影響

まず、被験者 A を対象にして音声を聴取し違和感のある部分を選び出した。被験者 A がオリジナルのセンサを用いて計測した 1 回目の音声は、口の中に異物があるような印象が感じられた。特に、(1)「し」、(2)「じ」、(3)「した」、(4)ラ行の音節に違和感が感じられた。例えば、「むかし」、「おじいさん」、「きました」、「しばかりに」、の下線部である。これらはいずれも調音時に舌背や舌尖を口蓋や歯茎に接触もしくは接近させるものである。2 回目の音声では 1 回目で感じられた違和感は減少し、3 回目の音声ではほとんど問題がないといえるまでに減少した。

これに対して、被験者 A が改良後のセンサを用いて計測した場合には、1 回目から発話への影響が感じられなかった。

次に、被験者 B を対象にして音声を聴取し違和感のある部分を選び出した。被験者 B は被験者 A と比較してセンサ装着による影響が小さかった。1 回目と 2 回目の音声において、(1)「し」と (2)ラ行の音節に発話しにくそうな印象が若干感じられたが、3 回目の音声には違和感は感じられなかった。

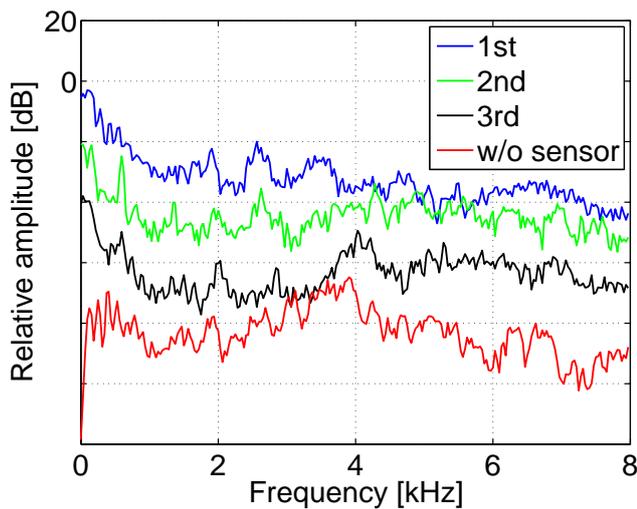
被験者 B が改良後のセンサを用いて計測した場合には、1 回目の音声において「し」に若干の違和感が感じられたものの、2 回目以降は発話への影響が感じられなかった。

#### 4.2.3 対数スペクトル上の影響

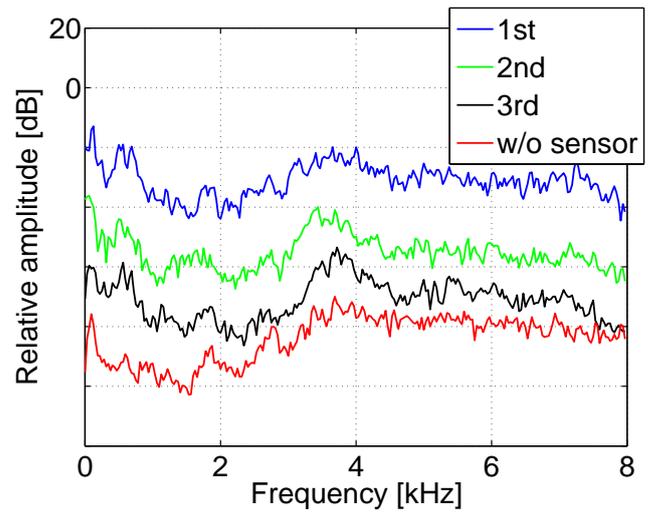
被験者 A がオリジナルのセンサを装着して発話した「むかし」の [j] の対数スペクトルを図 3(a) に示す。比較のため、センサを装着せずに発話した音声の対数スペクトルも示している。なお、各対数スペクトルには 20 dB ずつバイアスを加えている。

図 3(a) によると、1 回目、2 回目の音声の対数スペクトルはセンサ未装着時のものと大きく異なっている。高域に向かって振幅が減衰しており、[j] よりむしろ [h] に近い音になっていることがわかる。3 回目には約 4 kHz 以上の振幅が増大し、本来の [j] の対数スペクトルに近づいているが、それでもセンサ未装着時の音声の対数スペクトルとの間に差異がある。

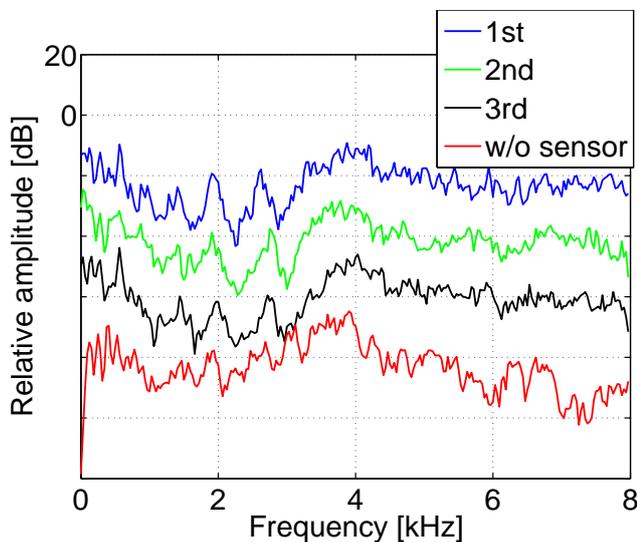
被験者 A が改良後のセンサを装着して発話した「むかし」の [j] の対数スペクトルを図 3(b) に示す。図 3(a) と同様に、センサを装着せずに発話した音声の対数スペクトル



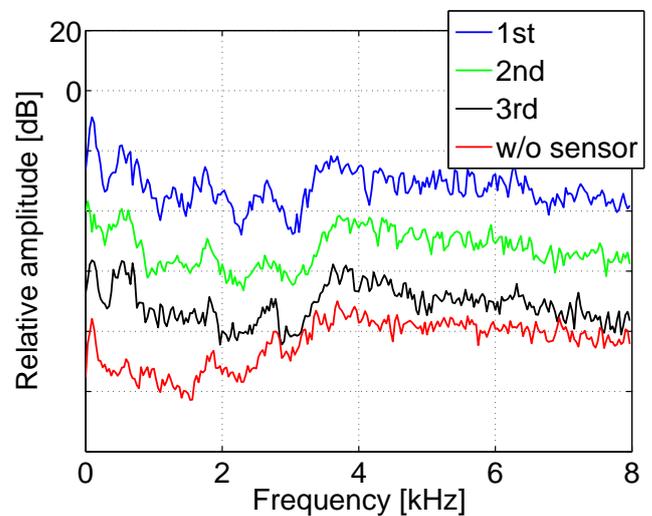
(a) オリジナルのセンサ



(a) オリジナルのセンサ



(b) 改良後のセンサ



(b) 改良後のセンサ

図 3 被験者 A が (a) オリジナル, (b) 改良後のセンサを装着して発話した「むかし」の [j] の対数スペクトル。青: 1 回目の音声, 緑: 2 回目の音声, 黒: 3 回目の音声, 赤: センサー装着なしの音声。各対数スペクトルには 20 dB ずつバイアスを加えている。

も示している。改良後のセンサを用いた場合、1 回目から 3 回目の対数スペクトルの概形にほとんど違いがない。また、いずれの場合も 4 kHz 付近にピークをもつハイパス特性を示している。これはセンサ未装着時の音声と同じ傾向である。

次に、被験者 B がオリジナルのセンサを装着して発話した「むかし」の [j] の対数スペクトルを図 4(a) に示す。被験者 B は被験者 A よりもセンサ装着による影響が小さい。被験者 B の対数スペクトルは、1 回目から 4 kHz 付近にピークをもっているが、その振幅は 2 回目以降より小さい。2 回目以降はセンサ未装着時の対数スペクトルとほぼ同じ形状であった。

被験者 B が改良後のセンサを装着して発話した「むか

図 4 被験者 B が (a) オリジナル, (b) 改良後のセンサを装着して発話した「むかし」の [j] の対数スペクトル。青: 1 回目の音声, 緑: 2 回目の音声, 黒: 3 回目の音声, 赤: センサー装着なしの音声。各対数スペクトルには 20 dB ずつバイアスを加えている。

し」の [j] の対数スペクトルを図 4(b) に示す。被験者 A と同様に、1 回目から 3 回目の対数スペクトルの概形にほとんど違いがない。

#### 4.2.4 調音運動への影響

被験者 A がオリジナル, 改良後のセンサを装着して発話した「むかし」の [j] 部のセンサ位置を図 5, 6 に示す。後者では上顎歯茎部 (UI) のセンサが不調であったため表示されていない。

オリジナルのセンサを装着した場合には、センサ貼り付けから時間が経つに従って T1 の動作範囲が広がっている。一方、改良後のセンサを装着した場合には、3 回の発話を通じて T1 の位置に大きな差異は見られない。前者の場合、時間の経過とともにセンサ装着状態で発話することに慣れていったこと、および自分の音声をフィードバックす

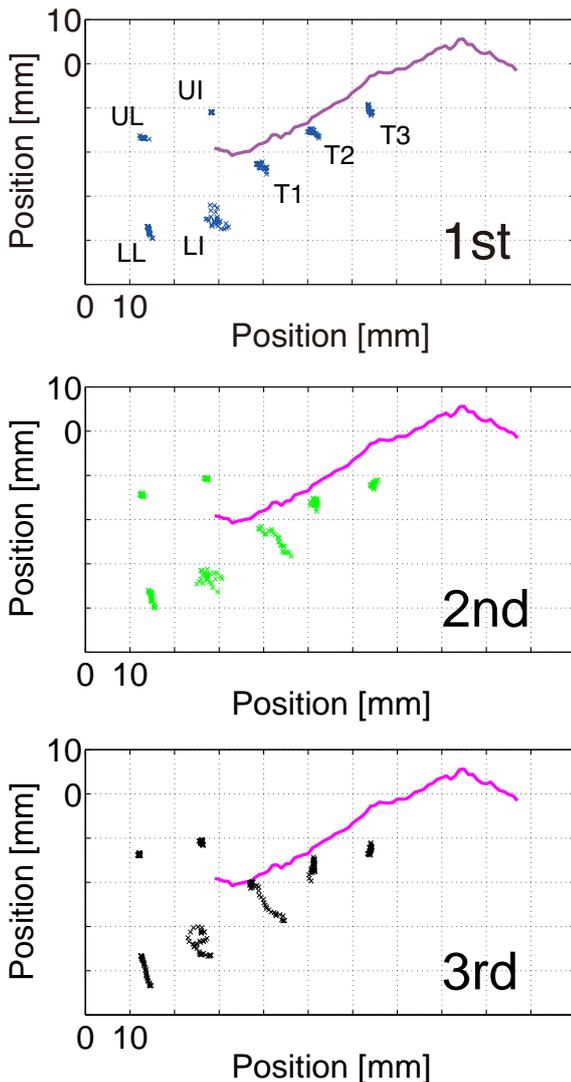


図5 被験者Aがオリジナルのセンサを装着して発話した「むかし」の「j」のセンサ位置。マゼンタの線は口蓋を表している。

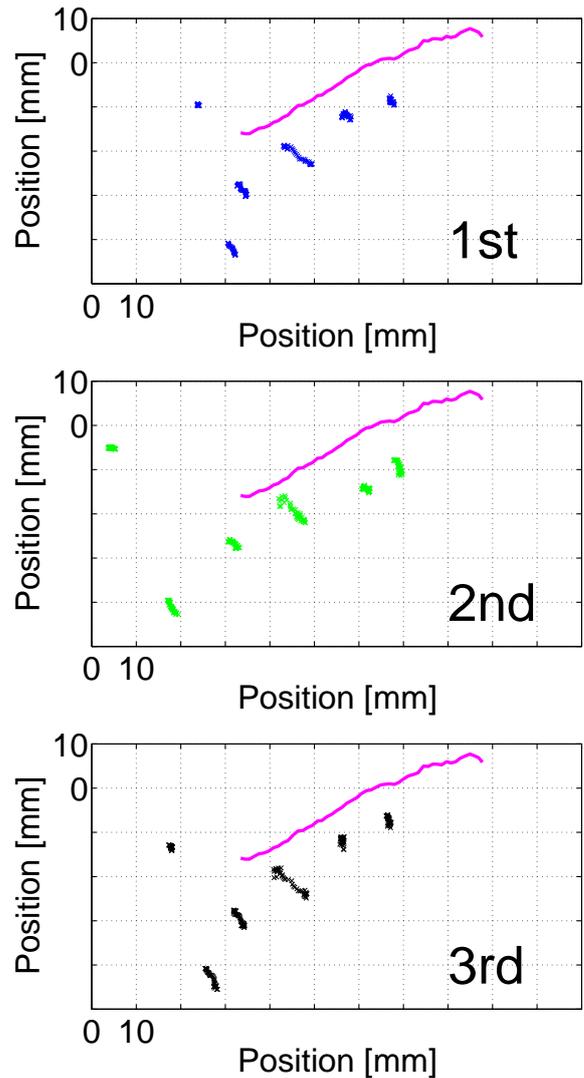


図6 被験者Aが改良後のセンサを装着して発話した「むかし」の「j」のセンサ位置。

ることによってセンサ装着状態でも未装着時と同様の音声  
を生成できるよう調音運動を調節したことが原因と考えら  
れる。一方、後者の場合は、センサのワイヤによる発話へ  
の影響が小さいため1回目から通常通りの発話ができたと  
解釈できる。

なお、これらの調音運動の結果は図3に示した被験者A  
の対数スペクトルと対応している。すなわち、前者の場合  
は3回の発話において調音位置のせばめが変化したため対  
数スペクトルが変化し、後者の場合は調音位置のせばめが  
ほとんど変化しなかったため対数スペクトルが変化しな  
かったものと推測される。

次に、被験者Bがオリジナル、改良後のセンサを装着し  
て発話した「むかし」の「j」部のセンサ位置を図7、8に示  
す。被験者Aの結果と同様に、後者では上顎歯茎部(UI)  
のセンサが不調であったため表示されていない。

被験者Bの場合は、オリジナル、改良後のセンサを装着  
した場合ともに3回の発話を通じてセンサ位置にほとんど

差異がない。これらの結果は、被験者Bの対数スペクトル  
が3回の発話を通じてほとんど変化しなかった図4の結果  
と対応している。

図7、8では口蓋とT1の距離に違いが見られる。前者で  
は約2mmであるのに対して、後者では約6mmである。  
被験者Aでも同様の傾向が見られる。この違いがセンサの  
ワイヤに由来するものなのか、計測上の問題によるものな  
のかは引き続き調査が必要である。

## 5. おわりに

本研究では、NDI社の調音運動観測用磁気センサシス  
テムWaveのセンサのワイヤを細く、柔軟性のあるものに  
交換し、さらにセンサと本体装置との接続部をコネクタに  
変更した。センサのワイヤの交換によって、発話への影響  
は軽減され、センサ装着直後から計測を行えるようになった。  
ただし、オリジナルのセンサが発話に及ぼす影響は被  
験者によって差があった。また、コネクタの改良によっ

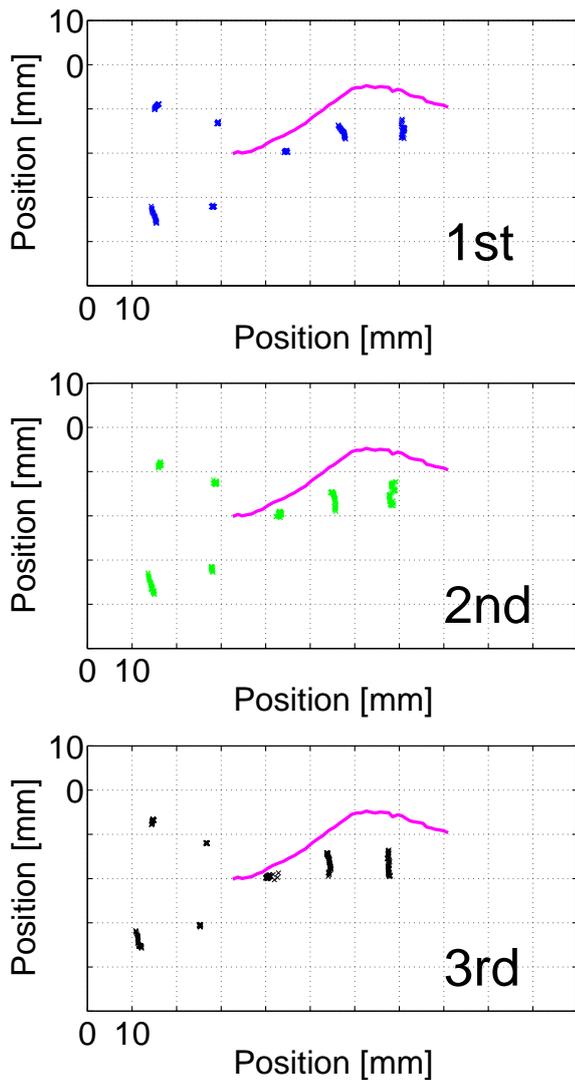


図 7 被験者 B がオリジナルのセンサを装着して発話した「むかし」の [j] のセンサ位置.

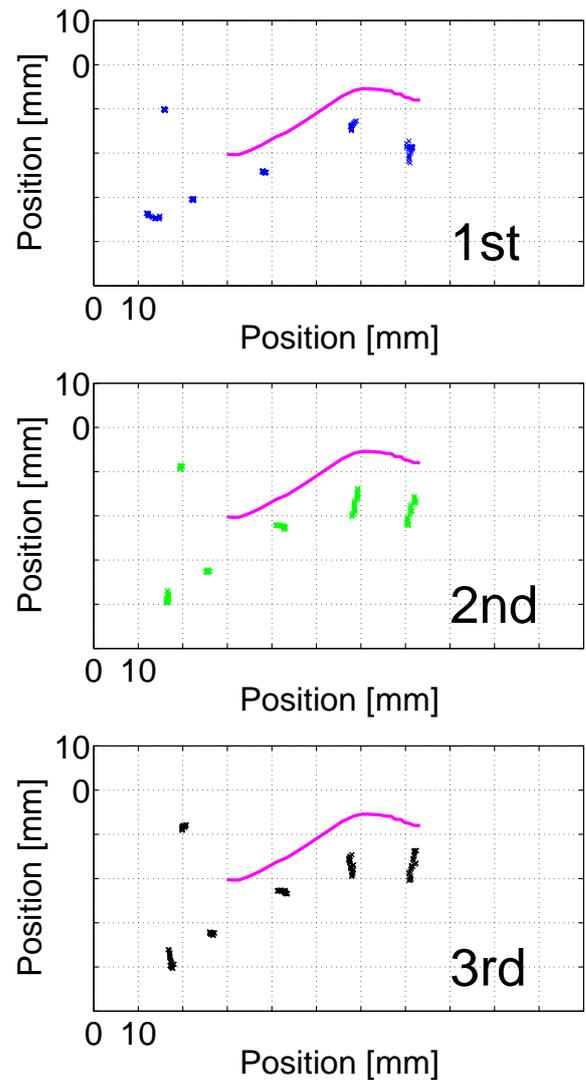


図 8 被験者 B が改良後のセンサを装着して発話した「むかし」の [j] のセンサ位置.

て、Wave を用いる実験の準備に要する手間が大幅に軽減された。

Wave は簡単に調音運動と音声を同時計測することができるシステムである。本研究で行ったセンサの改良が、音声生成のみならず、幅広い分野に貢献することを期待したい。

**謝辞** 本研究の一部は平成 24 年度科研費基盤研究 (B)(21300071), 同挑戦的萌芽研究 (24652085), 平成 24 年度私立大学等経常費補助金の支援を得て行われた。センサ改良にご協力いただきましたコバテル (株) 小林興弘氏, ATR-Promotions 正木信夫氏に深謝いたします。

#### 参考文献

- [1] 籾木時彦, 正木信夫, 元木邦俊, 松崎博季, 北村達也: 音声生成の計算モデルと可視化, コロナ社 (2010).
- [2] J.J. Berry: Accuracy of the NDI Wave speech research system, J. Speech, Language, and Hearing Research, 54, 1295–1301 (2010).

- [3] S. Oumi, L. Mangeonjean and I. Steiner: VisArtico: A visualization tool for articulatory data, Proc. of Interspeech2012 (2012).