

歌唱発声における発声難度の音高・音色依存性に関する 分析的検討

電気電子工学科 03-180480 小林海斗

2020/2/7

目次

第 1 章	序論	2
1.1	研究の背景	2
1.2	研究の目的	2
1.3	本論文の構成	2
第 2 章	関連研究	3
2.1	日本語単語における発声しやすさの自動評価	3
2.2	歌声特有の F_0 動的変動成分	3
第 3 章	理論的背景	6
3.1	母音とフォルマント周波数	6
第 4 章	音高・音色における発声難度の調査	8
4.1	実験条件	8
4.2	実験結果	9
第 5 章	分析結果	13
5.1	主観評価との比較	13
5.2	単音発声の安定性	14
5.3	F_0 の周波数変調	15
第 6 章	結論	18
6.1	総括	18
6.2	今後の課題	18
付録 A	他被験者の収録データ	20

第1章

序論

1.1 研究の背景

自然な歌唱は

従来の研究としては宋らの研究では調音運動に基づいて発声難度を求め、それを歌詞自動生成の指標として用いている [?]. また、佐藤らの研究では調音運動に加えて音響特徴量に着目し、日本語単語の発声しやすさを評価している [?]. しかしこれらは調音的な側面が強く、歌声の要素の一つである音高による影響が考慮されていない。

1.2 研究の目的

本研究では、音声合成や歌詞作成などへの応用のために発声難度の音高・音色依存性を分析することを目的とする。使用する音声データは被験者による母音に限定した歌声を収録したもの用いる。得られるデータに対して分析を行い、音高・音色依存性を探る。また、収録プロトコルを確立することで今後の発声難度分析の一助となることも本研究の目的である。

1.3 本論文の構成

本論文は全 6 章で構成される。第 2 章では、日本語単語における発声しやすさの自動評価に関する先行研究と、 F_0 動的変動成分に関する先行研究について述べる。第 3 章では、本研究で用いる F_0 抽出の理論的背景、母音と構音の関係について述べる。第 4 章では、収録実験の詳細と得られたデータに対する客観的評価について述べる。第 5 章では、得られたデータに対して行った分析、およびその分析から導かれる発声難度への影響要素について述べる。そして第 6 章では、結論と今後の課題について述べる。

第2章

関連研究

2.1 日本語単語における発声しやすさの自動評価

佐藤らは、日本語単語の発声難度を識別するための特微量として調音データから抽出した特微量と音声データから抽出した特微量に一定の有効性があることを示した [?].

2.1.1 調音運動及び音響特微量に着目した特微量の提案

発話の長さが異なる場合、フレーム分割による特微量抽出は次元が合わずそのまま比較できないため、ここでは表 2.1 の特微量を用いている。

2.1.2 作成データセットを用いた発声しやすさの識別

主観実験で得られたデータに対し、表 2.1 の特微量を用いてサポートベクターマシンによって分類を行う。

ラベルは 2 種類作成している。単語に付与された 5 段階スコアとともに、スコア 3.0 を境界として発声難度を定める方式 (with inntermediate: w/ mid) と、スコア 2.0 以下と 4.0 以上ののみを用いる方式 (without inntermediate: w/o mid) である。後者は中間スコアを含まず、主観評価がより安定しているものみを識別に用いているといえる。

これによる正解率、適合率、再現率、F 値を示したものを見たものを表 2.2 に示す。w/ mid では正解率は 70% 前後で特微量による大きな差はなかったが、w/o mid では全ての特微量で正解率が上がり、音響特微量に基づく cepstrum-based、mfc-based が特に高い正解率を示した。

2.2 歌声特有の F_0 動的変動成分

齋藤らは様々な歌声データの F_0 変化パターンを分析し、以下に示す 4 種類の F_0 動的変動成分が歌唱スタイルや歌唱者に関係なく存在することを明らかにした [?].

オーバーシュート (Overshoot) 滑らかな音高変化、およびその直後に目的音高を超える瞬時的な変動成分

表 2.1 発声しにくさの指標として機能することを期待する特微量

タグ	抽出する特微量	次元数
articulation-based1	12 次元の調音データの各次元の時系列 データに対する平均、ケプストラムの先頭 12 次元	156
articulation-based2	2 次元の調音データの各次元の時系列 データ、時系列データの一次微分、二次微分それぞれに対する平均、分散、最大値、最小値、最小二乗法によって一次関数、二次関数に近似した際の係数および近似誤差	396
cepstrum-based	音声データのケプストラム、 Δ 特微量、 $\Delta\Delta$ 特微量各 13 次元に対する平均、分散、最大値、最小値、最小二乗法によって一次関数、二次関数に近似した際の係数および近似誤差	429
mfcc-based	音声データの MFCC、 Δ 特微量、 $\Delta\Delta$ 特微量各 13 次元に対する平均、分散、最大値、最小値、最小二乗法によって一次関数、二次関数に近似した際の係数および近似誤差	429

表 2.2 データセットを用いた SVM による分類

ラベリング	特微量	正解率	適合率	再現率	F 値
w/ mid	articulation-based1	62.9%	38.7%	33.3%	35.8%
	articulation-based2	75.0%	59.5%	61.1%	60.3%
	cepstrum-based	72.4%	56.3%	50.0%	52.9%
	mfcc-based	70.7%	52.8%	52.8%	52.8%
w/o mid	articulation-based1	75.4%	47.1%	53.3%	50.0%
	articulation-based2	75.4%	47.3%	60.0%	52.9%
	cepstrum-based	90.8%	75.5%	86.7%	81.3%
	mfcc-based	81.5%	57.9%	73.3%	64.7%

ヴィブラート (Vibrato) 同一音高区間で観測される 4~8Hz の準周期的な変動成分

プレパレーション (Preparationn) 音高の変化直前に変化とは逆方向に触れる瞬時的な変動成分

微細振動 (Fine fluctuation) 発声区間全体に観測される 15~20Hz 程度の不規則で細かい変動成分

図 2.1 はアマチュア歌手による日本童謡「七つの子」の F_0 変化パターンおよび F_0 動的変動成分である。これから上記の微細振動を除く変動成分が生じていることが読み取れる。

このような F_0 動的変動成分は歌声を知覚する上で重要な役割を担っており、自然な歌声合成の要素の一つとして応用が可能であると考えられる。

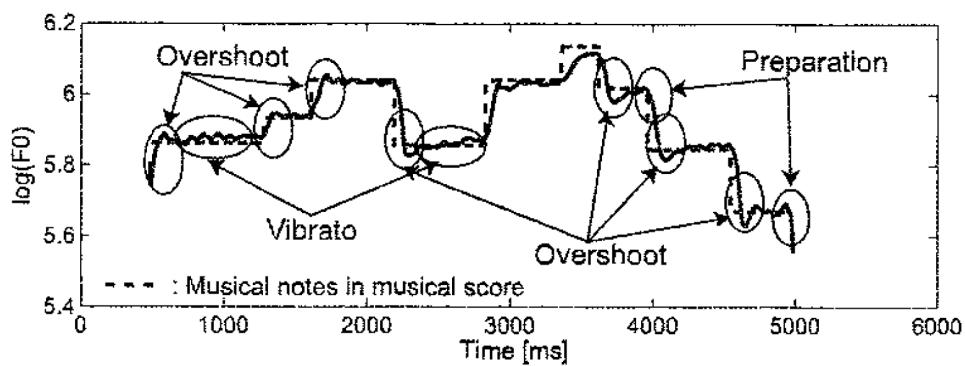


図 2.1 アマチュア歌唱者の歌声における F_0 動的変動成分の例 [?]

第3章

理論的背景

3.1 母音とフォルマント周波数

フォルマント周波数は母音や声質の決定において非常に重要である。フォルマント周波数は声道スペクトルのピークとなる共鳴周波数のことであり、低い方から順に第1フォルマント (F_1)、第2フォルマント (F_2)、と名付けられる。

調音器官の運動は、フォルマント周波数全てに影響を与える。第1 フォルマントは顎の開きに敏感で、開き具合が大きくなるほど第1 フォルマント周波数は上昇する。第2 フォルマントは特に舌の形状に大きく影響を受け、声道前方を狭めるときに第2 フォルマント周波数はもつとも高くなる。逆に、軟口蓋や咽頭部を狭める時には低くなる。第3 フォルマントは舌尖の位置、正確には前歯のすぐ後ろの空間に影響を受け、空間が大きいほど第3 フォルマント周波数は低くなる。

フォルマント周波数のうち特に F_1 、 F_2 は母音を特徴付ける重要なパラメータであり、これらの分布によって母音を分類することができる [?]. 図 3.1 は日本語 5 母音における F_1 、 F_2 を示したものである。

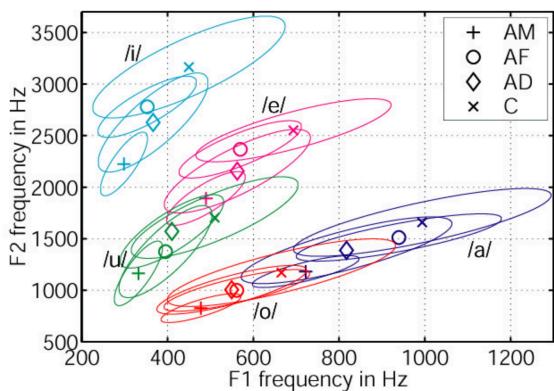


図 3.1 日本語 5 母音の E_1 – E_2 [?]

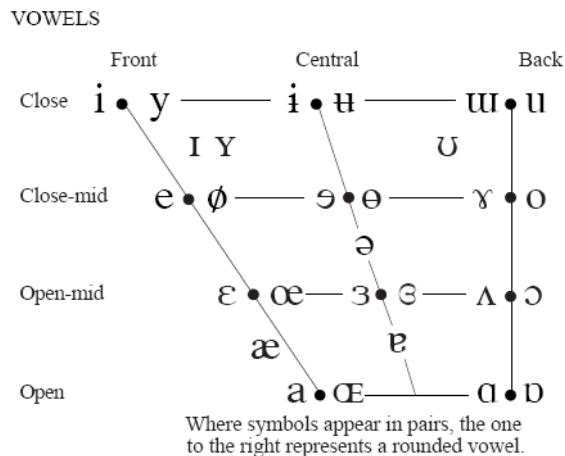


図 3.2 IPA 母音チャート

ここで AM は成人男性、AF は成人女性、AD は青年、C は子供である。一般に、平均的な声が高いほどフォルマント周波数も高い傾向にある。成人男性を例に挙げると、 F_1 は 250~1000 Hz、 F_2 は 600~2500 Hz、 F_3 は 1700~3500 Hz である[?]。

また、図 3.2 は IPA が定めた母音チャートである^{*1}。縦軸は口の開き具合に対応し、上に行くほど狭くなる。横軸は舌の前後位置に対応し、左に行くほど前寄りである。これは適切に反転と回転を行うことで図 3.1 と重なることが知られている。

^{*1} <https://www.internationalphoneticassociation.org/content/ipa-vowels>

第4章

音高・音色における発声難度の調査

歌唱能力は個人差に依るところが大きいが、全体としておおまかな発声難度の傾向はあると考えられる。一般には/i/の母音では高音域が出しにくいとされている。女性は男性と比較してフォルマント周波数が高く、音域も異なることから同時に傾向を掴むのは難しいと考え、ここでは男性に限定して調査を行う。男性の地声と裏声の声区の重複する区間は 200~350Hz (G3~F4) であるが、この声区の幅、境界は個人により大きく異なる。そのため、使用音域は男性の平均的な地声の範囲に収まるように設定した。

4.1 実験条件

被験者

実験は 20 代男性 4 人を対象として行った。

収録音声

発声は/a/、/i/、/u/、/e/、/o/を用いた。音高は平均律における F3 (174.6Hz)、A3 (220.0Hz)、C4 (261.6Hz)、A4 (440.0Hz) の 4 音を使用した。サンプリング周波数は 44100Hz で、モノラル形式で収録を行った。

実験内容

実験内容は以下に示す通りである。

共通項目

- ・全てテンポ 120 で行う。
- ・各実験はそれぞれ 2 回ずつ行う。
- ・初めに音源を流しながら発声練習する時間を設け、発声に慣れたことが確認でき次第収録を開始する。
- ・発声後に、発声難度を主観で評価する。この評価は「発声しにくい」「どちらとも言えない」「発声しやすい」の 3 段階で行う。
- ・収録と同時に、正面から発声時の口部の様子を録画する。

単音

8 拍 (4 秒) のカウントの後、/a/の発声で F3 の音高を 16 拍 (8 秒) 伸ばした後 8 拍 (4 秒) 休憩する。これを音高を A3、C4、F4 と変えて行う。発声を/i/、/u/、/e/、/o/に変えた場

合においても同様に行う。発声が 5 通り、音高が 4 通りで合計 20 回収録する。

上昇

4 拍（2 秒）のカウントの後、/a/ の発声で F3 の音高を 2 拍（1 秒）伸ばし、その後 /a/ の発声で A3 の音高を 2 拍（1 秒）伸ばす。これを、後半の音高 A3 を C4、F4 と変えて行う。前後の発声を /i/、/u/、/e/、/o/ に変えた場合においても同様に行う。発声が 25 通り、音高変化が 3 通りで合計 75 回収録する。

下降

4 拍（2 秒）のカウントの後、/a/ の発声で A3 の音高を 2 拍（1 秒）伸ばし、その後 /a/ の発声で F3 の音高を 2 拍（1 秒）伸ばす。これを、前半の音高 A3 を C4、F4 と変えて行う。前後の発声を /i/、/u/、/e/、/o/ に変えた場合においても同様に行う。発声が 25 通り、音高変化が 3 通りで合計 75 回収録する。

使用機材

使用機材を表 4.1 に示す。

表 4.1 使用機材

使用機材	型番
コンデンサマイク	SONY C-48
ヘッドフォン	SONY MDR-CD480
レコーダー	KORG MR-1000
カメラ	SONY FDR-AX45

4.2 実験結果

以下に被験者 1 の実験結果を示す。他の被験者の結果は付録に載せている。 F_0 抽出には音声合成システム WORLD [?] の Harvest [?] を用いた。Harvest は図 4.1 のように F_0 候補の推定部と推定された F_0 候補から軌跡を生成する部からなるピッチ抽出手法であり、他の手法と比較して耐雑音性が高いことで知られている。なお、 F_0 間隔は 1ms である。

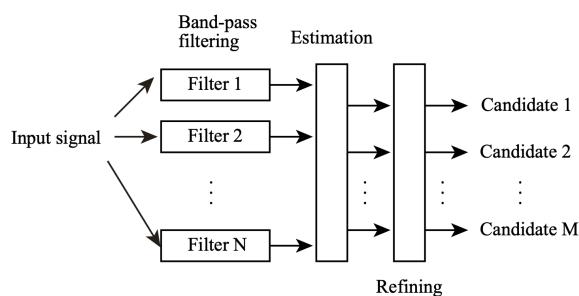


図 4.1 Harvest による F_0 候補推定の流れ [?]

4.2.1 単音

被験者 1 の単音発声の F_0 軌跡を図 4.2、4.3、4.4、4.5、4.6 に示す。歌い出しで音高が外れてしまっている場合でも、1 秒以内に正しい音高に近い音高で発声を修正できている。また、歌い終わりは息が持たなかつたために早めに切ってしまった音声が存在している。

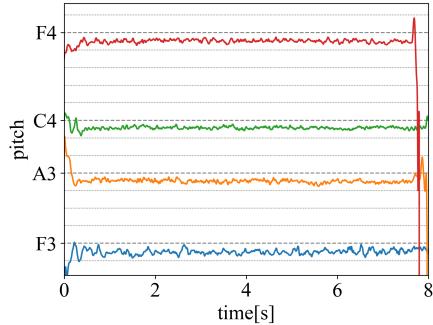


図 4.2 被験者 1 の/a/単音発声

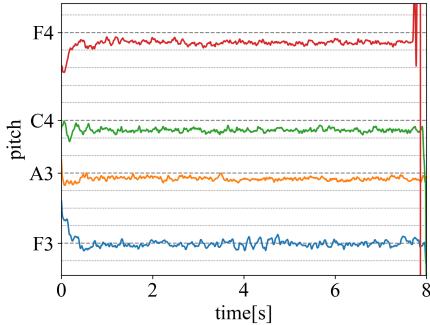


図 4.3 被験者 1 の/i/単音発声

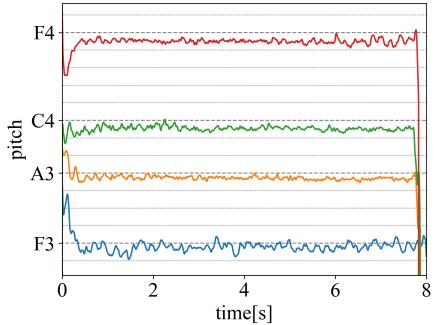


図 4.4 被験者 1 の/u/単音発声

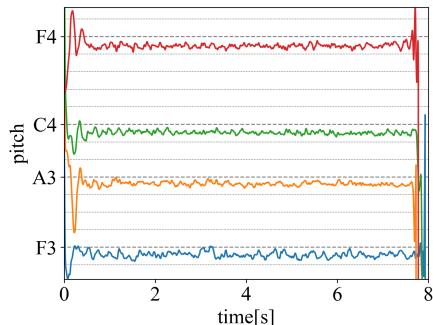


図 4.5 被験者 1 の/e/単音発声

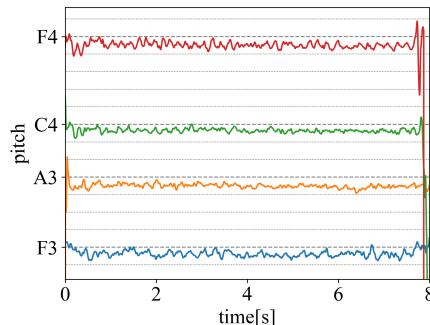


図 4.6 被験者 1 の/o/単音発声

4.2.2 上昇

被験者 1 の上昇音形発声の F_0 軌跡を図 4.7 に示す。左側が上昇前の母音、右側が上昇後の母音を表している。上昇前が/i/のグラフにおいて上昇時に値が暴れているが、これは収録時に被験者が意図せず音を切り無音が発生したことに起因する。一部の音声においてはオーバーシュートおよびプリパレーションの様子が見られる。

4.2.3 下降

被験者 1 の下降音形発声の F_0 軌跡を図 4.8 に示す。こちらは上昇と比べると意図しない無音は発声しにくくなっている。またプリパレーションはほとんど見られず、オーバーシュートが全体的に発生している。

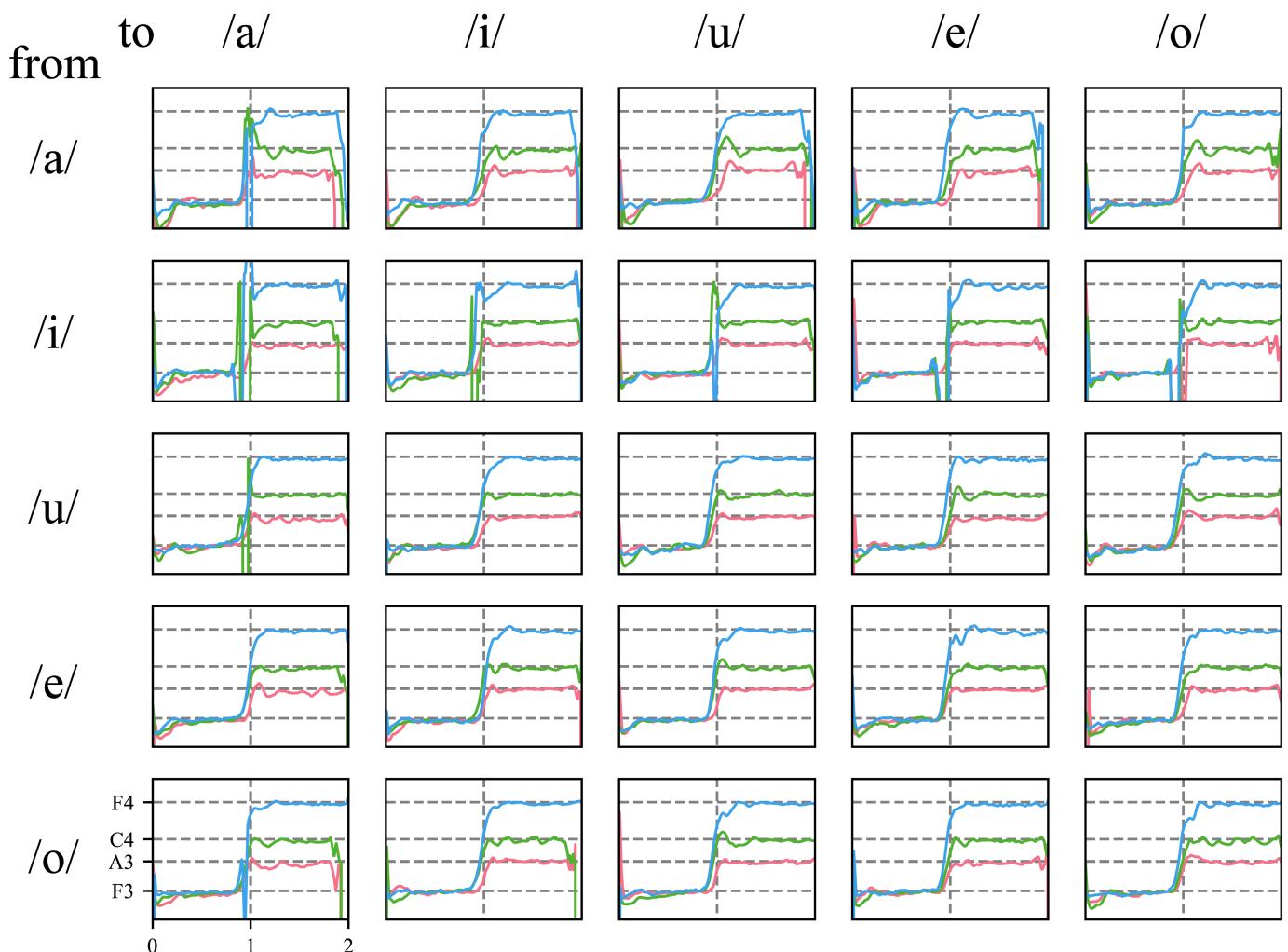


図 4.7 被験者 1 の上昇音形発声

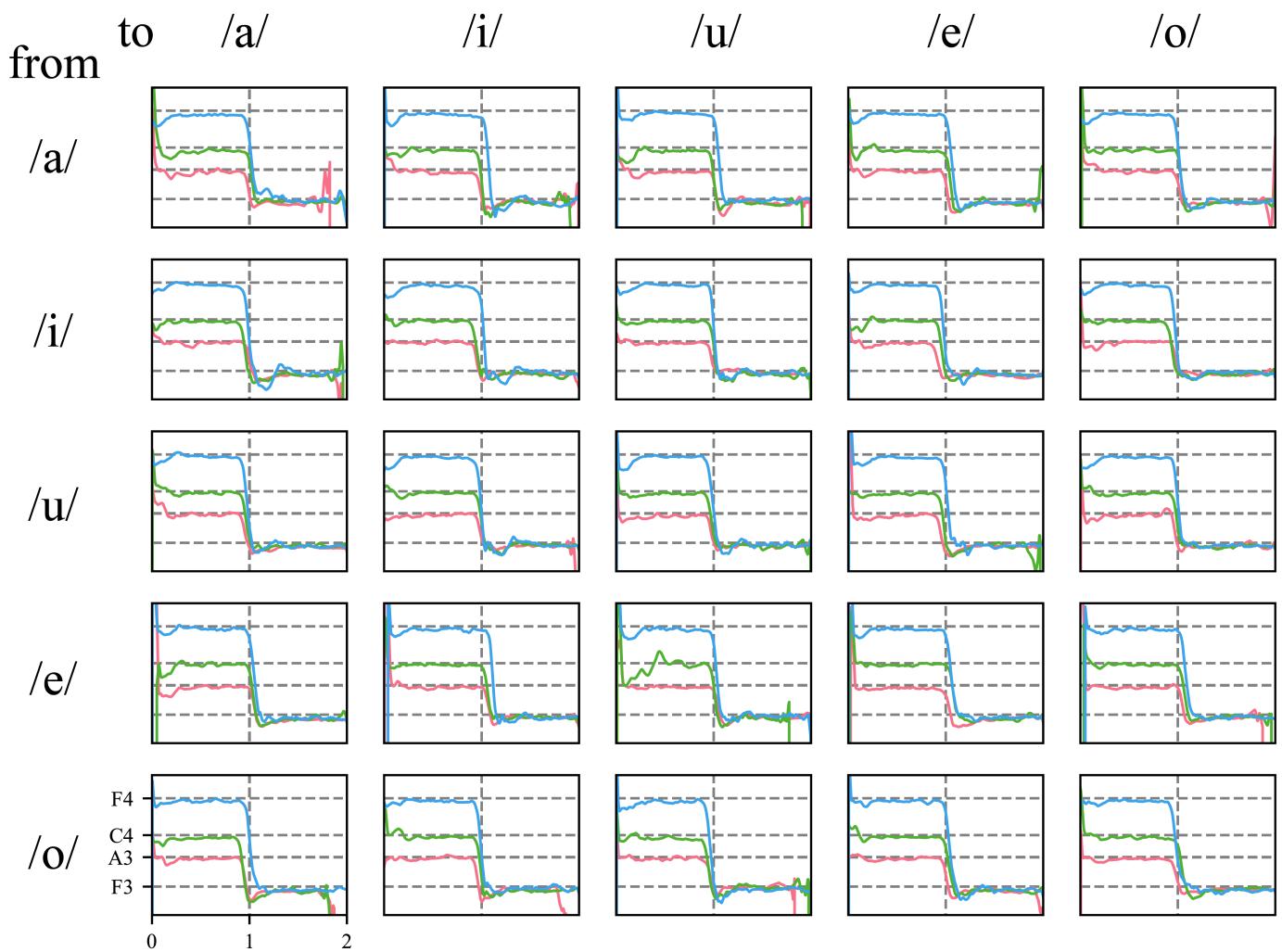


図 4.8 被験者 1 の下降音形発声

第5章

分析結果

前項で得られたグラフだけでは、発声難度との関連性が見出せない。そこで、前項のデータのうち特に単音発声の収録データとその発声難度の主観評価に着目して細かく分析を行い、発声難度との関連性を検討する。

5.1 主観評価との比較

図5.1に、被験者による主観評価と平均音高のずれを示した。

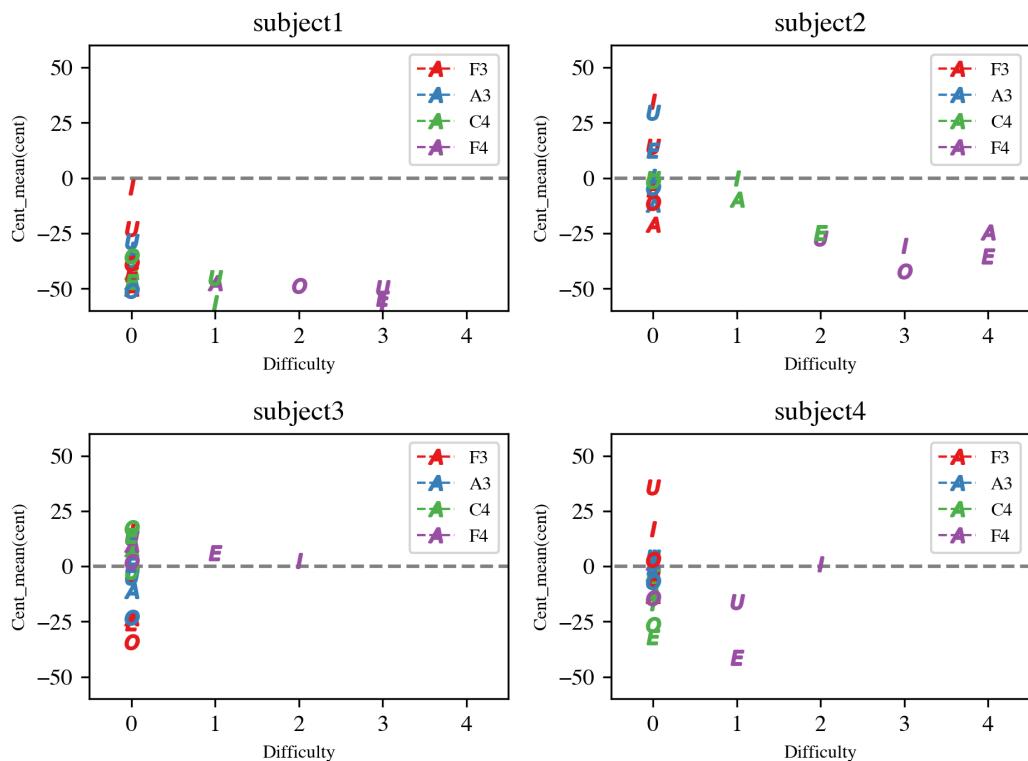


図5.1 主観評価と F_0 平均値の比較

横軸は「発声しやすい」を0点、「どちらとも言えない」を1点、「発声しにくい」を2点としたとき

の 2 回の評価の和で、数値が大きいほど発声しにくいことを表している。縦軸は基準音高からのずれをセントで表している。セントは音程を表す対数単位であり、12 平均律においてオクターブを 1200 等分したものが 1 セントである。すなわち 100 セントで 1 半音のずれを表す。

2 つの音 x と y の周波数が分かれれば、その間のセント値は以下のように求められる。

$$cent = 1200 \log_2 \frac{y}{x} \quad (5.1)$$

どの被験者においても F4 における/i/、/e/の母音で発声しにくく評価しており、音高は正確なピッチから最大で 50cent ほど低くなっている。これは/i/、/u/の母音が狭め母音であるという構音上の性質が影響していると考えられる。また、被験者 1、2 は C4 においても発声しにくさを感じており、被験者 3、4 と比較して声区変換が早期に発生していると考えられる。

5.2 単音発声の安定性

単音発声は 8 秒間同じ音高を発声し続けるため、後半に音程の安定性が崩れることがあった。この安定性を具体的に見るために、時間軸をいくつかの区間に分けて分析を行った。歌い出し、歌い終わりの影響を避けるため、初めと終わりのそれぞれ 1 秒間を除いた 6 秒間を 4 つの区間に分け、それぞれの区間ににおいて平均と標準偏差を求めた。この結果を図 5.2、5.3、5.4、5.5 に示した。

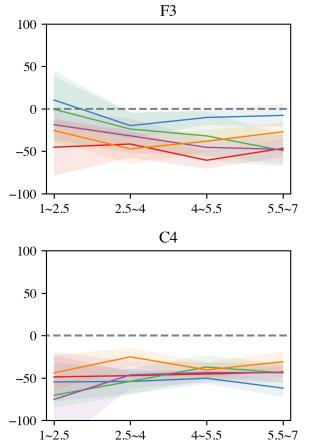


図 5.2 被験者 1 の平均値変化

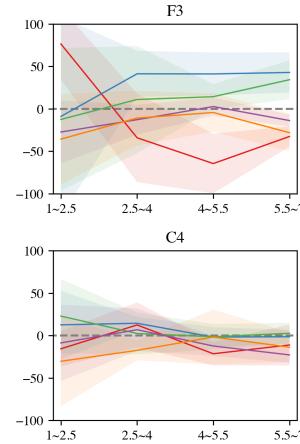
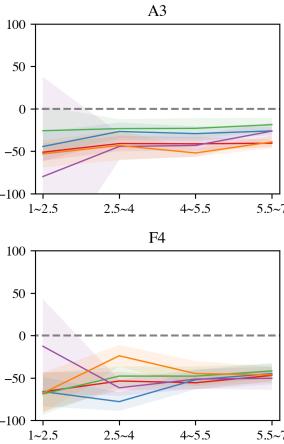


図 5.2 被験者 1 の平均値変化

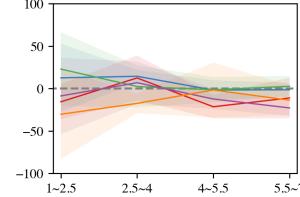
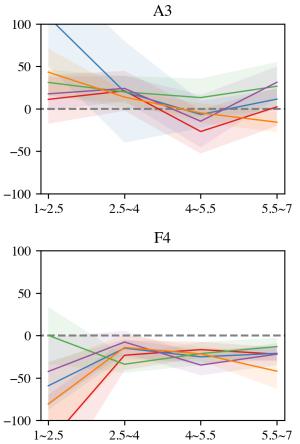


図 5.2 被験者 1 の平均値変化

これらを見ると、音高が上がるほど最終的な音のぶれが小さくなっていることがわかる。このことから、音高を合わせ続けることについては発声時の感覚に反し低音の方が難しいと考えられる。また、F3 において/i/、/u/の音高の平均が他の母音に比べて全体的に高くなってしまっており、低音域では/i/、/u/発声時における歌唱者の知覚する音高が他の母音と比べて相対的に高くなっていることを示唆している。これは/i/、/u/の F_1 が発声している音高と近いことが原因の一つと考えられる。

被験者 1 は音高の目標点が基準より全体的に低めであり、被験者の想定する正しい音高と実際の音高が異なる可能性がある。被験者 2 は高音でも標準偏差が大きいままであり、地声の音域外の音高を無理に発声しようとした結果不安定になったと考えられる。また被験者 3、4 は被験者 1、2 と比べると標準偏差が全体的に小さい。このように、被験者ごとに個人の特性が強く出ていると考えられる。

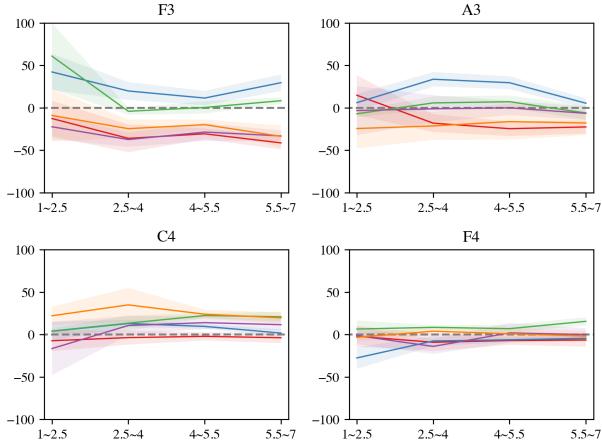


図 5.4 被験者 3 の平均値変化

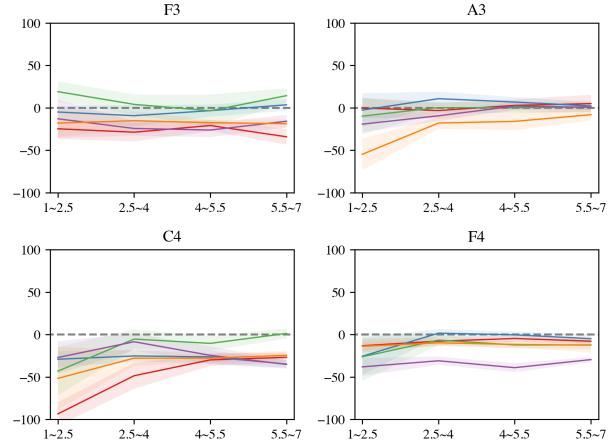


図 5.5 被験者 4 の平均値変化

5.3 F_0 の周波数変調

続いて、発声の不安定性がヴィブラートのように F_0 に現れている可能性から F_0 動的変動成分に着目する。歌い出しと歌い終わりを除く 6 秒間の F_0 軌跡を波形としてフーリエ変換を行い、どのような成分が含まれるかを確かめた。詳細な条件は表 5.1 の通りである。

表 5.1 フーリエ変換の詳細

データ長	6001 点
フレームレート	1000Hz
窓長	500 点
シフト長	50 点

なお、窓長で切り出したデータに対して、平均が 0 になるように値を移動してからハミング窓をかけた。得られたモジュレーションの全体平均をとったものが図 5.6、5.7、5.8、5.9 である。

被験者 3 と 4、特に被験者 4 に強く見られる現象として、/i/、/u/の狭め母音に対して 80Hz 周辺でピークが生じている。この原因の一つとして、音源-フィルタ相互作用の影響が考えられる。/i/、/u/の F_1 は 300Hz 前後であり発声している音高 $F_3(174.6\text{Hz})$ と比較的近いことから、声帯音源の発声機構と声道の間に強い干渉が生じ [?]、音源-フィルタ相互作用によって基本周波数に変化が生じたと考えられる。[?]

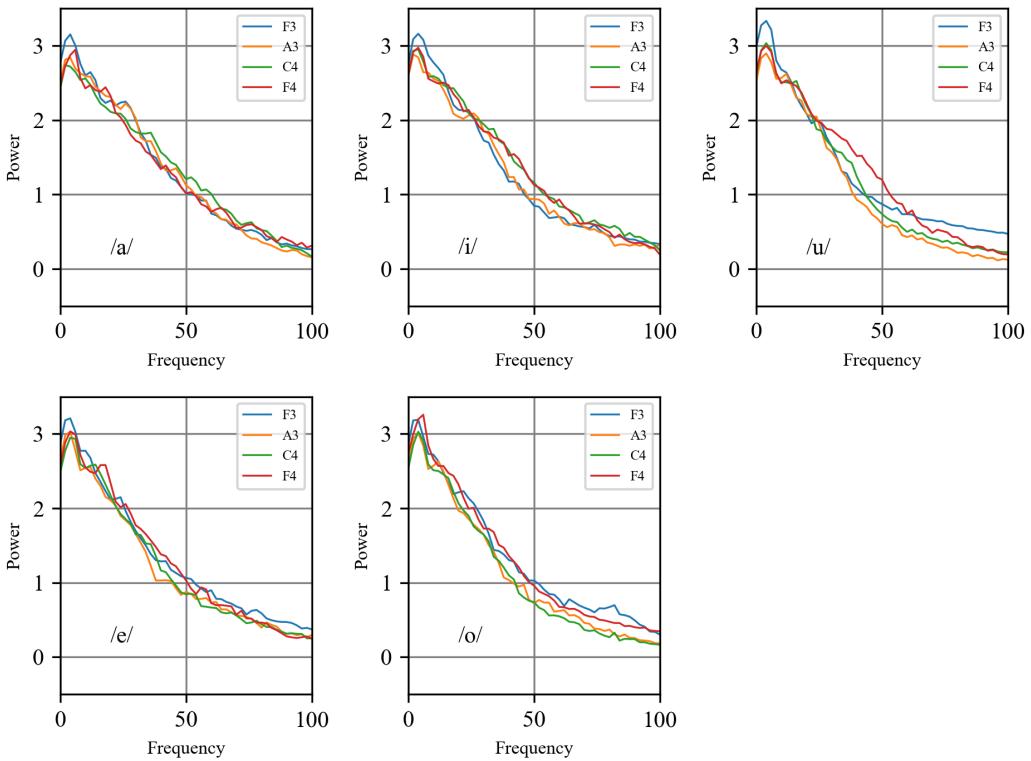


図 5.6 被験者 1 のモジュレーションの全体平均

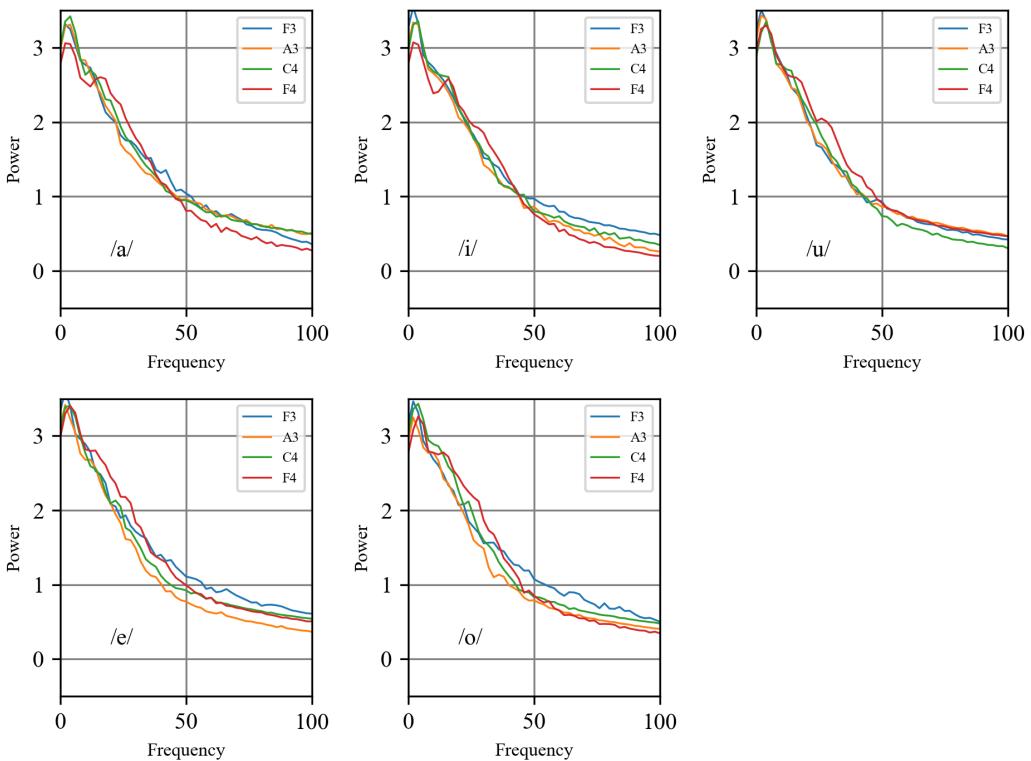


図 5.7 被験者 2 のモジュレーションの全体平均

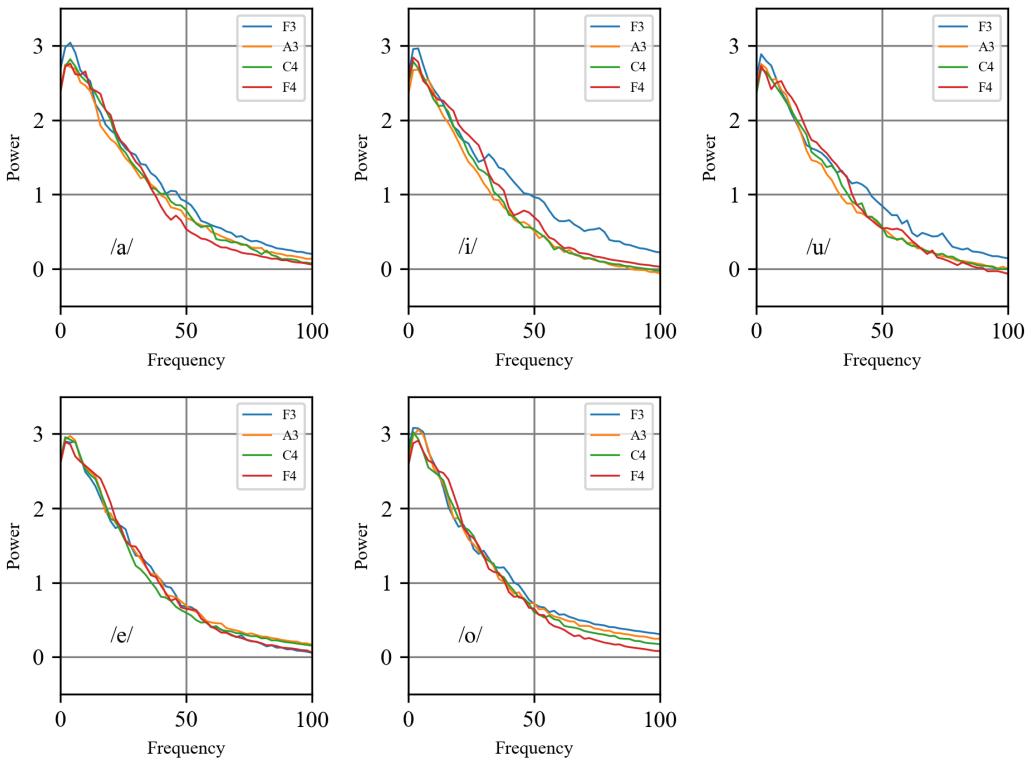


図 5.8 被験者 3 のモジュレーションの全体平均

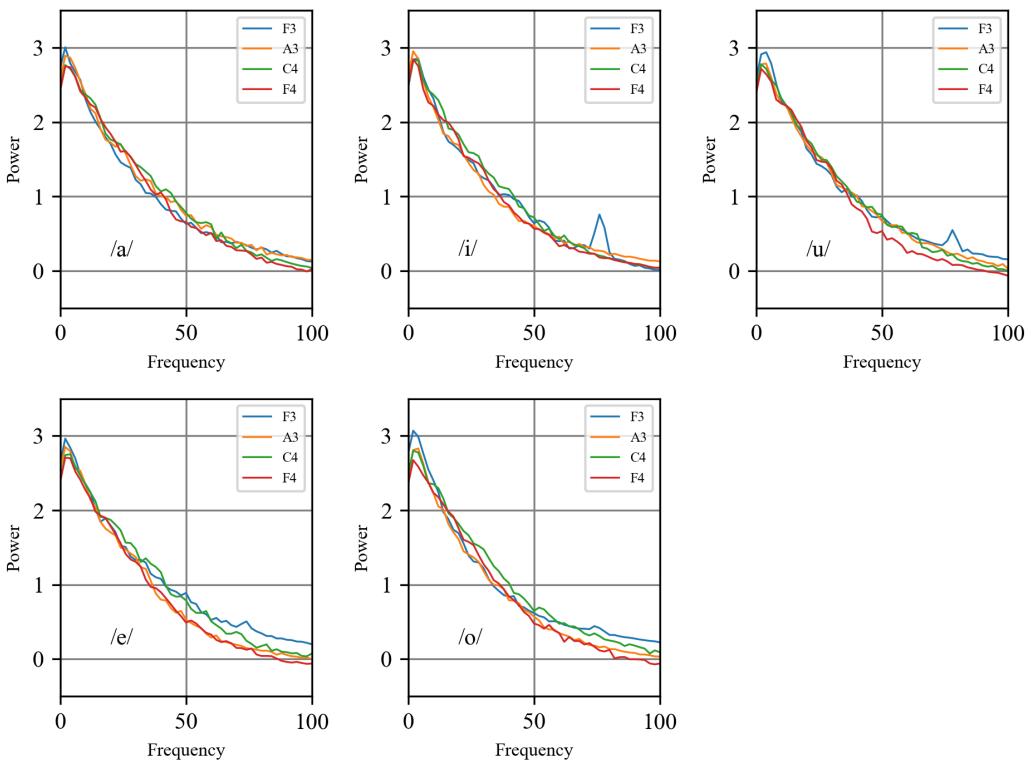


図 5.9 被験者 4 のモジュレーションの全体平均

第6章

結論

6.1 総括

本研究では、音高・音色ごとの発声難度の分析による様々な応用を目指し、これを実現するために被験者らの歌唱音声を収録して主に F_0 に関する分析を行った。

6.2 今後の課題

音高・音色ごとの発声しにくさを分析するために4人の被験者からデータを収録したが、個人性が強く全体に共通する特徴は多く見られなかった。この個人性は歌唱経験の差や歌い方の癖で一定の傾向があると考えられるため、被験者数を増やして同様の実験を行い、傾向ごとに発声難度の分析を行う必要がある。また、収録は短時間に連続して行ったため、疲労による影響が生じていると考えられる。十分な休憩を取るなどして各音ごとに出来る限り平等な条件で収録を行う必要がある。

また、それぞれの母音が正しいものとして議論を進めたが、実際は歌唱における母音の明瞭性の低下も考慮に入れる必要がある。聴取実験を行い、母音の明瞭性や発声の客観的な評価を行う必要があると考えられる。

今回の収録は母音のみに限定しておりメロディもごく簡単なものであったため、より音楽的な歌詞・メロディにおいても収録し、同様に分析を行う必要がある。

謝辞

本研究を進めるにあたって、齋藤大輔講師には指導教員としてテーマをいただいたほか、研究を通して指導していただきました。深く感謝申し上げます。峯松信明教授には、もう一人の指導教員として研究の方針について指導いただきました。深く感謝申し上げます。峯松・齋藤研究室の先輩方にも様々な助言をいただきました。深く感謝申し上げます。また、忙しい時期にもかかわらず実験に参加してくださいました皆様のおかげで研究を進めることができました。深く感謝申し上げます。最後に、私を支えてくれた家族と友人に深く感謝申し上げます。

2020年2月7日

小林 海斗

付録 A

他被験者の収録データ

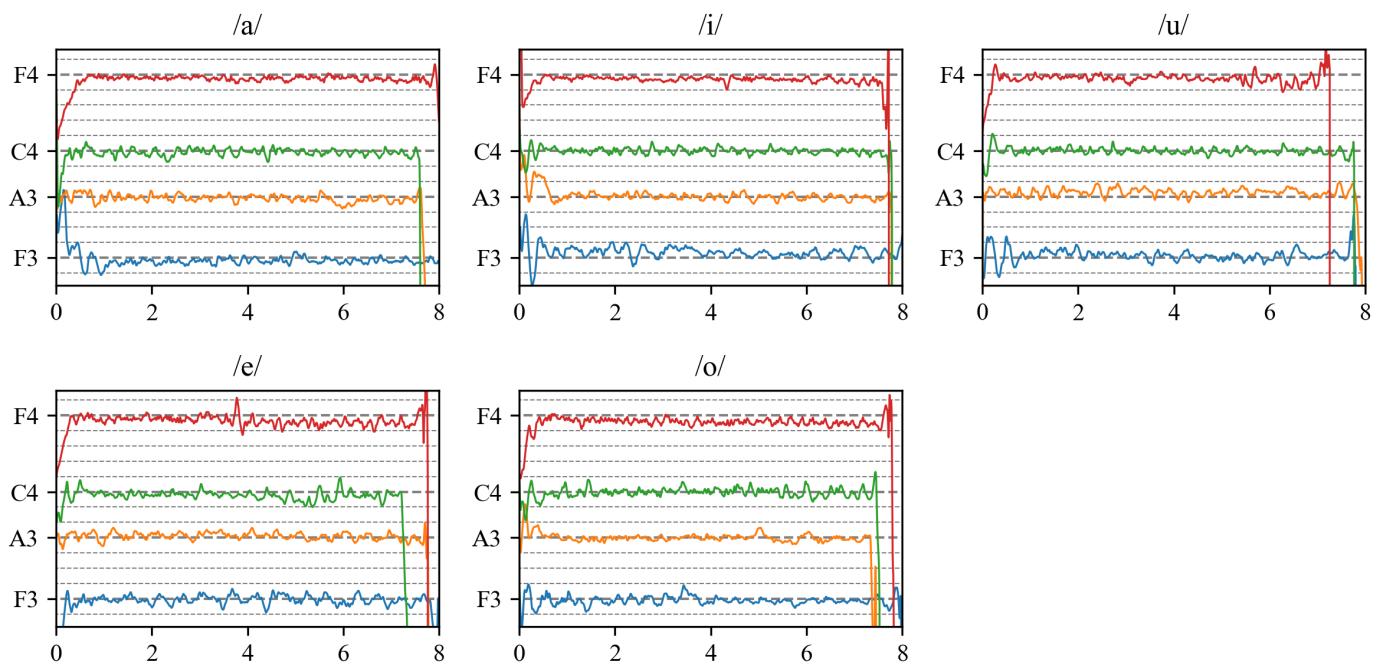


図 A.1 被験者 2 の単音発声

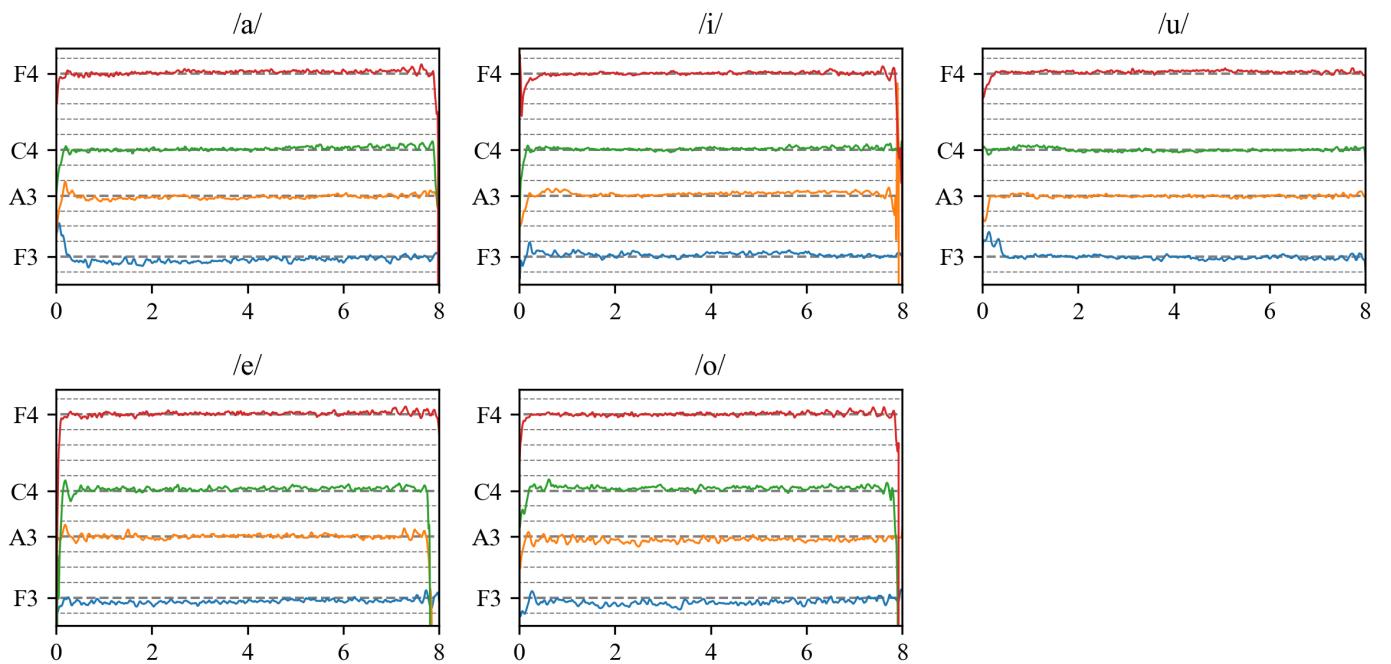


図 A.2 被験者 3 の単音発声

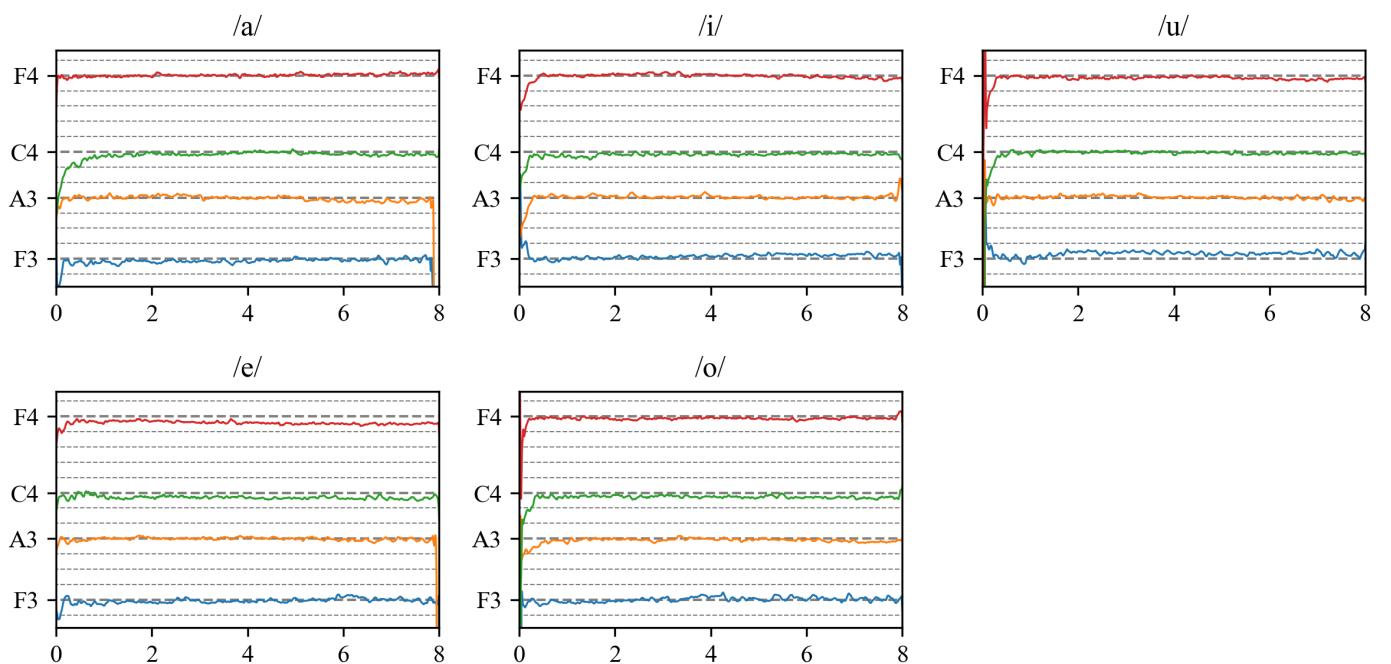


図 A.3 被験者 4 の単音発声

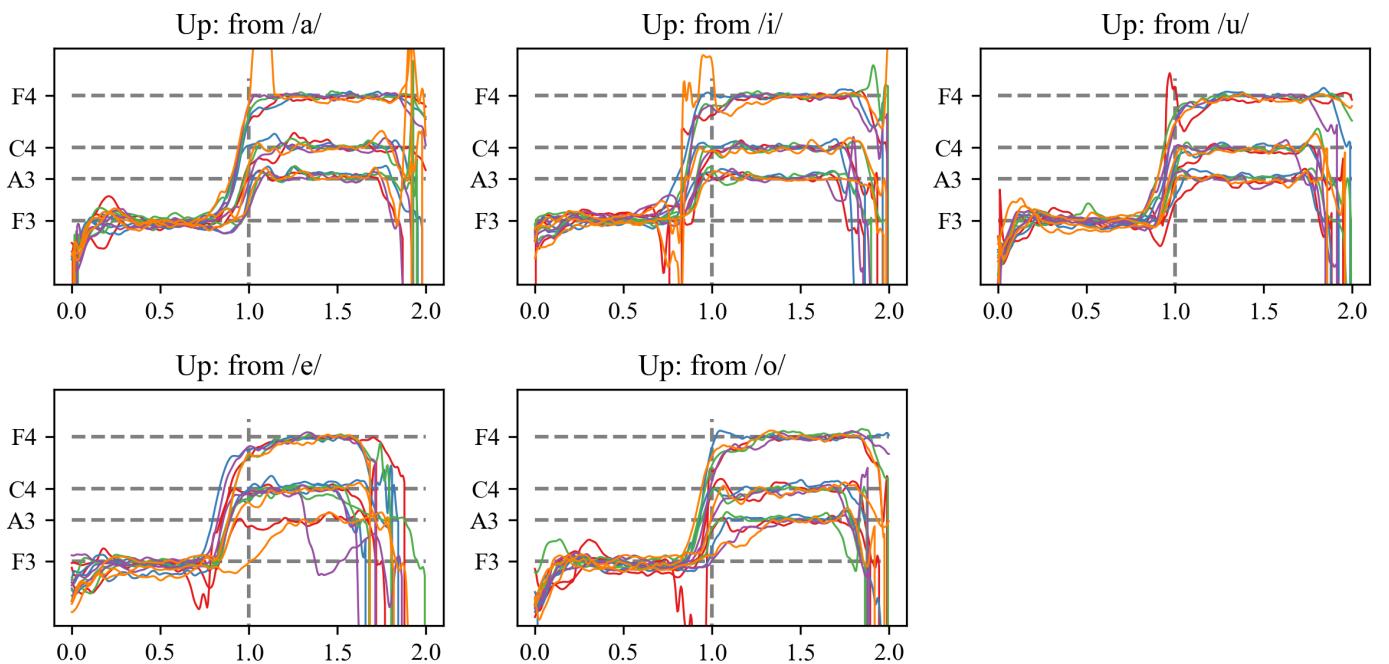


図 A.4 被験者 2 の上昇発声

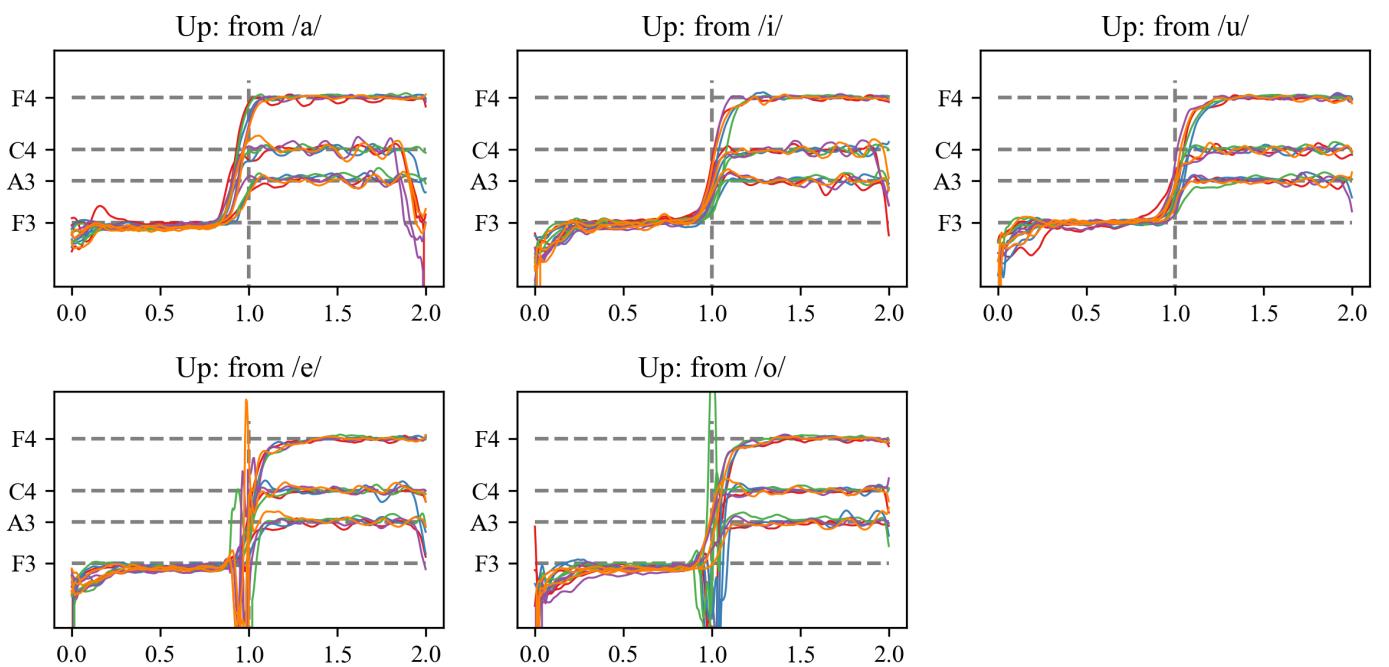


図 A.5 被験者 3 の上昇発声

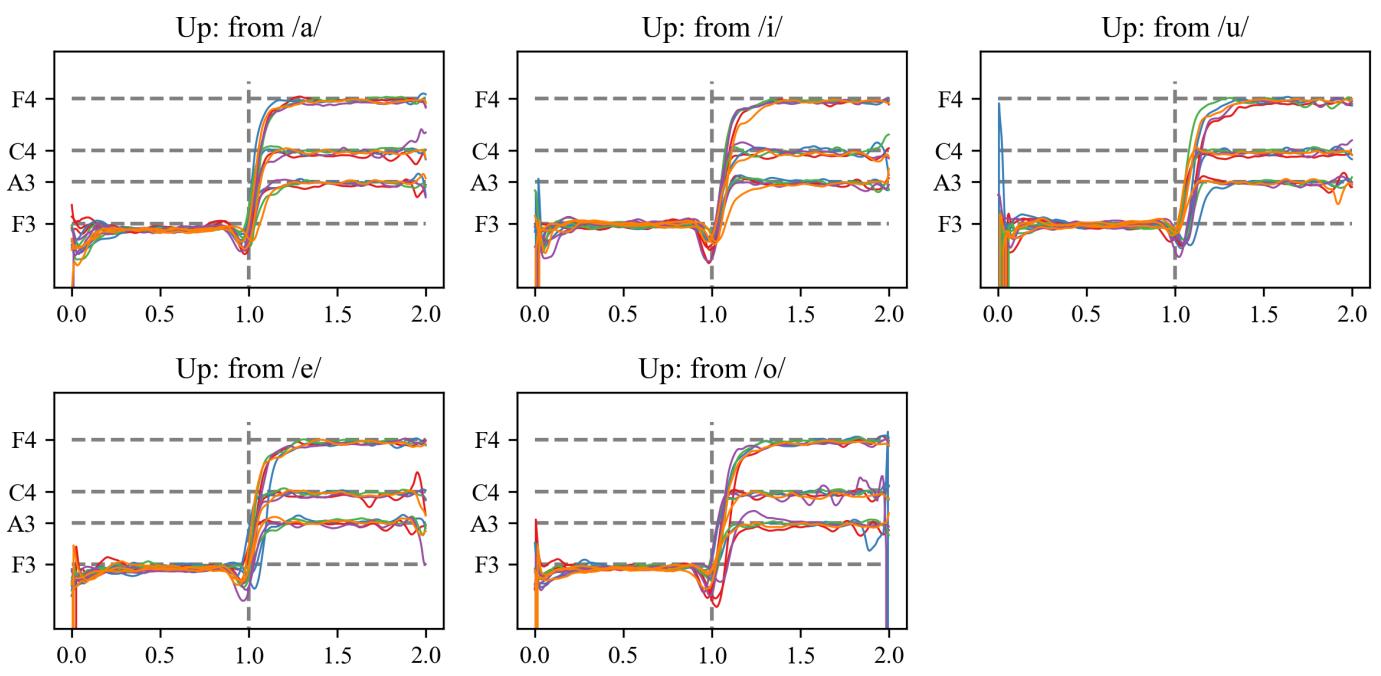


図 A.6 被験者 4 の上昇発声

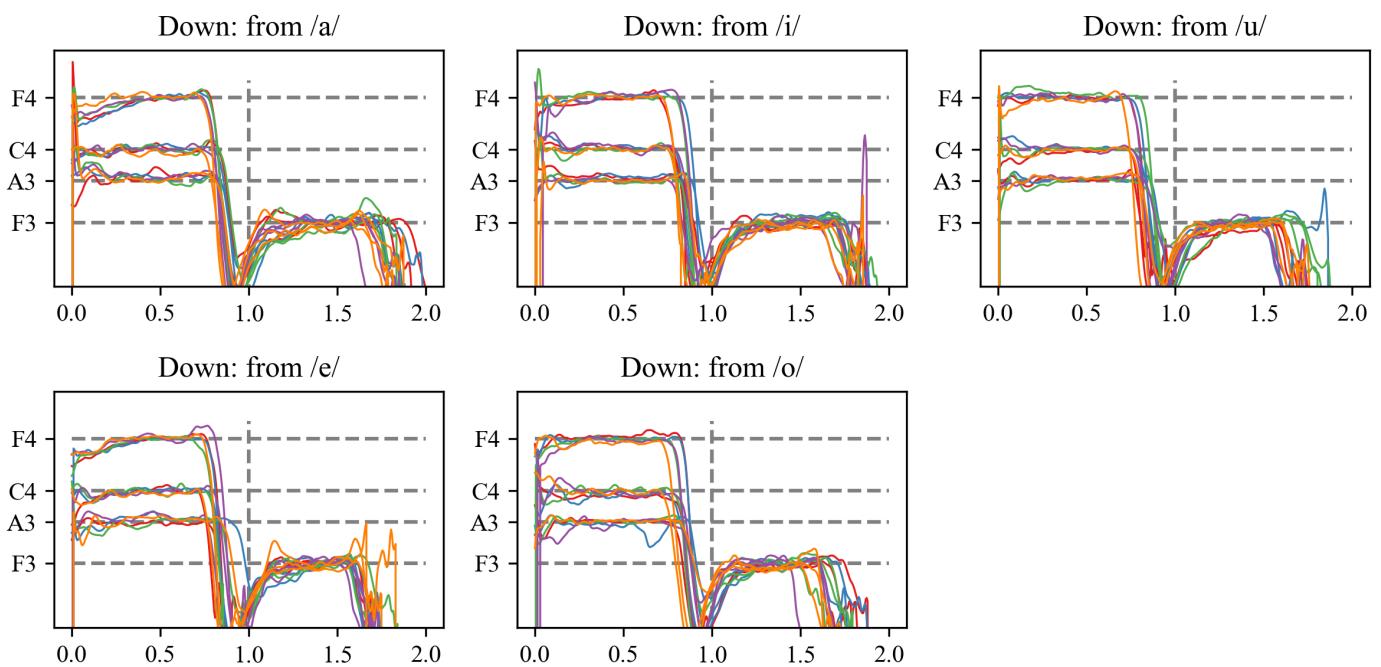


図 A.7 被験者 2 の下降発声

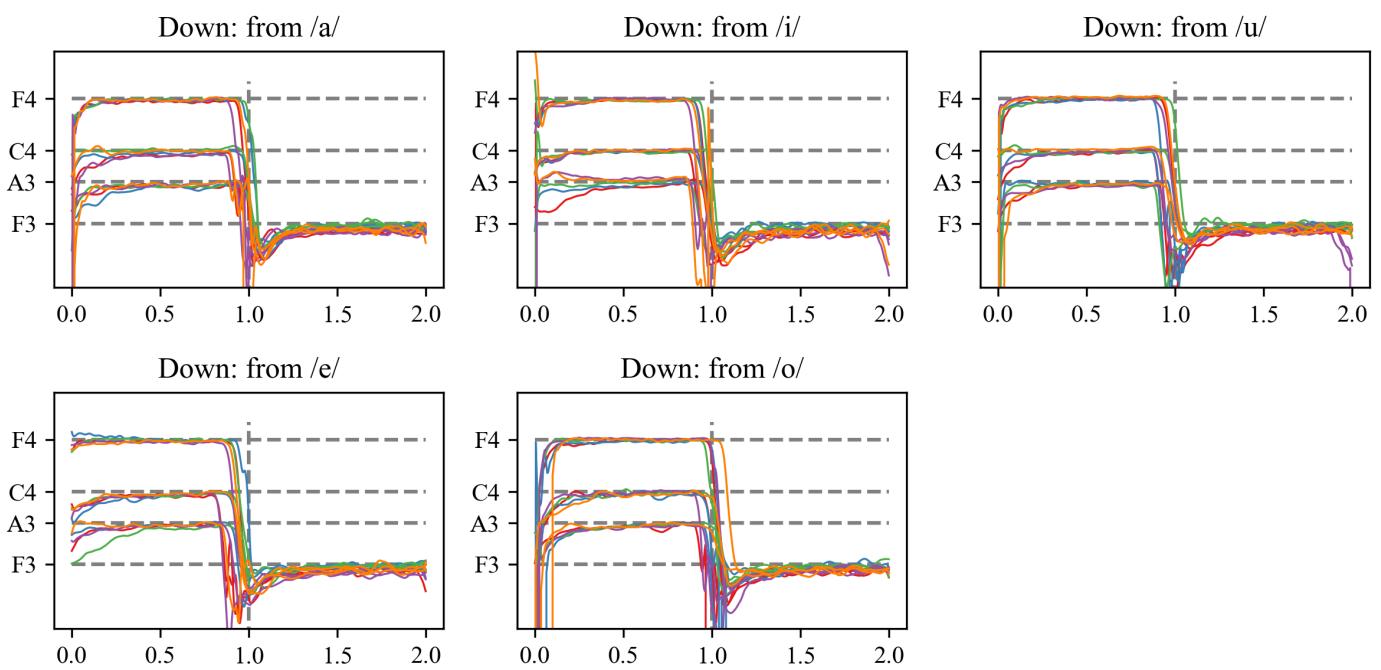


図 A.8 被験者 3 の下降発声

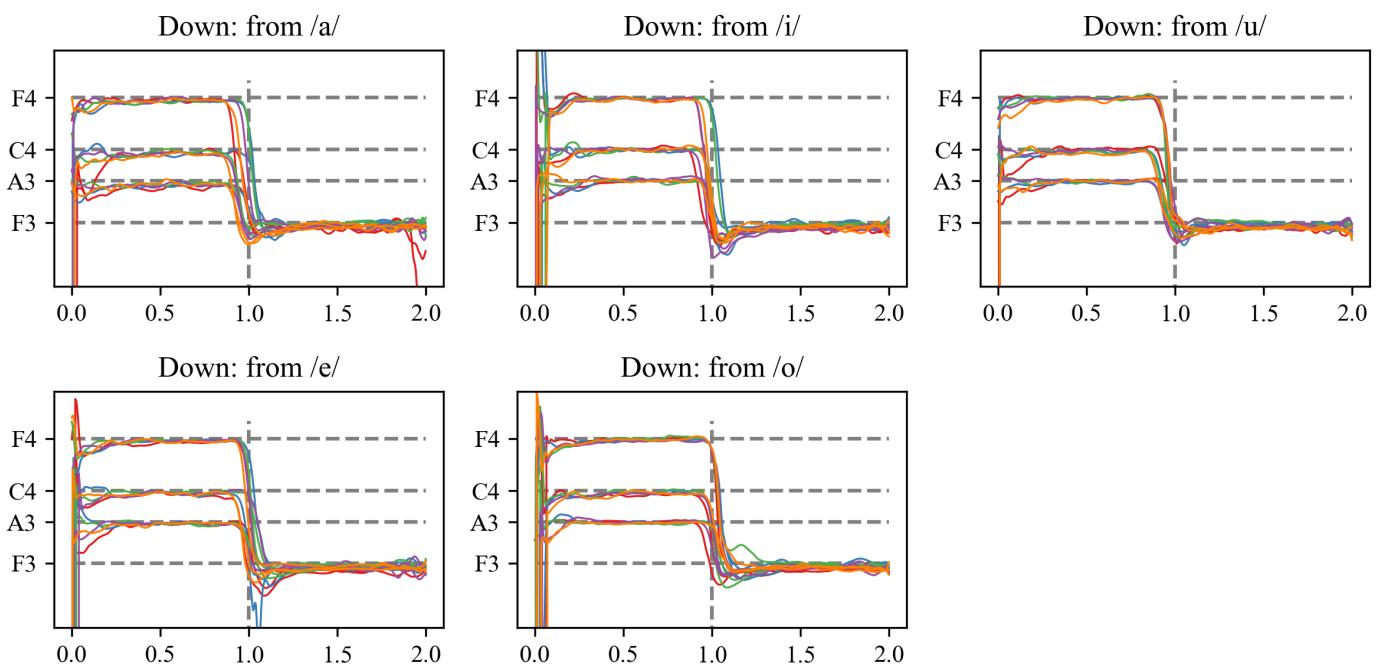


図 A.9 被験者 4 の下降発声