Vol.2021-MUS-131 No.22
Vol.2021-SLP-137 No.22
2021/6/18

ラーモニック減算法を用いたフォルマント周波数の自動推定法の検討

河口竜馬[1]　モクタリ パーハム[1]　森川大輔[1]

概要：フォルマント推定において、基本周波数(fo)が高い場合、従来のフォルマント分析では問題が生じやすい。この問題に対処するため、Zhang ら (2020) によってラーモニック減算法(RS-CEPS)が提案された。しかし、RS-CEPS を用いてフォルマントの自動推定を試みると、従来の方法と同様にfoが高い音声のとき、問題が起きる場合があった。そこで本報告では、RS-CEPS をフォルマントの自動推定に用いるために RS-CEPS のラーモニックの減算範囲について検討を行った。また、ラーモニックの減算範囲が調整可能な、フォルマント自動推定 GUI を開発した。

On automatic estimation of formant frequency using a rahmonic subtraction method

RYOMA KAWAGUCHI[1]　PARHAM MOKHTARI[1]　DAISUKE MORIKAWA[1]

1. はじめに

フォルマントはスペクトルの大まかな概形を表すスペクトル包絡の山のことであり, 声道の共振周波数に対応する. 周波数の低いものから順に第 1 フォルマント (F1),第 2 フォルマント (F2) …といい, 声の性差や音韻を特徴づけるパラメータの 1 つである[1][2]. フォルマントの推定は, 男性, 女性, 子供の声の違いや, はきはきした声, ぼそぼそした声などの合成音声の生成のため, 重要である[2]. しかし, 線形予測符号化 (LPC 分析法)やケプストラム分析法などの, 従来のフォルマント推定法は, 声帯の開閉周期によって決まる基本周波数 (fo) が 200 Hz 以上と高い場合, フォルマントを誤推定してしまうことが多い. この問題に対処するため, 様々なフォルマント推定法が提案されてきた. しかし, Shadle ら (2013) がフォルマントを半自動で推定する 5 つの方法と手動でフォルマントを抽出する方法でフォルマント推定精度を比較した結果, 半自動で推定する方法の中で最適な方法は見つからなかった[3]. そのため, 本研究では, Zhang ら (2020) によって提案された RS-CEPS (improved cepstral method using finer rahmonic subtraction)を用いた[4]. RS-CEPS は, ケプストラム上に現れるピークであるラーモニックを除去し, フォルマントを手動で推定することで, 従来のフォルマント推定法にみられたfoと高調波によるフォルマントの誤推定を抑えることができる手法である. 本研究では, 効率よくフォルマントを抽出するため, RS-CEPS と LPC 分析法を用いてフォルマントの自動推定を試みた. しかし, 自動推定を行った場合, 従来の方法と同様にfoが高い場合に問題があった. そこで本報告では, RS-CEPS をフォルマントの自動推定に用いるために, ラーモニックの減算範囲について検討を行った. また, ラーモニックの減算範囲の調整ができ, 自動でフォルマントを推定する GUI を開発した.

2. 実装方法

2.1 RS-CEPS

Zhang らによって提案された RS-CEPS[3]は, ケプストラム上に現れるピークである, ラーモニックを減算したスペクトル (RS-spectrum) から, スペクトル包絡を得て, フォルマントを推定する手法である. 代表的な従来のフォルマント推定法である LPC 分析法やケプストラム分析法は, fo が高い場合, foまたはその倍音をフォルマントとして誤推定することが多い. しかし, RS-CEPS はfoと高調波に対応する, ラーモニックを複数回減算し, foと高調波を除去した RS-spectrum を得ることで, foと高調波によるフォルマントの誤推定を抑えることができる. そのため, foが高い音声でも従来のフォルマント推定法よりも高い精度でフォルマントを抽出できると提案された. この,RS-spectrum は, Zhang らによると, 声帯の開閉や空気が声門を通るときに生じる声門部のランダムノイズである声門ノイズに基づくと考えられている[4].

Zhang らは, 従来の手法と RS-CEPS でスペクトル包絡と固体声管モデルの伝達関数より, フォルマント推定精度を比較したところ, RS-CEPS は従来の手法よりも正確にフォルマントを推定したことを確認した[4]. しかし, 比較の際に用いた音声は, Zhang らが製作したものであり, 3D プリンターで製作された固体声管モデルに音源を入力し, 声門ノイズを付与したものであった.

1 富山県立大学 工学部
Faculty of Engineering, Toyama Prefectural University

Vol.2021-MUS-131 No.22
Vol.2021-SLP-137 No.22
2021/6/18

2.2 アルゴリズム

RS-spectrum を得るため，音声信号に従来のケプストラム分析を行い，ケプストラム分析で求めたスペクトル包絡E_{cep}を求める．そして，以下のステップ i ~iv によって，ケプストラム上に現れるラーモニックを減算する．また，図 1 に対数スペクトル及びE_{cep}とラーモニックを減算する際に算出する，スペクトルS_{H}及びケプストラムC_{H}を示す．また，図 2 にラーモニックの減算範囲を示す．

i. 対数スペクトルlog$S$から，E_{cep}より相対レベルが高い対数スペクトルS_Hを得る．

ii. 元の対数スペクトルlog$S$およびS_{H}を離散コサイン変換 (DCT) し，ケプストラムC及び明確なラーモニックを有するケプストラムC_{H}を得る．

iii. 算出したC_{H}を2乗し，${C_{\mathrm{H}}}^2$からラーモニックの減算範囲を図 2 に示したように，左右の最も近いディップとする．そして，CからC_{H}を決定した範囲のみ減算し，ラーモニックを減算した新たなケプストラムC_{r}を得る．

iv. ラーモニックを減算したケプストラムC_{r}を逆離散コサイン変換 (IDCT) し，ラーモニック減算後のスペクトルS_{r}を得る．

得られたS_{r}を元の対数スペクトルlog$S$とし， ステップ i ~iv を繰り返し行うことでラーモニックを複数回除去する．ラーモニックの除去により，S_rは収束に近い値となる．本研究では，S_{r}のスペクトル距離が 0.01 dB 以下になるまでステップ i ~iv を繰り返した．このとき，ステップ i ~iv を繰り返す回数は約 6〜10 回である．

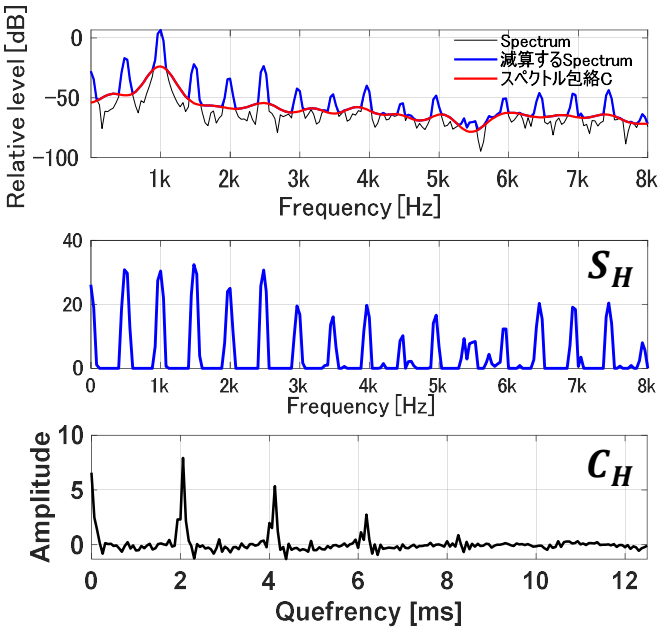

図 1: RS-CEPS によるフォルマント抽出の過程

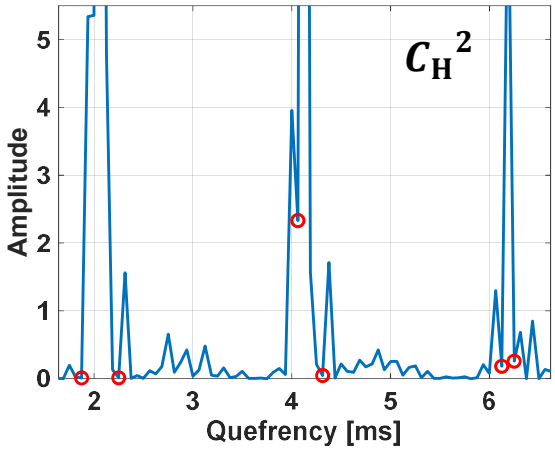

図 2: ラーモニックの減算範囲

3. RS-CEPS を用いたフォルマントの自動推定

3.1 方法

RS-CEPS では，ラーモニックを減算した結果のC_rにローパスリフタをかけて IDCT することでスペクトル包絡を求め，フォルマントを手動で抽出していた．本研究では,フォルマントを自動推定するために, 2.2 の手順で得られた，ラーモニックを減算したスペクトルS_rに LPC 分析を行う．そして，極の周波数Fを求め，帯域幅が 600 Hz 以下のFを周波数の低いものから順に 3 つ抽出した．

RS-CEPS を用いたフォルマントの自動推定を実装すると，音声信号により高調波が残留し，従来のフォルマント推定法と同様に，高調波をフォルマントとして誤推定することがあると判明した．そのため，従来と同様に誤推定していた場合，ラーモニックの減算する範囲を左右の最も近いディップよりも広く設定した．

3.2 結果

RS-CEPS を用いて，OPENGLOT Repository3 (Alku, 2019) に含まれる，声道モデルに，音源を入力して生成された，男性母音[u]の音声の[5]，フォルマントを自動推定した．この音声のfoは，420 Hz であった．真の F1~F3 と，従来のケプストラム分析，従来の減算範囲の RS-CEPS，拡大後の範囲の RS-CEPS で推定した F1~ F3 を表 1 に示す．表 1 の括弧内は真の値との差であり，減算範囲は，従来の範囲から ±0.5 ms と ±1 ms 拡大した．ただし，ここでいう真の F1~F3 は，正弦波スイープの周波数応答のデータから測定した値とした．図 3 に，ケプストラム分析法で求めたスペクトル包絡と，減算範囲を広げる前と広げた後のスペクトル包絡を示し，減算範囲を広げる前と広げた後の,RS-spectrum を図 4 に示す．

Vol.2021-MUS-131 No.22
Vol.2021-SLP-137 No.22
2021/6/18

表 1: 減算範囲によるフォルマント周波数の推定精度

	F1	F2	F3
真の値	390 Hz	810 Hz	2100 Hz
従来のケプストラム	400 Hz (+10 Hz)	880 Hz (+70 Hz)	1480 Hz (-520 Hz)
従来の範囲	460 Hz (+70 Hz)	870 Hz (+60 Hz)	2050 Hz (-50 Hz)
範囲拡大 (±0.5 ms)	390 Hz (+0 Hz)	820 Hz (+10 Hz)	2090 Hz (-10 Hz)
範囲拡大 (±1 ms)	410 Hz (+20 Hz)	900 Hz (+90 Hz)	2090 Hz (-10 Hz)

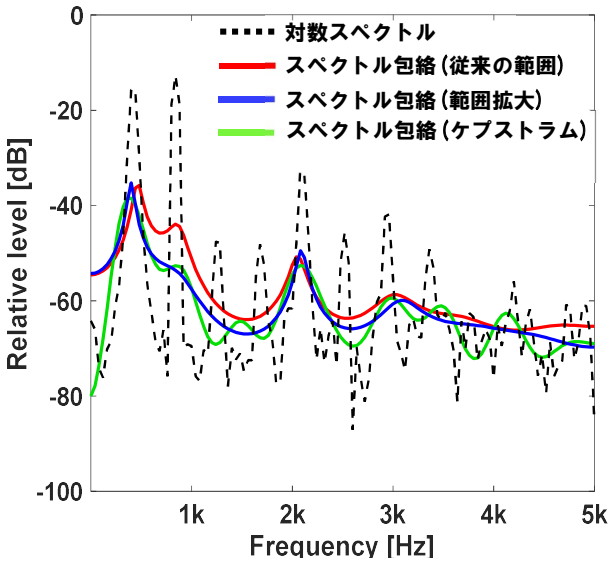

図 3: 減算範囲を調整したときのスペクトル包絡

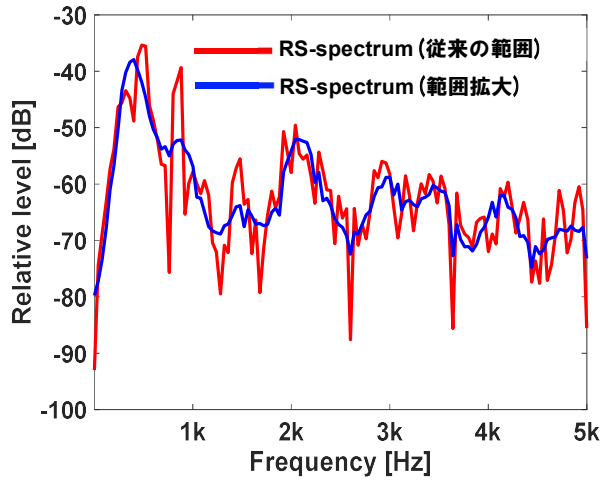

図 4: 改善した RS-CEPS で得た RS-spectrum

従来のラーモニックの減算範囲では, 2 次高調波を F2 として誤推定していたが, 表 1 より減算範囲を広げると, 従来の減算範囲より真のフォルマント周波数に近くなると判明した. また, 従来の減算範囲で RS-CEPS より求めたフォルマント周波数は, 最大で 70 Hz 真の値と差があった. 対して, ケプストラム分析より求めた F3 は, 真の値との差が 520 Hz となり, 大幅に誤推定していた. 真のフォルマント周波数に 1 番近い値になったのは, 減算範囲を±0.5 ms にしたときの RS-CEPS であった. 減算範囲を±1 ms にすると, F2 において 1 番誤差が大きくなった.

図 3 より, 減算範囲を広げることで高調波をさらに除去できていることがわかる. また, スペクトルが滑らかになった.

3.3 評価

自動化した RS-CEPS のフォルマント推定精度を従来の LPC 分析法と比較した. 比較の際に用いた音声データは, 3.1 と同様に OPENGLOT Repository3 を用いた[5]. 本評価では, 男性母音 3 種類, 女性母音 3 種類を用いてそれぞれの推定誤差を算出した. この音声は, サンプリング周波数 44.1 kHz で録音したものである. ただし, この音声は Zhang らが生成した音声とは異なり, 声門ノイズは付与されていない. Zhang らは敢えて声門ノイズを付与していたという点と, RS-CEPS で得る RS-spectrum は, 声門ノイズに基づくと考えられているという 2 点から, 本評価は Zhang らよりも厳しい評価であったと考える.

このデータベースには, 1 つの母音ごとにfoが 41 種類 (10 Hz ごとに 100~500 Hz)の音声データと, フォルマントの真の値を求めるための 80~7350 Hz の正弦波スイープの周波数応答のデータがある. このデータを用いて, 推定したフォルマント周波数$F_{est,i,n}$と測定値$F_{ref,i,n}$の各母音における推定誤差を算出した. ここで, $F_{est,in}$及び$F_{ref,in}$はfoがn $(n = 100,110,\ldots 500)$ Hz の音声データのi $(i = 1,2,3)$ 番目のフォルマント周波数を表す.

各母音においての誤差の算出方法は, Zhang らが RS-CEPS の評価の際に算出した方法と同様に, F1~F3 の測定値と推定値の誤差$d_{i,n}$をそれぞれ,

$$d_{i,n} = 100 \times \frac{|F_{est,i,n} - F_{ref,i,n}|}{F_{ref,i}} \ [\%] \qquad (1)$$

より求めた. そして, 41 種類全ての音声データの F1~F3 までの相対推定誤差の平均値Eは, 式(2)より算出した.

$$E = \frac{1}{41 \times 3} \sum_{n=1}^{41} \sum_{i=1}^{3} d_{i,n} \ [\%] \qquad (2)$$

各手法のEを求めた. 表 2 に従来の LPC 分析法, 従来の減算範囲の RS-CEPS, 音声データごとに減算範囲を調整した RS-CEPS のEを示す.

Vol.2021-MUS-131 No.22
Vol.2021-SLP-137 No.22
2021/6/18

表 2: 各手法のフォルマントの推定誤差

	従来の LPC	RS-CEPS (従来の範囲)	RS-CEPS (範囲拡大)
Male [a]	4.9 %	4.7 %	3.9 %
Male [u]	5.1 %	6.6 %	3.6 %
Male [i]	7.8 %	5.6 %	5.3 %
Female [a]	5.0 %	3.6 %	2.4 %
Female [i]	7.0 %	8.5 %	5.6 %
Female [e]	3.8 %	4.0 %	2.9 %

評価の結果, 声門ノイズが付与されていない音声でも, 6 種類全ての母音で従来の LPC 分析法よりも, 減算範囲を調整した RS-CEPS のほうがEが小さかった. また, 従来の減算範囲のとき LPC 分析よりも, Male [u], Female [i], Female [e]においてEが高かった.

4. フォルマント推定 GUI

フォルマントを効率よく推定するために, LPC 次数, ラーモニックの減算範囲等のパラメータを容易に変更しフォルマント推定を行う GUI (Graphical User Interface)を作成した. この GUI は改善した RS-CEPS または, 従来の LPC 分析法を用いてフォルマントを推定する. GUI の外観を図 5 に示す.

図 4 の(a)は, 音声波形である.分析を行いたい時間をマウスでクリックすることで, ユーザーの任意の時間を分析することができる.

(b)はスペクトログラムである. スペクトログラムを見ることで, フォルマントのおおよその周波数がわかるため推定したフォルマントが大幅に誤推定していないかを確認できる. 本 GUI では, スペクトログラムのフレームの長さを 5 ms,シフト幅は 1 ms としている. このスペクトログラム上に, ユーザーが指定したフレームの長さごとに, 極の周波数と帯域幅及びフォルマントを描画している. 青線は極の帯域幅および周波数を示している. 赤線, 黄線, 紫線はそれぞれ, 第 1, 第 2, 第 3 フォルマントを示しており, ユーザーが指定した帯域幅以下の極の周波数を低いものから順に第 1, 第 2, 第 3 フォルマントとした. ただし, 発話時のフレームを分析するために, 音声信号のエネルギーが高いフレーム間のみ, フォルマントを描画した.

(c)の座標軸はスペクトルである. 音声波形の黄色の縦線間のフレームを分析しており, 図 4 では 0.8 秒から 0.825 秒の間のスペクトルを表示している. また, スペクトル上の白の破線は極の周波数を示している.

(d)はラーモニックの減算前と減算後のスペクトル距離を示しており, スペクトル距離が 0.01 dB 以下になるまで減算を繰り返す.

赤枠内は, 分析及びパラメータの設定を行うことができる. 最適なパラメータは分析する音声信号に依存するため, GUI 上で変更可能なパラメータを設けた. 変更可能なパラメータを表 3 に示す. ラーモニックの減算範囲が 0 の状態とは, 左右の最も近いディップを左右境界としていることである.減算する点数を 1 点増やすごとに,$(1/\mathrm{Fs}) \times 2$ 秒だけ減算する範囲が広がる.

表 3: GUI 上で変更可能なパラメータ

録音のサンプリング周波数	16 kHz, 24 kHz, 48 kHz
ダウンサンプリング(Fs)	8 kHz, 10 kHz, 16 kHz
帯域幅	600 Hz, 800 Hz, 1000 Hz, 指定なし
録音時間	5 秒, 10 秒, 30 秒
フォルマント推定法	従来の LPC, RS-CEPS+LPC
減算範囲	任意の点 (0,1,2,…)

録音波形およびスペクトログラムの時間軸の範囲は, 1.5 秒間であり, slider1 を動かすことで任意の時間軸に変更可能である. slider2 はスペクトログラム上のカラーマップの範囲を設定しフォルマントを見やすくできる. また, スペクトログラム上の推定したフォルマントが誤っている場合,change_Fr を押し, 修正したいフォルマントを第 1～第 3 フォルマントの中から選択する. そして, スペクトログラム上の青色の極の周波数をクリックすることで, 極の中から手動でフォルマントを修正できる. その他のメニューの機能においては表 3 に従う.

5. 考察

減算範囲を適切に拡大すると, 真の値のフォルマント周波数に近づくが, 減算範囲を広げすぎると従来の減算範囲よりも, 真の値との差が大きくなった. これは, 高調波を除去しすぎることでスペクトル包絡の山が得られなくなったことが原因であると考える. 減算範囲を適切に調整した RS-CEPS は, 女性母音[i]のとき, Eは 5.6%であった.従来の LPC 分析法でも, 女性母音[e]のときは, Eが 3.8 %であったため, 推定精度をさらに向上させる必要があると考える. しかし, RS-CEPS は声門ノイズに基づくと考えられているため, 声門ノイズを付与した音声で評価を行えば, 推定精度はより高くなると考えられる.

構築した GUI の今後の課題は, 分析時間をさらに短くすることである. また, ラーモニックの減算範囲は, 音声信号によって最適な範囲は異なるため, 高調波が残留している場合, 自動で範囲を調整する機能をつける必要があると考える.

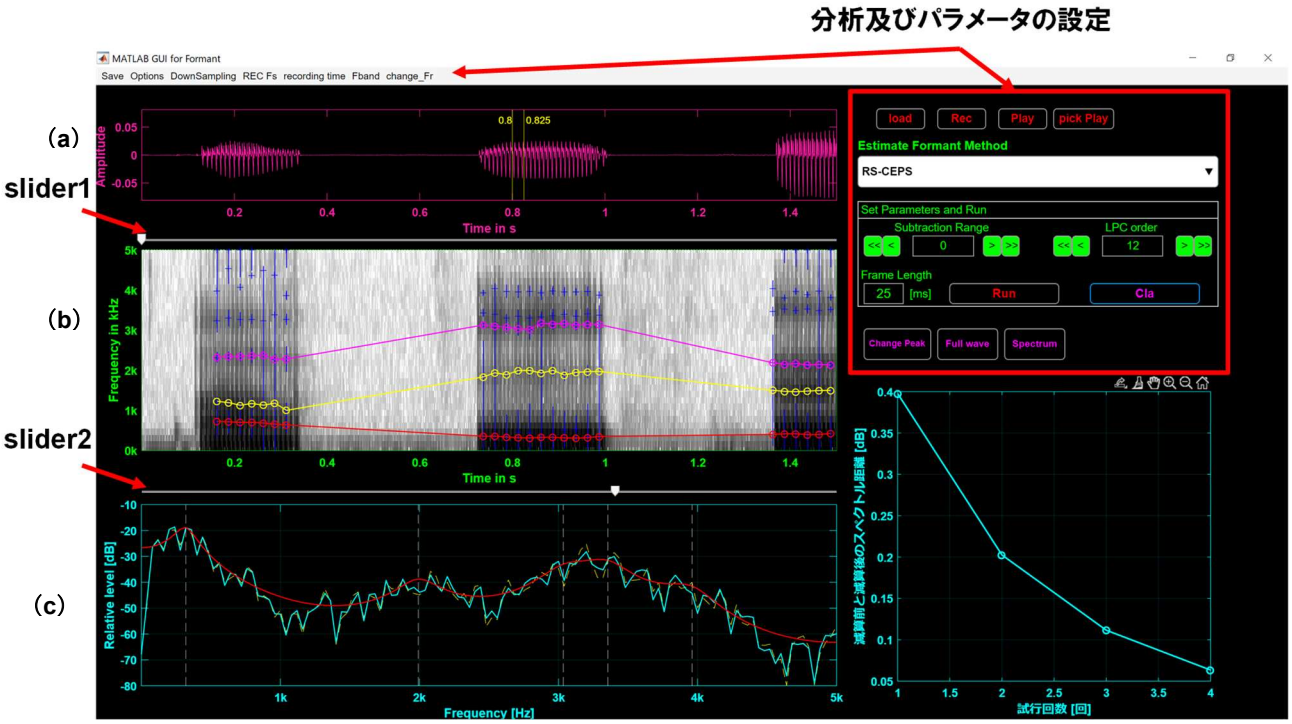

図 5: GUI の外観

6. まとめ

RS-CEPS を用いてフォルマントの自動推定をしたところ，従来のフォルマント推定法と同様の問題があった．そのため，ラーモニックの減算範囲を拡大し，foと高調波をさらに除去した．減算範囲を拡大しすぎると，従来の減算範囲よりも真の値との差が大きくなったが，適切に設定すると，真の値との差は±10 Hz 以内であった．従来の LPC 分析と減算範囲を調整した RS-CEPS でフォルマント推定精度を比較した．減算範囲を調整した RS-CEPS は，Zhang らの製作した音声でなくても，調査したすべての母音でフォルマント推定精度が向上した．また，従来の LPC 分析法または RS-CEPS により,フォルマント推定を行う GUI を構築したことにより，効率よくフォルマントを推定できるようになった．

謝辞

本研究の一部は，カシオ科学振興財団の助成金（第 38 回令和 2 年度の 25 番）の支援を受けた．

参考文献

[1] 平原 達也, “音と人間”, (コロナ社, 東京, 2013), p.169.
[2] 櫻庭 京子 *et al.*, “女性と判定される声の特徴,” 音声言語医学, 50 巻, 1 号, pp. 14-20, 2009.
[3] C H Shadle *et al.*, “Comparing measurement errors for formants in synthetic and natural vowels,” J. Acoust. Soc. Am., pp.713-727, 2016.
[4] Z.Zhang *et al.*, “Retrieving vocal-tract resonance and anti-：：resonance from high-pitched vowels using a rahmonic subtraction technique,” Proc. IEEE International Conference on Acoustics, Speech and Signal Processing pp.7359-7363, 2020
[5] P. Alku *et al.*, “OPENGLOT – An open environment for the evaluation of glottal inverse filtering,” Speech Communication, vol.107, pp.38-47, 2019