物理一班

声真似は本当に似ているのか？

～音声分析による解明～

関屋 良健　箱田 千陽　菅 英寛　五十嵐 凛

指導教員　佐々木 重宏

1,研究動機

私たちは声真似を聞いて、
「なぜ声真似は似ているように感じるのか？」「どの要素が似ているのか？」
という疑問を持った。
これらの疑問を解決するため、声真似を科学的に分析し、比較するために研究を始めた。

2,音声データについて

防音の放送室で指向性マイクを用いて音声を録音した。
内容は1年生の古典の教科書より『児のそら寝』中の僧の台詞、「いざかいもちひせん」を
今入先生の音声をオリジナルとし、横手高校1，2年生の生徒22人による声真似を録音した。
分析には音声分析ソフト「Audacity」、画像処理ソフト「imageJ」を使用。

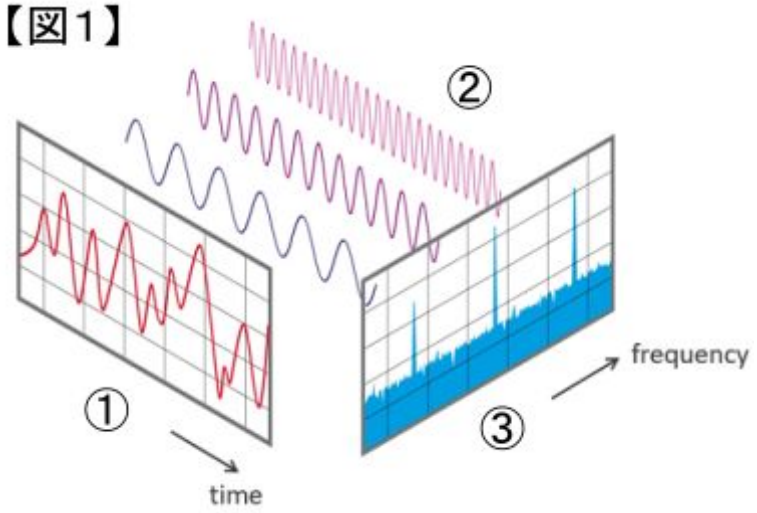
3,先行研究

先行研究について、インターネットで調べた。
音声分析には音色の違いが分かる波形の比較、周波数とその強さが分かるスペクトラムや
スペクトログラムの比較といった方法が用いられることが分かった。
また、波形からは音響的特徴が分かりづらいことも分かった。
実際に予備調査として波形の比較を行ったが、声真似の似ている、似ていないに関係すると思
われるような違いは見つけることができなかった。
そこで、周波数とその強さが関係するのではないかと考え、比較のために**フーリエ変換**を用い
ることにした。

4,フーリエ変換（図1）

フーリエ変換とは、与えられた音声の波形を**正弦波（sin波）に分解**し、それぞれの強さを周波
数ごとにまとめるというもの。複雑な波形を単純なsin波のみで表し、比較しやすくすること
で、波形からは分かりづらい音響的特徴を抽出することができる。

【図1】



5,スペクトラム分析

（**図1**）の③のグラフのようにフーリエ変換でできたグラフを**スペクトラム**という。
このグラフでは縦軸はdB、横軸は周波数をとる。
スペクトラムの特徴を見ることで、音声のどの周波数が強いのかを知ることができる。

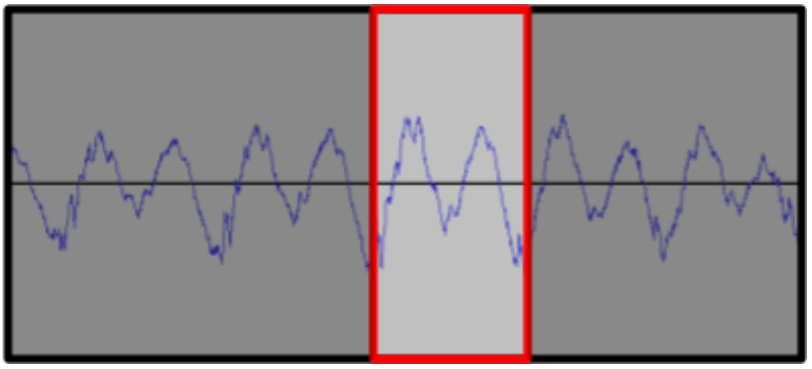
【仮説】

聞いて似ていると感じた音声のスペクトラムの概形は、オリジナルと比べて大きな違いがみられず、逆に聞いて似ていないと感じた音声のスペクトラムは概形に違いがみられるのではないか？

【検証】

まずAudacityを用いてオリジナル・似ている音声・似ていない音声のスペクトラムを表示する。そして概形を目視で比較し特徴を見る。この検証の際、話者ごとに話す速さや音の長さに違いがあり、時間をそろえての比較ができなかった。
そこで、「いざかいもちひせん」の最初の「い」の一波長のみを切り取って用いた。（**図2**）
一一波長は同じような波形が繰り返されていると十分に確認できるものを選んだ。

【図2】

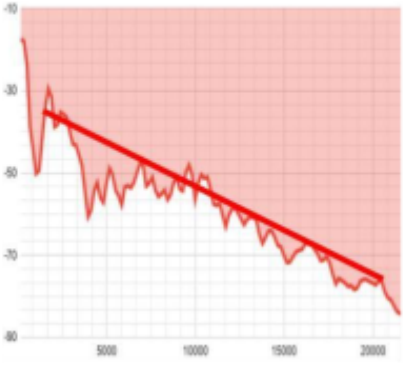


【結果（図3）】

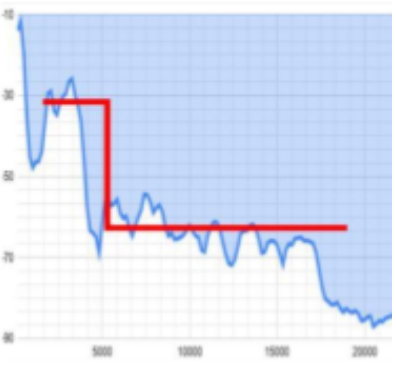
話者Bには5000Hz付近でのdBの値が大きく下がる、その後の傾斜が小さいなど、大きな違いがみられる。

【図3- オリジナル】

【図3- 話者A】



【図3- 話者B】



【考察】

結果から周波数は音声の似ている・似ていないに関係すると言えるのではないかと考えた。
しかし、目視での分析では細かい違いまでは見ることができず不正確で限界があった。
そこで、スペクトラム表示されたデータを**数値化**することで、具体的に比べることにした。

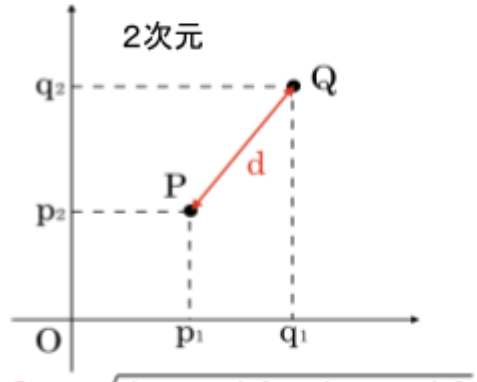
6,数値化のために

似ている・似ていないという感覚的な部分の数値化という難しい課題を解決するため秋田県立大学の高根先生に協力を仰いだ。
また、夏季休業中に参加した校外のプログラミング講習で、データの比較の方法として**ユークリッド距離**について知った。
また、これまで用いてきたdBは相対的な値で、また対数表示であるため負の数になってしまい、数値化をしての比較においては不便であるため、実際の音のエネルギーを示す電力比に変換して分析を行うことにした。

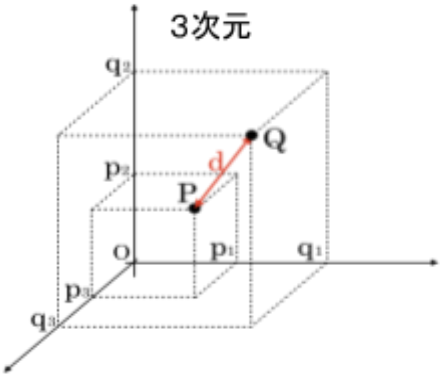
7,ユークリッド距離とは

ユークリッド距離とは、ある座標空間内にある2つの**データ同士の距離**のことで、軸の数が次元数を表す。2次元、3次元、・・・・・・、n次元と拡張しても、同様に求めることができる。（**図4**）

【図4】



$$d = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2}$$



$$d = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+(q_3-p_3)^2}$$

・
・
・
・
・
・

n次元

$$d = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+(q_3-p_3)^2+\cdots\cdots+(q_{n-1}-p_{n-1})^2+(q_n-p_n)^2}$$

8,ユークリッド距離による分析①

ここからは、先ほどのユークリッド距離を用いた分析を行う

【仮説】

似ている音声⇒ユークリッド距離の値が小さい
似ていない音声⇒ユークリッド距離の値が大きい

【検証方法】

実際のスプレッドシートの画面（**図5**）において、オリジナルの電力比Pn、話者Aの電力比Qn、QnとPnの差の2乗を出す。
その後、各周波数帯の差の2乗の合計の正の平方根をとり、ユークリッド距離を出す。
ここからの検証には話者C,D,Eを追加した。

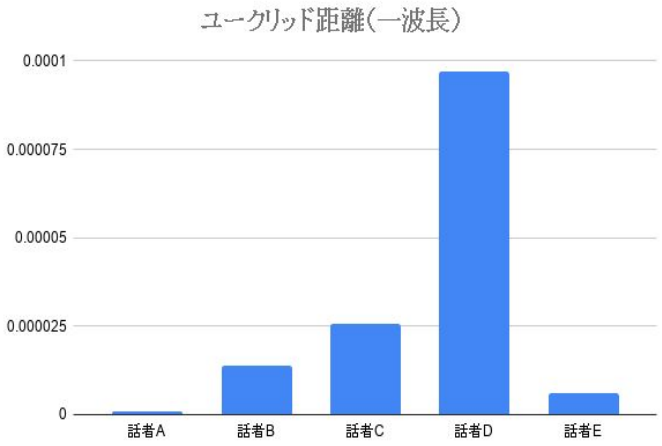
【図5】

n	周波数(Hz)	オリジナル		話者A		差の2乗 (Qn-Pn)
		レベル (dB)	電力比 Pn	レベル (dB)	電力比 Qn	
1	21.533203	-60.366875	0.000000918993	-61.10083	0.000000776098	0.000000000000020418937
2	43.066406	-65.101463	0.000000308925	-59.089783	0.000001233166	0.000000000000085422140694
3	64.599609	-66.34433	0.000000232042	-61.919102	0.000000642826	0.000000000000016873889988
4	86.132813	-60.827881	0.000000826441	-62.42807	0.000000571733	0.000000000000006487638376
5	107.666016	-42.596676	0.000054996164	-55.616276	0.000002743926	0.000000027302963941571
6	129.199219	-34.972786	0.000318215551	-34.896187	0.000323877889	0.000000000032062068362
7	150.732422	-30.61381	0.000868198437	-26.449314	0.002265002053	0.000001951060339814430
8	172.265625	-27.201756	0.001904690431	-23.800993	0.004167740785	0.000005121396906185820
9	193.798828	-26.297178	0.002345752566	-20.05262	0.009879569028	0.000056758390499776100
10	215.332031	-23.723352	0.004242919577	-21.505005	0.007071303865	0.000079997576816487900
11	236.865234	-24.055912	0.003930147048	-33.611923	0.000436319077	0.000012213822545954200
12	258.398438	-31.627167	0.000687516776	-42.643574	0.000054405474	0.000000400829920884277
1021	21985.40039	-93.170479	0.000000000481	-96.991707	0.000000000199	0.000000000000000000795
1022	22006.93359	-92.070282	0.000000000620	-97.928932	0.000000000161	0.000000000000000000211
1023	22028.4668	-91.758583	0.000000000667	-98.442497	0.000000000143	0.000000000000000000274

【結果（図6）】

話者A（最も似ている）
　⇒値がとても小さく望ましい結果
話者B（最も似ていない）
　⇒値が3番目に小さく望ましい結果
話者C（似ていない）
　⇒値が小さく望ましくない結果
話者D（似ている）
　⇒値が大きく望ましくない結果
話者E（似ていない）
　⇒値が小さく望ましくない結果
＊グラフは左から話者A～E

【図6】



【考察】

一波長を切り取っての分析では使用した部分が極端に短く、実際に聞いた時の結果と、**ずれ**が生じるのではないか。
また、実際の音声に含まれる、微妙な音程の変化や抑揚などを考慮しておらず、これもずれに関係しているのではないか。

9,検証方法の転換

一部のデータでは望ましい結果が得られ、結果に差を見ることができたためユークリッド距離による分析は有効だと判断した。より正確なデータを出すため、秋田県立大の高根先生からいただいたアドバイスをもとに**音素**に区切って比較を行うことにした。
音素とは、端的には音声をローマ字ごとに分けたもののことである。

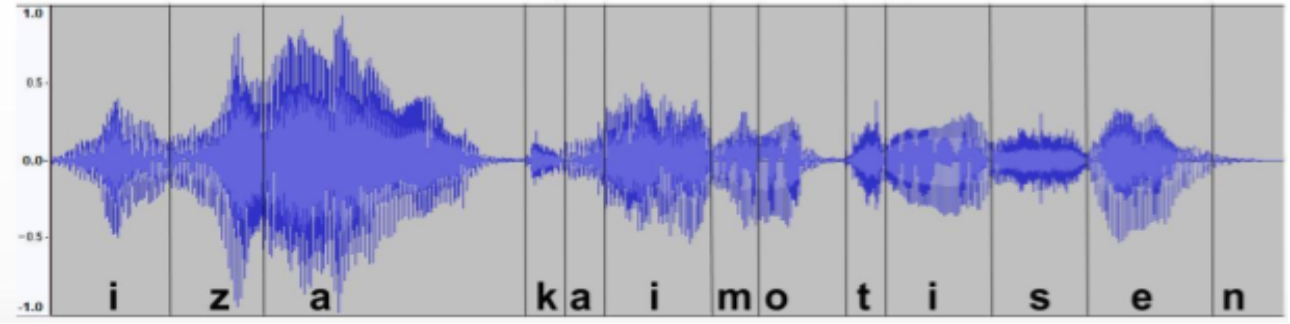
10,ユークリッド距離による分析②

同じ話者での音声に違いはないかを確かめたいと考えた私たちは、オリジナルの音声を複数回録音したものを使用して検証することにした。

【検証方法】

「いざかいもち(い)せん」を音素ごとに13個に区切り（**図7**）、それ以外は①と同様にしてユークリッド距離を求め、比較した。

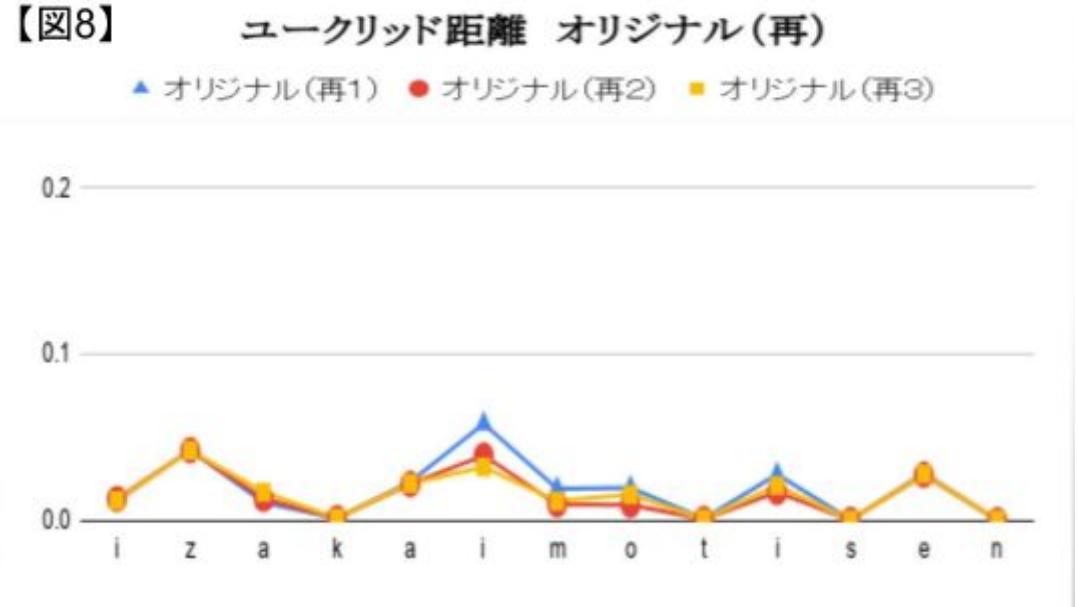
【図7】



【結果（図8）】

同じ話者中ではあまり大きな違いは見られなかった。
このことから、同一の話者中では音声にズレはないと言える。

【図8】



11,ユークリッド距離による分析③

オリジナルの音声と声まねの音声を実際に比較する

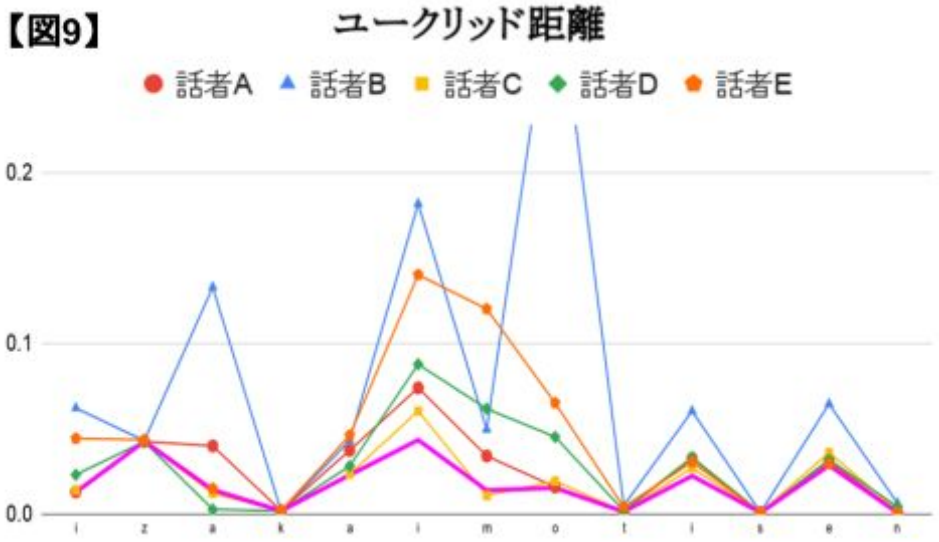
【検証方法】

②と同様に13個の音素に区切り、ユークリッド距離を求めて比較した。

【結果】

話者A（最も似ている）⇒総合的に見て値がとても小さく、データ的にも似ている
話者B（最も似ていない）⇒値が大きく、データ的にも似ていない
話者C（似ていない）⇒値がとても小さく、データ的にはよく似ている
話者D（似ている）⇒値が小さく、データ的にも似ている
話者E（似ていない）⇒値が大きく、データ的にも似ていない
全体的に見て、子音に大きな差は見られず、**母音に大きな差**が見られる。

【図9】



今回は特に、大きな差が見られる、母音の結果（**図10**）に特に注目した。

【図10】

結果（i）	
話者B	0.062060
話者E	0.044199
話者D	0.022902
話者C	0.013811
話者A	0.012893
平均	0.012559

結果（a）	
話者B	0.132917
話者A	0.039792
話者E	0.014795
平均	0.013962
話者C	0.011968
話者D	0.002591

結果（a）	
話者E	0.046024
話者B	0.042921
話者A	0.037223
話者D	0.028009
話者C	0.023307
平均	0.022743

結果（i）	
話者B	0.181831
話者E	0.140165
話者D	0.087691
話者A	0.073876
話者C	0.060254
平均	0.043338

結果（o）	
話者B	0.323880
話者E	0.064984
話者D	0.045106
話者C	0.019174
話者A	0.015731
平均	0.015194

結果（i）	
話者B	0.060250
話者D	0.033076
話者A	0.032499
話者E	0.030392
話者C	0.026865
平均	0.022221

結果（e）	
話者B	0.064425
話者C	0.035902
話者D	0.031781
話者A	0.031089
話者E	0.029242
平均	0.028308

【結果から】

話者A,Dは全体的に値が小さく、話者B,Eは全体的に値が大きくなったことから望ましい結果が得られた。しかし、話者Cは全体的に値が小さくなり、望ましくない結果となってしまった。
このように、実際の結果と分析の結果との**ずれ**が生じているデータが見られる。

12,ここまでの考察

以上の結果から、周波数とその強さはたしかに声真似の似ている・似ていないに関係すると言える。
しかし、今回の私たちの分析では、話者ごとの話す速さや音の長さをそろえることができず、**音素ごとの瞬間的な分析**しかできなかった。
よって実際に聴いた結果と分析結果との**ずれ**を解消することができなかったと考える。
以上の事から、人の声真似には単純に周波数やその強さと言った瞬間的な要素だけでなく、音素間の移り変わりなど**連続性のある時間的な要素**が関係しているのではないかと考えた。
また、声の出し方や抑揚などの表現上の特徴も関係しているのではないかと考えた。

13,スペクトログラム分析

時間的な要素を排除せずに分析する方法として、スペクトログラム分析に目をつけた。
スペクトログラムは、横軸⇒時間、縦軸⇒周波数をとり、音の強さを色で表したもの。
今回の場合、色は、強いところが白、弱いところが黒で表されていて、特に強いところを赤で強調した（図11）。これにより、時間的な比較が可能になる。

【図11】

【仮説】

オリジナルと似ているものは色の分布、特に赤い部分の分布がよりオリジナルに近いのではないかと考えた。

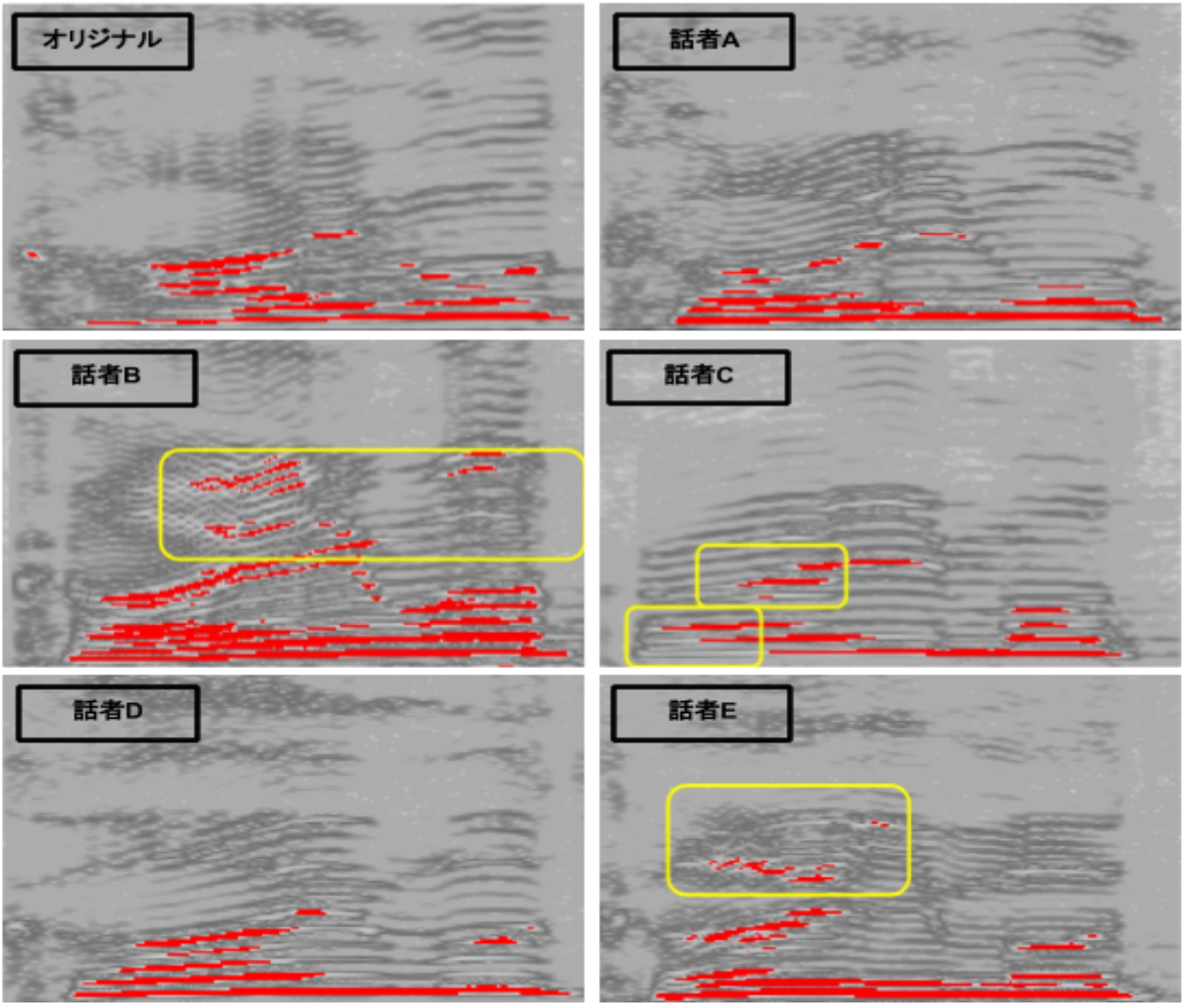
【検証方法】

音声をAudacityでスペクトログラム表示した後、それを目視で比較する。
分析には、「いざかいもちいせん」のうち、ユークリッド距離の分析の際に大きなずれが見られた「かいも」の部分を使用した。
さらに、特に大きな違いがみられるところを黄色の枠で囲った。

【結果（図12）】

話者A（最も似ている）⇒オリジナルとの大きな違いが見られる部分はみられない
話者B（最も似ていない）⇒黄色の部分にオリジナルとの違いがみられる
話者C（似ていない）⇒黄色の部分にオリジナルとの違いがみられる
話者D（似ている）⇒オリジナルと大きな違いがみられる部分はみられない
話者E（似ていない）⇒黄色の部分にオリジナルとの違いがみられる

【図12】



【結果から】

時間的な要素を含めた分析では望ましい結果になったので、**時間的な要素は重要**であると言える。

14,展望

スペクトログラムの分析の際、目視での比較に戻ってしまい、正確さに欠けると考えたので、やはり、**数値化**して具体的に比べられるようにしたい。
また、声真似をしてくれた生徒の普段の音声を録音して分析し、声真似の音声と比較して、どこを変化させているのかを調べると言った方法も有効なのではないかと考えた。
さらに、今あるデータについて、他の分析方法についても考えを深めていきたい。

15,まとめ・気付き

ずれや不明瞭な部分は見られたが、おおむね望ましい結果を得ることが出来た。
しかし、残念ながら声真似をはっきりと判別することまではできなかった。
これからは、様々な方法でより正確に、**数値化**を目標として研究を進めていきたい。
また、実際に、音素接続という方法を用いて実際に音声を出せるものをつくってみたところ、音声の出力は簡単なことではないということに気付いた。
現代の技術の再現はとても難しく、現代の音声技術のすごさを改めて感じるよい機会になった。

16,参考文献

・音声の音響分析のいろは
峯松 信明　東大 日本音声学会音声学普及委員会
https://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/nlp+slp/lecture-02.pdf
・フォルマントの見つけ方
http://www.akjava.com/english/detectformant.html
・画像比較
https://picdiff.tk/
・第６章　デシベル[dB]計算の応用編（電力比？電圧比？）
http://www.decibel.co.jp/what_db/what_db_page6.htm

17,謝辞

・指導教員　　佐々木　重宏　先生
・秋田県立大学　　高　根　昭一　先生
・横手高校国語科　　今　入　直樹　先生（オリジナルボイス）
・録音に協力していただいた放送部、横手高校1,2年生のみなさん
皆さんご協力ありがとうございました。