

スペクトルによる音声分析の研究

清泉女学院高等学校 高1 森 元乃

研究動機

現在の個人認証技術には悪用される可能性が多く、他人に認証されてしまうなどの危険が多く存在している。この問題を解決する手段として身体的な特徴を用いて個人が特定可能なバイOMETRICS認証が利用される機会が増えた。本研究は、バイOMETRICS認証の中でも音声認証に着目し、声紋から分かる情報を科学的に分析する。

音声データについて

静かな部屋(40dB以下)において、スマートフォンを用いて録音した。スペクトル測定には「オーディオ/スペクトルアナライザ」を使用した。

予備実験

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を見る波形の比較や、周波数とその時の強さが分かるスペクトルを比較するといった方法が用いられることが分かった。調査の中で波形からは音響的特徴が分かりづらいことも分かった。実際に事前調査として波形の比較を行ったが、相違点を見つけることが出来なかったと考える。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、スペクトルを作成する。

実験 スペクトルを用いた分析①

スペクトルは縦軸に大きさ【dB】、横軸に周波数【Hz】を取る。スペクトルの特徴を見ることで、音声のどの周波数が強いのかという事を知ることができる。

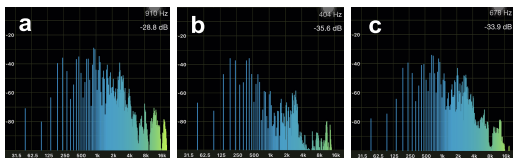
▶実験1

自分の声にアクセント^{a)}を加えた音声データと歌声^{b)}のスペクトルを作成する。使用する言葉はマザーグース『春風』より"大事な秘密"とする。複数の文字からスペクトルを作成する事は難しいため、最も変化がよくよく見ることの出来た"dajinahimutu"のaの母音のスペクトルを基準とした。

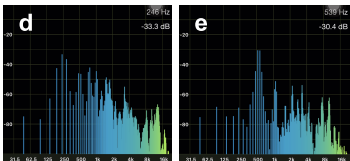
〈補足1〉

アクセント...標準語、関西弁、平坦の3種類を用いる
歌声...mid2C、hiCの2種類を用いる

▶結果1 (図1)



上に示した3つのスペクトルは左方から標準語、関西弁、平坦のものである。以上の3つを比較すると標準語に含まれる周波数が最も多いことが見て取れる。標準語と平坦のものに含まれる周波数とは異なる周波数が関西弁のものには見えて取ることが出来ると言える。



また歌声については、指定した音の周波数を最も多く含む、それを中心に山形になっている事が見て取れる。上記に示したスペクトルについて左方はmid2C、右方はhiCを示すが、mid2Cに含まれる周波数のほうが多く見ることが出来た。

〈補足2〉

実験1で作成したスペクトル(図1)について、実験2においては標準語、関西弁、平坦、mid2C、hiCの音声データをそれぞれa,b,c,d,eと表記する。

▶考察1

この実験ではスペクトルの見た目からその違いを見たものに過ぎない。得られたデータを数値化することで、具体的に比べることが出来るのではないかと考えた。

実験 スペクトルを用いた分析②

実験1においては、作成したスペクトルを見た目で判断したに過ぎなかった。そのためスペクトルに含まれる各周波数ごとに比較を行い、含まれる周波数とその割合を比較し、それぞれの音声と比較した。

▶実験2

実験1で用いた音声データと歌声のスペクトルに含まれる周波数をそれぞれ比較する。スペクトルはそれぞれ音量が異なる為一番多く含まれている周波数を基準に音量を揃え、その割合を表にして示す。

〈補足2〉

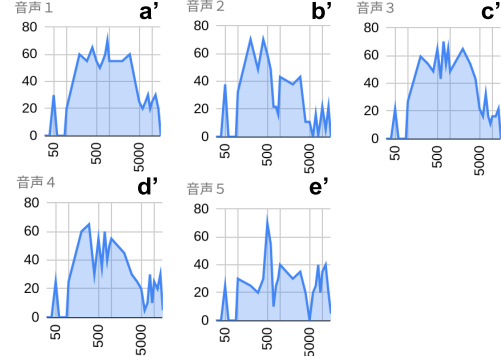
スペクトログラムで得られたデータから含まれる要素を抽出するにあたって、いくつかの基準となる要素を指定する。周波数の指定に際して、今回使用したアプリが31.5Hzから表示するものであったため、30Hzを起点とした。また使用したアプリにおいて、音程が高くなるほど要素の間隔が広がっていることを考慮して、基準を定めた。30Hzを起点として100Hzまでは10Hzずつ増加、100Hzから1000Hzまでは100Hzずつ増加、1000Hzから10000Hzまでは1000Hzずつ増加、10000Hzからは2000Hzずつ増加するものとする。終点は使用したアプリにおいて16000Hzまでを表示するものであったためである。

▶結果2 (表1)

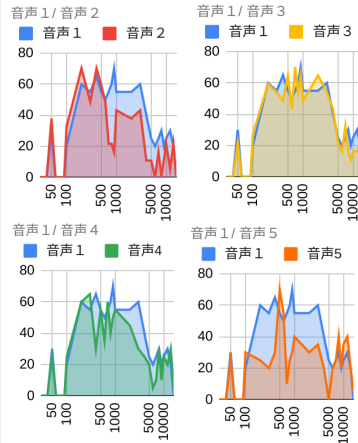
	1	2	2'	3	3'	4	4'	5	5'
30	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0
50	30	35.378	20	21.6	25	25	25	25	25
60	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0	0	0
100	20	30.324	25	27	25	25	30	30	30
200	60	65.702	55	59.4	60	60	65	65	65
300	55	45.486	50	54	55	55	60	60	60
400	65	65.702	45	48.6	30	30	30	30	30
500	55	55.594	60	64.8	55	55	70	70	70
600	50	45.486	40	43.2	35	35	55	55	55
700	55	20.216	65	70.2	60	60	10	10	10
800	60	20.216	50	54	40	40	25	25	25
900	70	15.162	60	64.8	50	50	30	30	30
1000	55	40.432	45	48.6	55	55	40	40	40
2000	55	35.378	60	64.8	45	45	30	30	30
3000	60	40.432	50	54	30	30	35	35	35
4000	40	10.108	40	43.2	25	25	20	20	20
5000	25	10.108	20	21.6	20	20	0	0	0
6000	20	0	15.162	5	5	5	20	20	20
7000	25	15.162	30	32.4	10	10	25	25	25
8000	30	10.108	15.162	30	30	40	40	40	40
9000	20	10.108	10.108	10	10	20	20	20	20
10000	25	20.216	15.162	25	25	35	35	35	35
12000	30	5.4	15.162	20	20	40	40	40	40
14000	20	20.216	20.216	30	30	20	20	20	20
16000	0	0	0	0	5	5	5	5	5

左の表は音声データと歌声のスペクトルの基準となる周波数を読み取り、音声1を基準に音量を調整したものの一覧である。この表はGoogleスプレッドシートで作成したもので、この表を基に以下のグラフを作成した。縦軸に大きさ【dB】横軸に周波数【Hz】を取り、スペクトルを簡略化したものに相当する。作成したグラフを以下に示す。

なお、作成した音声1～5のグラフをそれぞれa',b',c',d',e'と表記する。



(図3)



左方に示したのは、a'を基準としてb',c',d',e'を重ね合わせその違いについて示したものである。b'を比較したグラフでは、1000Hzから10000Hzの間に生じる差を見て取ることが出来る。一方でc'との比較を示したグラフでは、その差はほとんど見ることができず、500Hzから1000Hzの間において多少の差が見られることが分かった。次に音声データと歌声のスペクトルに生じた差について言及していく。d'との比較を示したグラフにおいては、3000Hzから5000Hzの間に見られる差のみが生じたのに対し、e'と比較したグラフでは、スペクトルに含まれる周波数については共通点が見られるものの、音量に大きな差が生じているということが分かる。

これらの相違点・共通点が生じた理由について、考察する。なお、a'とb'を比較したグラフをa'/b'とし、a'とc'のものをa'/c'、a'とd'のものをa'/d'、a'とe'のものをa'/e'とする。

▶考察

実験2の結果、a'/b'では大きな違いが見られたのに対し、a'/c'では僅かな差しか見られなかったことから、音声分析の技術は方言に対応することが難しいということが分かった。また肉声であるため合成音声と等しいとは言えないが、それに近い平坦なアクセントのものが標準語すなわち音声認証するとき使用するアクセントに類似しているということは、特定の人物の肉声に倣って作成された合成音声と一致してしまうことで、悪用されてしまう危険性を孕んでいるということが考えられる。またa'/d'にはほとんど違いが見られなかったのに対し、a'/e'では大きな違いが見られたのは、言うまでもなくその音程の差にある。mid2Cは比較的地声に近い高さであるのに対し、hiCは普段使うことは無い高さである。実験2では一番多く含まれていた周波数を基準にa'とその他の音声データを比較しているため、e'には他と異なる周波数が多く含まれていたことがこの結果を生んだと考えられる。さらにこのことを利用して、特定の周波数を音声認証に利用することができればいいと考えた。

課題と今後の展望

本研究を通して、バイOMETRICS認証にも用いられている音声認証技術を知るために複数の音声データの分析を行った。含まれる周波数の違いが与える影響を可視化することによって、その音声データを数値化し分析することは音声認識技術についてその仕組みを把握することに繋がったと考える。今後は本研究で分かった音声データに生じる特徴を基に、音声認識技術を開発し、その技術の悪用を防止するために本研究を活用していきたいと考える。

参考文献

- 1) 首都大学東京。「自分の声」と「他人の声」を識別する音響学的特徴を解明.公立大学法人首都大学東京.2013.
https://www.houjin-tmu.ac.jp/assets/library/2019/04/press_131024f245.pdf (参照 2023-01-31)
- 2) 峯松 信明.音声の音響分析の「いろは」.日本音声学会音声学普及委員会.2013.
<https://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/nlp+slp/lecture-02.pdf> (参照 2023-01-30)