

音声認識と音声学習

文学研究科英文学専攻博士後期課程満期退学

名倉 秀人

1. はじめに

AIは音声のビッグデータを数値として解析し、その傾向の確率で人の言葉をテキスト化している。音素を正確に捕らえているわけではなく、およその数値から推測しているのである。たとえば画像認識で「りんご」をAIが認識するときに、どのくらい球形に近いか、どのくらい赤いかを数値的に判断しているのと同じことだ。「だいたいこの音」という感覚をAIは学んでいる。だが、それ以前にAIはまず「音素」を学習している。どの音素の組み合わせが"pen"であって、"Ben"と違うかを習う。英語学習者もその音素がどのような数値で、どのような音の流れの傾向にあるかを知ることによって聴き易くなるのではないだろうか。音声学に音響学の要素を組み合わせ、数値化、グラフ化し、音声認識の仕組みを知れば、脳内で「音をテキスト化する」ことが可能になるかもしれない。その手法を教授できれば、英語教育にも役立つかもしれない。そう思い、音素をひとつひとつ画像化、数値化することにした。母音については昨年の紀要論文で述べたので、今回は子音について調査してみようと思う。結論は出なくても、まずはデータ化することに集中した。また、画像化・数値化した音声データを学生に示し、音声矯正のテストも行った。

2. 音声のサンプリング

映画やニュースなどから子音のみをサンプリングすることはかなり難しい。そこで、イギリス人の英語教師に手伝って頂いた。30代の女性で、普段からきれいな発音をしていると感じていたため、彼女にお願いをした。ネイティブスピーカーであっても、子音のみを純粋に発音することは難しく、その後ろや前に母音のような有声音が付き易い。そこを意識して発音してもらった。さらに、母音と、50個の短い単語をサンプリングした。マイクはSHURE製のダイナミック マイクロフォン BETA58Aという指向性の強いボーカル用マイクを使い、周囲の音が入らないよう、音楽スタジオで録音した。音量ができるだけ一定になるように気を付けた。リミッターやコンプレッサーなどのエフェクターは、自然な音ではなくなるため

使用していない。本来なら未圧縮のwavファイルで録音すべきなのだが、人間の可聴域以外を調査する必要はないので、動作の軽いmp3で録音した。約30分で作業は終わった。録音した音声は、著作権、著作隣接権に違反しないよう、この論文以外では用いないことを約束した。

3. 音声の数値化・画像化

音声の数値化と画像化にはPraatを用いる。Wavesurferなど、他にも音声分析に役立つソフトウェアは多数あるが、無料で動作が軽く、使い慣れているものを選んだ。設定はデフォルト。表示項目はSpectrum, Pitch, Intensity, Formant, Pulsesだが、Intensityは波形と一致するので表示しなかった。母音の場合は、第1、第2フォルマントの数値だけをサンプリングすればよかったのだが、子音はどの要素が音声の特性を表しているか予測できない。よって全ての項目で検証し、その差異を見出すことを目的とする。結果はPraatの機能のDrawを使うのではなく、WindowsのPrint Screenを使ってキャプチャーし、それをPhotoshopで編集する。Drawではひとつの画面に複数の項目を示すことが困難なためである。

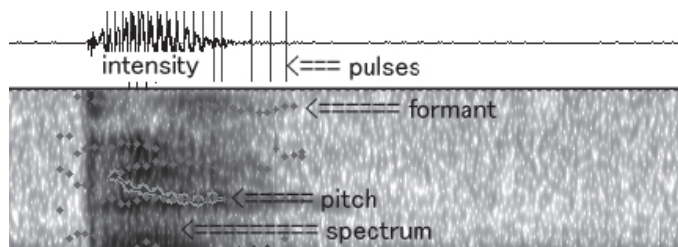
4. 分析の基準

子音はInternational Phonetic Alphabet (国際音声記号, IPA) に準拠して範疇化を行った。ただし、英語以外の音は使わないため、安井 (1992) の子音の分類表を使っていく。発音記号の表記フォントは、IPAに近いMicrosoft Sans Serifを使った。英語の子音は [p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, θ, ð, h, tʃ, ɔʃ, m, n, ŋ, l, j, w, r] の24個である。カテゴリーについては、英語表記は省略し、ここでは日本語表記で進めていく。まず、有声音 [b, d, g, v, z, ʒ, ð, ɔʃ, m, n, ŋ, l, j, w, r] の15個と無声音 [p, t, k, f, s, ʃ, θ, h, tʃ] の9個に分かれる。Praatでは有声音にはPulseが現れるため、区別し易い。さらにこれを調音様式による分類と、調音位置による分類を行う。調音様式による分類は、閉鎖音 [p, b, t, d, k, g]、摩擦音 [f, v, s, z, ʃ, ʒ, θ, ð, h]、側音 [l]、破擦音 [tʃ, ɔʃ]、鼻音 [m, n, ŋ]、半母音 [j, w, r] の6つのカテゴリーになる。サンプル提供者は英国人なので [r] は有聲摩擦音のカテゴリーに入るのだが、子音のみで録音したせいかアメリカ体系の音に近かったため、[r] を半母音のカテゴリーに回した。ここは次回録音するときに注意すべき点であり、今回の反省点でもある。調音位置による分類は、両唇音 [p, b, m, w]、唇歯音 [f, v]、舌歯音 [θ, ð]、歯茎音 [t, d, s, z, n, l, r]、硬口蓋歯茎音 [ʃ, ʒ, tʃ, ɔʃ]、硬口蓋音 [j]、軟口蓋音 [k, g, ŋ]、声門音 [h] の8つのカテゴリーになる。はじき音や咽頭音といった英語にはない音には触れないことにする。これをまとめた安井による分類表を引用しておく。初版では一部表記ミスと思われるものがあったので、筆者が修正した。

調音点 調音様式		唇音		舌頂音		舌背音			咽喉音
		両唇音	唇歯音	舌歯音	歯茎音	硬口蓋 歯茎音	硬口蓋音	軟口蓋音	声門音
閉鎖音	無声	p			t			k	
	有声	b			d			g	
摩擦音	無声		f	θ	s	ʃ			h
	有声		v	ð	z	ʒ			
側音	無声								
	有声				l				
破擦音	無声					tʃ			
	有声					dʒ			
鼻音	無声								
	有声	m			n			ŋ	
半母音	無声								
	有声	w				r	j		

5. 子音のフォルマント及び波形の画像

ここからPraatによって子音を画像化し、その特徴を探っていく。明確に差が出ないものについては、その原因を考察する。個人のデータであるため長さは意味をなさないのも、その形によってのみ判断する。「ある子音はもう一方の子音より長い」という相対的なデータは無駄ではないが、その秒数を計ることは今回の分析の場合判断基準にはなり得ない。



画像の上部はintensityを縦の振れ幅で表している。発音のエネルギーの推移を表し、有声音と無声音を共に持つ音は基本的に同じ形になることが予想される。（実際はそうではない場合もある）同時に上部には、有声音の場合pulsesが縦線となって現れる。pulsesは有声音と無声音を明確に区別する。画像の下部の黒の濃淡で表されている図はspectrum（spectrogram）で、エネルギーと音の高さの推移を表している。音声分析では、音の「大きさ」と「高さ」の違いを間違えないようにしなければならない。音が高くても、エネルギーが低い要素もあり得る。グラフが上部に集まっていれば、その部分は高い（周波数が高い）ことになる。色が濃ければ、その部分は強いことになる。spectrumは大きさと高さ（信号成分の強さと周波数）を時間系列で並べた三次元のグラフとなっている。音の高さはドッ

トが線でつながったグラフで表されるpitchでも確認できる。だが、それはその瞬間の総合的な高さである。音は、機械的に作った信号音でもない限り、ひとつの成分で出来ているわけではなく複数の成分で出来ている。特に人の声は、声帯を通った後様々な箇所反射し、最終的に固有の音波となって空気を伝わる。聞き手は耳から入って来た音波を瞬時に分析し、脳内で「意味」に変換する。音声器官 (the organs of speech) は、安井によれば13個ある。鼻腔 (nasal cavity)、上唇 (upper lip)、下唇 (lower lip)、上歯 (upper teeth)、下歯 (lower teeth)、歯茎 (alveolar ridge)、硬口蓋 (hard palate)、軟口蓋 (soft palate)、舌 (tongue)、口蓋垂 (uvula)、声帯 (vocal cords)、気管 (windpipe)、食道 (gullet) である。研究者や研究対象によってこの分類の仕方や数は違ってくるが、本稿ではこれを基準とする。これらを使って作られた人の声は、formantと呼ばれる波長の複数のピークを持つ。声帯を通った音源が音声器官の各部を通ることによって共鳴し、特定帯域ごとに倍音が増幅される。高い音だと感じる音でも、そこには少し低い音、さらに低い音が含まれていることになる。formantは人の声の場合、4つくらい見られる。数は決まっていない。3つの場合もある、5つの場合もある。低い方からF1, F2, F3...となっている。音声分析に使われるformantはおおよそF4までとなっている。母音はF1とF2の数値で舌の位置がわかるため、解析しやすい。筆者も前回の論文で分析したが、ほぼIPAの母音チャート通りの結果が出た。一方、子音は母音よりも複雑に音声器官が関わってくるため、単純に説明はできない。ひとつひとつ画像化し、その差異を調査しようと思う。

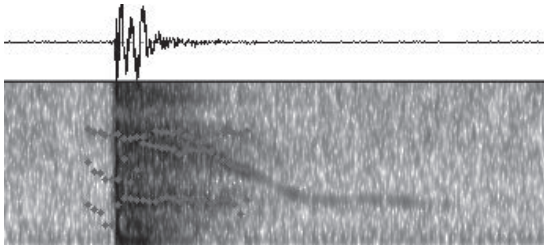
6. 有声音の子音の特徴

有声音の子音はpulsesが現れるため、Praatでの判別はしやすい。その他にもspectrogramに大きな特徴が見られる。Formantの最底部に強いエネルギーがある。これがF1になる場合もある、そうならない場合もある。低い位置のエネルギーは声帯の振動を表す。声帯振動には高い位置の倍音も含まれるはずである。ところが低いエネルギーの方だけが残っているのはなぜだろうか。川原 (2018) によれば、隣の部屋からの音楽がドラムのバスドラやベースのような低い音だけが響くのと同一原理であるとのことだ。ラジオでFMは周波数が高いため、トンネルに入ると受信し難くなるが、周波数の低いAMはそのまま聴けることがあるのもそのためだ。気管と閉じた声帯という閉鎖された空間では、低い周波数が目立つということだ。

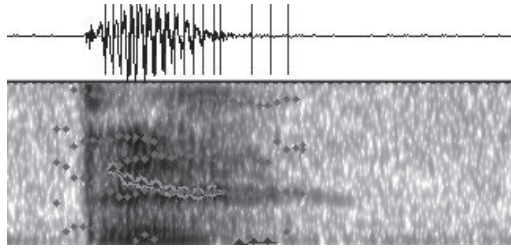
7. 閉鎖音

では、実際に子音のspectrogramを解析していってみよう。閉鎖音の6つの波形とspectrogramを並べる。同時にF1-F4の数値を計測し、表にする。小数点以下切り捨てで単位はHz。

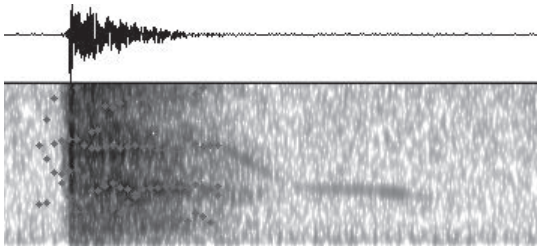
[p]



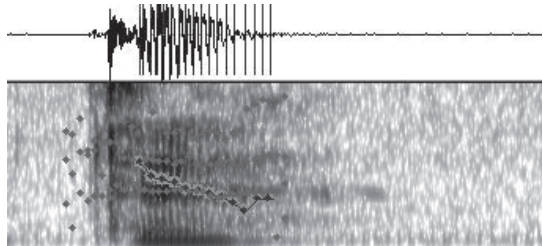
[b]



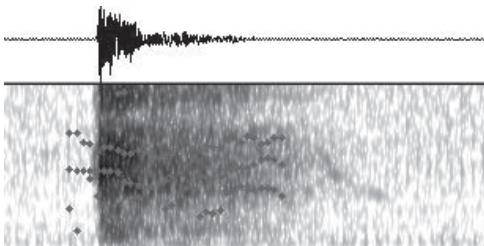
[t]



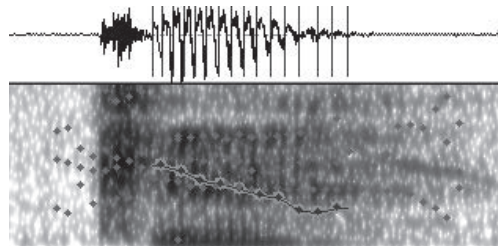
[d]



[k]



[g]

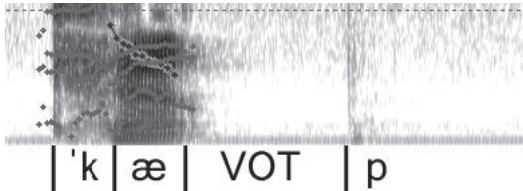


	F1	F2	F3	F4
[p]	1382	2944	4130	4866
[b]	1392	2853	4156	4750
[t]	1385	2697	3920	4953
[d]	1365	2656	3836	4768
[k]	1426	2717	3817	4901
[g]	1523	2784	3912	4833

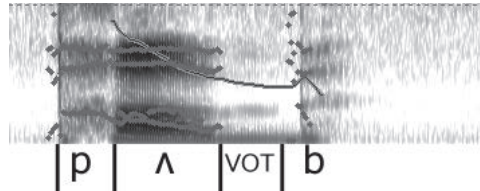
まず、閉鎖両唇音の [p]・[b] を比較する。閉鎖音は、閉じた状態から突然エネルギーを解放するため、音の発生開始からspectrogramが急に濃くなっている。[b] は有声音のため、低い位置にエネルギーがみられる。ここではそれがF1とはなっていない。無声音で口内に息を留めておくことはできるが、声帯を振動させながら口内にエネルギーを貯めておくことは難しい。「口の閉じの開放から声帯振動が始まるまでの時間」はVOT（Voice Onset Time；有聲開始時間）と定義されている。[p] のVOT、つまり唇を閉じて破裂するまでの時間は、[p] の子音を単独で発音させた場合spectrogramには表記されないが、[b]

より長い。証明するために前に母音を置き、cap [kæp] と pub [pʌb] のspectrogramの比較検証を行う。時間軸は同一スケールにした。

・ cap [kæp]



・ pub [pʌb]



capは母音の後に [p] が来ており、その前にかなり長時間の無音部分がある。pubのVOTは短い上に [ʌ] のF1を引っ張っている。完全な無音にはなっていないということもわかる。[b] は完全に破裂するまでに時間がかかるため、波形はなだらかに開始されている。有声音の [b]・[d]・[g] の破裂前の無声音部を見ると、長さが長くなっているのがわかる。これは舌が後ろに行き、鼻に抜ける確率が高くなるためである。

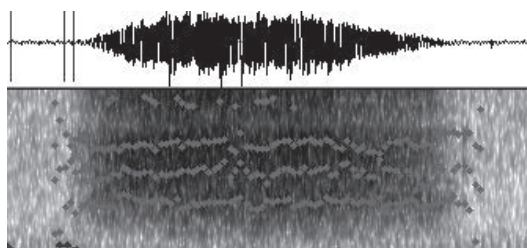
次に閉鎖歯茎音である [t]・[d] との比較をする。[p]・[b] と比較すると、よく似ているが、差異はある。閉鎖両唇音は唇が合わさってできる音である。唇は両方とも比較的柔らかい。閉鎖歯茎音は舌と（裏側の）歯茎によって作られる音である。舌は唇並みに柔らかいが、歯茎は硬い。硬ければ高い周波数の要素が増える。[t]・[d]の方が、F4の色がわずかに濃い。また、[t]・[d]の場合、唇は閉じてないので、[p]・[b]とは違って完全に閉鎖されてはいないため、VOTの部分にエネルギーが漏れる。波形の幅が全体的に [p]・[b]の方が緩く、[t]・[d]の方が密であることがわかる。

さらに閉塞軟口蓋音 [k]・[g] と比較する。これも軟口蓋とはいえ唇よりは硬いため、[p]・[b]より高い周波数の要素が現れる。また、F1の数値が高いということは、母音で言えば「あ」や [æ] のように、舌の位置が後ろにあることを表している。これは軟口蓋を使っている証拠であると言える。

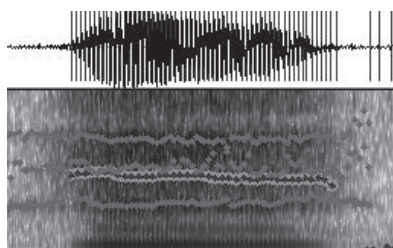
8. 摩擦音

摩擦音は [h] 以外、有声音と無声音があり、数も9種類と多い。そして高い周波数帯にエネルギーを持つ。歯や舌、歯茎などで非常に狭い間を空気が通るため、短い空間で共鳴し、周波数が高くなる。[h] だけは口の奥で広い範囲で摩擦を起こすため、周波数は少し低い。

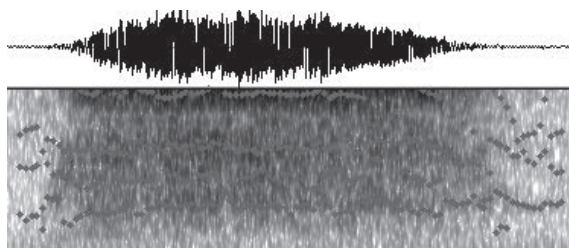
[f]



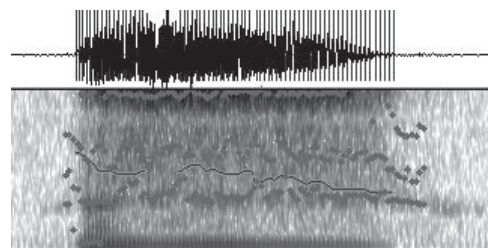
[v]



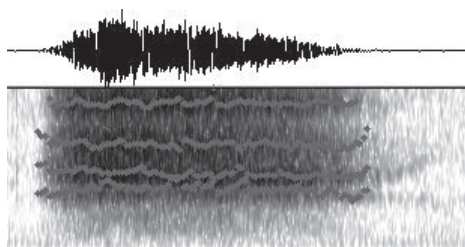
[s]



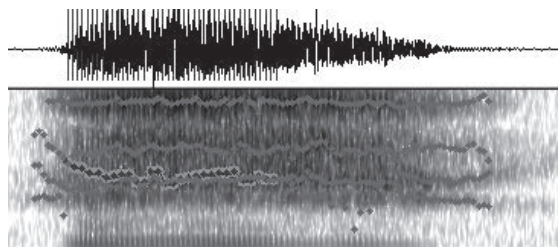
[z]



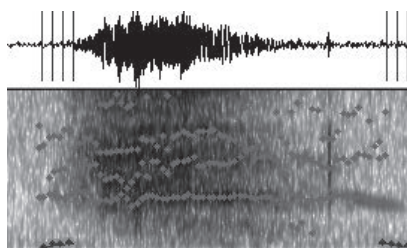
[ʃ]



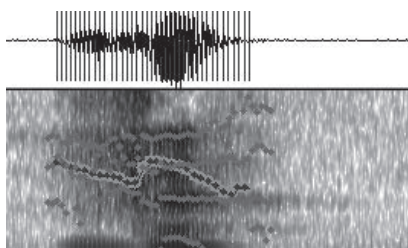
[ʒ]



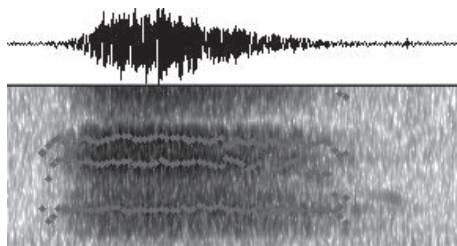
[θ]



[ð]



[h]

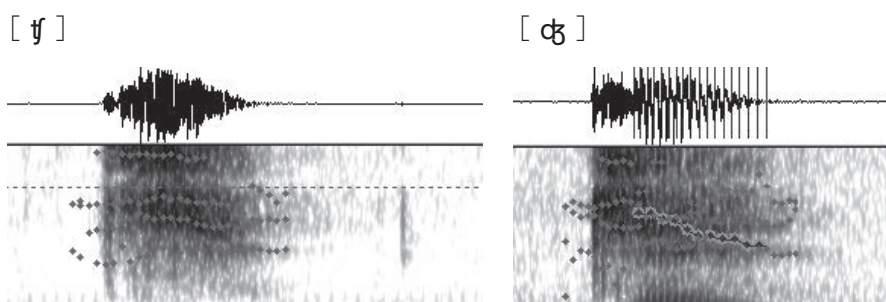


	F1	F2	F3	F4
[f]	1419	2699	3835	4997
[v]	1343	2715	3856	4777
[s]	1429	2637	3705	5204
[z]	1545	2961	4472	5317
[ʃ]	1632	2682	3845	4934
[ʒ]	1757	2849	4070	4966
[θ]	1521	2658	3764	4974
[ð]	1317	2623	3827	4527
[h]	1146	2668	3734	4894

閉鎖音と違って突然音量が上がるのではなく、徐々に大きくなり、徐々に小さくなっている。無声音で見ると、spectrogram全体が上部に集中している。またformantの推移が平行であることが特徴だ。そして閉鎖音よりも長い時間発音できる。例えば閉鎖音の [b] は、長く発音しようとする [bu:] と母音に変わってしまう。この中でも特徴的なのは [s] ・ [z] で、F4で5000Hzを超えている。歯茎音ではあるが、むしろ歯そのもので発音する傾向にあるため、硬い素材で鳴らすこととなり、高い倍音を持つ。木管楽器よりも金管楽器の方が耳に響く音を出すのと同じ原理だ。一方、[h] はF1が1146Hzと最も低い。咽喉音の声門音は声帯そのものを使う無声音で、口腔内に当たる摩擦音ではない。低い音の要素であるF1小さいことにより、高い音の要素が前に出るという傾向にあるようだ。また、唇音の唇歯音である [f] ・ [v] は閉鎖音と同じように下唇が動ため、「破裂音」に近い。完全閉鎖からの破裂ではなく、少しずつ息が漏れて行くため摩擦時間が長くなる。そして唇の振幅は [p] ・ [b] より大きい。そのため、発音に使うエネルギーが他の摩擦音より大きくなる。spectrogramが他の摩擦音より相対的に濃くなっている。余談だが、大学生時代に「Vで始まる単語はvigorとかvictoryとか『力』に関係する場合が多い」と習ったことを思い出した。硬口蓋歯茎音の [ʃ] ・ [ʒ] は、F1が1600Hz超えと最も高い。F1が高いということは、母音のformantチャートで見ると舌の位置が奥ということになる。硬口蓋の後ろの方で音を発声させているのがわかる。舌歯音である [θ] ・ [ð] は、数値的にも図表的にも、これらの中間に位置している。あえて言うなら、[s] ・ [z] の数値に近く、F4が低い。第二言語としての英語を、[θ] ・ [ð] に近い子音をもっていない私たち日本人が苦手とする理由は、このあたりにあるのかもしれない。

9. 破擦音

破擦音は破裂音ではあるが、解放時に摩擦音が生ずる子音である。[tʃ] ・ [dʒ] が英語の破擦音に当たる。

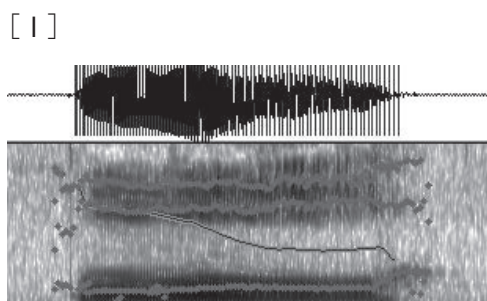


	F1	F2	F3	F4
[tʃ]	1514	2836	3719	5090
[tʃ]	1478	2862	3833	4861

spectrogramは摩擦音より閉鎖音に近い。突然開始されてはいるので、急に濃くなっている。伸ばすことができないという閉鎖音の特徴も持っている。一方F4の値が高いということは、摩擦音に近い。spectrogramも高い音を示している。閉鎖音・摩擦音との差は明らかである。

10. 側音

側音は歯茎音のみで、英語では有声音の [l] のみである。この音素のみを聴くと、ほぼ母音に感じた。母音の場合、舌は口内で必ず浮いているが、[l] は歯茎音のため歯の裏側の歯茎に舌先が着いている。空気は下の両側から出て行く。



	F1	F2	F3	F4
[l]	801	2795	4049	4600

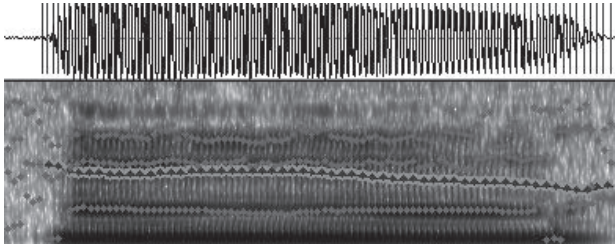
F1が今までの子音と比較してかなり低い。母音のformantチャートで言えば、舌の位置が前にあることになる。F2も2795と高めだ。F2が高いということは、舌の位置が高いということだ。歯の裏側に付いているなら、舌の位置は前で高いのは当然のことだ。一方F4は小さい。音としては低めになる。

11. 鼻音

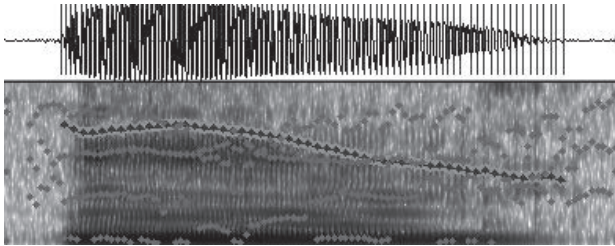
鼻音は口から空気が出ることはなく、鼻を通して発音される。[m][n][ŋ] の三つで、いずれも有声音。両唇音、歯茎音、軟口蓋音という区分で、唇や歯茎、軟口蓋から反射し、

振動は鼻に抜ける。唇で反射する場合は長い距離を、軟口蓋での反射は短い距離を空気は移動する。ところが、この距離だけではformantの高さを決定できない。硬さの要素も複雑に絡んでくる。

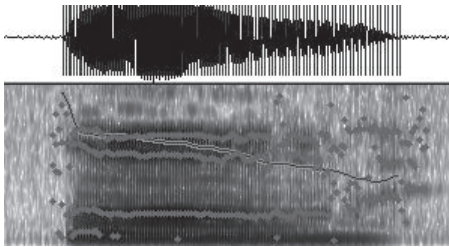
[m]



[n]



[ŋ]



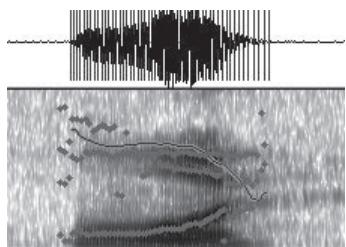
	F1	F2	F3	F4
[m]	1073	2680	3715	4750
[n]	985	2509	3661	4424
[ŋ]	1114	2764	3910	4538

F1はかなり低く、F4も低めだ。特に [n] [ŋ] のF4二つは子音の中で最も低い部類に入る。鼻腔の管は長いため、低い音が共鳴するのである。formantをみても、高い波長の色は薄い。また、藤村（2007）によれば、「口腔内にエネルギーが閉じ込められることによってある特定の周波数が弱められる」とある。その部分をantiformantsと言う。[n] と [ŋ] の差は、今回の実験で作ったformantの図ではわかり難いが、[m] [n] [ŋ] の順で、白い部分が低くなっている。その原理は、筆者が数学的知識に乏しく、理解できなかったが、現象は確かに存在した。

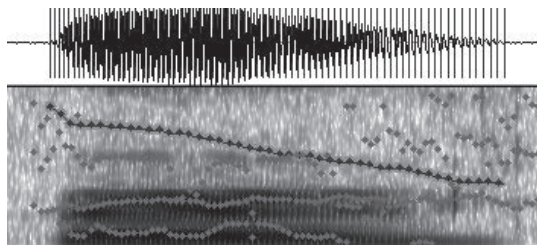
12. 半母音

半母音は、母音に続く語頭では子音としての役割をするが、間や最後に来ると母音的な役割をすることがある。辞書の発音記号上では、yearは [jɪə] と子音になるが、fireはfireは [ˈfaɪə] で母音化している。しかし、日本語でも「ファイヤー」と言うように、微妙な立ち位置にある。fewは [fjuː] で [j] をいわゆる「渡り音」として使っている。半母音の子音部分のみを発音してもらうことは、ネイティブスピーカーでも難しい。データは母音とともに発音されていると考えるべきであろう。種類は [w] [r] [j] の3つ。有声音で、それぞれ両唇音、硬口蓋歯茎音、硬口蓋音となる。

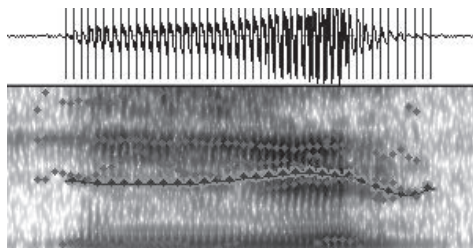
[w]



[r]



[j]



	F1	F2	F3	F4
[w]	1208	2749	3983	4762
[r]	912	2131	3551	4802
[j]	1119	2733	3844	4920

まず、私たち日本人の苦手な [ɹ] と [r] をformantの数値で比較してみよう。

	F1	F2	F3	F4
[l]	801	2795	4049	4600

両音とも母音に近いので、F1とF2で比較できる。F1は両方とも小さいので、舌が前方にあることがわかる。歯の裏に付いている [l] の方が前方にある。F2は [l] が高く、[r] は子音中最も低い。舌先は上がっているが、全体的には丸まっている分沈み込む。特徴的なのは [r] のF3だ。これも子音中最も低い。F3はF2と緩やかに比例していることがわかっており、F2で沈み込んだ舌がF3に影響を及ぼし、それが顕著に表れている音が [r] なのである。[j] は波形を見てもわかるように、途中で音が大きくなっている。ここに境目があると考え、[j] を前半と後半に分けてformantを計測し直した。

	F1	F2	F3	F4
[j] front	820	3008	3921	5013
[j] back	1539	2784	3553	3678

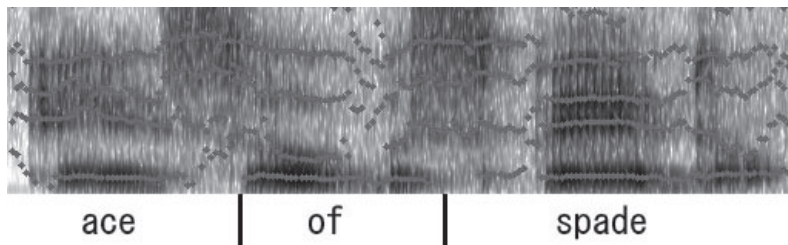
[j] frontのF1の数値が非常に低く、F2の数値が非常に高い。これは母音の [i] と同じ組み合わせで、舌が前方上方にあることを示している。それが [j] backの数値まで戻るということは [j] は [iə] の二重母音とほぼ同じということになる。[w] も、両唇音で唇をすぼめるため、どうしてもその後に開いてしまい、二重母音のようになってしまう。数値的には納得できるものは出なかったが、[w] [j] は子音の要素が少ない音であることがわかった。

13. 音声画像を使った発音矯正

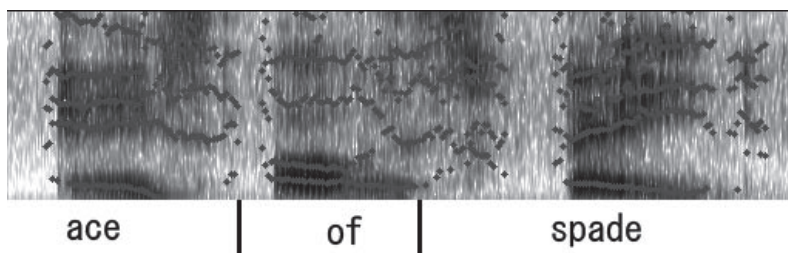
音声の画像化・数値化にはもうひとつ目標がある。それは、学生の発音矯正である。小学校からネイティブスピーカーと触れ合い、生の英語を学んできたわりには、発音がよくない学生が多いと感じている。では、私たち日本人教師が発音を教授するために何ができるだろうか。それは、今学生らが使っている英語の発音と、実際の発音が違うということ、数値や画像で確認させることではないだろうか。「何となく」ではなく、口の動きはこう、舌の位置はここと、具体的に見せることが必要となる。そこで、学生に協力して頂いて、発音した音声データを集めた。課題は以下の通り。

- A. “ace of spade”.を発音して録音し、その音声データをmanabaにアップロードして下さい。

今回の狙いは二重母音の修正である。子音の判定は学生の発音環境が一定ではなく、結果が上手く出なかった。教室で録音を行うと、どうしても周りの音が入ってしまうため、各自家でスマートフォンを使って録音させた。manabaとはweb上の学習支援システムである。学生からのファイルを集め、集計、配布できる。スマートフォンによって、音声ファイルの種類はwavであったりaccであったりするが、mp3に指定すると混乱が起きそうだったので、受け取ったファイルをmp3化した。集まったデータは、ノイズが入って使えなかったものを除いて、非英語科の学生25人分。学生は戸惑うことなくアップロードしてくれた。ほとんどの学生がace [**ers**] とspades [**sperid**] の二重母音を発音せず、長母音で発音していた。[**v**] は意外に正確に発音し、12人が正確に発音していた。その音声をPraatにかけ、画像を各学生にmanabaを使って渡した。以下の図はある学生の発音のformantである。



aceもspadeもF1とF2が平行である。これは長母音であることがわかる。これを一緒に見ながら、画像の持つ意味を説明し、二重母音というものの存在を教え、訓練し、再度提出するように言った。結果が以下の図である。



aceは変わらず長母音のままだったが、spadeは明らかに変わっていた。母音部分のF1が下降気味になり、F2は大きく上がっているのがわかる。この形は [**e**] から [**i**] へ口が移行したことを意味する。英語を専門としない学生にとって、この進歩は大きい。

14. 今後の課題

子音は子音だけで独立しておらず、必ず母音及び他の子音と繋がっている。その場合の変化はどうなるのか。また、単語と単語が繋がった時の変化はどうなるのか。そして最終的

には、formantを見ただけで何を言っているかがわかるようにしていきたい。発音矯正に関しては、録音環境が大事だということを改めて感じた。大人数で行うのではなく、まずは少ない人数で確実に録音できる環境で行うべきであった。今後も学生の協力を得て、発音矯正法を考えて行きたいと思っている。

参考文献

- 藤村靖 2007. 『音声科学原論』 岩波書店
- 川原繁人 2018. 『ビジュアル音声学』 pp144-171. 三省堂
- 北原真冬, 田嶋圭一, 田中邦佳 2017. 『音声学を学ぶ人のためのPraat入門』 ひつじ書房
- 名倉秀人 2019. 「AIと言語学」 東洋大学大学院紀要 第55集
- 篠田浩一 2017. 『機械学習プロフェッショナルシリーズ：音声認識』 講談社
- 安井泉 1992. 『音声学』 pp. 34-42. 開拓社

Voice recognition and learning English sounds

NAKURA, Hideto

Abstract

Speech recognition systems are created from "Big Data" of speech. With recognition of phonemes alone, the systems cannot deal with connection, disappearance, or weakening of sounds. It also models the patterns of collocations. This process is similar to a pattern that people learn language. Then, can we incorporate that method into English phonetic learning?

We, non-native speakers, will be able to listen to English if we can change English speech into texts accurately, like AI. If the sound which you are listening to is logically analyzed, expressed numerically, and visualized as a graph rather than intuition, we can understand it. That means "We can listen to English".

So, I tried to visualize English sounds using "Praat" and make "formants". Formants means what the sounds are. Last year I analyzed the vowels of English. This time, I attacked all of 24 consonants of English. The sounds of them are more complicated than vowels. And I tried to let my students who are not study English mainly set right their pronunciations.