5Q-09

英語発音学習のための音声の画像化
(Speech Visualization for English Pronunciartion Learning)

鈴木　里穂†　　山本　博史‡
近畿大学総合理工学研究科†　　近畿大学理工学部

1. 序論

近年の国際化の影響で，日本人にとって英語の習得は必須となりつつある．しかしながら，音節が母音で終わる開音節からなる日本語話者にとって，音節が子音で終わり得る閉音節であり，かつ子音連続を伴う英語の発声習得は，容易とは言いがたい．特に音声の場合は模範発声と自分の発声を同時に聞いて比較することができないことが，発声習得の妨げの原因の一つとなっていると考えられる．この問題の解決手段として，発音博士[1]，英語の会[2]等がある．英語博士はネイティブの手本となる発音を聞いた後にまねをして発声することで，手本と実際に発声した発がどう違うのかを発音記号で表し学習を補助するものであり，英語の会は発音記号を上手く発音できる舌の位置を図で示し，実際に動画を見ながら発音練習をするものである．しかしながら，これらでは，日本語に現れない子音連続等の特定の音素の組み合わせには対応できていない．そこで，本研究では，発声を画像化することによって，両者を直接比較可能にし，発声習得の一助とすることを試みる．この方法では，模範発声と自分の発声の差を直接比較することができ，かつ発声のどの部分がどれくらい模範発声と違うのかを目視できるようになることが期待できる．

2. 研究方針

音声は空気を媒体とする波であり，その構成要素としては，振幅，周波数，音色がある．人間は，声帯を振動させることで音を出し，それを口の動きで加工することによって音声を出す．その結果，音声には声帯の情報と口の動きの情報が含まれていることになる．従って，音声から声帯の情報を取り除くことによって，ケプストラムと呼ばれる口の動きの情報を得ることができる．振幅，周波数，音色を画像化する手法としては，フーリェ変換によるペクトラムの抽出が知られているが，本研究ではその内，口の動きの情報である，ケプストラムを画像化する．ケプストラムの抽出手順は以下の通りである．

- フレームデータを切り出す
- フレームの Wave データの平均値減算を行う
- 高域強調を行う
- 窓がけを行う
- 高速フーリエ変換によりスペクトラムを求める
- log パワースペクトラムを求める
- フィルタバンクのエネルギーに逆フーリェ変換を施してケプストラムを算出する

続いてこの log パワーペクトラムから声帯の情報を取り除く．声帯からは特定の周波数しか出ず、それが図の赤のカーブの山の頂点となる．この周波数の音は，口の中を通る時に，歪み，強調，抑制され，最終的に図１に示すような形となる．どの部分が強調>，抑制されるかは口の動きで決まるため，口の動きの情報は，この赤いカーブの概形である，青いカーブで示されることになる．赤いカーブの山の間隔は，概形にくらべ間隔が短いため、この log パワースペクトルをさらにフーリェ変換し，高周波数成分を取り除いた上で>，逆フーリェ変換を行うことで，ペクトラムの概形である青いカーブであるケプストラムを得ることができる．

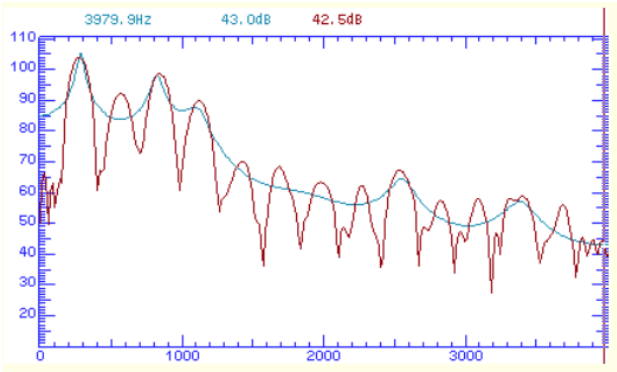

図１　スペクトラムの例

ケプストラムを抽出することによって，個人性の多くの部分を取り除くことができるが，各音（音素）の時間長，高さの違いが吸収できていない．そこで，模範発声と比較対象音声に対して，音声認識をかけることで，各音素の長さを得，時間補正をかける．

†Graduate School of Science and Engineering, Kinki University
‡Department of Science and Engineering, Kinki University

3. 実験

3.1 実験条件

ケプストラムの抽出条件は以下に示す通りである.

- フレーム長さ：20ms
- 入力音声の周波数：1600Hz
- フレームシフト幅：10ms
- ケプストラム数：26
- フィルタバンク数：20

実験に用いた音声は，模範発声として英語教員5 名，利用者として学生4 名を対象とし，各10 単語を対象とした．またノイズを軽減するために人がいな所で録音を行った．音声のサンプリング周波数は1,600Hz，フレーム周期10ms，ケプストラムの代表点は26 点ある．また，音声認識にはatrasr[3]を用いた．また，発声対象の単語は，spring, world,athlete, barbeque, birthday, business, clothes, February, liver, sixth, third, squirrel の12 単語とした．選定の理由は，日本語にない子音連続やth あるいは子音で終わる単語であることである．また，利用者と模範発声の比較を行う際には，5 名の英語教員のケプストラムのうち，利用者のそれに最も近い物を提示することとした.

3.2 実験結果

図 2 に、利用者の「 Spring 」のケプストラムを画像化したものを示す．図は白黒で表示してあるが，実際はサーモグラフィーでカラー化されている．ケプストラムの値は，それぞれの次元で 0 から 255 の値になるように補正をかけている.

図２　利用者の Spring のケプストラム

続いて図３に模範発声のケプストラムを，図４に利用者のケプストラムに対して，模範発声に合わせた時間補正をかけた物を示す．両者を比較すると，後半の ring の部分は良好であるが，前半の子音連続の sp の部分にはっきりと差が見られるのがわかる．また，図５と図６は「World」に対する模範発声と利用者のケプストラムであり，利用者の単語末尾には模範発声には見られない母音の挿入が起こっていることがわかる.

4. まとめと今後の課題

音声を画像化することによって，模範発声との直接比較が可能になった．今後の課題としては，今回は高さ補正を見送ったがあげられる．また，音声の

図３　模範発声の Spring のケプストラム

図４　利用者の Spring のケプストラム（時間補正後）

図５　模範発声の World のケプストラム

図６　利用者の World のケプストラム

高さ方向の補正をかけることによって，模範発声と利用者の発声を重ね合わせて，違いを強調させること，さらに違いの部分が口のどのような動きと対応しているかを示すことによって，より能動的な学習ができると期待される.

参考文献

[1] http://kccn.konan-u.ac.jp/ilc/english/index.html
[2]http://www.koto.co.jp/products/seephony.html
[3] http://www.atr.jp/topics/press_04092702_j.html