様式 C-19

科学研究費補助金研究成果報告書

平成２１年５月８日現在

研究種目：若手研究(B)
研究期間：2007～2008
課題番号：19700242
研究課題名（和文）　フォルマント理論を凌駕する音声知覚モデルの構築

研究課題名（英文）　A speech perception model integrating formant theory and whole-spectrum model.

研究代表者
伊藤　仁（ITO MASASHI）
東北大学・大学院工学研究科・助教
研究者番号：00436164

研究成果の概要：
合成母音を用いた知覚実験と，成人男女と 6～12 歳の子供を含む様々な話者の母音の音響分析の結果に基づいて，スペクトル全体形状に基づく母音知覚モデルを提案した。このモデルは，話者による音声の音響特性のばらつきを効率よく抑制し，母音の音韻性を高い精度で識別できる点に特徴がある。実音声を用いた性能評価実験により，モデルの識別性能が，手動で最適なフォルマント周波数を決定した場合と同等以上であることが示された。

交付額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	1,100,000	0	1,100,000
2008 年度	1,000,000	300,000	1,300,000
年度			
年度			
年度			
総　計	2,100,000	300,000	2,400,000

研究分野：総合領域
科研費の分科・細目：情報学・認知心理学
キーワード：認知科学，実験心理学，音声学，画像,文章,音声等認識

１．研究開始当初の背景

Peterson と Barney(1967)は，多数の話者の米語母音を音響的に分析し，母音が声道伝達関数の共振周波数(フォルマント周波数)により特徴付けられることを見出した。また Klatt(1982)は合成音声を用いた知覚実験の結果から，知覚される母音が主にフォルマント周波数により決定付けられることを示した。これらの知見から，人間は低次の 2 つか 3 つのフォルマント周波数に基づいて母音を知覚するという，所謂フォルマント理論が提案された。

フォルマント理論は，母音の調音と音響特性を整合的に説明できるため広く受け入れられてきたが，知覚のモデルとしてはいくつかの問題がある。もっとも重要な問題は，フォルマント周波数の抽出メカニズムである。自然音声の短時間スペクトルには声道伝達特性以外に，声帯振動の情報も畳み込まれているため，このスペクトルからフォルマント周波数を一意に決定することは容易ではない。この問題は，特に話者が子供の場合や二つのフォルマントピークが近接する場合などに顕著になるが，我々の知覚特性がこのような音声に対して劣化するという報告はなされていない。

研究代表者ら(2001)は，フォルマント理論の基礎となった知覚実験において，刺激が全極型の音声合成器で生成された点に注目した。全極型の合成器では，フォルマント周波数の変化に伴ってスペクトル全体の形状も変化するため，この形状自体が知覚の手がかりとなっていた可能性が否定できないからである。この仮説を検証するために，スペクトル全体形状とフォルマント周波数をパラメータとした合成音声を用いて知覚実験を行った。その結果，フォルマントピークを抑圧しても，スペクトル全体形状が維持されていれば知覚される母音は大きく変化しない事，また母音の前舌/後舌の知覚の手がかりとして，スペクトルの低周波数に対する高周波数の振幅比が，フォルマント周波数と同等以上の有効性を持つ事が明らかになった。

この結果は，いくつかの研究で追試されており，少なくとも定常母音の知覚においてはフォルマント周波数が唯一無二の手掛かりではないことが確認されている。またスペクトル全体形状の振幅比は，フォルマント周波数と比べて抽出が容易であり，音の周波数を内耳で空間に展開する聴覚系の生理学機構との整合性も高い。しかし，話者により音響特性の異なる自然音声に対して，この振幅比を計算するための具体的な周波数範囲や，前舌/後舌以外の特徴量(広母音/狭母音)をスペクトル形状からどのように評価するかについては，十分に調べられていない。

２．研究の目的

本研究課題では，母音知覚においてスペクトル全体の形状がフォルマント周波数と同等以上に重要であるという知見に注目し，フォルマント理論に関する実験と著者らの実験結果を統一的に説明できる音声知覚モデルを構築することを目的とする。

３．研究の方法

上記の目的を達成するために，本研究課題では２種類のアプローチを用いた。ひとつは，著者らの先行研究と同じ合成音声を用いた心理物理実験である。音響的な特徴量を制御した刺激に対する被験者の知覚特性を調べることで，知覚モデルの大まかな枠組みを定めることができる。ここでは上述した知見に基づく作業仮説を設定し，その妥当性を評価するための知覚実験を行った。

また，もうひとつの重要なアプローチは，自然発話の音響分析である。音響特性の異なる多数の話者の音声信号を分析することで，母音を特徴付ける音響パラメータを抽出した。さらに，この様にして得られた特徴パラメータを，知覚実験から得られた枠組みと統合し，母音の知覚モデルを構成した。モデルの妥当性は，自然発話音声を用いた自動識別実験により得られる識別性能(識別率)を，従来のフォルマントモデルと比較することで，定量的に評価した。

４．研究成果

(1) 母音の知覚実験

著者らの先行研究では，母音の前舌/後舌の知覚の手掛かりとして，スペクトルの低周波数成分に対する高周波数成分の振幅比が重要であることが示された。この実験では振幅比として刺激音の第１フォルマントピークに対する第３フォルマントピークの相対振幅を用いていたが，厳密にこの範囲の周波数成分が知覚を決定づけるか否かは十分に調べられていなかった。そこで，図１のような合成母音刺激を用いて知覚実験を行った。刺激は，通常の全極型の合成母音(control)，この母音から第２フォルマントピークを取り去った刺激(no-peak)，フォルマントは変えずに振幅比を一定とした刺激(no-raise)，ピークを取り去り振幅比も一定とした刺激(neither)の 4 種類である。被験者に，これらの刺激をランダムな順序で呈示し，日本語５母音のどれに最も近いか回答させた。

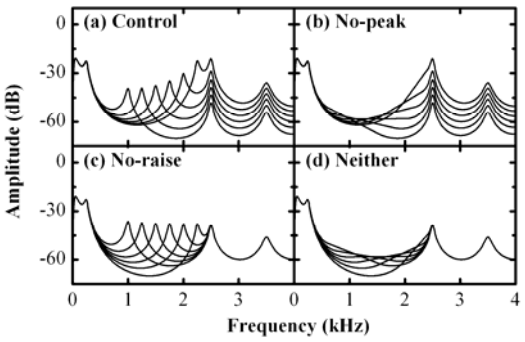

図 1. 知覚実験の刺激

図２に，８名の被験者から得られた母音の平均認識率を示す。刺激の種類に関わらず，これらは前舌母音の/i/か，後舌母音の/u/と知覚された。No-raise 刺激においては，どちらの母音が知覚されるかを決定するのは第２フォルマント周波数であり，認識率のパタンは Control 刺激に非常に近かった。この結果は，従来のフォルマント理論と整合するものである。また No-peak 刺激では，振幅比に応じて知覚される母音が変化し，これは著者らの先行研究の結果と整合するものであった。だが，フォルマント周波数も振幅比も特徴量として含まない Neither 刺激に対しても，知覚される母音は系統的に変化した。これは，従来無視されてきたスペクトルの微妙な形状(図 1d)が，母音知覚に影響し得ることを示す結果である。これらの結果から，前舌/後舌母音の知覚においては，特に第２フォルマント周波数近傍のスペクトル形

状が重要な役割を果たすことが明らかになった。

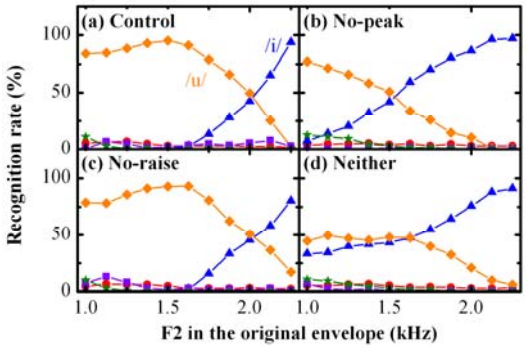

図 2. 知覚実験の結果

(2) 母音の音響分析

次に,様々な話者が発話した母音の音響特性を分析し，母音知覚を決定付けるスペクトル形状について調べた。対象とした母音は,成人男性 20 名，成人女性 20 名，子供 20 名が発話した日本語 5 母音 300 サンプルである。図 3 に，第 1,2 フォルマント周波数に基づく各母音の分布を示す。話者による音響特性のばらつきを抑制するため，各話者が発話した 5 つの母音の第 1,2 フォルマント周波数の幾何平均 GF を用いて，フォルマント周波数を正規化している。なお，フォルマント周波数は線形予測分析の結果に基づいて，最適値を手動で決定した。図の横軸が，GF に対する第 1 フォルマント周波数の比に，また縦軸が GF に対する第 2 フォルマント周波数の比にそれぞれ対応し，各母音の分布はマハラノビス距離が 1 となる楕円で表している。図から，横軸が狭母音/広母音に，また縦軸が前舌/後舌母音に大まかに対応し，フォルマント表現が調音と密接に関係することが確認できる。このフォルマント表現を用いた母音の識別率は 97.0 %であった。

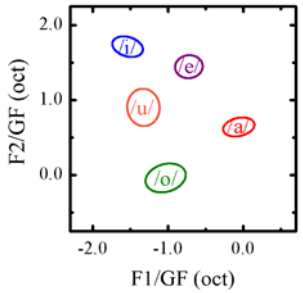

図 3. 母音のフォルマント表現

これらの母音データを，スペクトル全体形状に基づいて分析した。まず Pols ら(1967)と同様，1/3 oct フィルタ群を用いてスペクトル形状を多次元のベクトル空間に射影した。さらに，主成分分析を用いてこの特徴空間を 3 次元に圧縮した。なお，フォルマント分析の場合と同様，話者による音響特性のばらつきを抑制するために，1/3 oct フィルタ群の中心周波数は，各話者の全発話に基づいて設定した。得られた 3 次元の主成分空間における母音の分布を図 4 に示す。図の 3 つの軸(α, β, γ)は，第 1～3 主成分を回転して得られたものである。図から，α－β 平面における母音の分布は，図 3 のフォルマント表現による分布と良く似ていることが確認できる。しかし，この 2 次元を特徴量とした場合の母音識別率は 88.0 %であり，同じ条件のフォルマント表現より低かった。同様の結果は 1/3 oct フィルタ群を用いたこれまでの研究で得られており，スペクトル形状の単純な線形結合だけでは，フォルマント周波数に匹敵する効果的な特徴量を得ることが困難であることが明らかになった。

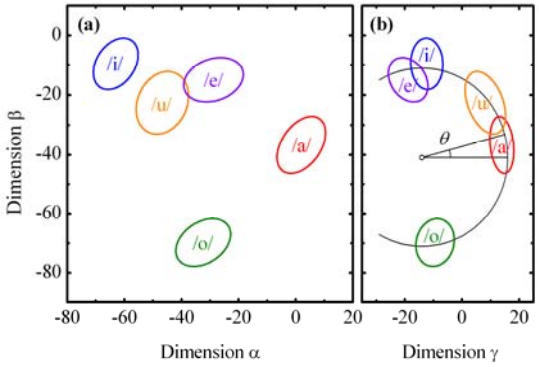

図 4. スペクトル形状による母音の分布①

図 4 のγ－β 平面に注目すると，全母音がほぼひとつの円周上に分布することが分かる。これは，この平面に極座標変換を導入することで，特徴次元をさらに圧縮できることを意味する。図 5 に，この変換を用いた母音の分布を示す。図 5a の縦軸が極座標の角度 θ に，図 5b の横軸が極座標の半径 ρ にそれぞれ対応する。得られたα－θ 平面における母音の分布は，図 3 のフォルマント表現により近いものになった。従って，スペクトル形状に基づく特徴量を用いた場合でも，調音と対応するパラメータを抽出することが可能であると言える。またα－θ 表現による母音の識別率は 97.7 %であり，同じ条件のフォルマント表現と同等以上の性能であった。

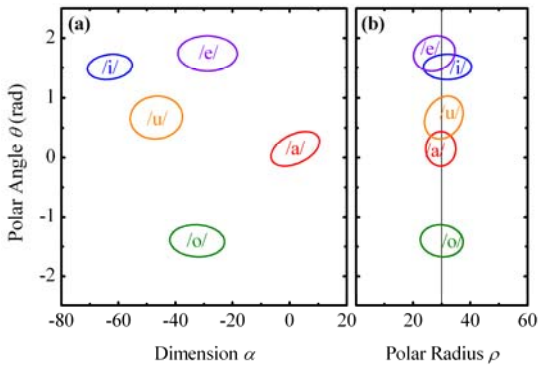

図 5. スペクトル形状による母音の分布②

(3) 母音の知覚モデル

これらの結果から，スペクトル形状に基づく母音の知覚モデルをまとめると，図 6 のようになる。まず入力母音のスペクトル形状を，1/3 oct フィルタ群により多次元ベクトルに変換する。次に，この形状ベクトルと，α，β，γ に対応する 3 つの特徴ベクトルとの内積を計算し，α－β－γ 空間における位置を得る。1/3 oct フィルタ群の中心周波数は，β－γ 平面で図 4b に示した円との距離が最小となるよう適切に調整する。なお，ここで得られる中心周波数は，話者の声道長を表すパラメータであり，話者認識の特徴量として利用することが可能である。

この最適中心周波数において，α－β－γ 空間の位置を極座標変換により α－θ 平面に射影する。ここで得られた α が広/狭母音の弁別に必要な特徴量を，また θ が前舌/後舌母音の識別に必要な特徴量をそれぞれ表す。例えば，α－θ 平面に射影された位置を，図 5a の各母音の分布と比較することで，入力母音の音韻性を決定する。上記の処理は，一次聴神経の応答と単純なシナプス結合だけで実現できると考えられ，生物学的な妥当性は高いと言える。また，従来の実験結果を少なくとも定性的には説明できる点に特徴がある。

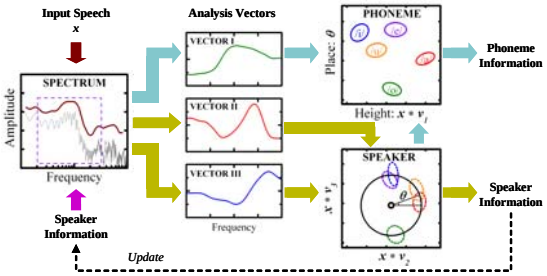

図 6. スペクトル形状による母音知覚モデル

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）査読有

[1] Ito, M., and Yano, M. (2007). “Sinusoidal modeling for nonstationary voiced speech based on a local vector transform,” J. Acoust. Soc. Am. 121(3), 1731-1741.

〔学会発表〕（計 5 件）

[1] 伊藤 仁, 伊藤 彰則, 矢野 雅文 (2009). “スペクトル全体形状モデルに基づく連続母音の音響特性,” 日本音響学会春季研究発表会, 東京(2009/3/17).
[2] 伊藤 仁, 小原 桂二, 伊藤 彰則, 矢野 雅文 (2008). “正弦波モデルに基づく非定常音声の分析と変調,” 日本音響学会秋季研究発表会, 福岡(2008/9/12).
[3] 伊藤 仁, 小原 桂二, 伊藤 彰則, 矢野 雅文 (2008). “正弦波モデルに基づく高品質音声変調の検討,” 応用音響研究会, 仙台(2008/8/4).
[4] 伊藤 仁, 矢野 雅文 (2008). “話速変換音声の知覚的自然性に関する検討,” 応用音響研究会, 東京(2008/3/7).
[5] Ito, M. and Yano, M. (2007). “Articulatory feature estimation for nonstationary vowels based on a local vector coding,” 19th international congress on acoustics, Madrid (2007/9/6).

６．研究組織

(1)研究代表者

伊藤 仁（ITO MASASHI）
東北大学・大学院工学研究科・助教
研究者番号：00436164

(2)研究分担者

なし

(3)連携研究者

なし