様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　24 年　6 月　16 日現在

機関番号：31303
研究種目：若手研究（Ｂ）
研究期間：2009～2011
課題番号：21700282
研究課題名（和文）　動的な話者適応性を考慮した統一的な音声知覚モデルの研究

研究課題名（英文）　A perceptual model of speech based on real-time speaker adaptation

研究代表者
伊藤　仁　（ITO MASASHI）
東北工業大学・知能エレクトロニクス学科・講師
研究者番号：00436164

研究成果の概要（和文）：
自然音声を用いた知覚実験により、異なる種類の母音の話者が 8 割以上の正答率で知覚されることを明らかにした。また成人 632 名の音声信号の音響分析において、正弦波モデルに基づく新たな高精度分析手法を開発した。さらに母音スペクトルのコサイン変換により得られる係数の 2 次結合が有効な特徴量であることが分かった。この結果は、母音知覚の手がかりが聴覚末梢系における 2 段のシナプス結合により実現される可能性を示唆する。

研究成果の概要（英文）：
Perceptual experiments indicated that speakers of different vowels could be correctly identified with accuracy of more than 80 %. Analyzing speech signals uttered by 632 speakers, a new analysis method was proposed on the basis of the sinusoidal representation of speech signal. Further, cosine expansion of speech spectra and the quadratic combination of their coefficients were shown to be effective features for vowel perception. The result supports the hypothesis that perceptual features for vowel might be extracted by two-step synaptic combination in auditory periphery.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	1,000,000	300,000	1,300,000
2010 年度	1,500,000	450,000	1,950,000
2011 年度	800,000	240,000	1,040,000
年度			
年度			
総　計	3,300,000	990,000	4,290,000

研究分野：総合領域
科研費の分科・細目：情報学・認知科学
キーワード：音声，話者適応，認知モデル，知覚実験

１．研究開始当初の背景

千葉と梶山の先駆的な研究を端緒とし、人間の音声知覚メカニズムの解明を目指す研究が続けられてきた。特に母音に関しては、知覚の手がかりが声道伝達特性の共振周波数であると考えるフォルマントモデルと、音声スペクトル全体の形状が知覚を決定付けると考えるスペクトル全体形状モデルの二種類が提案されている。

音声の生成、音響、知覚を統一的に説明できるフォルマントモデルは、広く受け入れられてきた有力な理論だが、研究代表者らが行った心理物理実験により知覚モデルとして不十分であることが明らかになった。この結

果は、他研究機関による追試でも忠実に再現されている。一方、後者の全体形状モデルは、音声スペクトルの形状を多次元の特徴量で表現するが、フォルマントモデルと同じ二次元では十分な識別性能が得られないこと、また調音など一般的な特徴との対応が見出し難いことから、音声の知覚モデルとしてやはり不完全であることが指摘されている。

研究代表者は、平成 19～20 年の文部科学省科学研究費補助金(若手 B：#197200242)において、単一話者の音声が全体形状を表す多次元空間の中で連続的な曲面上に分布することに着目し、この曲面から調音に対応する二次元の特徴量を抽出することに成功した。また話者の個人差の主要因が声道長であることから、対数周波数スペクトル上で特徴量を抽出するための周波数範囲を平行移動させることで、話者正規化が可能であることを見出した。この特徴量に基づく母音知覚モデルは、フォルマントモデルでは説明できない知覚現象を再現でき、かつ上で述べた全体形状モデルの問題点を解決できる可能性がある。

２．研究の目的

本研究課題では、前述した特徴抽出法に基づいて母音知覚モデルを構築し、人間の音声知覚過程の中で重要な役割を担う話者に対する動的な適応性の基本原理を解明することを目的とする。

これまでの知見から、物理的に同一の音声刺激が前後の文脈により異なる音韻と知覚され得ることが知られている。この文脈効果のメカニズムは、フォルマントモデルとスペクトル全体形状モデルどちらを用いても十分に説明することができない。その原因は、個人性と音韻性の同時推定の難しさにあると考えられる。通常人間は、入力された音声から、その音韻性と個人性を同時に決定する。これらの特性は互いを規定する拘束条件であり、一方が得られれば他方は付随的に定まるが、実際の知覚ではこれら両方が未定である場合が多く、聴覚系がこの問題をどのようにして解決しているか明らかではない。

そこで、音声の個人性と音韻性を競合させた刺激を用いた知覚実験を行い、この問題を解決する手法を検討する。これまでフォルマントモデル、スペクトル全体形状モデル双方に関して、肯定的・否定的な多数の知見が報告されている。本研究課題では、これらの知見全てを整合的に説明できる統一的な知覚モデルの実現を目指す。

３．研究の方法

本研究課題は、まず多数の話者が発話した音声を分析し、スペクトル全体形状に基づく特徴量を抽出する。その後、この特徴量をパラメータとして刺激音声を作成し、心理物理実験により話者適応のメカニズムについて調べる。

話者適応のメカニズムについて効率的に検討するためには、入力として複数の音素を含む連続音声信号を用いることが有効である。この様な音声信号を高い精度で分析する手法のひとつとして、本研究課題では研究代表者らが開発した Local Vector Transform (LVT)を用いる。LVT は入力信号を振幅と周波数が滑らかに時間変化する正弦波成分の和として近似する正弦波モデルであり、音声の基本周波数やフォルマント周波数が時間変化する場合でも高精度の分析ができる点に特徴がある。この手法を用いて、多数の話者が発話した音声信号を分析し、スペクトル形状を少ないパラメータで効率的に表現する手法を検討する。

次に、上記の分析結果に基づいて知覚実験を行う。まず大まかな傾向を把握するために、予備実験として母音知覚実験を行う。ここでは、被験者が未知の話者の音声の個人性をどの程度正確に把握できるか定量化する。刺激は、母音３つを継続的に接続した３つ組（AB-X）で、A と B の音韻は等しく話者は異なる。また母音 X の話者は A か B いずれかに等しいものとする。被験者は、例えば /a/-/a/-/i/のような刺激を聞いて、最後の母音と同じ話者の母音が１番目か２番目かを答える。

最後に、連続音声を用いてこれと同様の実験を行う。刺激は、/aia/-/aia/-/eoe/の様な連続母音３つ組とし、音響分析で得られたスペクトル形状パラメータを制御して、話者性と音韻性が競合するように設計する。被験者は、最後の音声と同じ話者の音声が１番目か２番目かを回答し、かつ３番目の刺激が音韻として何と聞こえたかも答える。得られた実験結果を解析し、定量的な知覚モデルを構築する。

４．研究成果

まず音響分析を行うために、LVT に基づく正弦波モデルを改良し、有声音声信号を 20 dB 以上の高い S/N で分析する手法を開発した。この手法は、LVT で得られた第一調波成分の瞬時位相を用いて入力信号の時間軸を変換するもので、音声信号の時間変化特性を広い周波数帯域で高精度に分析することができた。この成果は雑誌論文[1-3]に発表し、2009 年度の FIT 論文賞、及び 2010 年度の石田記念財団研究奨励賞を受賞した。

また知覚実験に関しては、3 種類の予備実験を行い、その結果を国内及び国際学会で発表した（学会発表[2-7]）。特に研究方法で述べた ABX 法を用いた予備実験では、人間が母音の話者を 80～95 %の正答率で正しく判定す

ることができることを明らかにした。この結果は、我々の聴覚系に入力された音声だけでなく、その話者が他の音韻を発話した場合にどのような性質を持つか予測する話者モデルが存在することを示唆している。

しかし、予備実験で得られた知覚特性は、話者、母音、被験者間のばらつきが大きく、一般的な性質を導出することはできなかった。AB-X 法では、刺激の総数を現実的な規模に抑えるために、話者の数を少なくせざるを得ない。この様な制限のもとでばらつきの少ない知覚特性を得るためには、話者として用いる音声サンプルを慎重に選択する必要がある。

そこで、成人男性 313 名、成人女性 319 名が発話した大量の音声サンプルを分析し、これらのパラメータから知覚に関係する本質的な特徴量を抽出する手法について検討した。いくつかの手法を比較した結果、入力信号の 5 kHz までの短時間フーリエスペクトルをコサイン展開し、その係数を 2 次結合による 3 つの特徴量が、フォルマント周波数とも整合する有効な特徴量であることを見出した（学会発表[1]）。この結果は、聴覚末梢系で周波数チャネルごとに分解された多次元の情報から、母音の知覚に対応する数次元の情報が、理論的には 2 段階のシナプス結合で実現できる可能性を示唆するものである。

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 3 件）

[1] 伊藤 仁, 伊藤 彰則 (2010). “局所変化率変換と時間軸変換に基づく有声音声の正弦波モデル,” 電子情報通信学会論文誌 Vol. J93-D(9), pp.1745-1754.
[2] Ito, M., Ohara, K., Ito, A., and Yano, M. (2009). “A source-filter separation for non-stationary voiced speech based on sinusoidal representation,” Acoustical Science and Technology Vol.31(2), pp.181-184.
[3] 伊藤 仁, 伊藤 彰則 (2009). “局所変化率変換に基づく有声音声の正弦波モデル,” 第 8 回情報科学技術フォーラム講演論文集 Vol2, 43-48.

〔学会発表〕（計 7 件）

[1] 伊藤 仁, 蒔苗 久則 (2012). “ケプストラム係数を用いた母音のフォルマント分析,” 日本音響学会 2012 年春季研究発表会講演論文集 pp.393-394.
[2] 伊藤 仁, 小原 桂二, 伊藤 彰則, 矢野 雅文 (2010). “フォルマントとスペクトル全体形状を統合した母音知覚モデルの検討,” 日本音響学会 2010 年春季研究発表会講演論文集 1-7-10.
[3] 小原 桂二, 伊藤 仁, 矢野 雅文 (2010). “フォルマントピークとスペクトル傾きが母音知覚に及ぼす影響,” 日本音響学 2010 年春季研究発表会講演論文集 1-7-9.
[4] 岩佐 尚輝, 亀井 大陸, 伊藤 仁 (2011). “話者認識における母音の音韻性の影響,” 平成 23 年東北地区若手研究者研究発表会 p.125-126.
[5] Ito, M., Ohara, K., Ito, A., and Yano, M. (2010). “An effect of formant amplitude in vowel perception,” Interspeech 2010 (Makuhari), Wed-S3-P3-10. pp.161.
[6] Ito, M., Ohara, K., Ito, A. and Yano, M. (2009). “Relative importance of formant and whole-spectral cues for vowel perception,” Interspeech 2009 (Brighton), Mon-S2-P1-1.
[7] 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 2009 年春季研究発表会講演論文集 1-6-15.

〔図書〕（計 0 件）

〔産業財産権〕
○出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者

伊藤　仁（ITO MASASHI）
東北工業大学・知能エレクトロニクス学科・講師
研究者番号：00436164

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：