

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 25 年 5 月 31 日現在

機関番号：14701

研究種目：基盤研究(B)

研究期間：2009～2012

課題番号：21300069

研究課題名（和文） 音声知覚の基盤となる聴覚特性と計算理論の研究

研究課題名（英文） Experimental and theoretical studies on the auditory system for speech perception

研究代表者

入野 俊夫 (IRINO TOSHIO)

和歌山大学・システム工学部・教授

研究者番号：20346331

研究成果の概要（和文）：

音声知覚の基盤となる聴覚特性を明確にし、数理的な理論の構築／検証を行った。1) 寸法・形状知覚：発声方法による寸法弁別閾の違いが無いことや時間特性を明確にした。2) 聴覚フィルタ特性／難聴者・健聴者の聴覚特性：聴覚フィルタの周波数選択性や圧縮特性の同時測定と、模擬難聴を実現できる枠組みを世界に先駆けて開発した。3) 機能的磁気共鳴像 (fMRI) 実験：音声からの寸法知覚の情報処理の座に関して知見を得た。4) 音声知覚モデル化／音声・音響処理：理論的な背景をもとに話者の声道長推定が精度良くできることを示した。また、知覚的音響処理の改善も行った。

研究成果の概要（英文）：

We performed the experimental and computational studies on the auditory system which provides the basis of speech perception. 1) Perception of size and shape from speech sounds: We found the size JNDs are almost the same for voiced and whispered speech and sufficiently small even for short sounds. 2) Auditory filter characteristics and perception by hearing impaired and normal hearing: We proposed a new method to measure auditory filter selectivity and compression simultaneously. We developed a signal processing technique to simulate hearing impairment within a unified gammachirp framework. 3) Functional MRI (fMRI) experiments: We studied the location of size processing in the auditory cortex. 4) Modeling of speech perception and signal processing: We demonstrated the gammachirp filterbank outperforms conventional mel-frequency filterbanks in the task of vocal tract length estimation. We also improved the STRAIGHT system for speech experiments.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	4,700,000	1,410,000	6,110,000
2010 年度	3,400,000	1,020,000	4,420,000
2011 年度	3,600,000	1,080,000	4,680,000
2012 年度	2,200,000	660,000	2,860,000
年度			
総計	13,900,000	4,170,000	18,070,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：聴覚計算理論・音声信号処理・寸法知覚・聴覚心理実験・神経科学・fMRI

1. 研究開始当初の背景

音声は、人と人とのコミュニケーションにおいてもっとも基本的で重要な手段である。音声を受容するためには、聴覚系が不可欠であることは言うまでもない。この聴覚系を理解するために、実験心理学や生理学の研究が着実に進められているが、それを数理的に定式化して理解する「計算論アプローチ」は遅々として進んでこなかった。この状況は、視覚の計算理論と実験的研究が相互作用起こし、強力に研究推進されるようになったのと対照的ではある。この状況を打開すべく、平成 15 年度より科研基盤 B「聴覚計算理論の構築とそれに基づく音信号処理の研究」（課題番号 15300061）を開始し、平成 18 年度より科研基盤 B「初期聴覚系における寸法・形状知覚理論の検証と応用に関する研究」（課題番号 18300060）と科研萌芽「音声知覚特性の解析に対する音声認識技術の適用」（課題番号 18652040）を進め、インパクトが大きい成果をあげてきた。それは、我々が独自開発して発展させ、多くの研究者に受け入れられている 2 つの only-one 理論/技術である、「ガンマチャープ聴覚フィルタを含めた寸法・形状知覚理論」と「高品質音声分析合成法 STRAIGHT」に関するものである。具体的には、(1) 動的圧縮型ガンマチャープの提案・(2) 周波数選択性/圧縮特性の定量化・(3) 聴覚計算理論とモデルの提案・(4) 寸法知覚実験・(5) STRAIGHT を用いたモーフィング処理・(6) 実時間 STRAIGHT・(7) TANDEM- STRAIGHT の提案等が挙げられる。これらの背景を本研究の基盤とし、以下の目的で述べる新規課題を提案した。

2. 研究の目的

音声知覚の基盤となる聴覚特性を数理的に定式化/モデル化することを目指し、以下の課題を検討することを目的とした。

(1) 寸法・形状知覚: 音源の寸法/形状知覚に関する実験的検討をさらに進め、その特性や有効範囲を明確にする。

(2) 聴覚フィルタ特性/難聴者・健聴者の聴覚特性: 聴覚末梢系のフィルタの周波数特性や圧縮特性から、音声知覚特性に至るまでの様々な段階での心理実験とモデルによる定量化を行う。さらに、(特に難聴者の) 心理実験の負担を減らすため、測定データ削減や精度向上を目指す。健聴者が難聴者の「聞こえにくさ」を体験できる模擬難聴の実現に向けた検討を行なう。

(3) 機能的磁気共鳴像 (fMRI) 実験: 音声知覚における寸法・形状分離抽出機能の脳内部位を明らかにし、聴覚モデルの制約条件として活用する。

(4) 音声知覚モデル化/音声・音響処理 以上の知見に基づき聴覚の計算理論/計算モデルを確立し、補聴器信号処理や知覚的音響処理に応用する。

3. 研究の方法

(1) 寸法・形状知覚: 高品質音声分析合成法 STRAIGHT によるモーフィング処理を利用した、音韻性判断や寸法判断の順応実験について検討する。さらに、音響的に特徴が大きく異なる通常発声 (有声) 音声とささやき声での寸法知覚特性の違いや共通性の検討や、寸法知覚の時間特性の検討を行うための実験を進める。これらの知見をまとめ、聴覚モデルの理論的な制約として用いることができるかを検討する。

(2) 聴覚フィルタ特性/難聴者・健聴者の聴覚特性: 独自技術の「圧縮型ガンマチャープ」を用いた同時適合法により、聴覚末梢系の周波数特性や圧縮特性を同時にかつ安定に推定する方法を確立する。さらに、測定を高精度でかつ短時間でいう簡易測定法の確立のために、感度解析を用いて測定点のデータ削減量と推定精度の関係を調査する。難聴者の「聞こえにくさ」を健聴者が体験できる模擬難聴を実現するための信号処理を検討する。また、様々な聴力特性を持つ人間の音声知覚特性を解明するために、劣化音声の認識特性に関して自動音声認識器 (ASR) との対比を行う。

(3)機能的磁気共鳴像 (fMRI) 実験： 寸法情報処理部位解明のため、前基盤研究 B の枠内で収集した fMRI 実験のデータの解析を継続して実施する。さらに、音声と非音声の相違と類似点を探る fMRI 実験を実施して分析を行う。両実験から得られた知見を音声の聴覚モデルの制約という立場から解釈を試みる。

(4)音声知覚モデル化／音声・音響処理

音声知覚を説明する聴覚モデルおよび、応用技術に関して検討を行う。

① 聴覚系の寸法知覚に関連したモデル／理論の妥当性や有効性を示す。特に「安定化ウェーブレット・メリン変換」の骨組みとなっている「スケール共変表現」や「ガンマチャープ聴覚フィルタバンク」(GCFB)の検証と、工学的応用を検討する。具体的には、話者の声道長(寸法)推定法や大人・子供の判別や年齢推定を、従来の特徴量(例えば良く使われる MFCC)と比較し有効性を示す。さらに、MRI 測定から得られた実際の声道長データと提案手法の推定値を対比することにより、物理的な背景との整合性を確認する。さらに、GCFBをC言語や並列化することにより処理速度の向上を図り、応用研究に供する。

② 音声分析合成法 STRAIGHT の合成音は極めて自然で、音声知覚心理実験に幅広く応用可能な知覚的音響処理となっている。しかし、しかし音声変換際に生じる品質劣化はまだ避けられていないため、さらなる改良を行う。また、実時間処理版や TANDEM 法に関してもさらなる高速化・高品質化を推進する。

4. 研究成果

主な発表論文等の論文番号を引用し記述する。雑誌論文は[1]…[4]、学会発表はイタリック [1]…[19]、図書はボールド [1]…[4]で示す。

(1)寸法・形状知覚：

通常発声の音声とささやき声における寸法・形状知覚に関して実験した結果を分析してまとめた。通常発声とささやき声で寸法弁別閾は約 5%で相違がないことがわかった。こ

れは音の大きさ(ラウドネス)の弁別閾約 10%よりもずっと感度が高いことを示している。また、通常発声はささやき声よりも明瞭度は通常高いものの、不自然な基本周波数と声道長の組み合わせでは逆転する場合もあることがわかった[1, 19, 2]。これは聴覚系での処理をモデル化する上で、重要な制約条件の知見が得られたことになる。

さらに、寸法知覚の時間特性に関して検討を進めた[4, 17]。この結果から、約 30ms ほどのごく短い刺激音でも、10%以上の弁別閾を得られることがわかった。この時間は、聴覚系の時間積分の時定数とほぼ同じである。すなわち時間積分の機構により安定して寸法判断ができていている可能性を示している。

高品質音声分析合成法 STRAIGHT によるモーフィング音声を利用した、音韻性判断や寸法判断の順応実験について検討した。当初の目論みとは異なり、順応は極めておこりにくいことがわかった。同じ刺激音が変化したと知覚されるのではなく、まったく違う音源からの情報が入ったと知覚されているものと考えられる。音声処理のトップダウン情報も利用した聴覚情景分析 [6]が行なわれているためと考えられる。発表には至らなかったが、理論の考察 [14, 4]や、実験刺激音を検討する機会となり、今後の進展につなげられるものであった。

(2)聴覚フィルタ特性／難聴者・健聴者の聴覚特性：

only-one 技術の圧縮型ガンマチャープ [12, 16, 1]を用いた同時適合法により、聴覚末梢系の周波数選択性や圧縮特性を同時にかつ安定に推定する方法を提案した[3, 4]。周波数選択性を計るために確立したノッチ雑音法に、非対称な音圧レベルを導入することにより圧縮特性もできるだけ正確に推定しようというものである。この結果、提案法の方がより妥当と思われる圧縮特性を推定できることがわかった。圧縮特性推定法として最近提案されている順向性マスキング法を用いた方法との比較が今後課題となった。

上記の実験は健聴者でも7時間以上かかり、高齢者や難聴者にとっては負担が大きすぎた。そこで、高精度でかつ短時間でいう簡易測定法の確立のために、感度解析を用いて測定点のデータ削減量と推定精度の関係を調査した。この結果、おおむね半分程度まで測定点を削減できそうであることがわかった[1]。今後、閾値推定の統計的手法の改良も加えてさらに改善をする必要があるが、難聴者の特性も測定できるようにするための一歩となったと考える。

上記の手法の提案と同時に、難聴者の「聞こえにくさ」を健聴者が体験できる模擬難聴のための信号処理を、GCFBによって実現できることを示した[3]。これは、周波数選択性や圧縮特性の推定と、信号処理を1つの枠組みで行なう世界で唯一の手法の提案で、画期的な成果である。

また、様々な聴力特性を持つ人間の音声知覚特性を解明するために、劣化音声の認識特性に関して自動音声認識器(ASR)との対比を行い、論文化を進めている。

(3) 機能的磁気共鳴像(fMRI)実験:

寸法情報処理部位解明のため、前基盤研究B(21300069)の枠内で収集したfMRI実験のデータの解析を継続して実施し、結果を国際会議で報告した[18]。計算理論[4]を裏付けるまでの精度はまだ得られていないものの、寸法処理の先行fMRI研究と整合性がある部分と異なる部分があることを明確にできた。さらに、音声と非音声の相違と類似点を探るfMRI実験を設計し、実施して分析を行なった[11]。こちらは解釈が難しく、今後さらなる検討が必要であることがわかった。

(4) 音声知覚モデル化/音声・音響処理:

音声知覚を説明する聴覚モデルおよび、応用技術に関して検討を行った。

① 話者の声道長(寸法)の推定に関して、ガンマチャープ聴覚フィルタバンク(GCFB)、ガンマトーン聴覚フィルタバンク(GTFB)、メル周波数フィルタバンク(MFFB、音声認識で最

も使われるMFCCを計算するためのフィルタバンク)を対比した。さらに、計算で利用する周波数範囲を明示的に制御して、最も推定精度が良い条件を探った。この結果、非線形の動的圧縮型のGCFBを用いて、周波数範囲を500Hz以上とする場合が最も推定精度が良いことがわかった[3, 5, 7, 8, 10]。さらに、MRI測定から得られた実際の声道長データと提案手法の推定値を対比も行い、両者の整合性の確認を行ない、妥当性を示した。

このことは、「安定化ウェーブレット・メル変換」の理論から予測される事柄に2つの点で整合性がある。1つめは、「スケール共変表現」が500Hz以上で最も有効に働くことで、もう1つはGCFBが最適フィルタ系をなして寸法変形に対して透明であることである。このような特性があるがゆえに、推定精度も良くなったと解釈できる。

さらにこの結果を受けて、GCFBを用いた音声からの大人・子供の自動判別や年齢推定を行なった。この場合、GCFB出力の時間変調成分を特徴量とすると、他の特徴量よりも良い推定精度が得られることがわかった[2]。

実際に工学応用する場合には、信号処理の速度も重要な要因となる。特に動的圧縮型のGCFBは、matlabを用いてサンプル点ごとに特性を更新するため、非常に計算速度が遅かった。この状況を改善するため、GCFBをC言語化して速度向上を目指した。完全C版でなくmatlabコンパイラ等も援用したもので約2倍の高速化は実現できた[13]。その後もアルゴリズム改良や再コンパイル等を試しておりさらなる高速化ができています。

② 音声分析合成法STRAIGHTの合成音は極めて自然で、音声知覚心理実験に幅広く応用可能な知覚的音響処理となっている[3]。しかし、しかし音声変換際に生じる品質劣化はまだ避けられていないため、さらなる改良を行った。特に、基本となるスペクトル形状や基本周波数の推定精度向上[2, 9]や、モーフィング特性の改善[15]等に関して、良い成果を挙げることができた。

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4件)

- [1] Toshio Irino, Yoshie Aoki, Hideki Kawahara, and Roy D. Patterson, "Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination," *Speech Commun.*, 54 (9), pp.998-1013, 2012. [doi:10.1016/j.specom.2012.04.002] (査読有)
- [2] 赤桐隼人, 森勢将雅, 入野俊夫, 河原英紀, "スペクトルピークを強調したF0 適応型スペクトル包絡抽出法の最適化と評価," *電子情報通信学会論文誌 A*, Vol.94-A, No. 8, pp.557-567, 2011. (査読有)
- [3] Erika Okamoto, Toshio Irino, Ryuichi Nishimura and Hideki Kawahara, "Evaluation of voice morphing using vocal tract length normalization based on auditory filterbank," *J. Signal Processing (信号処理)*, Vol.15, No. 4, pp.283-286, July, 2011. (Letter). (査読有)
- [4] Chihiro Takeshima, Minoru Tsuzaki, and Toshio Irino, "Perception of vowel sequence with varying speaker size," *Acoust. Sci. & Tech.* (ed. by the Acoustical Society of Japan), Vol. 31, No. 2, pp.156-164, Mar, 2010. [doi:10.1250/ast.31.156] (査読有)

[学会発表] (計118件)

(招待講演6件, 国際会議(査読有)22件, 国際会議(査読無)9件, 国内会議2009年度26件, 2010年度22件, 2011年度23件, 2012年度10件)

- [1] 深渡瀬 智史, 入野俊夫, 西村 竜一, 河原英紀, Roy D. Patterson, "ノッチ雑音マスキング法の測定点削減のための感度解析の改良," *日本音響学会：春季研究発表会講演論文集*, 2-Q-4, pp.609-610, 東京工科大, 八王子, 2013年3月13日~15日, (発表日3月14日).
- [2] Ryuichi Nisimura, Shoko Miyamori, Erika Okamoto, Hideki Kawahara, and Toshio Irino, "Detecting child speaker based on auditory feature vectors for VTL estimation," in *Proc. APSIPA 2012 #117*, Hollywood, California, 3-6 Dec., 2012. (発表日6 Dec.)
- [3] Toshio Irino, Tomofumi Fukawatase, Makoto Sakaguchi, Ryuichi Nisimura, Hideki Kawahara, and Roy D. Patterson, "Accurate estimation of compression in simultaneous masking enables the simulation of hearing

impairment for normal hearing listeners," 16th International Symposium on Hearing (ISH2012), St John's College, Cambridge UK, 23-27 July, 2012 (発表日23 July).

- [4] 入野俊夫 "内耳における圧縮特性とフィルタ特性の同時推定手法とその応用," *日本聴覚医学会, 第9回 内耳ひずみ研究会, 慶應大学病院, 東京, 2012年7月6日(招待講演)*
- [5] 入野俊夫, 河原英紀, "安定な声道長推定のための聴覚フィルタバンクとその理論," *日本音響学会：秋季研究発表会講演論文集*, pp.505-508, 島根大, 島根, 2011年9月20日~22日, (発表日9月22日). (招待講演)
- [6] 津崎 実, 入野俊夫, 竹島 千尋, 松井 淑恵, "寸法知覚を中心とした聴覚情景分析—物理世界と心理世界をつなぐ聴覚—," *日本音響学会：秋季研究発表会講演論文集*, pp.1437-1440, 島根大, 島根, 2011年9月20日~22日, (発表日9月21日). (招待講演)
- [7] 入野俊夫, "音声からの声道長推定における聴覚的ウェーブレット変換について," *平成23年度 数学・数理科学と諸科学・産業との連携研究ワークショップ「ウェーブレット理論と工学への応用」* (主催：文部科学省・大阪教育大), 大阪教育大, 大阪, 2011年9月12~13日(発表日9月12日)(招待講演)
- [8] Erika Okamoto, Toshio Irino, Ryuichi Nisimura, Hideki Kawahara, "Auditory filterbank improves voice morphing," in *Proc. Interspeech 2011*, pp.2517 - 2520, Florence, Italy, 27-31 Aug., 2011. (発表日30 Aug.)
- [9] Hideki Kawahara, Toshio Irino, and Masanori Morise, "An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction," *Proc. IEEE ICASSP 2011*, pp.5420 - 5423, Prague, Czech Republic, 22 - 27, May, 2011. (発表27 May 2011).
- [10] Erika Okamoto, Toshio Irino, Ryuichi Nisimura, Hideki Kawahara, "Evaluation of Voice Morphing Using Vocal Tract Length Normalization Based on Auditory Filterbank," in *Proc. 2011 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP'11)*, pp.187-190, Tianjin SaiXiang Hotel, Tianjin, China, 1-3 March, 2011. (発表日2 Mar.)
- [11] 塚田裕樹, 能田由紀子, 河原英紀, 入野俊夫, "単語の音節遷移情報の処理を担う脳領域のfMRIによる検討," *日本音響学会：聴覚研究会資料*, H-2010-154, Vol. 40, No.10, pp.851-856, かんぽの宿柳川, 福岡, 2010年

12月10日～11日(発表日12月11日)。(査読無)

[12] 入野俊夫, “はじめての聴覚フィルタ－心理物理実験デモで学ぶ聴覚フィルタ特性一,” 日本音響学会: 秋季研究発表会講演論文集, pp.1347 - 1348, 関西大学, 大阪, 2010年9月14日～16日(発表日9月16日)。(招待講演)

[13] Toshio Irino, Toru Takahashi, and Hideki Kawahara, “Demonstration of a C-implementation of the dynamic compressive gammachirp for machine hearing,” Auditory Features Workshop, Equipe Audition, DEC, Ecole normale supérieure, France, 1 & 3 Jun., 2010 (発表日1 Jun) (査読無)

[14] Roy D. Patterson, Tom C. Walter, Jessica Monaghan, Christian Feldbauer, and Toshio Irino, “Auditory speech processing for scale-shift covariance and its evaluation in automatic speech recognition,” Proc. IEEE International Symposium on Circuits and Systems (ISCAS) 2010, Paris, France, 30 May - 2 June, 2010. (発表日2 Jun 2010)

[15] Hideki Kawahara, Ryuichi Nishimura, Toshio Irino, Masanori Morise, Toru Takahashi, Hideki Banno, “High-quality and light-weight voice transformation enabling extrapolation without perceptual and objective breakdown,” Proc. IEEE ICASSP 2010, pp. 4818-4821, Dallas, Texas, USA, 14 - 19, Mar. 2010. (発表日19 Mar.)

[16] 入野俊夫, “聴覚フィルタの測定と定式化について,” 日本音響学会聴覚研究会資料, H-2009-73, Vol. 39, No. 6, pp.413 - 418, [聴覚研究会、レクチャー招待講演], 加太国民休暇村, 和歌山, 2009年10月9～10日, (発表日10月9日) (招待講演)

[17] Chihiro Takeshima, Minoru Tsuzaki, and Toshio Irino, “Influences of vowel duration on speaker-size estimation and discrimination,” Proc. Interspeech2009, pp.128 - 131, Brighton, UK., 6 - 10, Sept. 2009 (発表日7 Sept.)

[18] Toshio Irino, Yuki Tsukada, Yoshikazu Oya, Hideki Kawahara, and Roy D. Patterson, “Brain regions for auditory size processing of speech sounds,” Proc. Auditory Cortex 2009, Magdeburg, Germany, 29 Aug. - 2 Sept. 2009 (発表日30-31 Aug).

[19] Toshio Irino, Yoshie Aoki, Hideki Kawahara, and Roy D. Patterson, “Size Perception for acoustically scaled sounds of naturally pronounced and whispered words,” Proc. 15th International Symposium on Hearing (ISH2009), Salamanca,

Spain, 1 - 5 Jun. 2009 (発表日2 Jun).

[図書] (計4件)

[1] 森周司, 香田徹 (編), 香田徹, 日比野浩, 任書晃, 倉智嘉久, 入野俊夫, 鶴木祐史, 鈴木陽一, 牧勝弘, 津崎実, (共著), “聴覚モデル,” コロナ社, 東京, 233p., ISBN 978-4-339-01323-8, 2011年8月30日.

[2] Toshio Irino, Yoshie Aoki, Hideki Kawahara, and Roy D. Patterson, “Size Perception for acoustically scaled sounds of naturally pronounced and whispered words,” in “Neurophysiological Bases of Auditory Perception,” Enrique A. Lopez-Poveda, Alan R. Palmer, and Ray Meddis (Eds.), pp.235-243, Springer, LaVergne, TN USA, 644p., Apr., 2010. [ISBN 978-1-4419-5685-9, doi:10.1007/978-1-4419-5686-6]

[3] Hideki Kawahara, Masanori Morise, Toru Takahashi, Ryuichi Nishimura, Hideki Banno, Toshio Irino, “STRAIGHT, a framework for speech analysis, modification and synthesis,” in Computer Processing of Asian Spoken Languages,” Shuichi Itahashi and Chiu-yu Tseng (Eds.), pp.297-300, Consideration Books, Los Angeles, USA, ISBN 978-0-935047-72-1, 372p., Mar 2010.

[4] 入野俊夫, 河原英紀, “聴覚認知過程の数理,” in 数理科学辞典 (第2版) (広中平祐 他 編), pp. 1058-1061, 丸善, 東京, ISBN 978-4-621-08125-9, 1450p., 2009年12月26日.

[その他]

<http://www.wakayama-u.ac.jp/~irino/>

6. 研究組織

(1) 研究代表者

入野 俊夫 (IRINO TOSHIO)
和歌山大学・システム工学部・教授
研究者番号:20346331

(2) 研究分担者

河原 英紀 (KAWAHARA HIDEKI)
和歌山大学・システム工学部・教授
研究者番号:40294300
津崎 実 (TSUZAKI MINORU)
京都市立芸術大学・音楽学部・准教授
研究者番号:60155356
西村 竜一 (NISHIMURA RYUICHI)
和歌山大学・システム工学部・助教
研究者番号:00379611