

研究種目：基盤研究（C）  
 研究期間：2007～2008  
 課題番号：19509003  
 研究課題名（和文）音声映像合成モデルによる英語情動学習支援システム開発のための研究  
 研究課題名（英文）  
 A study concerning the development of a support system for studying English emotions using synthesized audio and audiovisual models

研究代表者  
 ヤーッコラ伊勢井 敏子 (Isei-Jaakkola TOSHIKO)  
 中部大学・人文学部・教授  
 研究者番号：00454364

研究成果の概要：

情動認知に関しては、文化差（同時に母語の言語系統の差）が関係し、また、情動・感情ごとに差異があるということも判明した。情動の認知結果と物理音響特性（距離）の相関に関しては、基本周波数よりも音圧の方がより相関が高いことも分かった。顔合成と音声との同期は極めて難しい課題であることがわかった。時間変化と音声の変化および顔の変化をすべて統合して合成することは現在の技術では不可能に近いことが判明した。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	2,200,000	0	2,200,000
2008年度	1,100,000	330,000	1,430,000
年度			
年度			
年度			
総計	3,300,000	330,000	3,630,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：英語、情動・感情、学習支援システム、音声映像、音声合成・顔合成、韻律、学習者

#### 1. 研究開始当初の背景

人間のコミュニケーションにおいて情動や感情は不可欠の要素である。情動や感情に関する研究は主に心理学分野を中心としていて、また母語話者が主たる対象となっている。音声工学の分野では感情音声の音響特徴

量等を統計処理することが多く、それも母語話者が対象になることが多い。加えて生理音声学では情動や感情の喉頭制御を観察する傾向がある。教育分野では様々な教材が使用されるが、ほとんど例外なく、情動や感情を伴った対話などを聴きまた発話の練習等を

する。しかし、世界中に英語学習者が大量に存在する一方、英語学習者が英語母語話者の発話だけでなく教材を見て聴いて、英語情動をどのように認知し学習しているかに焦点を置いた研究がなかった。音声学の世界においても、言語や意味的機能の観点からプロソディーやイントネーションのモデルが扱われてきたが、情動や感情に的を絞ってイントネーションのモデルが十分に研究されてきたとは言い難い。

そこで、本研究では、英語学習者が情動をどのように認知しているかを探ることにした。日本語母語話者の英語学習者（以下 JL2 と称する）の英語情動認知だけでは、それが普遍的な英語学習者の認知なのか、日本文化に基づいた言語固有の認知なのか検証ができない。そのため、JL2 以外の英語学習者も認知実験の対象とした。また、英語学習者による英語情動の認知と音響特微量（基本周波数および音圧）の関係も明らかではなかった。一つの理由は、音響特微量の計測が過去 30 年近く基本周波数のみを扱う傾向があったことも否めない。したがって、英語情動の認知と音響特微量の相関を見るために、基本周波数に加え、音圧も同様に取り扱う必要があった。情動認知は文化差に優位がないという先行研究がまだ主流である中で、その真偽を確認するため、異なった文化圏・異なった言語系統に属する英語学習者にとっての英語情動認知についても比較検証をする必要があった。さらに、声のみ聞いた場合と、顔の表情を伴った声、というふうに実験試料を分離して、表情の役割を明らかにする必要もあった。

## 2. 研究の目的

まず、英語情動・感情を日本人英語学習者がどのように認知するかを体系的に明らか

にする必要があった。そして、英語情動に関する英語学習者による認知実験と音響特微量の良好な相関を元に、英語情動の声と声付き表情の原型モデルを作ることと、さらに進んで音声合成・顔合成モデルを作ることによって、最終的に英語情動学習支援システムを推進するための基礎研究とすることが本研究の目的であった。

## 3. 研究の方法

英語学習者の英語情動認知の体系化のため、英語情動認知実験において、聴覚（音声）のみと視覚・聴覚（音声付表情映像）を提示し、比較検証をした。演劇専攻の学生をインフォーマントとして録音録画を行い、英語情動 18 種のデータベースを作った。音声および顔のすべての録音録画を分割し、非母語話者を対象とした認知実験の前にインフォーマントである英語母語話者の認知実験をおこなった。また、非母語話者に中国語話者（大学生）を選定し、認知実験を行った。結果を統計処理し、各情動の認知結果と物理音響特性の比較検証を行った。基本周波数および音圧の情動間の距離比較のため、実数値による面積および傾き（曲線パターン）を計算した。さらに、情動ごとの認知テスト結果と音響特性の相関を統計的に処理した。

## 4. 研究成果

本研究では、背景として、英語学習者が英語情動をどのように認知し学習しているかに焦点を置いた研究がなかったことから、(1) 英語学習者による英語情動の認知を体系化すること、情動は文化差よりも普遍的であるという研究に対して、(2) 英語情動認知における英語学習者の視点から文化差の有無を確認すること、情動認知と音響特微量との相関において基本周波数にのみ焦点が置か

れる傾向があったので、(3) 英語学習者の英語情動認知と音響特微量との相関において音圧との相関も見ることにした。最終的に、英語情動に関する英語学習者による認知実験と音響特微量の良好な相関を元に、英語情動の音声と表情を伴う音声（以下、表情）の原型モデルを作ることと、さらに進んで音声合成・顔合成モデルを作ることによって、最終的に英語情動学習支援システムを推進するための基礎研究とすることが本研究の目的であった。

方法として、複数の演劇専攻の英語母語話者（インフォーマント）に様々な情動をキーワードとして発話してもらい、音声と表情を同時録画し、音声と表情の2種のデータベースを作った。それを、インフォーマント及び英語学習者（2グループ）に提示し、認知実験を行った。認知結果を情動間で比較し、また、音響特微量（基本周波数と音圧）に基づき物理的距離を計算した。さらに、基本周波数と音圧それぞれに情動の認知との相関を見た。

結果、(1) 情動認知に関しては、文化差（同時に母語の言語系統の差）がある程度関係していることが明らかになった。その差は情動・感情ごとに差異があることも分かった。このことは研究代表者等の先行研究に照らしても、矛盾がないことが分かった。たとえば、英語話者やフィンランド語話者は皮肉等の否定的感情や情動に対して正解率が高いが、日本語話者はそれらより正解率が低い傾向がある。中国語話者の場合には英語話者やフィンランド語話者により近いことが分かった。また、声だけを聞いた認知結果と表情の認知結果を比較すると、英語話者、フィンランド語話者、中国語話者の場合には、表情の方が相当正解率が高いものの、日本語話者はむしろ音声だけと表情の間には大きな差

が出ない。さらに、情動ごとの回答のばらつきを見てみると、言語話者ごとに異なった反応が見られた。このばらつきには一定のパターンがあることも判明した。このような結果とこれまでの情動研究を比較すると、情動認知はどの言語話者にも普遍的とは言えず、文化差（母語話者の言語系統の差）が存在することが実証された。その一方、主張をする文化とそうでない文化によっても情動認知が影響されるらしいことが予測された。このことから、同じ英語という言語学習において、情動・感情学習に関しては、学習者の母語や属する文化背景を考慮する必要があることが示唆できる。これらをさらに実証するために、世界の他の地域の英語学習者に対しても同様の実験を行う必要があると思われる。本研究が目的とした英語情動認知体系化に関しても、各種認知実験を大量に必要がありそうである。

(2) 認知結果と物理音響特性の相関に関しては、基本周波数よりも音圧の方がより相関が高いことが明らかになった。中国語話者の認知結果と物理音響特性の相関に関しては、基本周波数よりも音圧の方がより相関が高いことが明らかになった。先行研究（英語話者・日本語話者・フィンランド語話者）と類似した結果がでた。このことは、情動や感情のイントネーションパターンやイントネーションモデルにおいても、過去あまり注視されなかった音圧を研究対象とするべきことがさらに強く示唆できる。

(3) 本研究の中で、フォーカスを持った情動とイントネーションの基本周波数および音圧の平均パターンモデルを表示することができた。ただし、情動の入った文は感情によって基本周波数および音圧のパターンについてインフォーマント間の差が出る場合があることが分かった。平均パターンモデ

ルを表示する場合にはこの点に関しての注意が必要であることが判明した。

(4) 顔合成と音声との同期は極めて難しい課題であることがわかった。時間変化と音声の変化および顔の変化をすべて統合して合成することは現在の技術では不可能に近いことが判明した。このことから、英語学習支援の方法として、画像処理や音声認識を取り込んだ方法を試した。一つは、3次元コンピュータグラフィックス技術を用いて顔のモデルを作成し、音素と同期させ構音(調音)運動を学習するシステム研究を開始した。これは子音学習を目的とする。もう一つの学習支援の方法として3次元可視化システムを導入し母音音素を学習するシステムを考案した。3次元母音図を表示し、これに音声認識を統合し同時音声同期ができるシステム設計をすることができた。

#### 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 16 件)

- ① Toshiko Isei-Jaakkola, Martti Vainio, and Keikichi Hirose, "Japanese prominence with different focus in affirmative sentences realised by Finnish L2," *Proceedings of International Conference on Speech and Computer*, Moscow, pp. 787-792, 10.2007. 査読有
- ② ヤーッコラ伊勢井敏子, 鈴木茂樹, 森島繁生, 広瀬啓吉, "英語音声教育のための3DCGによる舌の動きと音声のリンク開発の試み—語彙との同期—", 映像メディア学会技術報告, ITE Technical Report Vol.32, No.15, pp. 29-32, 2008.3. 査読有
- ③ ヤーッコラ伊勢井敏子, 中貴俊, "フォルマント母音図の3D可視化の試み", 日本音響学会春季大会講演論文集, pp. 435-536, 2008.3. 査読無
- ④ 鈴木茂樹, ヤーッコラ伊勢井敏子, 森島繁生, 広瀬啓吉, "英語流音学習のための音と同期した3DCGによる構音運動", 日本音響学会秋季大会講演論文集, pp. 339-340, 2008.9. 査読無
- ⑤ ヤーッコラ伊勢井敏子, 広瀬啓吉, 森島繁生, "英語情動文 "I love you" の中国語話

者による認知と音響特性相関 (1)", 日本音響学会秋季大会講演論文, pp. 331-332, 2008.9. 査読無

- ⑥ ヤーッコラ伊勢井敏子, 中貴俊, 広瀬啓吉, "目標言語母音達成のための3Dフォルマント母音図における母音二重表示", 日本音響学会秋季大会講演論文集, pp. 317-318, 2008.9. 査読無
- ⑦ ヤーッコラ伊勢井敏子, 鈴木茂樹, 広瀬啓吉, 森島繁生, "音と同期した3DCGを用いた日本人英語学習者に苦手な構音運動のためのトレーニングシステム—唇の突き出し", 日本音響学会秋季大会講演論文集, pp.369-370, 2009.03. 査読無
- ⑧ ヤーッコラ伊勢井敏子, 広瀬啓吉, 森島繁生, "英語情動文 "I love you" の中国語話者による認知と音響特性相関(2)", 日本音響学会秋季大会講演論文集, pp. 483-484, 2009.03. 査読無
- ⑨ 佐伯拓郎, 中貴俊, ヤーッコラ伊勢井敏子, 遠藤守, 山田雅之, 宮崎慎也, "3D フォルマント母音図における発声母音のリアルタイム可視化", 情報・システム講演論文集 1, p. 169, 2009.03. 査読無
- ⑩ Keikichi Hirose, Keiko Ochi, and Nobuaki Minematsu, "Corpus-based generation of prosodic features from text based on generation process model", *Proceedings Interspeech 2007*, Antwerp, August 27-30, WeD.O3-1, pp.1274-1277, 8.2007. 査読有
- ⑪ Keikichi Hirose, Qinghua Sun, and Nobuaki Minematsu, "Corpus-based synthesis of mandarin speech with  $F_0$  contours generated by superposing tone components on rule-generated phrase components", *Proceedings IEEE Workshop on Spoken Language Technology*, Goa, December 15-18, pp.33-36, 12.2008. 査読有
- ⑫ Keikichi Hirose, Keiko Ochi, and Nobuaki Minematsu, "Corpus-based generation of  $F_0$  contours of Japanese based on the generation process model and its control for prosodic focus," *Proceedings of 9<sup>th</sup> International Conference on Signal Processing (ICSP2008)*, Beijing, October 26-29, pp. 647-650, 10.2008. (Invited, Special Session)
- ⑬ Keikichi Hirose, "Speech prosody in spoken language technologies", *Journal of Signal Processing*, Vol.12, No.1, pp.7-16, 1.2008. (Invited Paper)
- ⑭ 森島繁生, "CGキャラクタの存在感", 日本バーチャルリアリティ学会誌, 特集 アンドロイドやエージェントに感じる人の存在感, vol.14 No.1 pp.23-28. 2009.3. (招待論文)
- ⑮ 森島繁生, "顔表情のCG合成と感動評価", 映像情報メディア学会誌, 小特集 今

"顔" が面白い ～顔の画像処理とその応用～, ISSN 1342-9607, Vol.62, No.12, pp.1923(20)-1927(23), 2008.12. (招待論文)

- ⑩ Shigeo Morishima, "Dive into the Movie-Audience-driven Immersive Experience in the Story-", IEICE TRANSACTIONS on Information and systems, Special Section on Human Communication III, Vol.E91-D, No.6, pp.1594-1603, 2008.6. (招待論文)

[学会発表] (計 8 件)

- ① Toshiko Isei-Jaakkola, "Long vowel quantity in perception in Japanese", Finnish and Czech, *5th Asia Pacific Conference on Speech, Language and Hearing, Program & Abstracts*, p.106, July 9-13, 2008. (The Univ. of Queensland, AUSTRALIA)
- ② Toshiko Isei-Jaakkola, "3D Representations of English Monophthongs, the 9th Joint Seminar on English Phonetics", March 25, 2008. (Seoul National Univ., SOUTH KOREA)
- ③ Toshiko Isei-Jaakkola and Takatoshi Naka, "A three-dimensional visualisation of monophthongs for L2", 1st International Conference on Literature, Languages and Linguistics, July 28-31, 2008. (Athens, GREECE)
- ④ Toshiko Isei-Jaakkola, Takatoshi Naka and Keikichi Hirose, "A multilingual 3D vowel chart for monophthongs based on formants", *Nordic Prosody X Abstracts*, p. 42, August 4-6, 2008. (Univ. of Helsinki, FINLAND)
- ⑤ Toshiko Isei-Jaakkola, Shigeki Suzuki, Shigeo Morishima and Keikichi Hirose, "A computer graphic three-dimensional tongue and lip movement synchronized with English fricatives for Japanese learners", 156th Meeting of Acoustical Society of America, *The Journal of Acoustical Society of America*, Vol. 124, No.4, Pt. 2 of 2, Oct, 2008, p.2594, November 10-14, 2008. (Florida, U.S.A.)
- ⑥ Keikichi Hirose, "Researches on speech prosody for advanced spoken language technologies", *Proceedings of International Symposium: Frontiers of Research on Speech and Music (FRSM-2008)*, Kolkata, pp. 40-45, February 20-21, 2008. (Invited Talk) (Kolkata, INDIA)
- ⑦ Keikichi Hirose, Qinghua Sun, and Nobuaki Minematsu "Generation of  $F_0$  contours for Mandarin speech in combination with rule-based and corpus-based methods", *Proceedings of 8<sup>th</sup> Phonetics Conference of China (PCC2008)/International Symposium on Phonetic Frontiers*, Beijing, 8 pages in

*CD-ROM Proceedings*, April 18-20, 2008. (Invited) (Beijing, CHINA)

- ⑧ Keikichi Hirose, Qinghua Sun, and Nobuaki Minematsu "Synthesis of  $F_0$  contours for Mandarin speech by superposing corpus-generated tone contours on rule-generated phrase components," *Proceedings of Speech Analysis, Synthesis and Recognition -Application in Systems for Homeland Security-*, Piechowice, Poland, September 8-12, 2008. (Plenary talk) (Piechowice, POLAND)

## 6. 研究組織

### (1) 研究代表者

ヤーッコラ伊勢井 敏子  
(JAAKKOLA-ISEI TOSHIKO)  
中部大学・人文学部・准教授  
研究者番号：00454364

### (2) 研究分担者

広瀬 啓吉 (HIROSE KEIKICHI)  
東京大学・情報理工学系研究科・教授  
研究者番号：50111472  
森島 繁生 (MORISHIMA SHIGEO)  
早稲田大学・理工学術院・教授  
研究者番号：10200411