様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月 ２６ 日現在

機関番号：２３９０１
研究種目：基盤研究(C)（一般）
研究期間：2015～2018
課題番号：１５Ｋ００４８７

研究課題名（和文）音声からの発話動作可視化技術に基づく発話訓練支援の研究

研究課題名（英文）Real-time Visualization of Pronunciation on an IPA Chart based on Articulatory Feature Extraction

研究代表者

入部　百合絵（Iribe, Yurie）

愛知県立大学・情報科学部・准教授

研究者番号：４０３９７５００

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：本研究では調音の正しい位置と学習者の位置を可視化する発音学習システムを開発した．調音様式（破裂音，摩擦音など）と調音部位（口唇，舌の位置など）を示す調音特徴を，音声から抽出することで発音動作を可視化する．母音音声の可視化は，舌の盛り上がる位置・場所や口の開け方を軸としたチャート上（矩形図）に学習者の音声をリアルタイムにプロットすることで実現した．子音に関しては調音様式や調音部位が簡単に確認できるチャート図を開発した．その結果，音声から英語音素の調音特徴を93%，日本語音素の調音特徴を96%の精度で抽出することを可能にした．また，音声をチャート上にプロットする精度は70%であることを確認した．

研究成果の学術的意義や社会的意義
グローバル化により語学に秀でた人材の育成が必要不可欠な中，本システムは発話訓練のための自主学習支援ツールとしての利用とその教育効果が期待できる．現在の語学教育は読むことと聴くことに偏重しており，近年は話すことの重要性が指摘されているが，話すことについてはマンツーマン教育以外によい方法がなかった．本システムはこうした状況を打破するに，大きな助けとなると期待できる．加えて，今回開発した発音学習システムは，多言語音声言語コーパスを利用することにより，容易に他の言語へ応用することが可能である．そのため，英語だけではなく留学生の日本語教育や構音障碍者の発話訓練にも応用できると考えられる．

研究成果の概要（英文）：We have been developing a pronunciation training system to evaluate and correct learner's pronunciation by extracting articulatory-features (AFs). In this paper, we propose a novel pronunciation training system that can plot the place and manner of articulation of learner's pronunciation on an International Phonetic Alphabet (IPA) chart in real time. First, the proposed system converts input speech into AF-sequences by using multi-layer neural networks. Then, the AF-sequences are converted into x-y coordinates and plotted on an IPA chart to show his/her articulation in real time. Lastly, we investigate plotting accuracy on the IPA chart through experimental evaluation. The correct rates of AFs to every English phoneme averages 93%, and that of unique Japanese phonemes averages about 96%. The results clearly show that the proposed system can extract AFs accurately. The accuracy to plot speech on the IPA chart confirmed that the average accuracy is over 70%.

研究分野：音声情報処理

キーワード：発音訓練　調音運動　IPAチャート

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

グローバル化の波により，日本社会も人的資源の流動性を高める必要に迫られている．課題の一つに，国際標準言語としての英語に対する教育改善があるが，発音を指導できる教師の数が絶対的に不足している．そこで，近年，音声認識技術をもとに学習者の発音誤りを指摘する自主学習用発音ソフトが開発されている．現在の発音学習ソフトは，誤った音素を指摘する機能や，発音の正しさをスコアリングする機能を有している．また，正しい音声と学習者の音声の波形やフォルマント情報を表示することでその違いを示すソフトも開発されている．

しかしながら，学習者はそれらを通して自身の発音が教師と異なっていることは認識できても，誤った調音動作を具体的にどのように修正すればよいのか分かりづらい．特に，波形パターンやフォルマントに関しては，音声の専門家でない限りどこをどのように矯正すればよいのか理解することは難しい．一方，動画ビデオを用いる方法もあるが，それらは予め用意した正しい調音動作に関するものであり，学習者の調音動作を含めて比較提示する研究は存在しない．一般的に，語学学習は face-to-face のように教師が学習者の調音動作を見ながら，調音の仕方（舌，口唇，顎，口蓋などの動き）を的確に指摘することで，正しい発音へと導く手法が採られる．このため，教師がいない環境では学習者の調音動作の誤りを的確に指摘し教示することが難しい．

２．研究の目的

本研究では，これまで開発した音声-調音特徴変換技術をベースに，この精緻化と，調音の正しい位置と学習者の位置を可視化する発音学習システムを開発する．調音特徴は，単音(phone)分類に用いられ，調音様式（破裂音，摩擦音，破擦音，鼻音，半母音，…）と調音部位（口唇，歯茎，口蓋，咽喉の位置(子音の場合)や，舌の最も盛り上る位置と口の開閉度(母音)）の諸属性から構成される特徴量で，国際音声記号(IPA)として世界共通に利用されているものである．

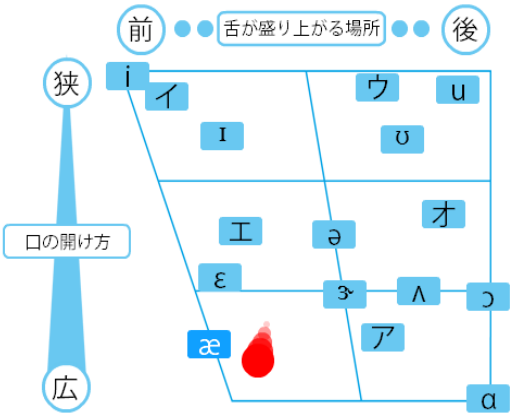

図１　母音のチャート図

本研究では，比較的単純な調音動作で発音する母音を図1のように舌の盛り上がる位置・場所や口の開け方を軸としたチャート上（矩形図）に学習者の音声をリアルタイムにプロットする．また，子音に関しても調音様式や調音位置が簡単に確認できるチャート図を開発する，図上の音素記号は正しい調音位置を，赤円は学習者の発音時の調音位置を示す．これにより，正しい調音と学習者の誤った調音の位置が容易に確認できる．このように調音動作を視覚的に直接観察することは教育効果が高いと言われている．

これらを実現するための具体的な研究目標は以下の 2 点である．

(1)　教師と学習者の音声から調音動作を数値化した調音特徴を直接抽出する（目標：日本語および英語音声に対する調音特徴の抽出精度 95%以上）

(2)　抽出した調音特徴をもとに，調音位置・様式を軸とした矩形図上に学習者の調音を的確にプロットする．

３．研究の方法

発話学習システムの全体構成を図2に示す．

（1）　**音声-調音特徴変換エンジン**

調音特徴（Articulatory Feature; AF）は，単音(phone)分類に用いられる調音様式（破裂音，摩擦音，破擦音，鼻音，半母音など）と調音部位（口唇，歯茎，口蓋，咽喉などの位置(子音の場合)や，舌の最も盛り上る位置と口の開閉度(母音)）の諸属性から構成される．音声から調音特徴を抽出できると，学習者の調音動作を推定でき，国際音声記号(IPA)の母音チャートや子音チャートへ直接プロットすることが可能になる．我々は，IPA から英語と日本語に関する部分(次元数: 28)を取り出した調音特徴セットを採用しており，46 英語音素と 13 日本語音素（英語と重複分は除く）を対象にしている．本研究は，日本人のための英語発話学習を想定しているが，日本人が英語を発話す

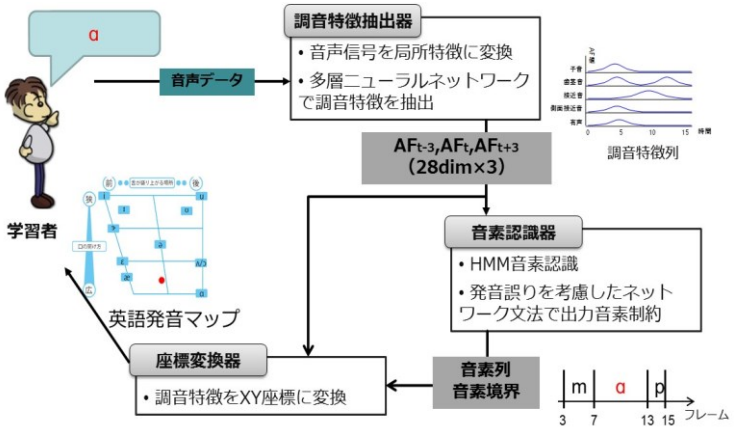

図２　発音学習システムの構成図

る際, 母国語の影響により, 英語音素にはない日本語音素混じりの英語発話を行う可能性が高い. そのため，英語音素と日本語音素を対象とする調音特徴セットを用意した．調音特徴抽出は，2段の多層ニューラルネットワーク(Multi-Layer Neural network)と直交化処理により行った．調音特徴系列は，離散的で音素列と一対一に対応することから，原初的言語情報と見做すことができる．MLNを通して音声に対応するAF 系列を抽出し，この系列データをx-y座標に変換することでチャート上に表示する.

(2) 発音マップシステム

単音および単語中の音素を母音・子音チャート図上にプロット表示する機能を開発した．具体的には，音声から抽出した調音特徴系列から母音および子音に各々関連する特徴量（母音であれば，舌の盛り上がる位置（前－中－後），口唇の高さ（広－半広－半狭－狭））を取り出し，チャート上の横軸（母音であれば舌の盛り上り位置），縦軸（母音であれは唇の高さ）に適した座標値に変換するための計算アルゴリズムを提案する．また，英語に関して言えば，日本語にはない英語独特の母音を習得するために tense(強母音の[i]など), round([u]), 及び rhoticity(car[kɑɚ]などの r 音)の 3 種の調音特徴についても，学習者に効果的に表示を返す.

図 2 に発音マップのシステム全体図を示す．システムは学習者の発声を検知すると調音特徴抽出部により 10ms 毎に 48 次元の調音特徴を抽出する．抽出された調音特徴から，母音発音マップにおいては母音に関する 10 次元の特徴列が座標変換器により 2 次元平面上の X,Y 座標に変換される．子音発音マップも同様に，子音に関係する 14 次元の特徴列が座標に変換される．このとき，HMM(Hidden Markov Model)から得られる音素継続長を用いて，調音特徴系列から次元毎に平均値を算出し，プロットに用いる．次に図 1 に示した母音発音マップについて詳述する．母音発音マップは IPA 母音チャートを模した梯形図に発音記号が配置され，口唇の開き具合を示すスケール（縦軸），舌の盛り上がる位置を示すスケール（横軸）をもとに，ユーザの調音位置を示す赤い光点が示されている．また，調音の軌跡が薄い赤円で表現されている.

以降に，音声から抽出された調音特徴を母音発音マップへ変換する手法について説明する．特徴列から X 座標への変換は式(1-1)，Y 座標への変換は式(1-2)によって行われる.

$$X = D_{width} \times (\alpha + \Delta_x) / 4 \qquad (1\text{-}1)$$

AF_{front}, $AF_{central}$, AF_{back} はそれぞれ調音特徴列の口唇の開き具合を示す「前舌」「中舌」「後舌」に対する特徴量を指している．D_{width} は発音マップの水平方向の長さである．α はマップを X 方向に特徴数で分割した幅を，Δ_x は隣接する特徴の特徴量の差分を指しており，それぞれ特徴量の大小関係により表 1 のように定義される.

$$Y = D_{height} \times (\beta + \Delta_y) / 6 \qquad (1\text{-}2)$$

AF_{close}, AF_{close_mid} AF_{open_mid} AF_{open} はそれぞれ調音特徴列中の舌の盛り上がる位置を示す「狭」「半狭」「半広」「広」に対する特徴量を示している．D_{height} は発音マップの垂直方向の長さである．βはマップをY方向に特徴数で分割した幅を，Δ_yは隣接する特徴の特徴量の差分を指しており，それぞれ最大の特徴とそれに隣接する特徴の大小関係により定義される．例えば，調音特徴が「$AF_{front} = 1.0$ かつ $AF_{central} = 0.7$」の場合，最大の特徴 AF_{front} に対し，隣接する $AF_{central}$ の数値も考慮することで，「前舌と中舌の中間よりも若干前方」といった詳細なプロットが可能となる．母音発音マップは台形状であるため座標をプロットする際は，適宜台形のスケールに合わせて座標を変換し, マップ上へプロットする．次に，子音発音マップの例を図 3 に示す．子音発音マップは，等分割された矩形図に発音記号が配置され，調音位置を示すスケール（横軸）と調音様式を示すスケール（縦軸）を軸に，ユーザの発音位置を示す赤い光点がプロットされている．マップ上に示される音素は日本人の発音において置換誤りが多く見られる英語子音の組み合せとした．即ち，日本人が誤り易い似通った音素対に焦点を当て，発音マップ上でその調音動作の違いを明確にしつつ，正しく矯正することを目指す．横軸と縦軸に用いる特徴は，学習の難しい似通った音素と対として，その音素間で異なる調音特徴を示すこととした．表 1 に音素の組み合わせと，横軸と縦軸に用いる特徴を示す. ①の/b/と/v/の発音マップの例を図 3 左に示す．図 3 に示すように，音素間で調音位置と調音様式が異なる場合はそれぞれを縦軸と横軸にとる．②の/r//l//ɾ/の発音マップの

表 1 日本人が誤りやすい英語子音の組み合わせとマップ上の調音特徴

	音素	横軸の特徴	縦軸の特徴
①	b	両唇	破裂
	v	歯唇	摩擦
②	t	歯茎	破裂
	tʃ	歯茎＋歯茎	破裂＋摩擦
③	l	側面接近	－
	r	接近	
	ɾ	弾き	
④	θ	歯	－
	s	歯茎	
	ʃ	後部歯茎	
⑤	ð	歯	－
	z	歯茎	
	ʒ	後部歯茎	
⑥	m	両唇	－
	n	歯茎	
	ŋ	軟口蓋	

例を図 3 右に示す．音素間で異なる特徴が調音位置と調音様式のいずれかのみの場合は，縦軸を省略したマップとなる．調音特徴列から X 座標と Y 座標への変換は，それぞれマップ上の音素数が 2 の場合は式(2-1)および式(2-2)，3 の場合は(2-3)および式(2-4)によって行われる．

$$X = D_{width} \times (1 + AF_{x1} - AF_{x0}) / 2 \quad (2\text{-}1)$$
$$Y = D_{height} \times (1 + AF_{y1} - AF_{y0}) / 2 \quad (2\text{-}2)$$

$$Y = D_{width} \times (\alpha + \Delta_x) / 4 \quad (2\text{-}3)$$
$$Y = D_{height} \times (\beta + \Delta_y) / 4 \quad (2\text{-}4)$$

AF_{x0}, AF_{x1} はそれぞれ調音特徴列中の横軸に配置する特徴，AF_{y0}, AF_{y1} はそれぞれ調音特徴列中の縦軸に配置する特徴の特徴量を示している．D_{width} は発音マップの水平方向の長さ，D_{height} は発音マップの垂直方向の長さである．α および β はマップを X，Y 方向に特徴数で分割した幅を，Δ_x と Δ_y はそれぞれ隣接する特徴の特徴量の差分を指している．本手法は調音位置と調音様式を独立に評価することができるため，例えば「舌の位置は正しいが，摩擦音がうまく発音できていない」というように，似通った音素群の調音の違いをより具体的かつ詳細に示すことができる．

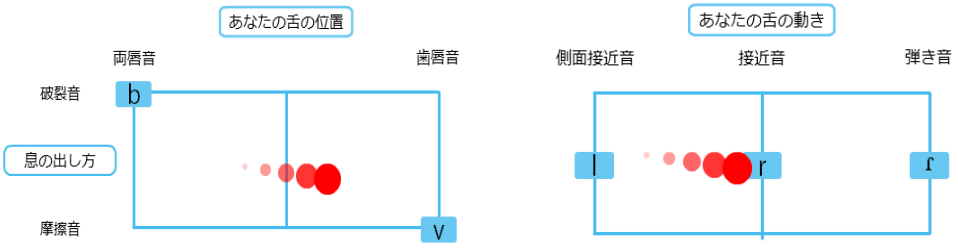

図 3 子音発音マップの例

4．研究成果

後述するプロット精度評価の前に，プロット精度に影響を及ぼす可能性の高い調音特徴の抽出精度を算出した．学習セットにより学習済みの MLN を用いて評価セットの音声データから抽出した調音特徴 28 次元に対して，正しく抽出できた特徴数をフレーム数×28 次元で除した抽出精度(AF-Correct Rate; AFCR)を計算した．

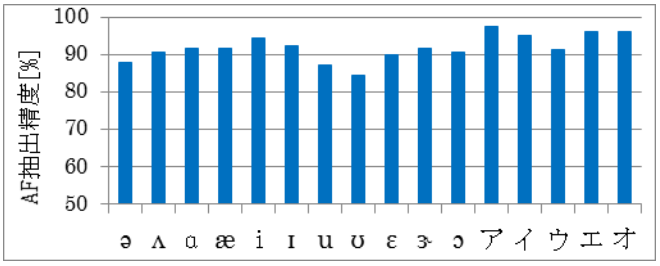

図 4 母音に対する調音特徴抽出精度

図 4 に母音毎の調音特徴の抽出精度を示す．英語全体では平均 93%，日本語全体では平均 96%の抽出精度が得られた．提案する英語母音発音マップは学習者の発音を 2 次元平面である IPA 母音チャート上にプロットし，マップ上の発音記号との相対的な位置から発音動作の違いを視覚的に教示するものである．従って，ネイティブ英語発音に近い発音がなされた場合は，その発音記号近傍にプロットされることが理想である．また，日本語と英語の母音の調音の違いを理解するために，誤って日本語母音を発音した場合は日本語の発音記号近傍にプロットされる必要がある．そこで，ネイティブ話者の母音音声からプロットされた座標と各母音の正解座標（IPA チャート上の座標）について次式に示す距離を算出し，プロット精度を評価した．

$$Dist_p = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i - x_{corr})^2 + (y_i - y_{corr})^2}$$

x_i, y_i は音素 p の音素継続長毎の X，Y 座標を指す．x_{corr}, y_{corr} は音素 p の正解座標の X，Y 座標を指し，MLN の教師信号を座標変換器に入力して得られる座標である．n は評価データ中に母音 p が出現する回数である．英語母音の調音特徴を抽出する MLN の学習に TIMIT の 2,600 文（男性話者 325 名）を使用し，日本語母音の調徴特徴を抽出する MLN の学習には CSJ コーパスの 22 時間分（男性話者 92 名）を利用した．プロット精度を評価するデータは TIMIT の 896 文（男性話者 112 名）と CSJ コーパスの 2.5 時間分（男性話者 10 名）を利用した．

図 4 に母音に対する調音特徴の抽出精度を示す．調音特徴の抽出精度の高かった/i/,/ɪ/,/イ/につ

いてはIPAチャートへのプロット精度も高く，調音特徴の抽出精度の低かった/u/,/ʊ/,/ウ/はプロット上でも正解座標からの距離が大きいことが分かった．このことから，調音特徴の抽出精度がプロット精度に大きく影響していることが確認できる．さらに本実験では話者によるプロットのばらつきを確認するために，TIMIT の各話者の出身地情報をもとに分けた 8 つのグループと CSJ の 10 話者に対してプロットを行った．話者グループ毎の平均座標を台形のマップに適応させる変換処理を行った後にプロットした結果を図 5 に示す．図中の破線で示す領域が IPA の母音図を模した領域である．幾つかの母音は正解座標から離れた位置にプロットされているが，どの音素も話者グループ間のばらつきが小さいことが分かる．さらに英語と日本語とで調音の異なる音素のプロットが正確に分離できていることが確認できる．

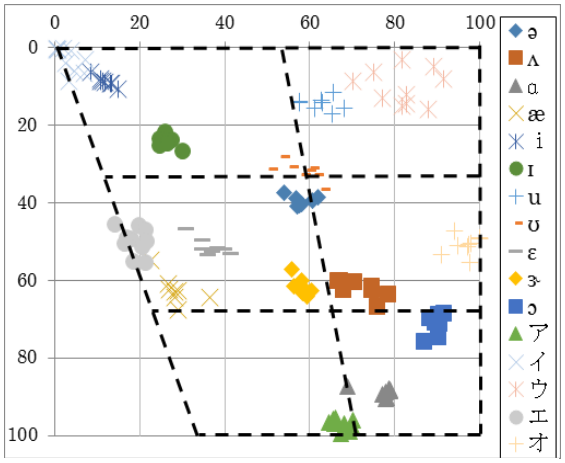

図 5 話者グループ毎のプロット例（母音）

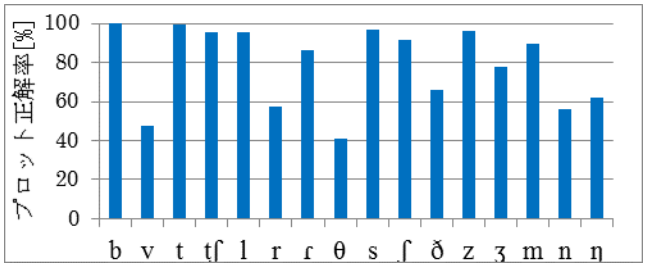

図 6 子音に対するプロット正解率

提案する子音発音マップは学習者の発音を，誤りやすい音素間の調音位置と調音様式の組み合わせからなる 2 次元平面上にプロットし，マップ上の発音記号との相対的な位置から発音動作の違いを視覚的に教示するものである．従って，ネイティブ英語発音に近い発音がなされた場合は，発音記号と同じ座標上にプロットされることが理想である．そこでネイティブ話者の子音発音が正しくプロットされた場合を正解として，プロット精度を評価した．ただし，母音発音マップと異なり，図 3 に示すように各音素の許容できる領域がマップ上の線分で明確に区切られているため，正解音素と同じ領域にプロットされた場合に正解とした．図 6 に子音発音マップの正解率を示す．/v/と/θ/を除いて，50%以上のプロット正解率が得られた．また，/θ/と/ð/はそれぞれ/s/と/ʃ/, /z/と/ʒ/に比べてプロット正解率が低い．歯音の調音特徴の抽出性能が低いことが分かっており，プロット正解率の低さの原因であると考えられる．有声・無声共に歯摩擦音は歯茎摩擦音に比べて弱いことが知られており，歯音の特徴を正確に抽出できる調音特徴抽出器の改良が必要である．さらに/v/のプロット精度が/b/に比べて低いため，摩擦音の調音位置の特徴抽出性能を改善することで，これらの音素のプロット正解率が向上すると考えられる．

５．主な発表論文等

〔雑誌論文〕（計 3 件）

(1) 入部 百合絵, 北岡 教英：音声認識にむけた超高齢者音声のコーパス構築, 日本音響学会誌, 査読無 vol. 73, pp. 303-310 (2017), https://doi.org/10.20697/jasj.73.5_303.
(2) Seng Kheang, Kouichi Katsurada, Yurie Iribe, Tsuneo Nitta, "Using Reversed Sequences and Grapheme Generation Rules to Extend the Feasibility of a Phoneme Transition Network-based Grapheme-to-Phoneme Conversion", IEICE Transactions, 査読有, Vol. E99-D No. 4 pp. 1182-1192 (2016), https://doi.org/10.1587/transinf.2015EDP7349.

〔学会発表〕（計 18 件）

(1) 梅村直人, 入部百合絵：調音可視化に向けた深層学習による音声からの口腔形状推定, 電子情報通信学会 東海支部 卒業研究発表会 (2019).
(2) 梅村直人, 入部百合絵：発音動作可視化を目的とした口腔形状の特徴量抽出, 第 81 回全国大会 6ZE-03 (2019).
(3) Norihide Kitaoka, Yurie Iribe, Hiromitsu Nishizaki, "Construction of a Corpus of Elderly Japanese Speech for Analysis and Recognition" Proc of. LREC2018 (2018).
(4) 平田 里佳, 入部 百合絵, 新田 恒雄：英語発話学習に向けた音素連結・脱落同化パターンの分析と検出, 日本音響学会 2018 春季研究発表会講演論文集, 1-Q-48 (2018).
(5) 川島 愛美, 入部 百合絵, 北岡 教英：高齢者の対話音声から抽出した言語的・音響的特徴に基づく認知症傾向の判別, 日本音響学会 2018 春季研究発表会講演論文集, 2-Q-36 (2018).
(6) 平田 里佳, 入部 百合絵, 新田 恒雄：英語発話学習に向けた音素連結・脱落同化パターンの分析と検出, 教育システム情報学会 東海支部学生研究発表会 (2018).
(7) Takuma Nakagawa, Ryota Nishimura, Yurie Iribe, Yoshio Ishiguro, Shin Ohsuga, Norihide Kitaoka, "A Human Machine Interface Framework for Autonomous Vehicle Control" Proc of. IEEE GCCE 2017 (2017).
(8) Yurie Iribe, Norihide Kitaoka, Shuhei Segawa, "Speech corpus spoken by young-old,

old-old and oldest-old Japanese″ Proc of. LREC2016 (2016).
(9) 入澤 浩太郎, 桂田 浩一, 新田 恒雄, 入部 百合絵：オートエンコーダと話者性変換ユニットを用いた声質変換法の提案 , 日本音響学会 2016 春季研究発表会講演論文集, 1-R-33 (2016).
(10) 石原 元気, 桂田 浩一, 新田 恒雄, 入部 百合絵：Suffix Array を用いた高速 STD システムにおけるリスコアリング法の検討, 日本音響学会 2016 春季研究発表会講演論文集, 2-R-12 (2016).
(11) Yurie Iribe, Norihide Kitaoka, Shuhei Segawa, ″DEVELOPMENT OF NEW SPEECH CORPUS FOR ELDERLY JAPANESE SPEECH RECOGNITION″ Proc of. O-COCOSDA 2015 (2015).
(12) Satoshi Tamura, Hiroshi Ninomiya, Norihide Kitaoka, Shin Osuga, Yurie Iribe, Kazuya Takeda, Satoru Hayamizu, ″INVESTIGATION OF DNN-BASED MODELING FOR AUDIO-VISUAL SPEECH RECOGNITION″ Proc of. MLSLP 2015 (2015).
(13) Hiroshi Ninomiya, Norihide Kitaoka, Satoshi Tamura, Yurie Iribe and Kazuya Takeda, ″Integration of Deep Bottleneck Features for Audio-Visual Speech Recognition″ Proc of. InterSpeech 2015 (2015).
(14) Seng Kheang, Kouichi Katsurada, Yurie Iribe and Tsuneo Nitta: ″Model Prioritization Voting Schemes for Phoneme Transition Network-based Grapheme-to-Phoneme Conversion″, Proc. of CIST2015, No.100 (2015).

〔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：
国内外の別：

○取得状況（計 1 件）
名称：発音動作可視化装置および発音学習装置
発明者：入部百合絵, 新田恒雄
権利者：国立大学法人 豊橋技術科学大学
種類：特許
番号：6206960
取得年：2017 年
国内外の別： 国内

〔その他〕
ホームページ等
http://www.ist.aichi-pu.ac.jp/~iribe/

６．研究組織
(1)研究分担者
研究分担者氏名：新田 恒雄
ローマ字氏名：Tsuneo Nitta
所属研究機関名：早稲田大学
部局名：グリーン・コンピューティング・システム研究機構
職名：招聘研究員
研究者番号（8 桁）：70314101
(2)研究協力者
研究協力者氏名：
ローマ字氏名：