

平成 30 年 6 月 19 日現在

機関番号：82626

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00262

研究課題名(和文) 異種音声単位と複数言語を用いた高分解能音声特徴空間の構築と応用の研究

研究課題名(英文) Research on construction and application of high discriminative speech feature space using heterogeneous speech units and multiple languages

研究代表者

李 時旭 (Lee, Shi-wook)

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：50415642

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、異種情報を用いて音声特徴空間上の識別能力を高めることにより音声認識の性能を向上させることを目的とする。

最近の深層学習技術では、殆どの音声認識システムが単一の音声単位を基に構築されるため、膨大なデータを用いても音声の多様性を十分にはモデル化できない問題が存在する。その解決策として、従来の文脈的拡張音声単位と完全に異なる時間的拡張の音声単位である音素片を導入し、異種性の高い音声特徴空間の構築を提案する。提案した異種音声単位による高分解能音声特徴空間が従来の生成モデルから最先端の深層学習モデルまで幅広い音声認識システム上で有効であることが、実応用の音声検索語検出タスクを対象に確認できた。

研究成果の概要(英文)：This research aims to improve speech recognition performance by enhanced discriminative ability on speech feature space using heterogeneous information.

Due to most speech recognition systems by recent deep learning techniques are constructed on the basis of a single speech unit, speech diversity cannot be sufficiently modeled even with enormous speech data. As a solution to the problem, we adopt a sub-phonetic segment unit which is a temporal extension speech unit and is completely different from the conventional contextual dependent speech unit.

We confirmed that the proposed high discriminative speech feature space based on heterogeneous speech units is effective on a wide range of speech recognition systems; from conventional generation models to leading-edge deep learning models.

研究分野：情報学

キーワード：音声情報処理 パターン認識 ヒューマンインタフェース 時系列解析 統計的パターン認識 情報検索 多変量解析 知能情報処理

1. 研究開始当初の背景

近年、深層学習(deep learning)の導入やクラウドコンピューティングの発展と共に、スマートフォンやスマートスピーカー等に音声認識機能が搭載され広く利用されるようになったが、本来の利便性に寄与する実使用には性能上いまだ一段の高度化が必要である。しかし、基本技術自体の改良開発は極めて困難であるのも現状である。このような状況の中で、音声信号処理や自然言語処理において複数システムの統合による性能向上の研究が最近の深層学習分野で活発的に展開されてきた。初期の具体例としては、米国の国立標準技術研究所(NIST)が複数研究機関の音声認識システムを統合した Recognizer Output Voting Error Reduction(ROVER, 1997年)があり、その後、統計的な機械翻訳や音声検索語検出(Spoken Term Detection: STD)においても研究開発が盛んになっている。また、海外においては IARPA's Babel project, DARPA's RATS program などの大規模のプロジェクトがあった。更に、深層学習の導入により、多種多様なシステムを統合するアンサンブル技術が積極的に取り込まれることから、世界的な研究動向としてもシステム統合が注目を集めている。これらの状況から、膨大な情報を同時に利用できる計算機の能力を利用し、より多くの多面的な情報を与え、それらの情報を最適に統合するアルゴリズムの開発が重要な課題となっている。

研究の社会的なニーズとして、高速インターネット網と大容量録画装置の普及に伴って、情報のパラダイムが従来のテキストからマルチメディアへ急速に拡大されつつある。一方、インターネットにおけるマルチメディアコンテンツの比重が著しく増大しているにも関わらず、動画や音声に関する実用的かつ有効な検索技術は確立しておらず、利便性が損なわれている。このように情報技術産業の発展から、最大のビックデータになったマルチメディアを有効的に利活用するためには、音声認識機能が重要な役割を果たす。

また、音声をインタフェースとして用いる様々なデバイス、スマートフォンやスマートスピーカー等がすでに製品化されているが、実際の使用状況を調査すると本来の利便性を発揮するまでには至っていない。今後、知能ロボットへの組み込みなど、障害者・高齢者を含む多くの一般ユーザの音声インタフェースへの期待は大きい。現在の製品の性能とユーザの想定にはギャップがある。ユーザインタフェース自体の研究も重要であるが、有用に使用されるためにはさらなる認識性能向上が必要不可欠である。以上のように、音声認識技術の更なる性能向上に向けた基礎研究や実応用システムの開発が急務であることが研究開始当初の研究背景であった。

2. 研究の目的

本研究では、異種情報を用いて音声特徴空間の分解能を極め、識別能力の高度化により音声認識技術の性能を向上させることを目的とする。

音声認識は、観測される音響信号から音声記述単位の音韻情報を抽出し、言語情報を用いて認識結果のテキストを生成する技術である。音響信号からの音韻情報抽出は、特徴抽出と音韻の音響モデル作成の二段階から構成される。特に、音響信号から抽出された特徴を音声特徴空間に写像する手法については識別モデルや生成モデルなどにより識別能力を高める研究が進められてきた。そのための有望な手法として、特徴空間の高分解能化が注目されている。本研究では、異種の音声記述単位が持つ特徴を利用し、複層的に音声特徴空間を生成することで特徴空間の分解能力を高める方法を研究・開発する。

現在の音声認識技術では、音素列の文脈的な拡張として文脈依存音素である triphone を用いて、高精度な特徴空間を構成する手法が一般的である。それに対して先行研究において、音声の時間的な分解能を拡張した音素片(Sub-Phonetic Segment: SPS)を開発して高い認識性能を実現して来た。それで、これら二つの異種単位の併用による多面的な解釈を合わせて、高分解能の音声特徴空間を構成する方法を開発する。図1に、特徴空間の構成を中心とする研究の構想を図解する。

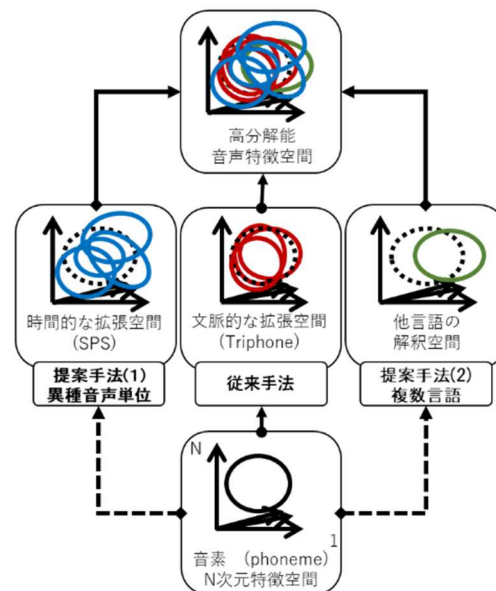


図1 特徴空間から図解する本研究の構想

3. 研究の方法

本研究では、高分解能の音声特徴空間を構築するため、二種類の音声記述単位である triphone (文脈的拡張と見做せる) と音素片 (SPS、時間的拡張) との最適な統合手法の研究を行う。まず、大規模な音声データを用いて triphone と音素片の音声特徴空間を生成し、

その音響的な尤度や事後確率に基づいて各々の音声特徴空間の最適化を行う。その後、統合のアルゴリズムを適用する。提案のアルゴリズムに対して、音声検索語検出タスク (Spoken Term Detection ; STD) の評価実験を通して評価する。

(1) 異種音声記述単位の統合

本手法では、性能向上のため異なるシステムによって生成された検出候補を統合する。システムのコアは、様々なレベル (フレーム、状態、または検出された候補) で統合される。最も単純なアプローチは、式 (1) のように N 個の複数のシステムの観測対数尤度の線形補間を用いてフレームレベルの結合を行う手法である。

$$\log p(\mathbf{o}_t|q) = \sum_{n=1}^N \alpha_n \log p_n(\mathbf{o}_t|q) \quad (1)$$

$$\text{, where } \sum_{n=1}^N \alpha_n = 1$$

ここで、 α_n は n 番目システムの補間重みであり、 $p(\mathbf{o}_t|q)$ は隠れマルコフモデル (Hidden Markov Model : HMM) の状態 q が与えられた場合の統合された尤度であり、 $p_n(\mathbf{o}_t|q)$ は n 番目システムの尤度である。異なるフレーム構成、ガウス混合分布 (Gaussian Mixture Model ; GMM) ベースのシステムの HMM 状態、DNN (Deep Neural Network) ベースのシステムの入力および出力レイヤなどに統一的にスコア融合を適用するために、各システムの検出候補リストに対してスコア融合を実行する。まず、検出されたリストは、検出区間の時間上の重複に基づいてシステム間で整列され、整列されたタームのスコアは N 個のシステムから式 (2) で統合される。

$$\hat{s}_d = \sum_{n=1}^N \alpha_n \cdot s_{d,n} \quad (2)$$

$$\text{where } \sum_{n=1}^N \alpha_n = 1$$

ここで、 d は統合される検索語であり、 $s_{d,n}$ は n 番目のシステムの検出された検索語 d のスコア、 \hat{s}_d は、検出された検索語 d の線形補間法により融合されたスコアである。検出された検索語がどのシステムのリストにも現れない場合、そのシステムのスコアは零とみなす。補間の重み α は、最高性能のために実証実験により最適化される。

(2) 音声検索語検出における事後確率によるスコアリング

フレームベースの動的時間伸縮法 (Dynamic Time Warping : DTW) による音声検索語検出では、まず、テキスト入力のクエリをサブワード状態列に変換する。そして、検索対象の音

声データベースはフレーム毎にサブワード状態の事後確率を bidirectional long short-term memory (LSTM) RNN から計算して置く。検出は式 (3) の DTW パスから計算されるクエリとの距離に近い順位で検出区間を列挙する。

$$G(q,r) = \min \left\{ \begin{array}{l} G(q,r-1) + D[(q,r)|(q,r-1)] \\ G(q-1,r-1) + D[(q,r)|(q-1,r-1)] \\ G(q-1,r) + D[(q,r)|(q-1,r)] \end{array} \right\} \quad (3)$$

ここで、 q はクエリの出力クラスであり、 r は音声データベースのフレームである。 $G(q,r)$ はクエリのクラス q と検出対象音声の r 番目フレームまでの蓄積される距離値を示す。 $G(q,r)$ はクエリの最終クラスで検出区間のフレーム数により正規化され、異なる長さの検出区間らのスコアとして使われる。

式 (3) の右辺の第 2 項の $D[\cdot]$ は、クエリのクラスとフレームとの距離値であり、本研究では式 (4) の bidirectional-LSTM-RNN のソフトマックス出力の事後確率を用いる。フレーム t で音響特徴 \mathbf{o}_t が観測された際、クラス q が出現される事後確率 $p(q|\mathbf{o}_t)$ を、式 (5) で計算する負の対数を局所距離として用いる。

$$p(q|\mathbf{o}_t) = \frac{\exp(x_q^t)}{\sum_{q_k \in Q} \exp(x_{q_k}^t)} \quad (4)$$

$$D = -\log(p(q|\mathbf{o}_t)) = -x_q^t + \log \left(\sum_{q_k \in Q} \exp(x_{q_k}^t) \right) \quad (5)$$

(3) 事後確率の統合

複数の多様なシステムの統合は単一システムより高性能である。しかし、その統合は実験的に決めることが多く、数理的な検証は稀である。本手法では、異なる条件下で学習された複数の bidirectional-LSTM-RNN からのクラス事後確率を統合する手法について検討する。

最近、大規模モデルや多数モデルのアンサンブルによる高い性能を一つの小規模モデルに伝え教え、効率的なシステムの構築が可能である蒸留 (distillation) 手法が報告されている。このアイディアに基づき、事後確率の統合により検出スコアの高精度化と STD タスクの性能向上を検証する。統合手法として、一般化ベキ平均 (generalized power mean) による事後確率の統合を検証する。 N 個のシステムによるクラスの事後確率、 $(p_1, \dots, p_N) \in [0,1]$ は正の実数であり、事後確率らのベキ平均は指数 k の基で式 (6) で計算される。

$$M_k(p_1, \dots, p_N) = \left(\frac{1}{N} \sum_{i=1}^N p_i^k \right)^{\frac{1}{k}} \quad (6)$$

ここで、ベキ指数 k は非零の実数であり、 $k = 1$ の M_1 は算術平均(arithmetic mean :AM)、 $k = 2$ の M_2 は平方平均(quadratic mean :QM)、 $k = -1$ の M_{-1} は調和平均(harmonic mean :HM)、 $k \rightarrow 0$ の $\lim_{k \rightarrow 0} M_k$ は幾何平均(geometric mean :GM)、そして $k \rightarrow +\infty$ の $\lim_{k \rightarrow \infty} M_k$ は最大値(MAX)である。この平均らは不等式 $M_k \leq M_l$ 、 $k < l$ を満たす。ここで等号が成立するのは $p_1 = p_2 = \dots = p_N$ のときに限る。このベキ平均らに加え、事後確率の総和(SUM)を用いることも試す。ここで、 $\sum_{i=1}^N p_i \geq \lim_{k \rightarrow \infty} M_k$ 、即ち $SUM \geq MAX$ の不等式が成り立つ。また、総和による統合では、統合後の事後確率の上限値を1に設定した。クラスの数が増え統合されるシステムの数より極端に多いため、上限値に近接するクラス数は稀であり、確率の平滑化による識別能力の劣化は生じ難いと考えられる。まとめとして、 $HM \leq GM \leq AM \leq QM \leq MAX \leq SUM$ の不等関係であるベキ平均と総和を用いる事後確率の統合手法に対して音声検索語検出タスクの比較実験を行う。

4. 研究成果

音声検索語検出タスクの評価実験は、NTCIR-10 SpokenDoc-2 TaskのSPDWS STDタスクを用いて、未知語を含む100個のテキストクエリを対象に行った。評価尺度としては検出率と正解率の調和平均の最大値を100個のクエリに対して平均をとったAMF(Average Maximum F-measure)を用いた。

(1) 異種音声記述単位の統合

平成27年度の初年度研究では、平成27年10月付きの追加採択から5ヶ月間の短い研究開始時期であるため、計算機資源の導入、データベースの準備、従来手法の確認と提案手法の予備実験などの研究環境整備を中心とした。まず、日本語の音声特徴空間を、従来の正規混合分布を用いる統計的な生成モデルから深層学習による識別モデルへ拡張構築し、予備実験を行った。また、本研究の提案である異種情報に基づく複数システムの統合によるさらなる性能向上も得られた。そのシステム統合による性能向上は、従来手法との比較実験から優れた手法であることを確認できた。これまで、多様な抽出からの音声特徴を統合する手法やデータ構成を変えて学習した複数のモデルの結果を統合する手法、文脈依存音素における状態共有の条件を変える手法などが提案されてきたが、いずれも実験的なアプローチであった。それに対して、本研究では、従来の文脈的拡張音声単位と完全に異なる時間的拡張の音声単位である音素片を導入し、異種情報間の異質性を高めることにより音声特徴空間の高分解能性を得ることができた。システム統合においては、統合する個別システムの正確さと多様性が、性能向上を高める必要十分条件であるといえる。

表1は、各検出セグメントのスコアを式(2)の線形補間法により融合した結果を示す。第3列の相関係数(Corr)と第5列の検出性能向上率(Imp)から、検出スコア間の相関が低い程、性能向上率が上がる逆比例の関係が確認できた。

表1 システム統合による性能向上

Sys. #1	Sys. #2	Corr	AMF	Imp.
MFCC_TRI	PLP_TRI	0.79	54.80	6.53
MFCC_SPS	PLP_SPS	0.81	52.28	5.59
MFCC_TRI	MFCC_SPS	0.65	60.77	18.14
PLP_TRI	PLP_SPS	0.64	57.97	17.09
MFCC_TRI	PLP_SPS	0.64	62.39	21.29
MFCC_SPS	PLP_TRI	0.64	57.18	17.08
MFCC_TRI_LM	PLP_TRI_LM	0.81	60.73	3.18
MFCC_SPS_LM	PLP_SPS_LM	0.87	61.95	2.74
MFCC_TRI_LM	MFCC_SPS_LM	0.65	68.50	13.59
PLP_TRI_LM	PLP_SPS_LM	0.63	66.28	11.88
MFCC_TRI_LM	PLP_SPS_LM	0.64	67.44	13.84
MFCC_SPS_LM	PLP_TRI_LM	0.64	67.52	11.97

(2) 音声検索語検出における事後確率によるスコアリング

平成28年度の研究は、異種音声単位を用いた高分解能の音声特徴空間を構築するために、異種単位のDNNモデル構築と事後処理の統合手法の研究を進めてきた。

まず、従来のGMMとの比較実験を行い、深層学習を用いることで基本性能を大きく高めることを確認した。その上、入力特徴空間と出力クラスの両方面に異なる定義を用いる提案の異種音声単位を導入することに成功した。高い性能と低い相関性で定義された異種性の定量的な数値とシステム統合による性能向上間の比例性を、実証実験から証明した。表2は、音素片をサブワードとしたサブワード認識の結果を対象にDTWマッチングを行った検出結果を示す。

表2 サブワード認識の結果に基づく性能

	AMF
8mixture GMM-HMM	60.30
5x2048 DNN-HMM	72.41

表3 事後確率の距離値に基づく各手法の性能

GMM	AMF
4 mixture	62.03
8 mixture	65.11
16 mixture	66.90
DNN	AMF
1x2048 HL	76.62
3x2048 HL	81.03
5x2048 HL	79.08
RNN-LSTM	AMF
1x1024 HL	82.13

表3は、音素片のHMM状態におけるフレーム毎の事後確率をDTWマッチングの距離として直接用いた場合の検出結果を示す。GMMでは混合数を、DNNでは隠れ層の条件を変えて実験を行った。全体を通してDNNはGMMより遥かに高い性能を示している。

表2と3の結果を比較すると、まず、全体を通してDNNはGMMより遥かに高い性能を示している。また、フレーム毎の事後確率を距離として用いることから検出の性能向上が得られた。最後に、前後5フレームに展開した1320次元の特徴を用いて時系列情報を処理したDNNの結果と比較すると、当該フレームの120次元のみを用いるRNN-LSTMでは一層の隠れ層でも高い性能が得られた。

(3) 事後確率の統合

平成29年度の研究は、複数システムの多様な出力を効果的に統合するアルゴリズムの開発に重点を置いた研究を行った。具体的には複数システムの統合における知識蒸留(knowledge distillation)の概念を導入した。異種音声単位の複数システムの事後確率を効果的に統合する手法を確立するため、ニューラルネットワークの最終層であるソフトマックスレイアの前後の段階でスコア融合を試した。従来手法の算術・幾何・調和平均や線形補間などと比較して事後確率の上限付き合計値によるスコアヒュージョンが最高性能である結果を得た。今後、高性能だが複雑であるシステムから有効的な情報のみを抽出して構成するシステムコンパクト化の研究において、本研究の上限付き合計値による統合手法を導入することが考えられる。評価尺度としてはmean Average Precision(mAP)を用いた。

表4 事後確率の統合による性能 ; mAP (%)

	2-way	4-way	6-way	8-way
HM	84.99	84.70	84.43	83.84
GM	85.34	85.38	85.50	84.58
AM	85.63	86.03	86.17	86.07
QM	85.81	86.22	86.45	86.25
MAX	86.11	86.96	87.20	86.99
SUM	86.68	87.72	87.80	87.62

表4では、事後確率の統合数や手法による性能として各条件下の最高値を示す。全体的に、統合するシステム数の増加とともに検出性能が向上する傾向が見られる。さらに、統合手法の比較ではベキ平均と総和の不等関係と同等の性能向上が確認できる。また、八つのサブシステムを統合する2-way、4-way、6-wayの組合せすべての ${}_8C_2=28$ 、 ${}_8C_4=70$ 、 ${}_8C_6=28$ 通りによる結果の分布を図2の箱ひげ図で比較する。図2から、2-way、4-way、6-wayの統合数条件において、 $HM \leq GM \leq AM \leq QM \leq MAX \leq SUM$ の不等式と一致する検出性能の向上傾向が確認できる。

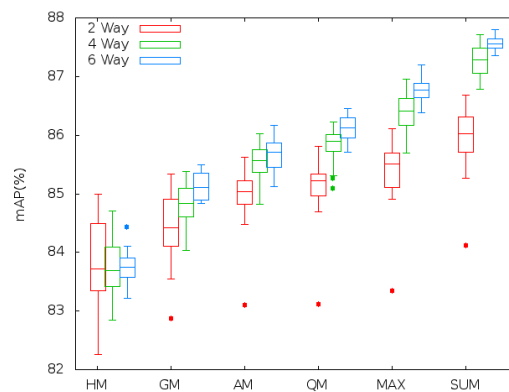


図2 各事後確率の統合手法と統合システム数による性能向上 ; (mAP (%))

<引用文献>

- ① Shi-wook Lee, Kazuyo Tanaka, Yoshiaki Itoh, “Generating complementary acoustic model spaces in DNN-based sequence-to-frame DTW scheme for out-of-vocabulary spoken term detection,” INTERSPEECH-2016, pp. 755-759, 2016-9, San Francisco, USA.
- ② 李時旭, 田中和世, 伊藤慶明, “音声検索語検出システムのスコアリングに関する実験的検討,” 日本音響学会春季研究発表会, pp. 181-182, (2017-3)
- ③ 李時旭, 田中和世, 伊藤慶明, “音声検索語検出の距離値における事後確率の統合,” 日本音響学会春季研究発表会, pp. 197-198, (2018-3)
- ④ Shi-wook Lee, Kazuyo Tanaka, Yoshiaki Itoh, “Empirical analysis of score fusion application to combined neural networks for open vocabulary spoken term detection,” INTERSPEECH-2018, Accepted, 2018-9, Hyderabad, India.

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) 伊藤慶明, 鳴海司朗, 大内一揮, 菅原翔太, 李時旭, “音声中の未知語の検索語検出における音節バイグラムのインデックス化方式,” 電子情報通信学会論文, D Vol. J99-D, No. 2, pp. 178-187 (2016-2).
- (2) 小嶋和徳, 紺野和磨, 田中和世, 李時旭, 伊藤慶明, “音声中の検索語検出における同文書内の高順位候補を利用したリスクスコアリング方式,” 電子情報通信学会論文, Vol. J100-D, No. 1, pp. 70-80 (2017-1).
- (3) 紺野良太, 李時旭, 田中和世, 小嶋和徳, 伊藤慶明, “音声中の検索語検出におけるDeep Neural Networkの出力確率を用いたリスクスコアリング手法の提案,” 電子情報通信学会論文, Vol. J100-D, No. 5, pp (2017-5).

[学会発表] (計 20 件)

- (1) Ryota Konno, Kazunori Kojima, Lee Shi-Wook, Kazuyo Tanaka, Yoshiaki Itoh, “Rescoring by a Deep Neural Network for Spoken Term Detection,” 4 pages, Asia-Pacific Signal and Information Processing Association APSIPA 2015-12.
- (2) 紺野良太, 李時旭, 田中和世, 小嶋和徳, 伊藤慶明, “STD におけるフレームレベル状態系列間照合による検索精度向上,” 日本音響学会秋季研究発表会, pp. 127-130, (2015-9).
- (3) 李時旭, 田中和世, 伊藤慶明, “確率分布間の距離近似と異種に基づく音声検索語検出システムの統合,” 日本音響学会秋季研究発表会, pp. 121-122, (2015-9).
- (4) Shi-wook Lee, Kazuyo Tanaka, Yoshiaki Itoh, “Generating complementary acoustic model spaces in DNN-based sequence-to-frame DTW scheme for out-of-vocabulary spoken term detection,” INTERSPEECH-2016, pp. 755-759, 2016-9, San Francisco, USA.
- (5) Masato Obara, Kazunori Kojima, Kazuyo Tanaka, Shi-wook Lee, Yoshiaki Itoh, “Rescoring by Combination of Posteriorgram Score and Subword-Matching Score for Use in Query-by-Example,” INTERSPEECH-2016, pp. 755-759, 2016-9, San Francisco, USA.
- (6) 伊藤慶明, 紺野良太, 小原真人, 李時旭, 田中 和世, “音声中の検索語検出の研究動向と DNN の導入事例,” 音声研究会・音声言語情報処理研究会合同研究会, 2016-7, 2016
- (7) 紺野良太, 李時旭, 田中 和世, 小嶋和徳, 伊藤慶明, “音声中の検索語検出における音響距離構築方式の検討,” 音声研究会・音声言語情報処理研究会合同研究会, 2016-7, 2016
- (8) 清水嘉乃, 岩崎瑛太郎, 李時旭, 田中和世, 小嶋和徳, 伊藤慶明, “STD における複数検索結果のスコア優先統合方式,” 日本音響学会秋季研究発表会, pp. 69-72, (2016-9)
- (9) 紺野良太, 李時旭, 田中和世, 小嶋和徳, 伊藤慶明, “サブワード/状態/フレーム照合スコアの統合による SQ-STD 検索精度向上,” 日本音響学会秋季研究発表会, pp. 73-76, (2016-9)
- (10) Yoshino Shimizu, Eitaro Iwasaki, Shi-wook Lee, Kazuyo Tanaka, Kazunori Kojima, Yoshiaki Itoh, “An integration method of multiple search results for spoken term detection,” 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, 2016-11, Hawaii, USA.
- (11) 李時旭, 田中和世, 伊藤慶明, “音声検索語検出システムのスコアリングに関する実験的検討,” 日本音響学会春季研究発表会, pp. 181-182, (2017-3)
- (12) 紺野良太, 小嶋和徳, 李時旭, 田中和世, 伊藤慶明, “SQ-STD における DNN 及び CTC 導

入方式の検討,” 日本音響学会春季研究発表会, pp. 183-186, (2017-3)

(13) 関恒平, 小嶋和徳, 李時旭, 田中和世, 伊藤慶明, “音声中の検索語検出における拗音及び長母音モデルの検討,” 日本音響学会春季研究発表会, pp. 187-188 (2017-3)

(14) Daisuke Kaneko, Ryota Konno, Kazunori Kojima, Kazuyo Tanaka, Shi-wook Lee, Yoshiaki Itoh, “Constructing acoustic distances between subwords and states obtained from a deep neural network for spoken term detection,” INTERSPEECH-2017, pp. 2879-2883, 2017-8, Stockholm, Sweden.

(15) Masato Obara, Munehiro Moriya, Ryota Konno, Kazunori Kojima, Kazuyo Tanaka, Shi-wook Lee, Yoshiaki Itoh, “Acceleration for Query-by-Example Using Posteriorgram of Deep Neural Network,” APSIPA ASC 2017, 4 pages, 2017-12, Kuala Lumpur, Malaysia

(16) Shi-wook Lee, Kazuyo Tanaka, Yoshiaki Itoh, “Empirical analysis of score fusion application to combined neural networks for open vocabulary spoken term detection,” INTERSPEECH-2018, Accepted, 2018-9, Hyderabad, India.

(17) 清水嘉乃, 李時旭, 田中和世, 小嶋和徳, 伊藤慶明, “音声中の検索語検出における Paragraph Vector を用いたスコアリング手法,” 日本音響学会秋季研究発表会, pp. 145-148, (2017-9)

(18) 李時旭, 田中和世, 伊藤慶明, “音声検索語検出の距離値における事後確率の統合,” 日本音響学会春季研究発表会, pp. 197-198, (2018-3)

(19) 小原真人, 小嶋和徳, 李時旭, 伊藤慶明, “音声中の音声検索語検出における Posteriorgram 照合の検索時間削減方式,” 日本音響学会春季研究発表会, pp. 83-96, (2018-3)

(20) 丹治遥, 小嶋和徳, 李時旭, 南條浩輝, 伊藤慶明, “音声中の検索語検出における最上位候補を含む講演及びその類似講演優先方式,” 日本音響学会春季研究発表会, pp. 185-186, (2018-3)

6. 研究組織

(1) 研究代表者

李 時旭 (LEE, Shi-wook)

国立研究開発法人産業技術総合研究所・知能システム研究部門・主任研究員
研究者番号：50415642

(2) 研究分担者

伊藤 慶明 (ITOH, Yoshiaki)

岩手県立大学・ソフトウェア情報学部・教授
研究者番号：90325928