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　５ 年　 ６ 月 ２８ 日現在

機関番号：１３８０１
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１４３１

研究課題名（和文）実環境音声認識のための深層学習と人手を併用する音声言語知識拡充フレームワーク

研究課題名（英文）A Spoken Language Knowledge Expansion Framework for Real-World Speech Recognition Using Deep Learning Technology and Human Collaboration

研究代表者
甲斐　充彦（Kai, Atsuhiko）

静岡大学・工学部・准教授

研究者番号：６０２８３４９６

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：自動音声認識（ASR）技術を長時間の自動字幕や検索等へ応用可能にするため、低コストで新しい用語等の音声言語知識の拡充を可能にするASR関連技術の開発を行なった。具体的には、リアルタイム出力可能なASRシステムを構築し、出力テキストを直接編集するのではなく修正語だけを入力する枠組みの半自動修正支援システムを実現した。修正語が録音中に現れた時刻を得るために用いる音声検索語検出技術では、かな読みを推論するEnd-to-end型ASRモデルを用いて誤認識箇所に多い未知語に対して検出精度を改善した。この他、雑音や複数話者が混在する音声を想定した話者分離・音声区間検出手法等の開発を行い、有効性を確認した。

研究成果の学術的意義や社会的意義
講義や会議などの長時間音声に対する自動音声認識（ASR）技術の適用において、近年のAI技術を用いた事例では、新しい語や話題を低コストで効率的に習得する手法が不足しており、全自動での字幕生成等では実用的な認識精度がまだ達成されていない。本研究課題では、リアルタイム性を重視して新しい語のテキスト情報のみを手動で提供する枠組みを提案し、ASR技術を基にした自動字幕や検索の精度を低コストで改善させる手法を提案した。これにより、ASR技術の応用可能性を一段と高められることを実証した。

研究成果の概要（英文）：In order to make automatic speech recognition (ASR) technology applicable to long-term automatic subtitling and retrieval, we developed ASR-related technology that enables the expansion of spoken language knowledge, such as new technical terms, at a low cost. Specifically, we constructed an ASR system that can output in real-time, and realized a semi-automatic correction support system in which users do not directly edit the output text, but only input the corrected words. Regarding the spoken term detection technique used to obtain the timing at which the corrected word appears in the recording, an end-to-end ASR model that infers the reading of speech was used to improve the detection accuracy for unknown words, which are common among misrecognized words. In addition, speaker separation and voice activity detection methods were developed for noisy and multi-speaker speech, and their effectiveness was confirmed.

研究分野：音声言語処理

キーワード：自動音声認識　深層学習　長時間録音　自動修正　音声検索語検出　読み推定　End-to-end型　リアルタイム

１．研究開始当初の背景

(1) 自動音声認識（Automatic Speech Recognition: ASR）技術の分野では、音響的または言語的な知識を機械的に学習する方法が主流であり、近年では深層学習を用いた ASR 技術が最も高い精度を得ている。講義や会議など長時間の音声をディジタル化して利活用する場合、例えば聴覚障碍者向けの情報保障としての字幕付与や、音声から発言内容の検索などにおいて、ASR 技術は重要な基盤技術となる。しかし、講義や会議などで想定される話題（ドメイン）や収録環境は様々であり、そのような新しい話題や環境等に対して ASR システムが高い認識精度を得るには、その対象ドメインに関わる音声とその書き起こしを別途準備してチューニングすることが必要となることが多い（図 1）。学習データを拡充するために音声の書き起こしを人手で得る場合には、通常は収録時間の 10 倍以上の時間的コストを必要とする。しかし、大学講義や社内会議などへの字幕付与や会議議事録作成を目的として ASR システムを用いることを考えると、個々の話題ごとの学習データを用意するために多大なコストをかけることは現実的ではない。一方で、講義の復習や会議議事録の検索のための書き起こし作成を考えると、必ずしも 100%の精度を求める必要はないと考えられ、作成コストの少なさを優先した自動字幕の修正や書き起こし作成支援の仕組みが望まれる。

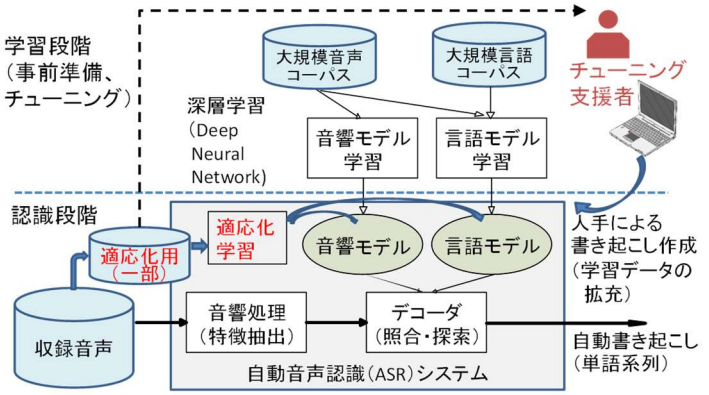

図 1 自動音声認識システムの構成とチューニング過程の例

(2) 近年では会議室や移動環境での収録音声のように周囲の雑音や残響の影響を含む実環境を対象とした音声認識タスクが設定され活発に研究されている。講義音声や会議音声などでは周囲の話者の音声が重なったり一繋がりになったりしてしまうことが多く、音声認識精度の大幅な低下につながることが報告されている。複数話者が存在する場合の発話時間の区分化については、「いつ、誰が話しているか」を同定する話者ダイアライゼーションのタスクが設定され研究されているが、十分に高い精度が得られていない。講義音声や会議音声では、講義する教員や注目する発言者だけに注目した音声分離が望まれるが、そのような目的での音声分離技術の適用効果についてはまだ十分に研究されていない。

２．研究の目的

(1) ASR システムが出力したテキスト（自動字幕の候補）に対して人手で修正を行う仕組みとして、テキストそのものを修正するのではなく、誤ったキーワード情報の入力のみで正しいテキストへ自動修正する仕組みを備えた半自動修正支援システムを開発する。そして、講義・講演時のリアルタイムでの字幕利用や、講義終了後の内容検索などの目的においてどの程度有効となるかを実験的に検証する。

(2) 本研究課題で開発する半自動修正支援システムの評価として、入力を求める修正対象語を重要な語句に限定した場合に、内容理解や検索精度の改善にどの程度効果があるか、書き起こし修正のコストがどの程度削減されるかを明らかにする。

(3) 注目する話者とそれ以外の話者の音声を分離するシステムを開発し、自動音声認識や音声検索の技術と併用することで、他の話者の音声が混入しやすい講義や会議音声を音声言語知識拡充に活用した場合の効果を明らかにする。

３．研究の方法

(1) ASR システムが出力する自動書き起こしテキストの誤りに対して、キーワード情報のみを人手で与える想定で ASR 出力テキストを半自動修正するシステムを開発する。その仕組みは、既に開発済みの音声検索語検出（Spoken Term Detection: STD）技術を用い、入力されたキーワードの音声区間推定を行い、ASR システムの中間出力情報であるラティスを自動的に操作することで実現する。

(2) 上記(1)の成果を踏まえ、書き起こし作成支援システムのユーザインタフェース（UI）を Web 技術をベースとして構築し、被験者による ASR テキストの修正活動のデータを収集する。その結果を分析し、書き起こし修正の効率や修正された書き起こしの内容理解度の評価と、これらの評

価結果を踏まえた上記の各要素技術の設計改善を繰り返す。

(3) 講義および会議音声を対象として、多人数の話者が混在する音声からの注目話者の分離技術を開発する。その仕組みは、注目話者の音声と他人の音声を人工的に重畳した音声から、音声分離モデルをディープニューラルネットワーク（DNN）で学習することによって実現する。講義や会議音声では、発言者がマイクに近い条件が一般的と考え、他人の音声との音量レベル差を考慮して注目話者の候補とする。そして、全体の仕組みとしては、これまで開発してきた DNN を用いた音声区間検出（Voice Activity Detection）システムを注目話者の区間候補の抽出のために利用し、音声分離モデルと併用する方法を基本とする。

４．研究成果

(1) 本課題の研究計画に従って、まずリアルタイムの自動字幕生成を想定した基盤 ASR システムを構築した。この基盤 ASR システムは、本研究課題で扱う ASR テキスト（自動字幕の候補）の半自動修正技術の開発を想定して、音響モデルとして高い精度が実現できるトライフォン単位 DNN-HMM（Deep Neural Network-Hidden Markov Model）モデルを用いた。これにより、対象とするトピックの語彙や言語モデルを容易に拡張可能な枠組みとした。そして、その基盤 ASR システムを対象として最初に挙げた研究課題を解決するために、次のような要素技術やシステムの開発を進めた。

1) ASR 出力に対する修正語のフィードバックによる半自動修正手法
2) 半自動修正手法の性能改善をもたらす未知語に頑健な音声検索語検出手法
3) 雑音下および複数話者混在の収録音声を想定した話者分離・音声区間検出手法
4) リアルタイムの ASR テキスト出力と修正語のフィードバック入力による半自動修正機能を備えた自動字幕支援システム

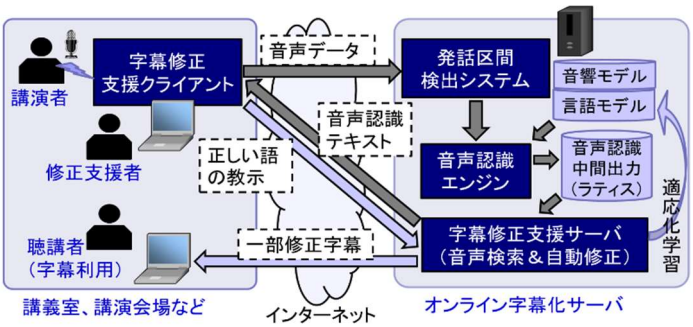

図 2 修正語のフィードバックによる自動修正を備えた自動字幕支援システム

リアルタイムの ASR テキスト出力と修正インタフェースを備えた基盤システム構成の概要を図 2 に示す[1]。上記の各要素のおもな成果について順に述べる。

(2) ASR 出力テキストの半自動修正手法の開発では、字幕修正を支援するユーザが字幕の切り替え単位に相当する無音区切り単位で ASR 出力の修正箇所の有無を確認し、誤認識箇所がある場合に正しい語句、すなわち修正語を手動でテキスト入力する想定とした。つまり、字幕が誤った語句の位置を指定せず、修正語のテキストのみ入力してもらう想定の仕様とした。そこで、本課題で開発する半自動修正手法は、まず無音区切りの音声区間単位の中で修正語が実際に発話された区間を推定する問題を考えた。誤認識を含んでいることが前提となる音声認識結果から修正語が実際に発話された区間を検出するために、既に開発済みであった音声検索語検出（Spoken Term Detection: STD）技術を用いた。用いた STD 技術は、音素単位の音響モデル（DNN-HMM）のモデルパラメータを利用し、音響的な類似度を考慮した照合を行うことにより、誤認識を含む ASR 出力テキストとの近似照合を考慮した。そして、ASR 出力テキストの半自動修正の仕組みとして、ASR 出力表現として単語レベルの曖昧さを表現する単語ラティスの情報を利用し、修正語を含む最尤候補を再探索する方法で ASR 出力テキストの自動修正出力を実現した。

(3) 前記の ASR 出力テキストの半自動修正手法では、修正語を人手でフィードバックしてもらう想定のため、ASR システムの単語辞書の拡充や、半自動修正で推定された修正後の ASR テキストを言語モデル適応に利用することで ASR システムのチューニングが可能である。そこで、半自動修正手法の単独の効果だけでなく、その半自動修正の結果を利用して言語モデル適応等を行い、チューニングした ASR システムによる再認識によって修正コストを削減できるかの評価を行った[2]。評価実験では、日本語話し言葉コーパス（CSJ）を使用した。本研究課題では講義のように専門語が多い音声の自動字幕を想定し、評価対象の講演としては CSJ のコア講演として定義される学会講演のうち、ASR システムにとっての未知語が多い 11 講演を選んだ。ASR システムの音響モデルは CSJ からコア講演を除く学会講演 910 講演（約 250 時間）、言語モデルはコア講演を除く学会 2362 講演で学習を行った（単語数:約 645 万語、語彙サイズ：約 6 万語）。音響モデルはトライフォン単位 DNN-HMM で、言語モデルは単語 3-gram で作成した。また、近年では言語モデルに対してもニューラルネットワークの有効性が示されており、N-gram 言語モデ

ルとの併用が一般的である。そこで、Long Short Term Memory を用いた言語モデル（LSTMLM）との併用の効果も比較した。比較評価の結果を表 1 に示す。LSTM 言語モデルを併用しない ASR システムでは、半自動修正手法のみを適用した場合に単語誤り率が 13.28%から 12.05%に改善されている。更に、修正語の辞書追加や修正語の形態素列による言語モデルの併用等、修正語のフィードバックの知識の活用による ASR システムによる再認識によって 10.34%まで改善された。ASR 出力テキストの複数候補を利用した LSTM 言語モデルとの併用においても改善が得られ、修正語のフィードバックの効果が同様に得られることが分かった。

表 1 音声認識出力の半自動修正とその結果を利用した言語適応の効果（単語誤り率%）

適応方法		LSTM 言語モデルの併用	
		なし	あり
適応前	ASR 出力	13.28	12.44
	字幕修正結果	12.05	-
適応後	一部修正字幕のみ	11.10	10.53
	+辞書追加	11.09	10.40
	+入力単語	10.72	10.07
	+文生成	10.51	9.88
	+入力単語+文生成	10.34	9.82

(4) 前記の ASR 出力テキストの半自動修正手法では、誤認識箇所に未知語が関係することが本質的に多いことから、音声検索語技術で未知語の影響を減らすことが必要であった。そこで、本研究課題では当時 ASR 技術の最先端のモデルとして注目されていた End-to-end 型音声認識モデルを、日本語特有の読みを推定するモデルとして応用した[3][4]。End-to-end 型モデルとしては、大語彙音声認識システムで高い精度を示している Transformer Encoder-Decoder の構成を採用した。表 2 に End-to-end 型 ASR モデルで採用した出力文字単位の定義の例を示す。一例として、ASR システム学習テキスト中に含まれていない単語（未知語）をクエリとして検索性能を比較した結果を図 3 に示す。図中の“DNN-HMM”の結果は、音響モデルとして DNN-HMM を採用し、音声認識結果から音素列を抽出してクエリと照合した場合、“E2E”の結果は End-to-end 型音声認識モデルを採用して出力文字列から音素列を抽出した場合である。比較手法は、かな漢字を出力文字として採用して予測された文字列から音素列に変換した場合（“(full-char) +G2P”の結果)、かな文字のみを採用して同様に変換した場合（“kana”の結果)、かな文字と名詞句区切りの記号を採用して同様に変換した場合（“kana+NPbm”）の結果を表している。これらの比較から、未知語となりやすい名詞句の区切り情報を付与した End-to-end 型音声認識モデルを採用することで、従来の方法よりクエリの検索精度（MAP 値）を大きく改善できることを示した。

表 2 End-to-end 型 ASR の出力文字の定義例

E2E (full-char)	機 能	と	し	て	は
Part of Speech	noun	postpositional particle	verb	postpositional particle	postpositional particle
E2E (kana)	キ ノ ウ	ト	シ	テ	ワ
E2E (kana+NPbm)	S_キ ノ ウ_E	ト	シ	テ	ワ

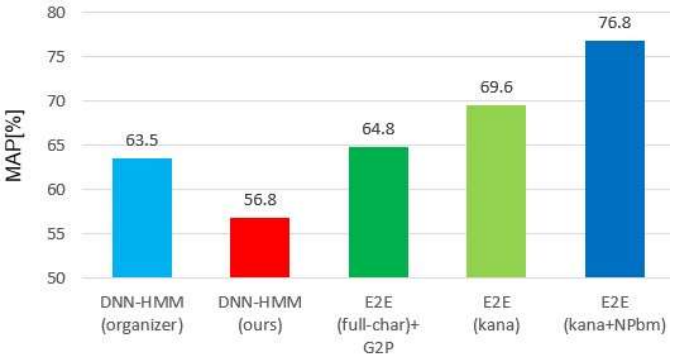

図 3 読み推定に用いる音声認識モデルの違いによる音声検索語検出の性能比較

(5) 雑音下および複数話者混在の収録音声を想定した話者分離手法の開発においては、注目話者の音声と他人の音声を人工的に重畳した音声から、音声分離モデルをディープニューラルネットワーク（DNN）で学習することによって実現した[5]。講義や会議音声では、発言者がマイクに近い条件が一般的と考え、他人の音声との音量レベル差を考慮して注目話者の候補とした。そして、注目する話者の音声特徴量と混入する話者の音声特徴量の両方を予測する特徴変換の DNN モデルを、人工的に 2 話者の音声を混合したデータをもとに学習した。一例として、講義室で講師の話者と受講生の話者がともにマイクを通して発声し、教室内の拡声器からの受講生の音声が講師のマイクに回り込み混入する場合を想定した録音データを日本語話し言葉コーパス(CSJ)の再生音声によって模擬して収録した。この再生音声のオリジナルの録音データに対して DNN-HMM 音声認識システムの単語誤り率が 15.2%であったのに対して、受講生の混入音声を想定した再収録音声では 32.75%に悪化した。しかし、提案したターゲット話者の音声特徴を分離する特徴変換の DNN を用いて、音声認識システムに入力する音声特徴量の話者分離を事前に行った場合、単語誤り率は 26.9%（変換モデルを話者適応すると 23.4%）まで改善した。また、会議音声のように周囲の雑音や他の話者の音声が混在しやすい長時間録音に対して、音声区間を頑

健に推定する音声区間検出（Voice Activity Detection: VAD）の方法の開発も進めた[6][7]。一つの方法としては、注目話者の話者特徴量を話者認識で有効な特徴表現として知られる i-vector で抽出し、CTC/Transformer Encoder-Decoder ハイブリッドの End-to-end 型 ASR モデルの入力に加えることで、暗黙的に注目話者の音声のみをデコードする仕組みを提案した[6]。その結果、提案手法は前後に他者音声を含むデータに対して、注目話者の音声認識性能を大きく改善できることを示した。一方、明示的に音声区間検出を行う方法としては、長時間録音と書き起こしによるアラインメント情報を利用し、人工的に雑音を加えたデータ拡張によって DNN モデルによる短時間単位の音声区間検出手法を提案した[7]。これらに方法によって、長時間収録音声に対して音声区間検出の誤りによる認識精度の低下を軽減できることを示した。

(6) リアルタイムの ASR テキスト出力と修正語のフィードバック入力による半自動修正機能を備えた自動字幕支援システムの開発を行った。リアルタイムの ASR テキスト出力のため、音響モデルのパラメータ数を軽減し高い認識精度を得る TDNN-HMM の音響モデルを用いた WFST デコーダを採用した。図 2 に示したように、音声入力側と ASR エンジン側をネットワークで接続するシステム構成を採用し、ASR 出力テキストの確認と修正語の入力などを Web ブラウザで実行するシステムを開発した。図 4 は実際の表示画面例であり、左側にリアルタイムで表示される ASR テキスト出力、右側に修正語（例では「異体仮名」を入力）を入力して半自動修正を実施するユーザインタフェースの内容が表示されている。ユーザが修正語を入力したことを想定したシミュレーション実験では、前記(3)に示したように STD による検出区間の精度が影響して単語誤り率の改善は 1%程度に留まっている。しかし、この実験やプロトタイプシステムでは前記(4)で示した STD 手法をまだ導入していないため、この導入によって更に改善を見込んでおり、それらを踏まえた被験者の評価実験を進める予定である（R5 年度学会発表予定）。また、近年、音声向けの基盤モデルとして注目される自己教師あり学習による事前学習モデルの応用で、少量の方言音声や実環境音声による音声認識モデルの適応性能が優れていることを確認しており[8][9]、これを導入することで提案システムの改善も図る見込みである。

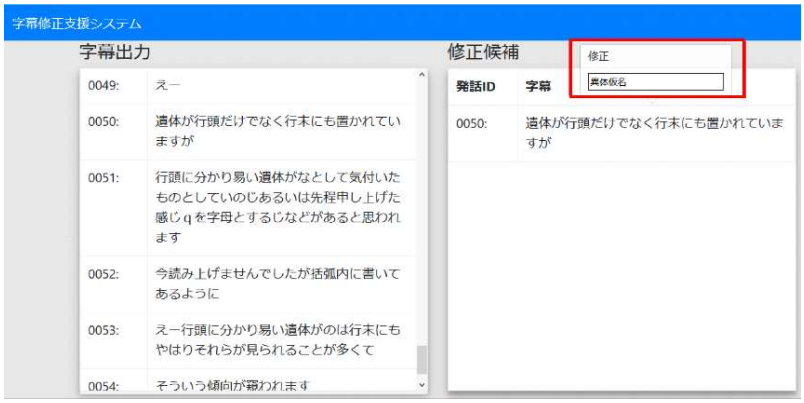

図 4 オンライン自動字幕化の支援を想定した字幕修正支援システムの画面出力例

＜引用文献＞

[1] 田宮, 寺田, 甲斐, “講義・講演の自動字幕システムを想定した低コストな半自動修正・適応手法”, 電子情報通信学会技術報告, vol. 117, no. 250, SP2017-50, pp. 89-94, 2017

[2] 寺田, 塚本, 甲斐, “講演音声認識の修正語のオンライン教示による半自動的な修正手法と語彙適応の併用の効果”, 日本音響学会 2019 年秋季研究発表会講演論文集, 1-P-12, pp. 859-862, 2019

[3] T. Kurokawa and A. Kai, “Robust Query-by-example Spoken Term Detection for Unknown Words Using Speech Retrieval-oriented E2E ASR Modeling”, Proc. GCCE2021, pp.342-343, 2021

[4] T. Kurokawa and A. Kai, “Retrieval-oriented E2E ASR Modeling for Improved Query-by-example Spoken Term Detection”, Proc. APSIPA ASC 2021, pp.1037-1042, 2021

[5] 脇屋, 福井, 甲斐, “講義音声認識のための回り込み音声の影響分析と DNN 音声分離モデルによる改善の一検討”, 日本音響学会 2019 年秋季研究発表会講演論文集, 1-P-11, pp. 855-858, 2019

[6] 大内, 甲斐, “End-to-end 音声認識における会議音声への適応および回り込み音声の影響軽減”, 電子情報通信学会技術報告, SP2019-60, pp.59-64, 2020

[7] R. Nahar, A. Kai, “Effect of Data Augmentation on DNN-Based VAD for Automatic Speech Recognition in Noisy Environment”, Proc. GCCE 2020, pp. 477-481, 2020

[8] 三輪, 甲斐, “自己教師有り学習モデル XLSR と日本語諸方言コーパスを利用した諸方言音声認識モデル”,電子情報通信学会技術報告, SP2022-63, pp. 141-146, 2023

[9] R. Nahar, R. Suzuki, A. Kai, “Domain Adaptation for Improving End-to-end ASR Performance of Classroom Speech with Variable Recording Condition”, IEICE Technical Report, SP2022-65, pp. 153-158, 2023

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 Nahar Raufun、Miwa Shogo、Kai Atsuhiko	４．巻 22
２．論文標題 Domain Adaptation with Augmented Data by Deep Neural Network Based Method Using Re-Recorded Speech for Automatic Speech Recognition in Real Environment	５．発行年 2022年
３．雑誌名 Sensors	６．最初と最後の頁 9945
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/s22249945	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計16件（うち招待講演　0件／うち国際学会　6件）

１．発表者名 R. Nahar, R. Suzuki, A. Kai
２．発表標題 Domain Adaptation for Improving End-to-end ASR Performance of Classroom Speech with Variable Recording Condition
３．学会等名 音声研究会
４．発表年 2023年

１．発表者名 三輪祥吾，甲斐充彦
２．発表標題 自己教師有り学習モデルXLSRと日本語諸方言コーパスを利用した諸方言音声認識モデル
３．学会等名 音声研究会
４．発表年 2023年

１．発表者名 Takumi Kurokawa, Atsuhiko Kai
２．発表標題 Robust Query-by-example Spoken Term Detection for Unknown Words Using Speech Retrieval-oriented E2E ASR Modeling
３．学会等名 IEEE 10th Global Conference on Consumer Electronics (GCCE2021)（国際学会）
４．発表年 2021年

１．発表者名
Takumi Kurokawa, Atsuhiko Kai

２．発表標題
Retrieval-oriented E2E ASR Modeling for Improved Query-by-example Spoken Term Detection

３．学会等名
Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC 2021)（国際学会）

４．発表年
2021年

１．発表者名
Nahar Raufun, Kai Atsuhiko

２．発表標題
Effect of Data Augmentation on DNN-Based VAD for Automatic Speech Recognition in Noisy Environment

３．学会等名
IEEE 9th Global Conference on Consumer Electronics (GCCE 2020)（国際学会）

４．発表年
2020年

１．発表者名
Takumi Kurokawa, Atsuhiko Kai, Hiroki Kondo

２．発表標題
Effects of End-to-end ASR and Score Fusion Model Learning for Improved Query-by-example Spoken Term Detection

３．学会等名
Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC 2020)（国際学会）

４．発表年
2020年

１．発表者名
Nahar Raufun, Kai Atsuhiko

２．発表標題
Efficient Channel Adaptation of ASR by DNN-based Data Augmentation using Re-recorded Paired data with Automatic Alignment Correction

３．学会等名
日本音響学会2021年春季研究発表会

４．発表年
2021年

１．発表者名
寺田侑司，塚本皓斗，甲斐充彦

２．発表標題
講演音声認識の修正語のオンライン教示による半自動的な修正手法と語彙適応の併用の効果

３．学会等名
日本音響学会2019年秋季研究発表会

４．発表年
2019年

１．発表者名
脇屋義也，福井明日香，甲斐充彦

２．発表標題
講義音声認識のための回り込み音声の影響分析とDNN音声分離モデルによる改善の一検討

３．学会等名
日本音響学会2019年秋季研究発表会

４．発表年
2019年

１．発表者名
川村智規，甲斐 充彦，中川 聖一

２．発表標題
CNNベース識別モデルによるF0推定と歌唱および読み上げ音声における評価

３．学会等名
第21回音声言語シンポジウム（情報処理学会音声言語情報処理研究会）

４．発表年
2019年

１．発表者名
大内一亜，甲斐充彦

２．発表標題
End-to-end 音声認識における会議音声への適応および回り込み音声の影響軽減

３．学会等名
電子情報通信学会音声研究会

４．発表年
2020年

１．発表者名
川村智規，甲斐 充彦，中川 聖一

２．発表標題
CNNベース識別モデルによるF0推定と伴奏重畳歌唱音声および雑音環境下読み上げ音声における評価

３．学会等名
日本音響学会2020年春季研究発表会

４．発表年
2020年

１．発表者名
Nahar Raufun、Kawai Takashi、Kai Atsuhiko

２．発表標題
Multi-Condition Training of Denoising Autoencoder by Augmenting Simulated Reverberant Speech Data

３．学会等名
2018 IEEE 7th Global Conference on Consumer Electronics (GCCE 2018)（国際学会）

４．発表年
2018年

１．発表者名
Kawamura Tomonori、Kai Atsuhiko、Nakagawa Seiichi

２．発表標題
Noise Robust Fundamental Frequency Estimation of Speech using CNN-based discriminative modeling

３．学会等名
5th. International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)（国際学会）

４．発表年
2018年

１．発表者名
近藤 宏樹，甲斐 充彦，大石 修司

２．発表標題
音声クエリからの音声検索語検出におけるスコア統合モデル学習の効果

３．学会等名
日本音響学会2018年秋季研究発表会

４．発表年
2018年

１．発表者名 川村智規，甲斐 充彦，中川 聖一
２．発表標題 CNN ベース識別モデルによる雑音に頑健な基本周波数の推定
３．学会等名 日本音響学会2018年秋季研究発表会
４．発表年 2018年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関