

令和 5 年 6 月 9 日現在

機関番号：16101

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K04311

研究課題名（和文）スムーズな対話のための対話テンポのリアルタイム制御に基づく音声対話システム

研究課題名（英文）Spoken dialog system based on real-time control of dialogue tempo for smooth dialog

研究代表者

西村 良太（NISHIMURA, Ryota）

徳島大学・大学院社会産業理工学研究部（理工学域）・講師

研究者番号：50635878

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、「リアルタイム制御可能な音声対話システムの構築」を行った。本システムはROSのアーキテクチャを採用しており、システム開発時の通信管理やデバッグを効率的に行えるようにした。またこのシステム内で用いるための応答タイミング制御モジュールである「相槌タイミング制御モジュール」の構築を行った。リアルタイム動作を実現させるために、単純な音響的情報のみを用いている。モデル構成も単純なLSTMのモデルとした。このモデルで、相槌タイミングをF値0.933という高精度で生成することが可能である。また、これらについてオープンソースとしてgithubで公開し、広く一般に利用可能とした。

研究成果の学術的意義や社会的意義

最近のChatGPTなどの大規模言語モデルの高精度化に伴い、音声対話システムに対しても、より高精度で自然な対話が期待され始めた。そして、自然に対話を行うためには対話のテンポが非常に重要であるものの、これまでの音声対話システムでは、設計上の問題で実現不可能であった。本研究の成果により、この問題が解決され、リアルタイムに制御が可能な音声対話システムを開発・動作させることが可能となる。高速に動作可能な相槌応答タイミング生成モデルも構築し、本システムに搭載したことから、音声対話システム開発者は、応答内容やその他の部分に注力してシステム開発することで、自然な音声対話システムを実現できる。

研究成果の概要（英文）：In this research, "construction of a real-time controllable spoken dialog system" was conducted. This system uses ROS architecture, which enables efficient communication management and debugging during system development. We also built the "back-channel timing control module," which is a response timing control module for use in this system. To achieve real-time operation, only simple acoustic information is used. The model is a simple LSTM model. This model is capable of generating the timing of the back-channel with an F-value as high as 0.933. The model is open-sourced on github and widely available to the public.

研究分野：音声対話システム

キーワード：音声対話システム タイミング テンポ 音声言語情報処理 深層学習 音声認識 リアルタイム制御
ROS

様式 C-19, F-19-1, Z-19 (共通)

1. 研究開始当初の背景

近年、スマートフォンの普及やその上で動作する音声認識ソフトウェアの精度の向上などにより、音声入力やその応用である音声対話エージェント (Apple Siri, Google Assistant, Amazon Alexa など) が普及しつつある。また、家庭用ロボット (ソフトバンク Pepper など) も一般販売され、今後広く一般家庭に普及することが見込まれる。しかし、現在一般に利用されている音声対話システムは一般ユーザが期待する高度な対話を実現することはできない。人間同士の対話のような自然でスムーズな対話の実現には程遠いのが現状である。スムーズな対話が行えない一つの要因としては、対話のテンポ・応答タイミングが考慮されていないことが原因であると考えられる。現在一般的に採用されているクライアント・サーバ型のシステム形態では、適切な応答タイミングに間に合わないほど、サーバとの通信に時間が掛かってしまうため、この問題への対処は難しい。応答タイミングに関する知見では、音声対話システムとの対話にてシステムの応答時間 (ポーズ長) が 2 秒を超えてしまうとそれだけで対話が破綻してしまうと言われており [①, ②], 応答に 4 秒以上 [③] かかる現状のクライアント・サーバ型の音声対話システムでは、テンポ・応答タイミングの良い対話の実現は不可能である。本研究では、近年の音声対話システムで軽視されがちな「対話のテンポ」を改善し、そこからスムーズな対話を行うことができる音声対話システムの実現を目指す。我々が扱うテンポとは、具体的には、応答タイミング (適切な間「ま」) や音の大きさ、高さ、速さなどの韻律情報のことを指す。

また、一般的な音声対話システムは、音声認識・言語理解・対話管理・応答生成・音声合成の 5 つのモジュールから構成されるが、それぞれが個別に研究・実装されていることから、モジュール間で情報を相互に共有することが困難である。この問題を解決しつつ音声対話システムを開発するためには、幅広い知識が必要であり (音声, 言語, 対話, これらを統合実装するソフトウェア技術ノウハウ等) 各分野の専門家が片手間で構築できるものではない。

2. 研究の目的

本研究では、人間同士の対話と同じような、スムーズな対話を実現するためには、何が必要かを明らかにする。また、申請者がこれまで様々な視点から音声対話システムを研究・開発して得られた知見を活かし、また近年劇的に発展している音声言語情報処理、機械学習技術を取り入れながら、テンポの良い対話が可能な音声対話システムの実現を目指す。

3. 研究の方法

本研究では、(1) テンポ制御可能な音声対話システムの構築 (2) 対話のテンポ制御のための相槌タイミング予測モデルの構築 (3) 研究成果の公開、の 3 つのテーマについて研究を実施した。以下にこれらの 3 つのテーマごとに詳細を述べる。

(1) テンポ制御可能な音声対話システムの構築

テンポの評価に際し、システムの高精度化が必要であるため、以下の要素の研究を行った。

音声認識では、深層学習を用いて、くだけた話し方にも対応できる音声認識モデルを構築した。また、リアルタイムの音声認識結果出力に対応させた。

応答生成では、自然な対話の事例を元にユーザ入力からシステム応答を生成するモデルを構築し、リアルタイムの応答準備・出力に対応させた。

各モジュールの統合では、ベースとなる音声対話システム構築・統合のための仕様策定、各モジュールの実装、リアルタイム動作・制御への対応を行い、これらを統合し、リアルタイム音声対話システムを構築した。

①リアルタイム制御可能な音声対話システム

システムがユーザの発話を遮ったり、ユーザの発話中にシステムが相槌を発話したりするような活発な対話を実現するためには、リアルタイムに応答を生成し、その応答を発言するタイミングを判断できる音声対話システムを構築する必要がある。現在普及しているシステムの多くはポーズ駆動型のシステムが多く、リアルタイム応答が可能なシステムとは設計方針が異なり、そのままでは活発な対話の実現は不可能である。より自然な対話を実現するためにはシステムの基本設計を変更し、再実装する必要がある。

我々は以前、ロボット用の音声対話システムを構築する機会があり、ロボット環境との親和性から ROS (Robot Operating System) を音声対話システムに導入した [④]。ロボット向けでない音声対話システムであっても、ROS を採用することで、開発・デバッグの効率化、各モジュールの再利用性の向上、他システムとの連携が可能となるため、音声対話システム開発において非常に有利である。

本研究で構築した音声対話システムでは、構成するモジュールが独立しており、それぞれが並列に動作している。それらをリアルタイムに制御し、モジュール間の通信を円滑に行わせるために ROS アーキテクチャを導入した。また、対話内容の表示や音声分析結果をグラフィカルに表示するためのダッシュボードを構築するとともに ROS に元々用意されている各種 GUI ツールも用いる。

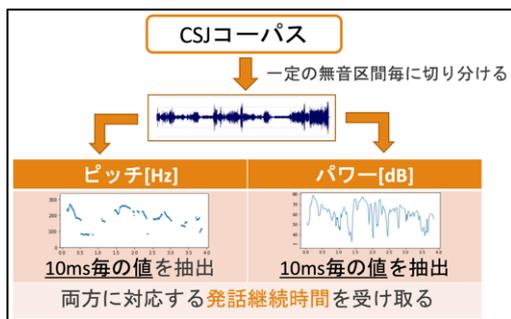


図 1：特微量抽出

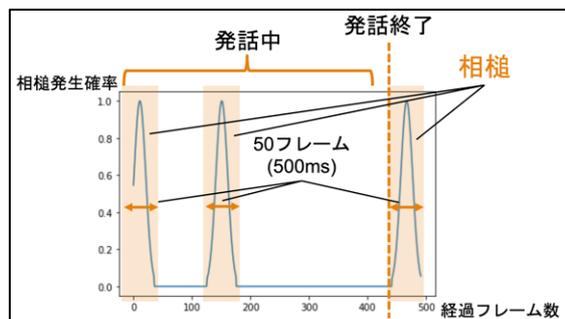


図 2：相槌確率の付与方法

(2) 対話のテンポ制御のための相槌タイミング予測モデルの構築

本研究では、リアルタイム下においてリズムカルで自然な相槌応答が可能な音声対話システムの実現に向けて、少ない特微量でも高い精度を誇る相槌タイミング予測モデルを提案する。モデルの学習・構築には深層学習技術を用いた。

先行研究では学習に用いる特微量として「ピッチ、パワー、発話開始からの経過時間」がタイミング予測において効果的であることが示されているので、本研究でもこれらの特微量を採用した。提案するモデルでは、長期的な時系列データの扱いに適している LSTM (Long Short-Term Memory) を用いてモデルを構築した。また、既存の音声対話コーパスに対して、相槌の定義を明確にしたうえで手作業にて相槌ラベリングを行った。

①モデル

本研究では深層学習モデルを構築して相槌タイミングの予測を行う。その際に使用するネットワークモデルとして LSTM を採用した。リアルタイム下で動作する音声対話システムに搭載するモデルとするため、計算負荷が軽く、さらに可能な限り長い期間で発話履歴を参照できるモデル構成が必要となる。LSTM であれば、計算不可を上げること無く相槌の予測の精度向上がより期待できるため採用した。

②対話コーパス

モデル学習に用いる音声データは、『日本語話し言葉コーパス (Corpus of Spontaneous Japanese:CSJ)』における対話音声データを使用した。各話者の音声に対して、一定の無音区間ごとに音声を切り分けたものを各ユーザの 1 発話分の音声とみなし、学習に使用する。切り分けた後の音声の長さは、すべて 4 秒から 10 秒の範囲に収まっていた。

③用いる特微量

学習に用いる特微量は、各発話音声に対して抽出した「ピッチ、パワー、発話開始からの経過時間」である。これらの値は、10 ms シフトで算出された値を用いる (図 1)。モデルには、これらの 3 つ組の値が逐次的に入力されていく。

④正解データへのラベリング規則

学習に用いるコーパスには、追加情報として相槌の有無の情報を付与した。付与する情報は確率値とし、相槌なしを 0、相槌ありを 1 とし、その間の実数値をとるラベリングデータとした。ラベル付与の例を図 2 に示す。この図の「発話中」がユーザの発話開始から発話終了の時間を表しており、ユーザの発話中にも相槌が打たれ得る範囲が 2 つ設定されている。そして、発話終了後にも 1 つ相槌が想定される箇所が設定されており、それぞれについて、正規分布のような傾きで相槌の確率値を各フレームに付与している。

⑤モデルの構造

モデルは入力が 128 ユニットの LSTM と、LSTM からの出力を 20%無効化する Dropout 層、そして 32 ユニットの全結合層で構成されている。Dropout 層はモデルの過学習を抑制するために追加した。モデルの学習については、損失関数は最小二乗誤差を使用し、最適化手法は Adam を用いて学習を行った。バッチサイズは 256、学習率は 0.001、エポック数は 100 に設定した。

(3) 研究成果の公開

本研究で得られた結果は論文として公開する。また、提案システムは、オープンソースとして公開する。音声対話システムの各モジュールのプログラムを github にて公開するとともに、合わせてテンポ制御モデルの公開を行い、対話テンポを考慮した音声対話システムの研究・開発・利用が行われやすい環境を提供する。

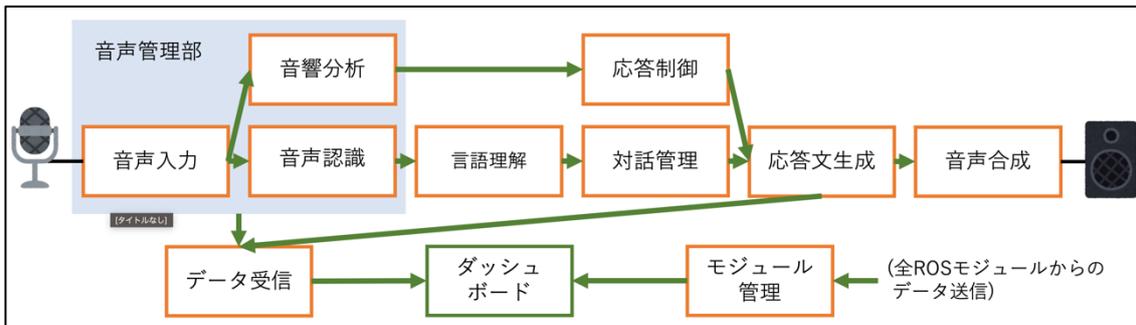


図 3：提案システムの概略図

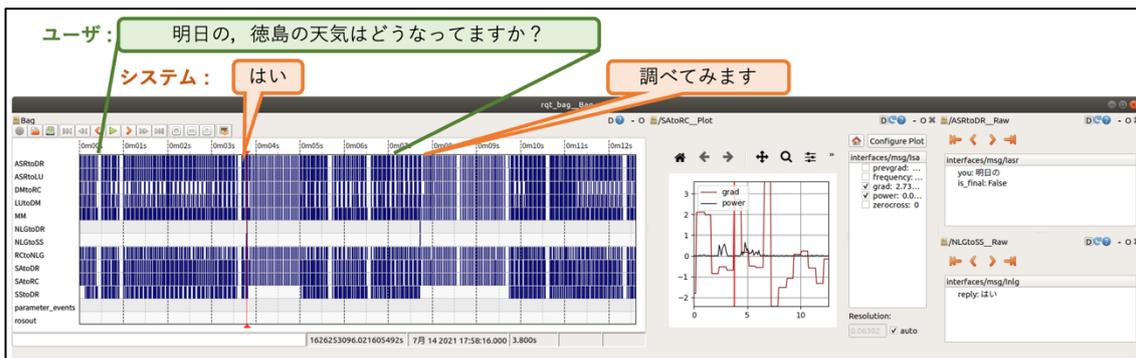


図 4：オーバーラップを含む対話の実験結果の分析（ROS: rqt_bag）

4. 研究成果

(1) テンポ制御可能な音声対話システムの構築

本研究で構築した音声対話システムの概略図を図 3 に示す。リアルタイム制御に対応し、すべてのモジュールが独立かつ並列に動作し、各モジュール間の入出力がどのモジュールからも参照できるシステムを構築することを目指した。提案システムは、音声対話システムの 5 つの基本モジュール(音声認識、言語理解、対話管理、応答生成、音声合成)に加え、音響分析、応答制御のモジュールを搭載している。音響分析モジュールは、入力音声进行分析して、パワーと Fo(基本周波数)と、それらの傾きを計算する。応答制御モジュールは、対話管理モジュールと音響分析モジュールからの出力を受けて、応答を生成するタイミングを決定する。

①音響分析：提案システムには、応答タイミングを決定するための特徴量を算出する音響解析モジュールを搭載した。音響分析には、Aubio ライブラリと、NumPy を使用した。ストリーミング音声認識では、マイクから音声を連続的に取得し、音声認識サーバに送信する。これと同時に、この音声データを音響解析モジュールに送信しリアルタイムに処理することで、音声の音響的な特徴を取得している。

②応答制御：応答制御モジュールは、得られた音響特徴量に基づき応答タイミングを決定する。従来のポーズ駆動型音声対話システムでは、ポーズを検出した後や音声認識結果を得た後に応答生成・合成を行っていたため、ユーザーの発話に対してオーバーラップしたり割り込んだりする応答は実現できなかった。しかし、提案システムでは、音響解析の結果、相槌応答を返すタイミングであると判断されれば、データ入力中であっても、任意のタイミングで相槌応答を返すことが可能である。提案システムで用いられている応答制御モデルは、申請者らがヒューリスティックに算出した閾値を用いて条件判断を行うシンプルなものである。

③音声認識：ここでは、マイクからサーバに送られたストリーミング音声をリアルタイムに逐次処理する。これにより、ユーザーが発話中であっても音声認識結果(途中結果)を得ることが可能となる。また、音声入力と同時に音声認識を行うため、発話終了後に音声処理・認識を開始する場合に比べ、発話終了から音声認識完了までの時間が短くなる。

④対話管理：ここでは、システムの応答行動が生成される。ChatGPT などの LLM (Large Language Model) を組み込むことも可能である。また、LLM を利用できないユーザー用に“A3RT Talk API”を利用可能にした。Talk API はチャットボットの作成時に使用され、チャットボットとユーザー間の対話を管理する。内部モデルでは、LSTM を用いて入力文から応答文を生成している。

⑤言語理解モジュール、応答生成モジュール：本システムの言語理解モジュールと応答生成モジュールは、対話管理モジュール内の A3RT Talk API が直接自然言語文を入出力しているため、当該モジュール内を自然言語文がそのまま通るだけであり、処理は行われていない。

⑥音声合成：音声合成は、応答生成の出力である応答文を入力とし、それをスピーカーから再生して出力している。

表 1：相槌予測の混同行列
 閾値:0.500 F 値:0.933

	予測：0	予測：1
正解：0	105650	73
正解：1	196	1881

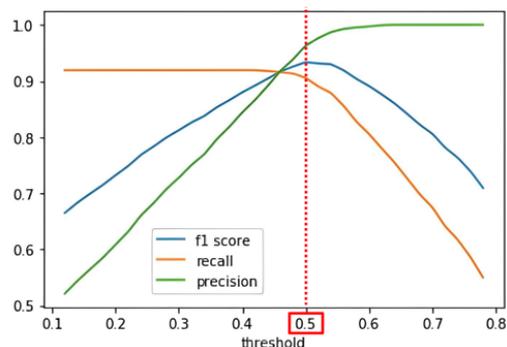


図 5：再現率，適合率，F 値のグラフ

⑦オーバーラップ対話を含む対話の実験と結果の分析

実際に音声対話システム全体を動作させ、ユーザ発話に対して途中でシステムからあいづちを打ち、ユーザ発話終了後に応答を返している例を図 4 に示す。発話時のシステム上に流れているデータは、rosviz にて記録し、図 4 に示すように、rqt_bag コマンドにて GUI 表示を行い、タイミングなどの確認を行った。

図の例では、ユーザが「明日の、徳島の天気はどうなってますか？」とシステムに問いかけ、システムが途中で「はい」とあいづちを打ち、その後に「調べてみます」と応答している。このように、ユーザの発話途中であっても、システムが応答可能である。

通常の一問一答の音声対話システムでは、ポーズ駆動により動作するものがほとんどであり、図に示すようにユーザの発話中に応答を返すことは不可能である。しかし提案システムでは、あいづちを含むシステム応答のタイミングを逐次的に決定しており、画一的な仕組みで、全体を制御し、応答生成することが可能となっている。

また、ROS アーキテクチャを導入したことにより、rqt_bag などの利便性の高い既存ツールを動作確認やデバッグなどに活用することができ、研究・開発を促進させることが可能となった。

(2) 対話のテンポ制御のための相槌タイミング予測モデルの構築

評価実験として閾値ごとにモデルの予測精度を計算した結果、閾値 0.50 のときに再現率 90.5%、適合率 96.2%、F 値 0.933 となり、提案モデルにおける最適な閾値が判明した (表 1, 図 5)。また、比較的単純な素性を入力値として用い、またモデル構成も単純な LSTM モデルを用いたが、高い判定精度を達成することができた。よって、本研究全体の主目的である、「リアルタイムに動作し、テンポ感の良い対話を実現する」ために必須となる、低計算コストにて動作する相槌タイミング予測モデルを構築できた。

(3) 研究成果の公開

①プログラムの公開

構築した音声対話システムは、Github の DiaROS リポジトリ (<https://github.com/kitaoka-lab/DiaROS>) にて公開した。ページ内の README.md を参照し、環境設定・インストールを行うことで、本研究の成果であるリアルタイム音声対話システムを簡単に導入することができる。また、音声対話システムと合わせて、相槌タイミング予測モデルも公開準備をしており追って公開予定である。

②ロボット実装

本研究成果の応用として、展示物案内音声対話ロボットの構築を行った [5]。

<引用文献>

- ① 志和敏之ら、「対話ロボットの反応時間と反応遅延時における間投詞の効果」、日本ロボット学会誌, Vol. 27, No. 1, pp. 87-95, 2009.
- ② 伊藤敏彦ら、「間投詞の働きの分析とシステム応答生成における間投詞の利用と評価」、日本音響学会誌, Vol. 55, No. 5, pp. 333-342, 1999.
- ③ 山本大介, 大浦圭一郎, 西村良太, 打矢隆弘, 内匠逸, 李晃伸, 徳田恵一, 「スマートフォン単体で動作する音声対話 3D エージェント「スマートメイちゃん」の開発」, 情報処理学会インタラクティブ 2013, pp. 675-680, 2013.
- ④ R. Nishimura, Y. Takase, Y. Nakano, 「Development environment of a spoken dialogue system based on PRINTEPS」, in 2016 IEEE 5th Global Conference on Consumer Electronics, pp. 1-2, 2016.
- ⑤ <https://www.jm.yamanashi.ac.jp/?p=2357>, 「案内ロボット『さとっちゃん』が甲府市の広報誌で紹介されました」, 山梨大学, 2022/01/28.

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 5件 / うち国際共著 0件 / うちオープンアクセス 5件）

1. 著者名 Nishimura Ryota, Mori Raita, Ohta Kengo, Kitaoka Norihide	4. 巻 37
2. 論文標題 Input Utterance Complementation Method by Anaphora Resolution for Spontaneous Utterances on Spoken Dialog Systems	5. 発行年 2022年
3. 雑誌名 Transactions of the Japanese Society for Artificial Intelligence	6. 最初と最後の頁 IDS~F_1-13
掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.37-3_IDS-F	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 -
1. 著者名 北岡 教英、西村 良太、太田 健吾	4. 巻 78
2. 論文標題 フォトリアルCGエージェントとのマルチモーダル対話	5. 発行年 2022年
3. 雑誌名 日本音響学会誌	6. 最初と最後の頁 257 ~ 264
掲載論文のDOI (デジタルオブジェクト識別子) 10.20697/jasj.78.5_257	査読の有無 無
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 -
1. 著者名 Fukuda Meiko, Nishimura Ryota, Nishizaki Hiromitsu, Horii Koharu, Iribe Yurie, Yamamoto Kazumasa, Kitaoka Norihide	4. 巻 77
2. 論文標題 A new speech corpus of super-elderly Japanese for acoustic modeling	5. 発行年 2023年
3. 雑誌名 Computer Speech & Language	6. 最初と最後の頁 101424 ~ 101424
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csl.2022.101424	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 -
1. 著者名 Ohta Kengo, Nishimura Ryota, Kitaoka Norihide	4. 巻 133
2. 論文標題 Response type selection for chat-like spoken dialog systems based on LSTM and multi-task learning	5. 発行年 2021年
3. 雑誌名 Speech Communication	6. 最初と最後の頁 23 ~ 30
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.specom.2021.07.003	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 -

1. 著者名 KITAOKA Norihide, SETO Eichi, NISHIMURA Ryota	4. 巻 E103.D
2. 論文標題 Example Phrase Adaptation Method for Customized, Example-Based Dialog System Using User Data and Distributed Word Representations	5. 発行年 2020年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 2332 ~ 2339
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2020EDP7066	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Chen Jiahao, Nishimura Ryota, Kitaoka Norihide	4. 巻 9
2. 論文標題 End-to-end recognition of streaming Japanese speech using CTC and local attention	5. 発行年 2020年
3. 雑誌名 APSIPA Transactions on Signal and Information Processing	6. 最初と最後の頁 1 ~ 7
掲載論文のDOI (デジタルオブジェクト識別子) 10.1017/ATSIP.2020.23	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21件 (うち招待講演 0件 / うち国際学会 4件)

1. 発表者名 Fukuda, M., Sugiyama, M., Nishimura, R., Kitaoka, N
2. 発表標題 A Corpus-based Analysis of Age-related Changes in the Acoustic Features of Elderly to Super Elderly Speech
3. 学会等名 The 0-COCOSDA 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 大須賀晋, 田中五大, 鍋倉彩那, 中野涼太, 渡邊凌太, 石川友香, 石川晃之, 中村晃一, 藤井裕也, 堀内颯太, 東中竜一郎, 西村良太, 太田健吾, 北岡教英
2. 発表標題 EMOtive A.I. "Saya"
3. 学会等名 人工知能学会研究会資料 言語・音声理解と対話処理研究会
4. 発表年 2022年

1. 発表者名 福村考洋, 西村良太, 北岡教英
2. 発表標題 BERT による雑談対話話題拡張
3. 学会等名 令和4年度 電気・電子・情報関係学会 四国支部連合大会 講演論文集 (Web会議)
4. 発表年 2022年

1. 発表者名 鳥井章宏, 西村良太, 北岡教英
2. 発表標題 音声対話システムにおける対話破綻検出器の構築
3. 学会等名 令和4年度 電気・電子・情報関係学会 四国支部連合大会 講演論文集 (Web会議)
4. 発表年 2022年

1. 発表者名 福田芽衣子, 杉山雅和, 西村良太, 入部百合絵, 山本一公, 北岡教英
2. 発表標題 超高齢者コーパスとS-JNAS を用いた高齢者音声の音響的特徴の分析
3. 学会等名 日本音響学会講演集 (2022年秋) (Web会議)
4. 発表年 2022年

1. 発表者名 森大輝, 太田健吾, 西村良太, 北岡教英
2. 発表標題 ドメイン外音響情報で補強した Encoder-Decoder 音声認識モデルの設計
3. 学会等名 日本音響学会講演集 (2022年秋) (Web会議)
4. 発表年 2022年

1. 発表者名 堀井こはる, 福田芽衣子, 太田健吾, 西村良太, 小川厚徳, 北岡教英
2. 発表標題 End-to-End 非流暢整形音声認識システムの対話音声による評価
3. 学会等名 日本音響学会講論集(2022年秋)(Web会議)
4. 発表年 2022年

1. 発表者名 杉山雅和, 太田健吾, 西村良太, 北岡教英
2. 発表標題 割り込み発話にも対応可能なリアルタイム話者交替システム
3. 学会等名 日本音響学会講論集(2023春)(Web会議)
4. 発表年 2023年

1. 発表者名 伊達龍斗, 太田健吾, 西村良太, 北岡教英
2. 発表標題 深層学習による口唇情報を用いた雑音下での音声認識精度の改善
3. 学会等名 日本音響学会講論集(2023春)(Web会議)
4. 発表年 2023年

1. 発表者名 堀井こはる, 太田健吾, 西村良太, 小川厚徳, 北岡教英
2. 発表標題 自発的発話認識のためのBERTによる非流暢文生成に基づく言語モデリング
3. 学会等名 日本音響学会講論集(2023春)(Web会議)
4. 発表年 2023年

1. 発表者名 Mori Daiki、Ohta Kengo、Nishimura Ryota、Ogawa Atsunori、Kitaoka Norihide
2. 発表標題 Advanced language model fusion method for encoder-decoder model in Japanese speech recognition
3. 学会等名 APSIPA Annual Summit and Conference 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Horii Koharu、Fukuda Meiko、Ohta Kengo、Nishimura Ryota、Ogawa Atsunori、Kitaoka Norihide
2. 発表標題 End-to-End Spontaneous Speech Recognition Using Hesitation Labeling
3. 学会等名 APSIPA Annual Summit and Conference 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 西村 良太、森 貴大、北岡 教英
2. 発表標題 ROSを利用したリアルタイム制御が可能な音声対話システムの構築
3. 学会等名 日本音響学会研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 森 大輝、太田 健吾、西村 良太、小川 厚徳、北岡 教英
2. 発表標題 タスク外音響情報を付加した End-to-End 音声認識モデルの設計
3. 学会等名 日本音響学会講演論文集
4. 発表年 2021年

1. 発表者名 堀井 こはる、福田 芽衣子、太田 健吾、西村 良太、小川 厚徳、北岡 教英
2. 発表標題 非流暢ラベルを用いた言い淀み整形End-to-End音声認識
3. 学会等名 日本音響学会講演論文集
4. 発表年 2021年

1. 発表者名 Chen Jiahao、Nishimura Ryota、Kitaoka Norihide
2. 発表標題 E2E Streaming Speech Recognition Using CTC and Local Attention
3. 学会等名 Nonlinear Circuits, Communications and Signal Processing 2020 (NCSP'20) (国際学会)
4. 発表年 2020年

1. 発表者名 太田健吾、西村良太、北岡教英
2. 発表標題 アクセント句を考慮した日本語End-to-End音声合成サーバの構築
3. 学会等名 日本音響学会講演論集(2021年春)
4. 発表年 2021年

1. 発表者名 小橋優矢、西村良太、北岡教英
2. 発表標題 Sequence-to-Sequence modelを用いた話し言葉音声認識用言語モデルのための書き言葉から話し言葉へのテキスト変換
3. 学会等名 日本音響学会
4. 発表年 2019年

1. 発表者名 森雷太, 西村良太, 北岡教英
2. 発表標題 自由発話に対応した照応解析を備えた音声対話システム
3. 学会等名 日本音響学会
4. 発表年 2020年

1. 発表者名 小橋優矢, 西村良太, 北岡教英
2. 発表標題 書き言葉から話し言葉へのテキスト変換を用いた話し言葉音声認識用言語モデルの評価
3. 学会等名 日本音響学会
4. 発表年 2020年

1. 発表者名 陳家浩, 西村良太, 北岡教英
2. 発表標題 Uni-directional LSTMとLocal Attentionを用いたストリーミング音声認識
3. 学会等名 日本音響学会
4. 発表年 2020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
研究 分担者	山本 一公 (YAMAMOTO Kazumasa) (40324230)	中部大学・工学部・教授 (33910)	

6. 研究組織（つづき）

	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
研究 分 担 者	西崎 博光 (NISHIZAKI Hiromitsu) (40362082)	山梨大学・大学院総合研究部・教授 (13501)	

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