

令和 4 年 6 月 9 日現在

機関番号：32689

研究種目：基盤研究(A)（一般）

研究期間：2018～2020

課題番号：18H04128

研究課題名（和文）音声メディアを利用した情報伝達における相互行為の時間構造的特徴と伝達効率の関係

研究課題名（英文）Temporal structural modeling of conversational interaction and efficient information transfer using speech media.

研究代表者

小林 哲則（Kobayashi, Tetsunori）

早稲田大学・理工学術院・教授

研究者番号：30162001

交付決定額（研究期間全体）：（直接経費） 34,400,000円

研究成果の概要（和文）：音声メディアでまとまった情報を効率的に伝達するにあたっては、情報伝達に会話的要素を取り入れた上で、相互行為のリズムを保証することが重要である。本研究では、文書を要約・断片化して順次伝える傍らで、随時ユーザの反応をモニタし、反応に応じて要約時に削減した情報を復元・提示することに特徴を持つ、会話型情報伝達システムをベースシステムとして採用したうえで、これに会話相互行為の時間構造に関する制約をモデル化して組み込んだ。また、システムの重要な要素技術として、低遅延音声認識技術、表現豊かな音声合成技術、パラ言語理解技術を取り上げ、独自性の高い高性能な技術を実現した。

研究成果の学術的意義や社会的意義

音声は利用負荷が小さいメディアであり、これを利用した情報アクセス技術確立することが古くから望まれてきたが、従来システムが扱うものはたかだか数文の短い文章に限られていた。大量の情報を音声メディアで円滑に伝える場合、情報を伝える傍らで適宜質問を受け、これに回答するといった処理を、リズム良くこなす必要があるが、これまで会話のリズムに関する研究は皆無であった。本研究によって、音声メディアにおける利便性の高い情報伝達の基礎が初めて築かれた。また、従来、情報検索・質疑応答（Pull）中心に進められてきた情報行動研究に、情報提供（Push）と取得（Pull）の密な融合という新たな視点を与えることができた。

研究成果の概要（英文）： In order to efficiently convey mass information via voice media, it is important to incorporate conversational elements into the information transmission and to guarantee the rhythm of the interaction.

Here, we have modeled the constraints on the temporal structure of conversational interaction that form the basis for realizing rhythmic conversation and incorporated the model into our information delivery system. The system has the ability to monitor the user's response at any time while delivering a summarized document, and to restore and present information that was reduced during summarization in response to the user's response. These features achieved efficient document transmission through spoken conversation.

In addition, as important elemental technologies of the system, low-latency speech recognition technology, expressive speech synthesis technology, and paralinguistic understanding technology were developed to enhance the performance of the system.

研究分野：知覚情報システム学

キーワード：音声会話システム 低遅延音声認識 表現豊かな音声合成 パラ言語理解 発話タイミング推定

1. 研究開始当初の背景

音声は利用負荷が小さいメディアであり、これを利用した情報アクセス技術を確立することが古くから望まれてきた。長年の研究成果により、Apple Siri、Amazon Alexaなどが実用化されたが、それが伝えるのはたかだか数文の短い文章に限られていた。一方、WEBでの文書検索を思い浮かべれば容易に想像できるように、少なくとも数段落から構成される文書の情報を得たいとするニーズは大きく、機能の拡張が求められていた。

こうした大量の情報を音声メディアで伝える場合、音声の持つ揮発性・同時性という性質（送った情報は瞬時に消えるため、伝える行為と受け取る行為を同時に行う必要がある）が問題となる。すなわち、受け手が一度に受理可能な情報の量には限りがあるため、伝え手はこの限界を考慮しながら、効率的に伝える工夫が必要となる。筆者らは、当時、WEB記事を要約して主要部を抜き出し、これを断片化して順次伝える傍らで、随時ユーザの反応をモニタし、反応に応じて要約時に削減した情報を復元して提示する枠組みを実装することによって、ユーザにとって有用な情報を過不足なく伝えることを試みていた。しかし、予備実験の結果、提案する枠組みによって、システムとユーザのやり取りがリズムよく噛み合っているときには効率の良い情報伝達が促進されるが、一旦リズムが崩れるとユーザからの反応は乏しくなり、情報復元する機能が働くなくなって、十分な情報伝達ができなくなることが判明した。すなわち、音声メディアでまとまった情報を効率的に伝達するにあたっては、相互行為のリズムを保證することが重要な鍵を握ることが示唆されていた。

2. 研究の目的

音声会話における情報伝達の効率と会話のリズムとは深い相関があるという仮説のもとで、リズムの良い会話を実現する枠組みを実現する。システムとユーザが行う相互行為の時間構造に関する制約（相互行為を構成する基本行為の遷移パターンと、その遷移時間に関する制約）をモデル化し、これを会話システムに組み込むことを試みる。より具体的には、ウェブ記事から動的に作成されるシナリオに基づいて会話を進めることに特徴を持つ、即時応答性に優れた音声会話型情報伝達システムをベースシステムとして採用した上で、これに対話行為の時間制御機能を組込む。

また、低遅延音声認識技術、表現豊かな音声合成技術、パラ言語理解技術などを開発し、ベースシステムの基本機能を強化する。これらによって、「送受信の同時性」と「揮発性」を特徴とする音声メディアにおける利便性の高い情報伝達の基礎を築くとともに、従来、情報検索・質疑応答（Pull）中心に進められてきた情報行動研究に、情報提供（Push）と取得（Pull）の密な融合という新たな視点を与える。

3. 研究の方法

上記目的のために、(1) 会話の基本行為の遷移パターンとその時間制約のモデル化、(2) リズム形成のための会話システムの要素技術、(3) 会話のリズムを阻害する要因の分析、の3つのサブテーマを選定し、検討を行った。

(1) 会話の基本行為の遷移パターンとその時間制約のモデル化（小林・藤江担当）

会話のリズム形成の根源となる発話タイミングの決定の問題を扱うもので、本研究の要となる。観測される音響、画像情報を入力として、システムがどのタイミングで発話をするべきかを、DNNを用いてモデル化する。特に、発話タイミングと対話行為の系列との関連性に注目した上で、対話行為の推定を補助タスクとして組み込むことの効果を調べる。

(2) リズム形成のための会話システムの要素技術

低遅延音声認識技術（小林担当）

会話のリズム形成に必須となる、即応性の向上について扱う。トリガアテンション型のストリーミング音声認識を採用した上で、エンコーダ部にMask-CTCの事前学習を導入する。先行文脈情報に鋭敏な特徴抽出を可能にすることで、低遅延とするために先読み時間を短くしたときでも、高い精度を保つことを試みる。

表現豊かな音声合成技術（徳田担当）

会話のリズム形成の阻害要因として知られるシステム発話への違和感の軽減を、特に発話表現の豊かさに注目して行う。センテンスコードと感情コードを用いて発話表現を自動分類しながら音声合成用モデルを学習する手法に、Variational Auto-Encoder構造、感情コードのアンニリング、因子分析構造などを導入することを試みる。

パラ言語理解技術（小林・藤江担当）

簡素な表現で文脈依存の意図を表現する会話独特の言語現象を扱う。基本周期抽出処理を陽に行うことなく韻律的特徴を捉え、意図抽出する方法を提案するとともに、実際に会話システムを利用したとき得られる話者の短い反応を集めた大規模なデータベースを収集して、このモデルを学習することを試みる。

(3) 会話のリズムを阻害する要因の分析（森担当）

人同士の対話にのみ存在する話し方の特徴を人為的に統制した条件で、会話を収録し、聞き手反応の出現傾向への影響を分析する。

4. 研究成果

(1) 会話の基本行為の遷移パターンと、その時間制約のモデル化

対話行為の系列を考慮しながら、システムの発話タイミングを決定するしくみについて検討した。会話参加者の音声と顔画像データを入力として、音響データの分析フレームに同期して、その時刻にシステムが発話すべきか否かを判定する仕組みになっている。提案するアーキテクチャを図1に示す。

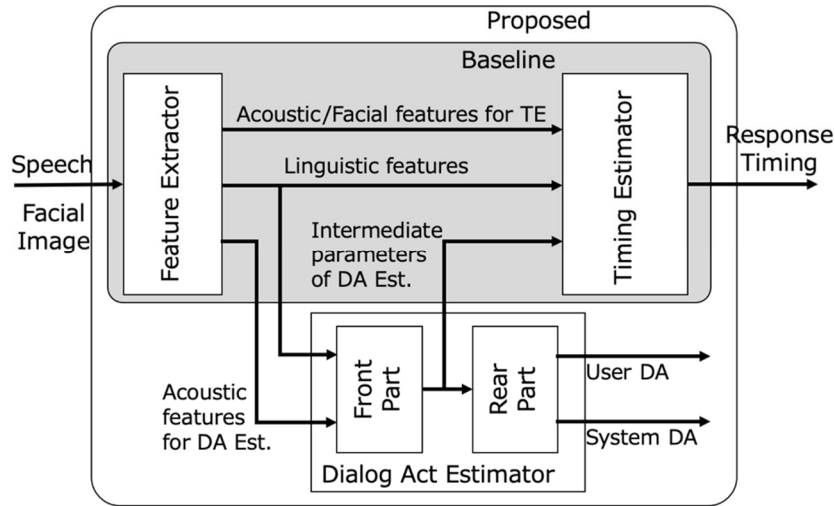


図1. 発話タイミング推定器の構造

重要な点は、補助タスクとして直前のユーザ発話と、それに続くシステム発話の対話行為の推定が組み込むことによってまれていること、Timing EstimatorはTransformerを用いた自己アテンション機構によって実現できることが挙げられる。これらの構造によって、対話行為のシーケンスを考慮して発話タイミングが決定できることが期待できる。

学習・評価用のデータベースとしては、様々なものを試したが、ここでは HARPERVALLEYBANK corpus という、人同士の疑似会話コーパスを用いたときの結果を示す。このデータベースは、59人の話者による約23時間分のデータで、画像データはない。

結果を表1に示す。表中、Baselineは対話行為を利用しない場合であり、システム構成としては図1における色づけした部分にあたる。評価は、 T ms Precision (P_{50})、 T ms Recall (R_{50})、 T ms F値 (F_{50}) による。ここで T ms F値等は、テストデータとのタイミングの誤差が T ms 以内のとき、正解とみなしてF値等を計算する方法である。

表1. 発話タイミング推定における、言語情報利用の効果.

	T=25ms			T=50ms		
	Precision	Recall	F	Precision	Recall	F
Baseline	0.351	0.367	0.359	0.609	0.637	0.623
Proposed (対話行為利用)	0.407	0.421	0.414	0.664	0.642	0.632
対話行為の正解を利用 (参考)	0.410	0.413	0.411	0.627	0.669	0.640

表1.から、対話行為を補助タスクとして組み込んだ提案手法の優位性が確認できた。対話行為の系列を考慮することの効果は正解の判定基準を厳密化したときに、より明確であること、および、対話行為の正解を用いた場合に大凡匹敵する結果となっていることも確認できた。

(2) リズム形成のための会話システムの要素技術

低遅延音声認識技術

Mask-CTC とトリガー型アテンション機構を組み合わせ、低遅延で高性能な音声認識 (ASR) システムを構築することを試みた。

CTC スパイクをトリガーとして自己回帰復号を行うトリガー型アテンション機構は、ストリーミング ASR において有効である。しかし、性能の要である CTC 出力に基づくアライメント推定精度を高く保つためには、長い先読みが必要でありその分レイテンシが大きくなる。この問題を回避するために、Mask-CTC を用いて事前学習した特徴抽出部をシステムに組み込むことを試みた。

Mask-CTC は、筆者らのグループが、ストリーミング ASR とは別に提案したものであるが、学習時に部分的に情報を隠蔽 (Mask) することの効果によって、特徴表現が周囲情報の予測能力を持つことになり、ストリーミング ASR とも相性が良い。

WSJ データを用いた実験の結果を表 2 に示す。Mask-CTC を事前学習に用いた提案方式が、従来方式に比べ、すべてのレイテンシにおいて、高い性能を示した。特に、160ms という低レイテンシの場合には、単語誤り率で 6.9 ポイント減と、大きな効果があることを確認した。

表 2. Mask-CTC による事前学習の有無による、トリガ型アテンション・ストリーミング ASR の単語誤り率の比較。データは WSJ eval92。

	Latency [ms]			
	160	320	480	640
Baseline (CTC-attention)	28.2	22.4	18.9	17.0
Proposed (Enhanced CTC-attention)	21.3	15.5	14.3	14.1

表現豊かな音声合成技術

1) センテンスコードを用いた発話表現のモデル化

様々な発話表現を含む音声データから音声合成用モデルを学習する際に、センテンスコードを利用することで、発話表現を自動分類しながら音声合成用モデルを学習する方法を提案した。センテンスコードとは、発声毎に別々に与えられた乱数で、発話者が意識せずに行った発話毎の表現をモデル化するためのものである。センテンスコードを用いて発話表現を自動分類しながら音声合成用モデルを学習することによって、指定した発話表現を実現可能となり、従来法から合成音声の品質を劣化させることなく表現性を大幅に改善することが可能となった。

2) センテンスコードと感情コードを用いた発話表現のモデル化

様々な発話表現を含む音声データから音声合成用モデルを学習する際に、センテンスコードに加えて、感情コードを利用する手法を提案した。発話者が発声毎に指定された発話表現 (ここでは感情表現を例とした) に基づいて発声した場合には、その発話表現の情報はモデル学習に役立つはずであるが、指定された発話表現の中でも発話毎の表現の変動があり、それらを吸収しながら学習を行う必要がある。この問題を解決するため、センテンスコードと感情コードをあわせて用いることにより、発話表現を自動分類しながら音声合成用モデルを学習する手法を提案し、合成音声の表現性を改善することを示した。

3) 因子分析に基づいた隠れセミマルコフモデルの構造を利用した音声合成

隠れセミマルコフモデル (HSMM) の構造を導入した構造化アテンション機構に基づく音声合成を拡張し、音響特徴量・音素継続長・発話スタイルを同時モデリング可能な手法を提案した。合成音声の自然性に関する受聴試験に基づく評価結果から、2000 文弱の比較的少量の学習データでも高品質なシステムが構築できること、従来手法に対して高品質な合成音声生成されることを確認した。また、感情表現に関する主観評価試験においても、従来手法に比べ、はっきりとした感情表現が可能であることがわかった。

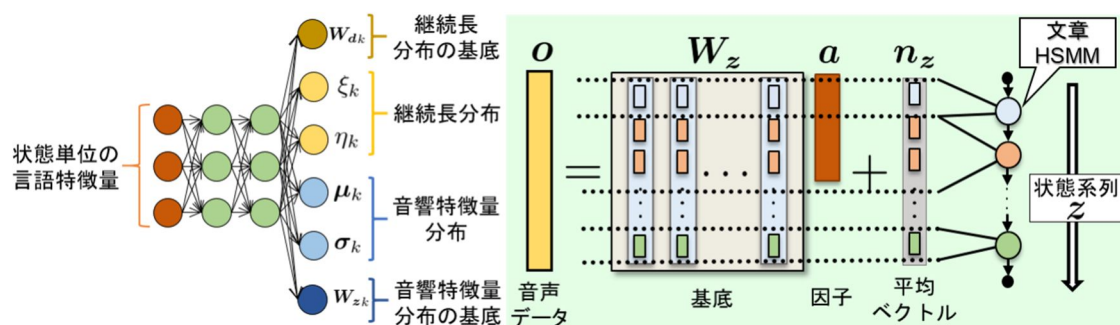


図 2. 隠れセミマルコフモデルに基づいたニューラルネットワーク (左) と因子分析構造 (右)。

4) 周期・非周期成分の分離に基づくニューラルボコーダによる音声波形のモデル化

明示的な周期信号と非周期信号を入力とし、並列型や直列型のモデル構造を用いることで、周期・非周期成分が自動的に分離される形で音声波形をモデル化できる手法を提案した。サンプリング周波数が 48kHz の音声波形を、実時間よりも高速に、自然音声と同等の品質で生成可能なことを確認するとともに、合成学習データの範囲外のピッチを持った音声波形の生成も可能であることを示した。また、受聴試験による評価試験の結果から、並列型のモデル構造を持ったボコーダは、事前に音声波形を周期・非周期成分に分離しモデル化した場合に対して、より適切に周期・非周期成分を分離しながら音声波形をモデル化していることが示された。

パラ言語理解技術

会話においては、一単語ないし二単語といった非常に短い文で構成される応答 (以下 短応答)

によって、会話進行上重要な情報を表現することが多く、これを正しく理解することがテンポの良い会話の実現に重要である。ここでは、実際の会話システムの利用時におけるユーザの短応答を収集し、その意図を自動的に解析する方法について検討した。

まず、2060 件の会話セッションを行い、そこで生じた 15604 件の短応答を収集し、このデータについて、12 人のアノテータが聴取実験を通じてラベル付与を行った。ここでラベルは、発話者が短応答を、どのような意図で行うか、あるいは会話進行上どのような役割を持つかについて予め検討・分類した結果を用いた。次に、出現頻度が高かった「疑問があること(Q)」、「興味がないこと(D)」、「情報補足が必要であること(R)」を表す短応答を選び、発話意図の認識実験を行った。認識器は、LSTM に基づくモデルを音声と顔画像情報を入力として構成し、これを作成したデータベースを用いて学習した。Q,D,R のそれぞれについて、97%、77%、76%の識別精度を得ることができた。

(3) 会話のリズムを阻害する要因の分析

人同士の対話にのみ存在する話し方の特徴を人為的に統制した条件での対話を収録し、聞き手反応の出現傾向への影響を分析した。以下の 3 条件で、システム役の人間(説明者)がネットニュース記事の内容をユーザ役の実験参加者(被説明者)に説明する実験を行った。

- HNVL 条件 説明者は記事の内容を暗記した上でアドリブで説明
- HVL 条件 説明者は HNVL 条件で収録した対話の書き起こしを元に作成した発話計画に従って説明
- HL 条件 説明者は人手で作成した発話計画に従って説明

実験は 26 名の被説明者を対象に被験者間計画(3 条件の 1 つを無作為に割り当て)で実施した。

1 記事の説明を通しての実験参加者の聞き手反応の回数の分布を図 3 に示す。HL 条件での聞き手反応の回数は平均 42.8 回であり、HNVL 条件での平均 115.5 回に比べ少なくなっている。すなわち、多くの音声対話システムと同様に人手で作成した発話計画に従って説明すると、ユーザは聞き手反応をしなくなる。一方、HVL 条件での聞き手反応の回数は平均 130.0 回であり、HNVL 条件との有意差はない。すなわち、アドリブで説明した際の書き起こしを発話テキストとして説明した場合には、人手で作成した発話テキストを用いる場合と異なり、ユーザの聞き手反応は減少しない。

この実験のフォローアップとして、HNVL 条件での説明者の発話を録音しておき、人間が発話するかわりにタブレットの操作により発話を再生することで、人同士の対話における発話の非決定的・動的な側面(L)を人為的に取り除いた実験も実施したが、そこでも自発音声の言語的特徴(V)の消失が聞き手反応の減少と関連するという一貫した結果が出ている。

以上の結果より、断片的な発話、文法の破格といった自発音声の言語的特徴を持つシステム発話がユーザの聞き手反応を引き出すために重要であることがわかった。

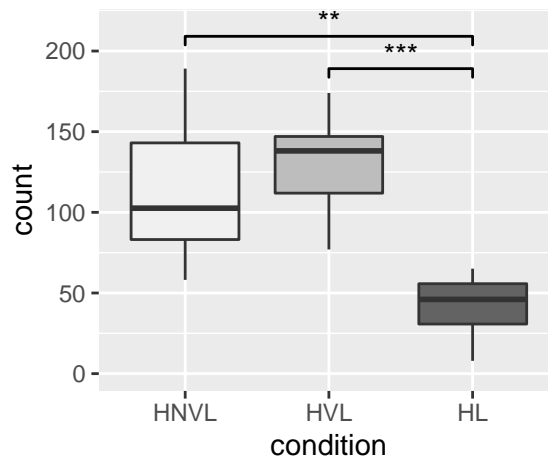


図 3. 説明者の話し方の違いによる聞き手反応の回数の変化。

5. 主な発表論文等

〔雑誌論文〕 計29件（うち査読付論文 28件 / うち国際共著 2件 / うちオープンアクセス 16件）

1. 著者名 Yukiya Hono, Shinji Takaki, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda	4. 巻 9
2. 論文標題 PeriodNet: A Non-Autoregressive Raw Waveform Generative Model With a Structure Separating Periodic and Aperiodic Components	5. 発行年 2021年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 137599-137612
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2021.3118033	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 森 大毅	4. 巻 78
2. 論文標題 対話システムはどのように話すべきか: 実際の会話データに基づく話し言葉の合成	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 283-288
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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 Takahisa Iizuka, Hiroki Mori	4. 巻 43
2. 論文標題 Comparison of machine learning algorithms and acoustic features in emotion recognition from spontaneous speech	5. 発行年 2022年
3. 雑誌名 Acoustical Science and Technology	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 Yosuke Higuchi, Naohiro Tawara, Tetsunori Kobayashi, Tetsuji Ogawa,	4. 巻 -
2. 論文標題 Speaker adversarial training of DPGMM-based feature extractor for zero-resource languages,	5. 発行年 2019年
3. 雑誌名 Proc. Interspeech 2019	6. 最初と最後の頁 266-270
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2019-2052	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 Hiroaki Takatsu, Katsuya Yokoyama, Yoichi Matsuyama, Hiroshi Honda, Shinya Fujie, Tetsunori Kobayash	4. 巻 -
2. 論文標題 Recognition of Intentions of Users' Short Responses for Conversational News Delivery System	5. 発行年 2019年
3. 雑誌名 Proc. Interspeech 2019	6. 最初と最後の頁 1193-1197
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2019-2121	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 Keiichiro Oura, Kazuhiro Nakamura, Kei Hashimoto, Yoshihiko Nankaku, and Keiichi Tokuda,	4. 巻 -
2. 論文標題 Deep neural network based real-time speech vocoder with periodic and aperiodic inputs	5. 発行年 2019年
3. 雑誌名 10th ISCA Speech Synthesis Workshop (SSW10)	6. 最初と最後の頁 13-18
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/SSW.2019-32	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 Takato Fujimoto, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda	4. 巻 -
2. 論文標題 Impacts of input linguistic feature representation on Japanese end-to-end speech synthesis	5. 発行年 2019年
3. 雑誌名 10th ISCA Speech Synthesis Workshop (SSW10)	6. 最初と最後の頁 166-171
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/SSW.2019-30	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 Motoki Shimada, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda	4. 巻 -
2. 論文標題 Low computational cost speech synthesis based on deep neural networks using hidden semi-Markov model structures	5. 発行年 2019年
3. 雑誌名 10th ISCA Speech Synthesis Workshop (SSW10)	6. 最初と最後の頁 177-182
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/SSW.2019-32	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 Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda	4. 巻 -
2. 論文標題 Speaker-dependent WaveNet-based delay-free ADPCM speech coding	5. 発行年 2019年
3. 雑誌名 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)	6. 最初と最後の頁 7145-7149
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICASSP.2019.8682264	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 Yukiya Hono, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda	4. 巻 -
2. 論文標題 Singing voice synthesis based on generative adversarial networks	5. 発行年 2019年
3. 雑誌名 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)	6. 最初と最後の頁 6955-6959
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICASSP.2019.8683154	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 Kazuki Kaya and Hiroki Mori	4. 巻 E102-D
2. 論文標題 Effectiveness of speech mode adaptation for improving dialogue speech synthesis	5. 発行年 2019年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 2064-2066
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2019EDL8024	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 Hiroki Mori, Tomohiro Nagata, and Yoshiko Arimoto	4. 巻 -
2. 論文標題 Conversational and social laughter synthesis with WaveNet	5. 発行年 2019年
3. 雑誌名 Proc. Interspeech 2019	6. 最初と最後の頁 520-523
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2019-2131	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 高津 弘明, 福岡 維新, 藤江 真也, 岩田和彦, 小林 哲則	4. 巻 34(2)
2. 論文標題 会話によるニュース記事伝達のための音声合成	5. 発行年 2019年
3. 雑誌名 人工知能学会論文誌	6. 最初と最後の頁 B-165 1-15
掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.B-165	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 岩田和彦, 小林哲則	4. 巻 J102-D(6)
2. 論文標題 対話音声合成の表現力向上に向けた文末音調の制御による付加的なニュアンスの表現に関する実験的検討	5. 発行年 2019年
3. 雑誌名 電子情報通信学会論文誌D	6. 最初と最後の頁 442-453
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2018JDP7055	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 Fujie Shinya, Katayama Hayato, Sakuma Jin, Kobayashi Tetsunori	4. 巻 -
2. 論文標題 Timing Generating Networks: Neural Network Based Precise Turn-Taking Timing Prediction in Multiparty Conversation	5. 発行年 2021年
3. 雑誌名 Interspeech 2021	6. 最初と最後の頁 3226-3230
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2021-874	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 Higuchi Yosuke, Inaguma Hirofumi, Watanabe Shinji, Ogawa Tetsuji, Kobayashi Tetsunori	4. 巻 -
2. 論文標題 Improved Mask-CTC for Non-Autoregressive End-to-End ASR	5. 発行年 2021年
3. 雑誌名 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	6. 最初と最後の頁 8363-8367
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICASSP39728.2021.9414198	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 該当する

1. 著者名 Takatsu Hiroaki, Okuda Mayu, Matsuyama Yoichi, Honda Hiroshi, Fujie Shinya, Kobayashi Tetsunori	4. 巻 -
2. 論文標題 Personalized Extractive Summarization for a News Dialogue System	5. 発行年 2021年
3. 雑誌名 2021 IEEE Spoken Language Technology Workshop	6. 最初と最後の頁 37-45
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/SLT48900.2021.9383568	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 Higuchi Yosuke, Watanabe Shinji, Chen Nanxin, Ogawa Tetsuji, Kobayashi Tetsunori	4. 巻 -
2. 論文標題 Mask CTC: Non-Autoregressive End-to-End ASR with CTC and Mask Predict	5. 発行年 2020年
3. 雑誌名 Interspeech 2020	6. 最初と最後の頁 3655-3659
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2020-2404	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 該当する

1. 著者名 Fujimoto Takato, Takaki Shinji, Hashimoto Kei, Oura Keiichiro, Nankaku Yoshihiko, Tokuda Keiichi	4. 巻 -
2. 論文標題 Semi-Supervised Learning Based on Hierarchical Generative Models for End-to-End Speech Synthesis	5. 発行年 2020年
3. 雑誌名 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	6. 最初と最後の頁 7644-7648
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICASSP40776.2020.9054466	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 Hono Yukiya, Tsuboi Kazuna, Sawada Kei, Hashimoto Kei, Oura Keiichiro, Nankaku Yoshihiko, Tokuda Keiichi	4. 巻 -
2. 論文標題 Hierarchical Multi-Grained Generative Model for Expressive Speech Synthesis	5. 発行年 2020年
3. 雑誌名 Interspeech 2020	6. 最初と最後の頁 3441-3445
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2020-2477	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 Fujimoto Takato, Takaki Shinji, Hashimoto Kei, Oura Keiichiro, Nankaku Yoshihiko, Tokuda Keiichi	4. 巻 -
2. 論文標題 Semi-Supervised Learning Based on Hierarchical Generative Models for End-to-End Speech Synthesis	5. 発行年 2020年
3. 雑誌名 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	6. 最初と最後の頁 7644 - 7648
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICASSP40776.2020.9054466	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 Yokoyama Katsuya, Takatsu Hiroaki, Honda Hiroshi, Fujie Shinya, Kobayashi Tetsunori	4. 巻 -
2. 論文標題 Investigation of Users' Short Responses in Actual Conversation System and Automatic Recognition of their Intentions	5. 発行年 2018年
3. 雑誌名 2018 IEEE Spoken Language Technology Workshop (SLT)	6. 最初と最後の頁 934-940
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/SLT.2018.8639523	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 Fujimoto Takato, Yoshimura Takenori, Hashimoto Kei, Oura Keiichiro, Nankaku Yoshihiko, Tokuda Keiichi	4. 巻 -
2. 論文標題 Speech Synthesis Using WaveNet Vocoder Based on Periodic/Aperiodic Decomposition	5. 発行年 2018年
3. 雑誌名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018)	6. 最初と最後の頁 644-648
掲載論文のDOI (デジタルオブジェクト識別子) 10.23919/APSIPA.2018.8659541	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 Nakao Kento, Hashimoto Kei, Oura Keiichiro, Nankaku Yoshihiko, Tokuda Keiichi	4. 巻 -
2. 論文標題 Speaker Adaptation for Speech Synthesis Based on Deep Neural Networks Using Hidden Semi-Markov Model Structures	5. 発行年 2018年
3. 雑誌名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018)	6. 最初と最後の頁 638-643
掲載論文のDOI (デジタルオブジェクト識別子) 10.23919/APSIPA.2018.8659791	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 Huaibo Zhao, Yosuke Higuchi, Tetsuji Ogawa, Tetsunori Kobayashi	4. 巻 -
2. 論文標題 An investigation of enhancing CTC model for triggered attention-based streaming ASR	5. 発行年 2021年
3. 雑誌名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2021 (APSIPA2021)	6. 最初と最後の頁 477-483
掲載論文のDOI (デジタルオブジェクト識別子) なし	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73件 (うち招待講演 4件 / うち国際学会 3件)

1. 発表者名 佐久間仁, 藤江真也, 小林哲則
2. 発表標題 Self-Attention を用いた多人数会話向け発話タイミング推定
3. 学会等名 人工知能学会 第93回 言語・音声理解と対話処理研究会
4. 発表年 2021年

1. 発表者名 趙懷博, 樋口陽祐, 小林哲則, 小川哲司
2. 発表標題 Triggered attention型ストリーミング音声認識におけるMask-CTCを用いた事前学習
3. 学会等名 情報処理学会研究報告 (SLP)
4. 発表年 2021年

1. 発表者名 片山颯人, 藤江真也, 佐久間仁, 松山洋一, 小林哲則
2. 発表標題 Timing Generating Networks: 会話の文脈を考慮したターンテイキングのタイミング推定
3. 学会等名 人工知能学会 第90回 言語・音声理解と対話処理研究会
4. 発表年 2020年

1. 発表者名 Hayato Katayama, Shinya Fujie and Tetsunori Kobayashi
2. 発表標題 End-to-middle training based action generation for multi-party conversation robot
3. 学会等名 10th International Workshop on Spoken Dialogue Systems Technology (IWSDS) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Keiichi Tokuda
2. 発表標題 Statistical approach to speech synthesis: past, present and future
3. 学会等名 Interspeech 2019 (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Kei Sawada, Takenori Yoshimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, and Keiichi Tokuda
2. 発表標題 The NITech text-to-speech system for the Blizzard Challenge 2018
3. 学会等名 Blizzard Challenge 2018 Workshop (国際学会)
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
研究分担者	藤江 真也 (Fujie Sinya) (00367062)	千葉工業大学・先進工学部・准教授 (32503)	

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	森 大毅 (Mori Hiroki) (10302184)	宇都宮大学・工学部・准教授 (12201)	
研究分担者	徳田 恵一 (Tokuda Keiichi) (20217483)	名古屋工業大学・工学(系)研究科(研究院)・教授 (13903)	

	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
連携研究者	林 良彦 (HAYASHI YOSHIHIKO) (80379156)	早稲田大学・理工学術院・教授 (32689)	
連携研究者	小川 哲司 (OGAWA TETSUJI) (70386598)	早稲田大学・理工学術院・教授 (32689)	
連携研究者	南角 吉彦 (NANKAKU YOSHIHIKO) (80397497)	名古屋工業大学・工学(系)研究科(研究院)・准教授 (13903)	
連携研究者	大浦 圭一郎 (OOURA KEIICHIROU) (20588579)	名古屋工業大学・工学(系)研究科(研究院)・研究員 (13903)	
連携研究者	橋本 佳 (HASHIMOTO KEI) (10635907)	名古屋工業大学・工学(系)研究科(研究院)・准教授 (13903)	

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
---------	---------