

ニュース記事を話題にできる音声対話システム 検証のための実験環境

Experimental Environment for Verification of Speech-based News Delivery System

本田 裕¹

Hiroshi HONDA

松山 洋一²

Yoichi MATSUYAMA

高津 弘明²

Hiroaki TAKATSU

小林 哲則²

Tetsunori KOBAYASHI

要 旨

Web上に配信されているニュース記事を話題として、情報伝達をしながら、ユーザの発話へ多様に応答できる独自の音声対話システムを開発した。従来の音声対話システムでは一問一答型が主流であり、ニュース記事のようなまとまった情報の伝達には適していない。また、ニュース記事の読み上げシステムの場合は、ユーザが質問をはさむことができない。そこで、会話用要約技術、パラ言語認識器、間の調整、音声合成器といった音声対話システムに関する新たな要素技術を開発し、Web上にリアルタイムに配信されているニュース記事を話題にして日本語で音声対話が可能な実験システムを構築した。これにより、さまざまな利用シーンでシステムの検証を容易におこなうことができるようになった。今後、本システムを実用化することにより、運転中にドライバの話し相手となるような音声対話が可能となる。

ABSTRACT

An original spoken dialogue system was developed that is capable of diverse responses to user utterances while communicating information using news articles distributed on the Internet. Most conventional spoken dialogue systems are question-and-answer types, and are not suited to communicating large amounts of information such as news articles. In addition, users are unable to interrupt news article read-aloud systems with questions. Therefore, new key technologies related to spoken dialogue systems were developed, including conversation summarization technology, a paralinguistic recognizer, pause adjustment, and a speech synthesizer, and an experimental system was constructed that is capable of spoken dialogue in Japanese using news articles distributed in real time on the Internet. This enabled easy system verification in various usage scenarios. Practical application of this system will enable spoken dialogue in the manner of a conversation partner for drivers while driving.

1. まえがき

自動車業界はグローバルで事業環境が急激に変化しており、CASE(コネクテッド、自動運転、シェアリング、電動化)をはじめとする新技術や新サービスの開発が急速に進んでいる。特に、コネクテッド技術の進化により、ドライバは車内にいながらさまざまなサービスを楽しむよう

になりつつある。近年、AIスピーカのように音声対話が可能な車内向けシステムの普及が進んでいる。その一例としてHonda Personal Assistantがあり、ドライバの発話から意図を理解し、音楽再生やルート検索、マニュアルの提示といった機能を提供している。

このような車内での音声対話システムは、コネクテッド技術の進展により進化しており、これまで車載器で処理を

¹ 本田技研工業株式会社

² 早稲田大学、博士(工学)

していた音声認識や言語理解などがクラウド上で実行されるようになった。その結果、システムの性能が飛躍的に高まり、音楽配信や地理情報などのクラウドサービスとの連携も広がっている。

また、今後の自動運転の普及により、車内での時間の使い方も変わろうとしている。これまでドライバは運転に集中しなければならなかった。しかし、必要なときだけ運転をおこなうのであれば、移動中に自由に使える時間が増え、外部とのコミュニケーションに対するニーズも高まる。その過程にある現在の自動運転レベルでは、ドライバは常に前方を注視しなければならないという制約がある。そのため、音声だけでコミュニケーションをおこなえる、音声対話システムの価値は高いと考えられる。このように、CASE時代においては運転中にドライバの話し相手となるような音声対話システムへの要望が高まると思われる。

そこで、我々は Web 上でリアルタイムに配信されているニュース記事について、情報伝達をしながら、ユーザの発話に応じた多様な応答をおこなうことができる独自の音声対話システム⁽¹⁾の開発をおこなってきた。

さらに本研究では、このシステムの検証のために構築したユーザがタブレット端末を介して音声対話を可能とする実験環境について紹介する。

2. 音声対話システム

2.1. 従来型の音声対話システム

AI スピーカをはじめとする従来型の音声対話システムは一问一答や読み上げが主流である。

一问一答の場合、ユーザが「今日の天気は？」と発話すると、システムは「今日の東京の天気は晴れです。」といったように一つの問いに一つの答えを返す。この手法では、システムから自発的な発話はおこなわないため、ユーザは自由に対話をおこなうことはできない。

また、読み上げの場合は、ユーザが「今日のニュースは？」と発話すると、システムが「第 19 戦メキシコ GP の決勝レースが 10 月 29 日午後 1 時 10 分よりおこなわれました。午前中は青空が広がった天候も…」のようにニュース記事を読み上げる。よって、この手法でも、ユーザが質問を発話しても、システムは回答をおこなわないため、ユーザは対話をおこなうことはできない。

このような従来型音声対話システムは、一般に Fig. 1 に示すように音声認識器、対話管理部、音声合成器で構成されている。まず、ユーザ発話の音声が入力されると音声認識器によりテキストへ変換される。続いて、変換されたテキストは対話管理部に渡される。対話管理部は、シナリオのデータベースを参照し、ユーザ発話の意図に応じた応答文を生成する。最後に、生成された応答文は音声合成器に渡され、システム発話の音声が入力される。

2.2. 提案する音声対話システム

そこで、提案する音声対話システムでは、Web 上に配信されているニュース記事を話題として、ユーザとこのシステムとの自由な対話を実現するものである。これにより、システムが「第 19 戦メキシコ GP の決勝レースがおこなわれたんだって。」と発話した際に、ユーザが「いつ？」と質問をすると、システムは「10 月 29 日午後 1 時 10 分だよ。」といった応答をおこなうことが可能となる。

従来システムの技術では、システムが伝達中の情報についてユーザが質問をおこなっても応答をすることができなかった。また、ユーザからの質問を受けてから応答処理をおこなっているため、即応性の高い対話をおこなうことができなかった。本システムは Fig. 2 に示すように対話前処理部と対話処理部で構成されている。対話前処理部では、Web 上のニュース記事を収集して解析することで、ユーザに伝達する発話計画データベースやユーザからの質問に対する応答データベースを事前に作成する。

この対話前処理部と対話処理部により、従来のシステムでは容易ではなかった情報伝達中の質問応答と即応性の高い対話を実現した。

2.2.1. 対話前処理部

対話前処理部では、会話所要約技術を使用してニュース記事からシナリオと質問応答を生成する。会話所要約技術は Fig. 2 中のシナリオ生成部と応答生成部で実現される。本システムに入力されるニュース記事は書き言葉で表現されており、音声による伝達には適していない。そこで、対話前処理部では、ニュース記事を会話用に要約し、話し言葉に変換する。要約された文は発話計画データベースに格納し、要約に用いられなかった文は応答データベースに格納する。また、Wikipedia などの Web 上の情報を用いて、ニュース記事から抽出した単語の定義を生成し、質問応答に用いる。

2.2.2. 対話処理部

対話処理部では、発話計画データベースの情報を音声合成し、ユーザに伝達する。また、ユーザ発話を認識すると、発話内容に応じて応答データベースの情報を利用して応答をおこなう。

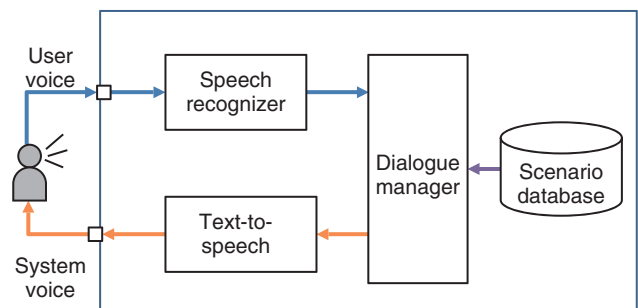


Fig. 1 Architecture of conventional system

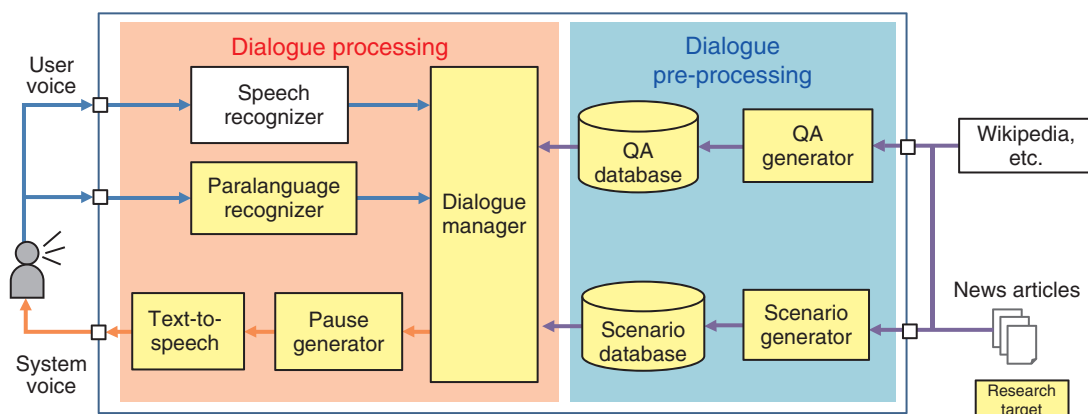


Fig. 2 Architecture of proposed system

対話処理部は Fig. 2 中のパラ言語認識器を備えており、発話意図に応じてシステムアクションをおこなう。パラ言語とは、イントネーション、リズム、ポーズといった周辺言語情報である。

例えばシステムが「第 19 戦メキシコ GP の決勝レースがおこなわれたんだって。」と伝えた際に、ユーザが「メキシコ GP?」と発話した場合、従来の音声認識のみでは「メキシコ GP」というテキスト情報しか取得できないため、ユーザの意図を解釈できない。6 種類のパラ言語を認識することで、ユーザ意図が疑問なのか反復要求なのか識別する。文の語尾のイントネーションが上がっているような場合、システムはユーザが疑問を問いかけていると判断し、「メキシコ GP っていうのは…」と質問応答をおこなうことができる。文の語尾のイントネーションが下がっている場合、ユーザが同意して欲しいと判断し、「そう。メキシコ GP」と反復をおこなう。

また、対話処理部には、Fig. 2 中の間の調整部を備えており、システム発話の間の調整をおこなうことで、ユーザが質問しやすく、理解しやすい会話のリズムを実現した。本システムでは、発話内の文節間および発話間の間を推定して、システム発話に挿入する。これにより、ニュース記事のようなまとまった情報を伝達する場合でも、ユーザが理解しやすく、発話中でも割り込みやすい間を実現した。

さらに、対話処理部では、ニュース記事のようなまとまった量の情報を伝えるため、談話構造上の発話の役割に応じたメリハリのある話し方が可能な音声合成器を搭載した。

3. 要素技術

3.1. 会話用要約技術

ニュース記事から会話用の要約を生成するため、深層学習を用いた要約モデル⁽²⁾を開発した。

3.1.1. 単語重要度推定モデル

ニュース記事から要約を作成するために、記事から重

要文を抽出した。重要文抽出のモデルには最大被覆モデル^{(3), (4)}を用いた。最大被覆モデルでは、制限の文数に収まる範囲で重要単語をできるだけ多く被覆する文集合を抽出した。記事内の重要単語を抽出するため、従来は分類や回帰に使用される機械学習の手法である random forest によって推定した値を用いていた。今回は重要単語の抽出に深層学習を適用した。

深層学習を用いた単語重要度推定モデルを Fig. 3 に示す。本モデルは、BERT⁽⁵⁾を用いており、文脈を考慮して単語の重要度を推定する。BERT とは、大量の文書データを用いて事前に学習することで、タスクの性能改善を図る自然言語処理モデルである。学習データは、ニュース記事 100 本について、3 人の作業者が要約したものを使用した。モデルに入力する補助情報には、以下 12 種類の情報を用いた。

- (1) 品詞大分類、品詞細分類、活用形、活用型、カテゴリ、ドメインといった形態素情報
- (2) 単語の出現頻度を表す指標値 (Term Frequency: TF)
- (3) 単語の希少性を表す指標値 (Inverse Document Frequency: IDF)

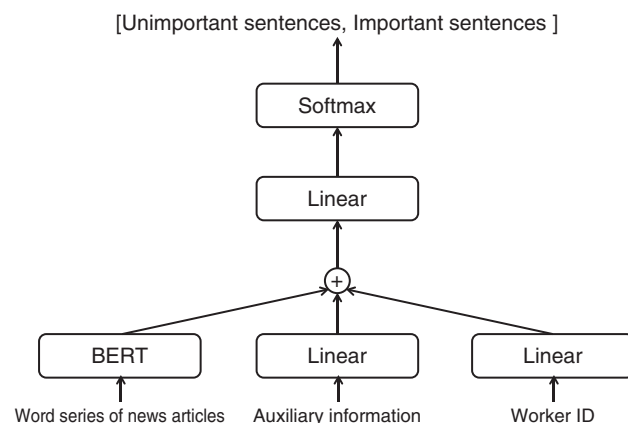


Fig. 3 Word importance estimation model

- (4) TF と IDF の積 (TF-IDF)
- (5) 単語がタイトルに含まれているかどうか
- (6) 括弧内の単語かどうか
- (7) 記事における文と段落の位置
- (8) 固有表現クラス
- (9) 係り受けの種類
- (10) 係り受け木 (文節の係り受け関係を表す木) の深さ
- (11) 係り元の文節数
- (12) 文頭からの文節の位置

作業員 ID は、複数の作業員のデータを混ぜて学習させるためのもので、作業員情報を 1-of-k で区別して与えた。1-of-k とは、一つの成分が 1 で残りの成分がすべて 0 のベクトル表現である。

モデルの評価結果を Fig. 4 に示す。評価尺度にはテキスト要約の評価値の一つである ROUGE-1 を用いた。ROUGE-1 は 1.0 に近いほど良い評価を表す。BERT による事前学習なし、補助情報なしで単一作業員のデータを用いて学習したモデルを SEM1、これに事前学習をおこなったモデルを SEM1_pre、さらに補助情報を加えて学習したモデルを SEM1+_pre、作業員 2 人のデータを用いて学習したモデルを SEM2+_pre とした。比較のため、random forest の結果を RF、作業員の結果を Human、ランダムに文を抽出した結果を Random、記事の先頭から K 文抽出したときの結果を LEAD として算出した。実験の結果、提案手法である SEM2+_pre の結果が最も良く、従来手法の random forest と比べて高い性能を達成した。

3.1.2. 文圧縮モデル

ニュース記事は書き言葉で記述されているため、一文に含まれる情報量が多い。そこで、一文に含まれる情報量を削減し、文圧縮をおこなうことで、話し言葉で伝達するのに適した形に変換した。文圧縮は文節レベルの系列ラベリング問題として定式化した。系列ラベリング問題とは、系列の各要素に適切なラベル列を付与する問題である。本研究では、文節単位で発話に含めるか否かをラベル付けた。

深層学習を用いた文圧縮モデルを Fig. 5 に示す。本モデルは、BERT を用いた。BERT で得られた単語の埋め込み表現を self-attention⁽⁶⁾ によって文節ごとに組み上げてい

る点が、単語重要度推定モデルと異なる。学習データは、単語重要度推定モデルと同じくニュース記事 100 本について、3 人の作業員が要約したものを使用した。モデルに入力する補助情報は、単語重要度推定モデルで使った補助情報から文と段落の位置を除いたものとした。

モデルにテストデータを入力し、正解率 (Accuracy) で評価した結果を Fig. 6 に示す。BERT の事前学習なし、補助情報なしで単一作業員のデータを用いて学習したモデルを SCM1、これに事前学習をおこなったモデルを SCM1_pre とした。さらに補助情報を加えて学習したモデルを SCM1+_pre、作業員 2 人のデータを用いて学習したモデルを SCM2+_pre とした。比較のため、補助情報と文節を構成する単語出現頻度のベクトル (bag of words) を素性として、時系列データを扱える深層学習モデルの一つである双方向 LSTM を用いて文圧縮した結果を BiLSTM として算出した。また、作業員の結果を Human、ランダムに文節を選択したときの結果を Random、文圧縮をおこなわなかったときの結果を NoCompression として算出した。実験の結果、提案手法である SCM2+_pre の結果が最も良く、BiLSTM と比べて高い性能を達成した。

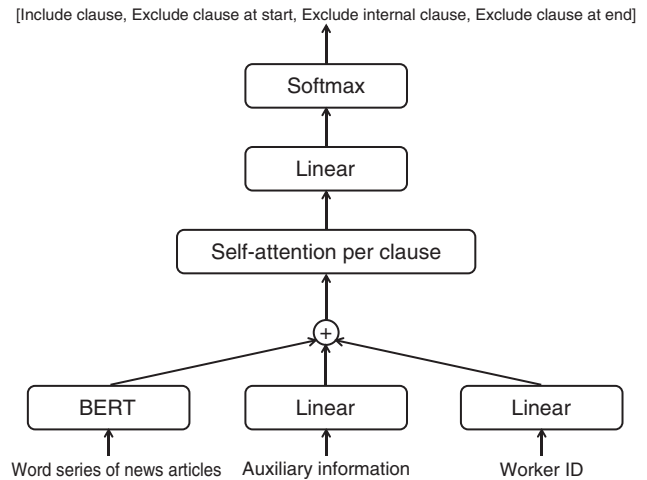


Fig. 5 Sentence compression model

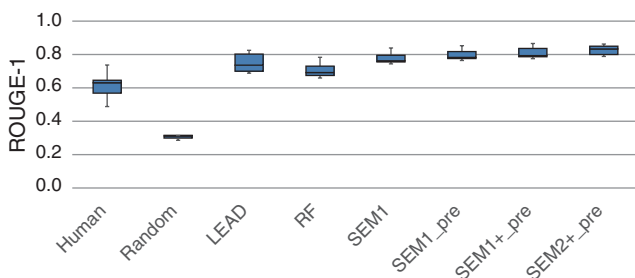


Fig. 4 Evaluation results of important sentence extraction

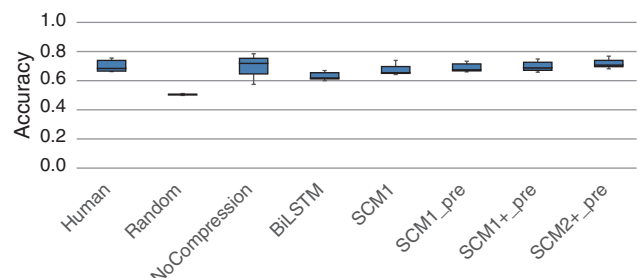


Fig. 6 Evaluation results of sentence compression

Table 1 User intention types, their effects, and corresponding system actions

Effects	User intention	System action
Increasing information	Question	Answer
	Request supplement	Provide explanations
	Request repeat	Repeat
Decreasing information	Disinterest	Change topic
	Already known	Omit details
Avoiding overlaps	Wait request	Listen

Table 2 Discrimination results of intention recognition

		With BERT pre-training With auxiliary information			With BERT pre-training Without auxiliary information		Without BERT pre-training Without auxiliary information	
		P	P+U	P+U+S	P+U	P+U+S	P+U	P+U+S
Question	Majority vote	0.920	0.958	0.962	0.957	0.961	0.940	0.951
	At least one vote	0.920	0.957	0.961	0.953	0.956	0.923	0.938
Request supplement	Majority vote	0.836	0.911	0.913	0.901	0.906	0.895	0.896
	At least one vote	0.827	0.882	0.887	0.881	0.887	0.847	0.865
Request repeat	Majority vote	-	-	-	-	-	-	-
	At least one vote	0.700	0.826	0.832	0.795	0.805	0.784	0.805
Disinterest	Majority vote	0.893	0.913	0.908	0.903	0.898	0.898	0.897
	At least one vote	0.797	0.811	0.805	0.810	0.804	0.804	0.803
Already known	Majority vote	0.818	0.939	0.939	0.939	0.939	0.939	0.939
	At least one vote	0.702	0.825	0.831	0.817	0.828	0.806	0.817
Wait request	Majority vote	0.716	0.885	0.865	0.885	0.858	0.845	0.838
	At least one vote	0.614	0.688	0.692	0.688	0.689	0.634	0.672

3.2. パラ言語認識器

ユーザの発話意図に応じたシステム応答を実現するため、深層学習を用いてユーザの発話意図認識モデルを生成し、パラ言語認識器⁽⁷⁾⁻⁽¹¹⁾を作成した。

ユーザの発話意図を認識するモデルを Fig. 7 に示す。本モデルは時系列の音声を入力すると現時点で予測される発話意図のラベルを出力する。学習には本システムとの約 2000 回の対話から作成した発話意図データセットを用いた。本データセットは、ユーザ発話に対して Table 1 に示すユーザ発話意図のラベル付けをおこなうことで作成した。

まず、音声を短い時間で切り出してスペクトログラムを生成し、CNN Autoencoder に入力した。CNN Autoencoder とは、教師なしで特徴量を抽出可能な深層学習モデルである。CNN Autoencoder の中間層から得られる韻律特徴量を識別部 (P) に入力することで、発話意図のラベルを推定できる。ユーザ発話の音声認識結果や直前のシステム発話を識別部 (L) に入力することで、最終的なラベルを推定する。識別部 (L) は BERT による事前学習を適用した。

本モデルによる発話意図の認識結果を Table 2 に示す。P はユーザ発話の韻律のみを用いて推定した結果、P+U は

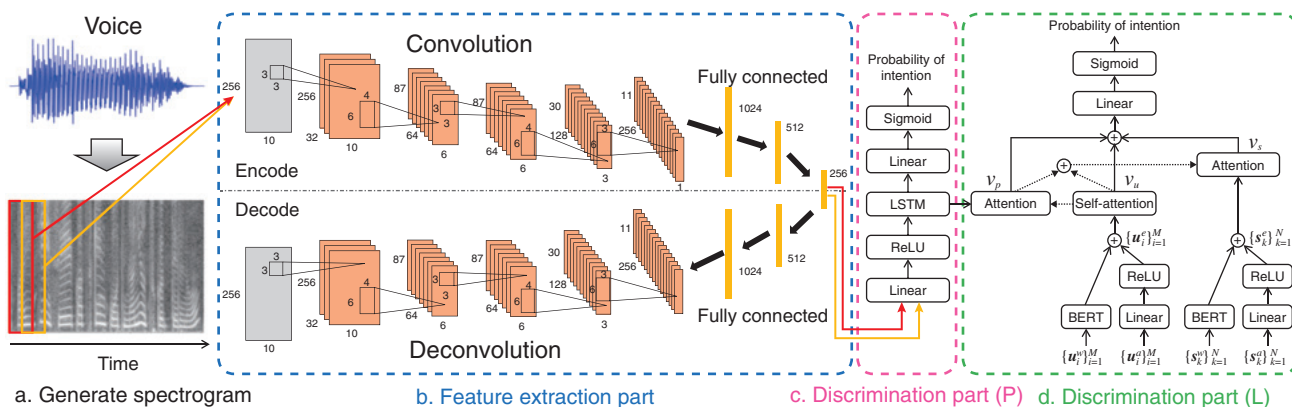


Fig. 7 Intention recognition model

ユーザ発話の韻律とその音声認識結果を用いて推定した結果、そして、P+U+Sはユーザ発話の韻律とその音声認識結果に加えてシステム発話を用いて推定した結果である。ユーザ発話の韻律に加えてユーザ発話の音声認識結果を入力することですべての意図の精度が高まった。BERTによる事前学習をおこなうことによっても精度が高まった。

3.3. 間の調整

深層学習を用いてシステム発話の間（ポーズ）を推定⁽¹²⁾することで、ユーザが質問しやすく、理解しやすい間を実現する機能を開発した。

3.3.1. 発話内文節間ポーズ推定モデル

発話内の文節と文節の間のポーズを推定するモデルをFig. 8に示す。本モデルは双方向LSTMを用いており、前後のLSTMの状態に基づいて文節間のポーズを決める。モデルにはLSTMにより文節の文字を入力した。

モデルの学習には要約、口語化したニュース記事100本を声優に発話させて作成したポーズ長タグ付きコーパスを使用した。収録にあたり、声優には以下の3点について注意するよう指示した。

- (1) 聞き手に理解させるつもりで発話すること
- (2) 相手に理解させるために重要な箇所と重要でない箇所をメリハリつけて発話すること
- (3) 発話中に質問が割り込まれる可能性を意識し、質問が予想される場所に自然な間を入れること

本モデルによるポーズの推定値と正解の誤差のヒストグラムをFig. 9に示す。0.2秒未満の誤差の割合は77%であった。

3.3.2. 発話間ポーズ推定モデル

発話と発話の間のポーズを推定するモデルをFig. 10に示す。現在の発話と次の発話の情報を用いて発話間のポーズを推定した。発話間のポーズは特徴量に、発話内文節間ポーズ推定モデルのLSTMの最後の隠れ層の値を用い、線形回帰モデルの一つであるBayesian ridgeモデルによ

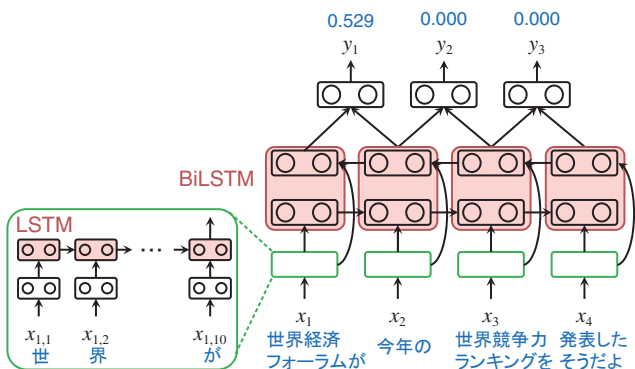


Fig. 8 Pause estimation model between phrases in utterance

て推定した。学習には発話内文節間ポーズ推定モデルと同様にポーズ長タグ付きコーパスを使用した。本モデルによるポーズの推定値と正解の誤差のヒストグラムをFig. 11に示す。0.5秒未満の誤差の割合は75%であった。

3.4. 音声合成器

ニュース記事のようなまとまった量の情報を伝えるため、発話の談話構造上の役割を利用して、メリハリのある音声合成器⁽¹³⁾を開発した。発話文を「発話節」という新規

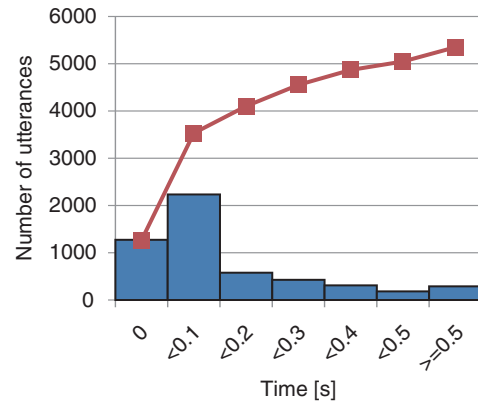


Fig. 9 Pause error between phrases in utterance

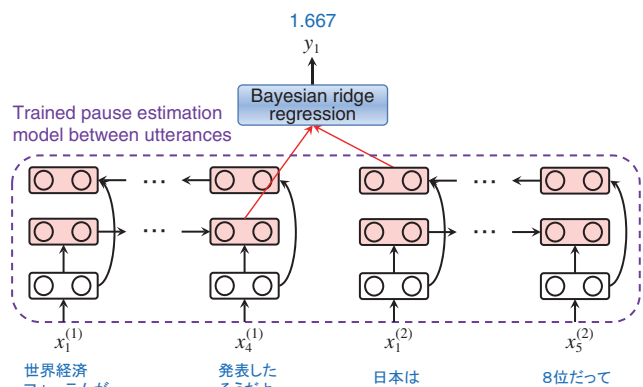


Fig. 10 Pause estimation model between utterances

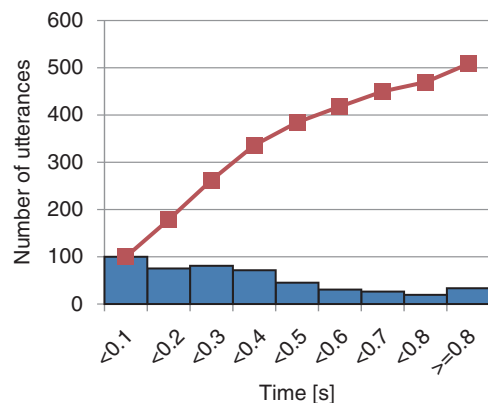


Fig. 11 Pause error between utterances

情報が複数含まれず、一息で発話する単位に分割する。その上で、発話節の役割を最も伝えたい内容を含む文(核)、核を伝えるためのリード文(前置き)、情報を補足する文(補足)の三つに分類し、これらを音声合成の補助情報として利用する。これにより、発話系列全体で韻律や間の制御を可能とした。発話節の役割の例を Fig. 12 に示す。音声合成の基盤システムには深層学習に基づく音声合成システムである Merlin⁽¹⁴⁾を採用した。

3.4.1. 音声合成モデル

深層学習による音声合成モデルを Fig. 13 に示す。本モデルでは、入力に言語特徴やフレーム特徴に加えて発話節の位置と役割を表す特徴(発話系列特徴)を与える。発話節の位置は、現在の発話が段落内の何番目であることを示す情報である。

音声合成モデルの学習には、ニュース記事 39 本を声優が読み上げたデータセットを使用した。読み上げ原稿にはあらかじめ発話節の役割を示してあり、声優には以下の 3 点について指示をおこなった。

- (1) 核の役割の内容を聞き手にはっきり伝わるように話すこと

Paragraph	Position	Role	Content of spoken phrase
1	1	Front satellite	宇宙を観測しているロシアの電波望遠鏡があるんだけど
	2	Nucleus	それが正体不明の「強い信号」を検知したっていつて注目を集めてるんだって
	3	Rear satellite	世界中の天文学者に衝撃が走っているらしいよ
2	1	Nucleus	信号はヘラクレス座の近くの恒星から届いたんだって
	2	Nucleus	その星は地球から約95光年離れていて
	3	Rear satellite	惑星も持っているみたい
3	1	Front satellite	地球外文明の進歩度を表す基準っていうのがあって
	2	Front satellite	それと照らし合わせたんだけど
	3	Nucleus	地球よりはるかに進歩した文明の可能性があるんだって
	4	Rear satellite	まだ確実なことはわからないんだけどね

Fig. 12 Example of discourse structure

Table 3 Comparison of performances with and without utterance sequence features

	F0-RMSE	dur-RMSE	MCD	BAD
Base line	36.420	4.779	5.064	0.660
Proposed model	35.435	4.768	5.047	0.659

- (2) 聞き手がうなずきながら聴いていることを想定し、途中で大きく間を空けることなく話すこと
- (3) 日常会話のような口調で聞き手にニュースを伝えるつもりで話すこと

10 分割交差検定で音声合成モデルの出力と元データの音響特徴量の誤差を計算した結果を Table 3 に示す。F0-RMSE, dur-RMSE はそれぞれ基本周波数、音素継続長の平均二乗誤差平方根を表している。MCD, BAD はそれぞれメルケプストラムひずみ、帯域ごとの非周期性指標のひずみを表している。メルケプストラムひずみとは、元データと音声合成モデルの出力の差異を測る客観指標値である。発話系列特徴を加えることにより、ベースラインに比べて基本周波数、音素継続長の誤差が小さくなった。メルケプストラム、非周期性指標についてもわずかに改善がみられた。

4. 実験システム

走行中の車両内で、多くのユーザがさまざまなシーンで利用して本システム評価をおこなえるようにするため、2.2. 節で述べた提案システムをもとに、タブレット端末を介してユーザが音声対話可能な実験システムを構築した。

実験システムの構成を Fig. 14 に示す。Web 上のニュース記事の収集と解析は対話前処理サーバでおこなう。対話前処理サーバは Web 上から定期的にニュース記事を収集することで、ユーザにリアルタイムに情報を提供できる。

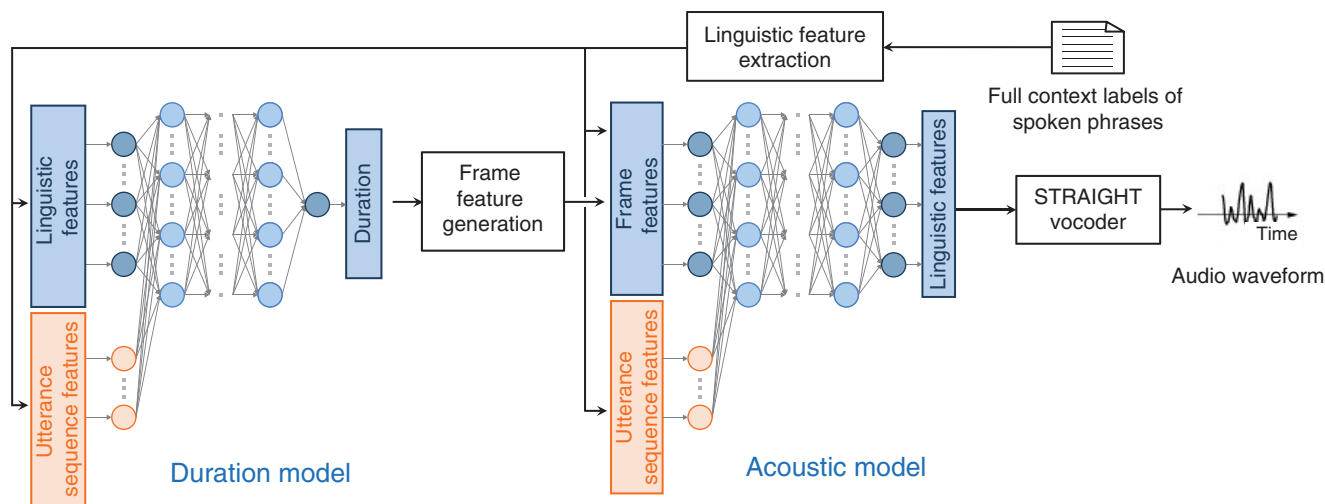


Fig. 13 DNN speech synthesis model considering utterance sequence features

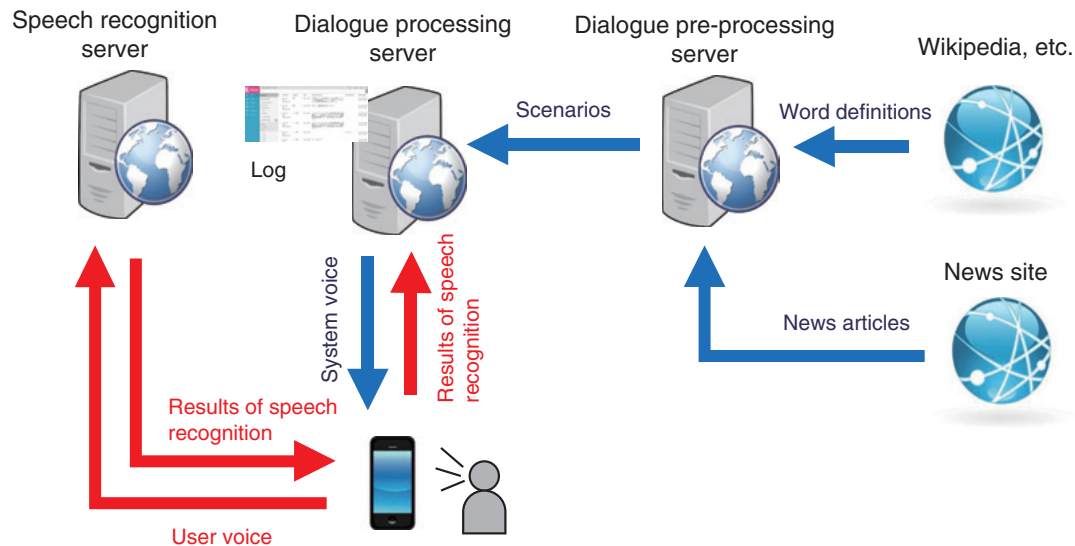


Fig. 14 Experimental system

発話計画の音声合成などは対話処理サーバがおこない、ユーザはタブレット端末を介して対話をする。音声認識には Google 社が提供する音声認識クラウドサービスである Google Speech API を使用した。

実験システムの利用例を Fig. 15 に示す。本実験システムは、日本語のニュース記事を用いており、日本語で音声対話ができる。ユーザはタブレット端末が備えるスピーカとマイクを利用して音声対話をおこなう。ユーザは Web に配信されている旬のニュース記事を知ることができ、疑問に思ったことをシステムに質問することができるようになった。

このシステムを活用することで、利用意欲に対する主観評価や、情報伝達効率などの測定ができる。今後このように、音声対話システムの有効性を検証していく所存である。



Fig. 15 Example of using experimental system

5. まとめ

会話用要約、パラ言語認識、問の調整および音声合成といった音声対話システムの要素技術開発に取り組んだ。また、それらの技術を組み込んでタブレット端末を介して対話可能なシステムを構築した。

従来型の一問一答型のシステムやニュース記事読み上げシステムに比べて、本システムは情報伝達中の質問応答と即応性の高い対話が可能となる。

今後は、本システムを実際に車内で使用した際の評価や日本語以外の言語へ対応していきたい。

参考文献

- (1) 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図性の異なる多様な情報行動を可能とする音声対話システム, 人工知能学会論文誌, Vol. 33, No. 1, p. 1-24, (2018)
- (2) 高津弘明, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための要約, 言語処理学会第 25 回年次大会発表論文集, p. 518-521, (2019)
- (3) Gillick, D., Favre, B., Hakkani-Tür, D.: The ICSI Summarization System at TAC 2008, Text Analysis Conference (TAC) 2008, (2008)
- (4) Takamura, H., Okumura, M.: Text Summarization Model based on Maximum Coverage Problem and its Variant, Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, p. 781-789, (2009)

- (5) Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of NAACL-HLT 2019, p. 4171-4186, (2019)
- (6) Lin, Z., Feng, M., dos Santos, C. N., Yu, M., Xiang, B., Zhou, B., Bengio, Y.: A Structured Self-attentive Sentence Embedding, Proceedings of the 5th International Conference on Learning Representations, p. 1-15, (2017)
- (7) Takatsu, H., Yokoyama, K., Matsuyama, Y., Honda, H., Fujie, S., Kobayashi, T.: Recognition of Intentions of Users' Short Responses for Conversational News Delivery System, Proceedings of the 20th Annual Conference of the International Speech Communication Association, p. 1193-1197, (2019)
- (8) Yokoyama, K., Takatsu, H., Honda, H., Fujie, S., Kobayashi, T.: Investigation of Users' Short Responses in Actual Conversation System and Automatic Recognition of their Intentions, Proceedings of the 2018 IEEE Workshop on Spoken Language Technology, p. 934-940, (2018)
- (9) 横山勝矢, 高津弘明, 本田裕, 藤江真也, 小林哲則: 会話におけるニュース記事伝達のための発話意図の分類と認識, 第 123 回音声言語情報処理研究会, 2018-SLP-123 巻, 5 号, p. 1-8, (2018)
- (10) 高津弘明, 横山勝矢, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための発話意図理解, 人工知能学会第 32 回全国大会論文集, 4Pin1-29, p. 1-4, (2018)
- (11) 横山勝矢, 高津弘明, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための発話意図分類とデータベースの構築, 人工知能学会第 32 回全国大会論文集, 3Pin1-23, p. 1-4, (2018)
- (12) 高津弘明, 横山勝矢, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための間の調整, 言語処理学会第 24 回年次大会発表論文集, p. 212-215, (2018)
- (13) 高津弘明, 福岡維新, 藤江真也, 岩田和彦, 小林哲則: 会話によるニュース記事伝達のための音声合成, 人工知能学会論文誌, Vol. 34, No. 2, p. 1-15, (2019)
- (14) Wu, Z., Swietojanski, P., Veaux, C., Renals, S., King, S.: A study of speaker adaptation for DNN-based speech synthesis, Proceedings of the 16th Annual Conference of the International Speech Communication Association, p. 879-883, (2015)

■ 著 者 ■



本田 裕



高津 弘明



松山 洋一



小林 哲則