

機関番号：94301
研究種目：若手研究（A）
研究期間：平成20年度～平成22年度
課題番号：20680011
研究課題名（和文）
発話音声に伴う頭部動作および表情と言語・パラ言語情報との関連構造の構築
研究課題名（英文）
Relationship between paralinguistic information and head motion/facial expression
研究代表者
石井 カルロス寿憲（CARLOS TOSHINORI ISHI）
株式会社国際電気通信基礎技術研究所・知能ロボティクス研究所・研究員
研究者番号：30418529

研究成果の概要（和文）：

本研究課題では、対話コミュニケーションにおいて、発話音声とそれに伴う頭部動作や表情の視覚的情報との関連構造を提案し、言語情報と韻律・声質特徴からなる発話スタイルの組み合わせにより、談話機能や発話意図などを介して頭部動作や表情の視覚的情報と関連付ける構造を解明した。主な成果として、「氣息音」と「りきみ」の声質が対話でもたらず機能を明らかにし、談話機能と頭部動作との関連を解明した結果、発話に伴う頷きと首かしげの動作生成モデルを提案し、有効性を評価した。

研究成果の概要（英文）：

The present research aimed on clarifying the relationship between speaking styles including prosodic and voice quality features, paralinguistic information such as dialogue acts, intentions and emotions, and visual modalities like head motion and facial expression. As main results, we clarified the paralinguistic functions carried by breathy/whispery voices and “rikimi” voices, we clarified the relations between dialogue acts and head motion, and proposed and evaluated a model for generating nod and head tilt motion from the dialogue acts.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
H20 年度	7,500,000	2,250,000	9,750,000
H21 年度	6,200,000	1,860,000	8,060,000
H22 年度	2,800,000	840,000	3,640,000
総計	16,500,000	4,950,000	21,450,000

研究分野：総合領域

科研費の分科・細目：情報学，知覚情報処理・知能ロボティクス

キーワード：音声情報処理，非言語情報

1. 研究開始当初の背景

近年、メディアはマルチモーダル化してきており、対話音声システムにおいて単に音声を伝達するだけでなく、ロボットや人間の姿を持つエージェントが情報を提供するインタフェースが増えてきている。

人が音声対話を介してコミュニケーションを行う際、発話音声に伴う頭部動作や表情などの視覚的情報も話者の意図や感情などのパラ言語情報を伝達することが多く、対話

の相互理解やインタラクションの自然さ、円滑さに影響する。

しかしながら、研究開始当初は、発話音声に伴う頭部動作や表情とパラ言語情報との関連は解明されていなかった。

2. 研究の目的

本研究課題では、対話コミュニケーションにおいて、発話音声とそれに伴う頭部動作や表情の視覚的情報との関連構造を提案し、言

語情報と韻律・声質特徴からなる発話スタイルの組み合わせにより、談話機能や発話意図などを介して頭部動作や表情の視覚的情報と関連付ける構造を構築することを目的としている。

3. 研究の方法

(1) マルチモーダル（音声、EGG，モーション・データ）のデータベース構築

分析用のデータとして、自然対話におけるさまざまな発話スタイル及び頭部動作・表情のノンバーバル情報を含んだデータベースを構築した。声質の分析に重要となる声帯振動を測定するため、EGG (Electroglottograph) 装置を用いた。頭部動作・表情のデータ収集には、複数のビデオカメラと、被験者の頭部に取り付けるマーカーの動作を検出可能なモーションキャプチャシステムを用いた。

(2) 言語・非言語情報のラベリング作業および動作パラメータの抽出

収集したデータベースの発話区間切り出し、書き起こし、およびラベリングを行った。

発話のラベルとしては、音調や句境界の情報を表現したラベルを付与し、それぞれのアクセント句もしくは韻律句において談話機能や発話意図・態度・感情のラベルを付与した。ラベルの項目としては、談話機能として「発話権を保持、発話権の譲渡、相槌」など、意図・態度・感情としては「肯定、否定、質問、驚き、感心、意外、怒り、楽しい、落胆」などが挙げられる。また、被験者間の一致度により、ラベルを統一した。

頭部動作に関しては、頭部の回転角度で表現した。表情や口唇動作とは独立した頭部と鼻と顎に貼り付けたマーカーを基に、SVD法 (Singular Value Decomposition) を用いて回転行列を求め、3次元の軸における回転角度により、頭部動作を表現した。

(3) 話者による発話スタイルや動作の違いの分析

得られた音声データにおいて、まず韻律及び声質に関連した音響特徴抽出を行った。

同じ態度や感情の表現でも、音声にも動作にも、話者によってその表現の仕方や癖などがあるため、話者による発話スタイルや動作の違いを調べた。

(4) 言語情報・パラ言語情報と視覚的情報の関連構造の分析および構築

同じ頭部動作、もしくは同じ表情でも、発話によって異なった意図や感情を示す。ここでは、そういった冗長性を明らかにし、談話機能や発話意図と介したモデル化を提案した。発話に含まれる言語情報と発話スタイル

によって伝達される意図や態度や感情などのパラ言語情報と、視覚的情報の頭部動作や表情との関連を調べた。ラベル化されたカテゴリと発話スタイルのパラメータとの関連を調べた。その一方、頭部動作においては、頷きや首振りや首かしげなどのイベントのモデルを構築し、そのイベントを生成する確立モデルを提案した。

(5) 言語情報と発話スタイルを考慮した談話機能の自動抽出

提案した関連構造を実現化するため、さまざまな要素を自動化する必要がある。発話音声に含まれるさまざまな談話機能や発話意図・態度・感情などの情報を抽出するため、まずその基本単位となる句境界を検出する必要がある。

(6) 発話音声に伴う頭部動作・表情の自動生成および実証実験による評価

ここでは、提案した構造を利用し、音声に含まれる言語情報と発話スタイルの情報を入力とし、その発話に伴う頭部動作・表情の視覚的情報を出力とする応用を検証する。ここでは (E) で構築したモデルを使用して、頭部動作・表情のパラメータを本研究室に既存のアンドロイドロボットのアクチュエータの制御値にマッピングし、動作を生成した。頭部動作・表情パラメータを求めるには処理時間がかかるため、遠隔操作においては、遅延を加えて音声と動作を同期させた。

4. 研究成果

(1) マルチモーダルのデータベース構築および言語・非言語情報のラベリング作業

まず分析用のデータとして、自然対話におけるさまざまな発話スタイル及び頭部動作・表情のノンバーバル情報を含んだマルチモーダル（音声、EGG，モーションキャプチャ・データ）のデータを収集した。収集したデータの同期、発話区間切り出し、書き起こし、および談話機能や発話意図・態度・感情のラベリングの作業を実施した。頭部と鼻と耳に貼り付けたマーカーを基に、SVD法などを用いて回転行列を求め、3次元の軸における回転角度により、頭部動作を表現した。得られた音声データにおいて、音響特徴や頭部動作に関連する話者の依存性を調べた。年齢や性別による違いは長時間に渡る平均的な特徴で表現可能であり、「驚き」「感心」などの心的態度や感情表現は比較的短い区間で特徴の変化が起きることが確認された。また、対話相手との関係および話題の明暗によって、話者内でも頭部動作や韻律特徴に著しい違いが観られた。

談話機能として、以下のラベルセットが得られ、これを後続の頭部動作との対応の分析

に用いた。

- k (keep) : 発話権の保持 (強い句境界)
- k2 (keep) : 弱い句境界 (発話権の保持)
- k3 (keep) : 発話末を伸ばし、発話の途中であることを表現 (発話権の保持)
- f (filler) : 「えっと」「あー」など、考え中であることを表現する感動詞
- f2 (conjunctions) : 「じゃ」などの接続詞 (短いフィラーとして捉えられる)
- g (give) : 対話相手への発話権の譲渡
- q (question) : 発話権の譲渡 (対話相手に応答を求める場合)
- bc (backchannels) : 「うん」「はい」などの相槌を表現する感動詞
- su (surprise/admiration) : 「えー!」「へー」など、驚きや感心などの感情を表現する感動詞
- dn (denial) : 「いいえ」「ううん」などの否定を表現する感動詞

(2) 頭部動作の分析および生成モデルの提案

頭部動作に関しては、複数の話者のデータを解析することにより、対話相手との関係および興味の度合いなどによって、話者内でも頭部動作の頻度が変わり、韻律特徴にも著しい違いが観られた。また、分析結果を基に、発話に含まれる談話機能を用いて発話に伴う傾き動作を生成するモデルを提案し、図1に示す2種類の人型ロボットの頭部を制御する手法を構築・評価した。傾き生成の規則として、談話機能が“k”、“g”、“bc”、“q”つまり強い句境界の場合、発話の最後の音節に傾きを生成するというものであり、H20年度に作成した人同士の会話データベースの分析に基づく。

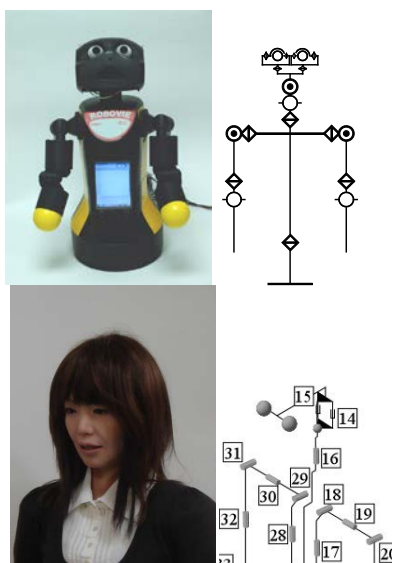


図1. 傾き生成モデルを評価した人型ロボット Robovie Mini-R2 およびアンドロイド

Repliee Q2

評価用の動作生成として、モーションキャプチャで測定したオリジナルの動作 (ORIGINAL)、オリジナルを1秒ずらしたものの (SHIFTED)、傾きを生成したモデル (MODEL I) と、傾きプラス発話中に頭を少し上げる動作を加えたもの (MODEL II) を比較した。図2に主観的評価結果を示す。結果より、両タイプのロボットで、提案手法 (MODEL II) が他の動作よりも自然な動作として知覚された。

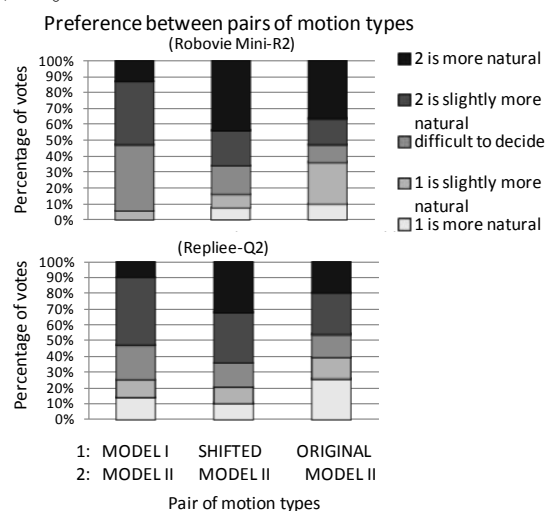


図2. 傾き生成モデルの種類の主観的評価: Robovie Mini-R2 (上図) および Repliee Q2 (下図)

頭部動作生成においては、傾き生成に加え、首傾げの生成モデルも提案し、ロボットに適用して評価実験を行い、提案モデルの有効性を示した。

(3) 発話スタイルとパラ言語情報の関連

初年度に作成したデータベースを用いて、さまざまな音響分析、発話に含まれる (言葉の意味が持つ) 言語情報と発話スタイルによって伝達される意図や態度や感情などのパラ言語情報と、視覚的情報の頭部動作や表情との関連を調べた。

音響分析に関しては、氣息音と「りきみ」を含む声質の音響特徴分析、およびこれらの声質がもたらすパラ言語情報に関して検証した。

① 氣息音

氣息音を含んだ発話では、データベースに出現した氣息音発話が多量にパラ言語情報の種類を以下のようにまとめた。

A) 強調、注意の促し、真の気持ちの表現

発話内で強調された単語やフレーズに、氣息音が頻繁に生じ、聞き手の注意を促す効果

があると考えられる。強調には、高く、強く、長く発声する方法があるが、これらの音の変化を伴わない場合の氣息音のみでも、強調が表現可能である効果が観られた。氣息音は、フレーズ頭あるいはアクセントが高の部分に生じやすい傾向が観られたが、フレーズ全体に伴う場合も観られた。

形容詞や形容動詞の場合、「強調」とともに、話者が「真の気持ち」を伝えている（気持ちがこもっている）印象を与える発話が多く出現した。例えば、「これは太いよ」という発話に氣息音が伴う場合、「これは本当に太いよ」というような表現となる。

B) 興奮・興奮の再現

「強調」とともに、「興奮」とラベル付与された発話もあった。発話スタイルは強調の場合と似ているが、発話を通して氣息音区間がより頻繁に生じる。発話内で通常発声と交互に氣息音が生じることが観られた。

また、対話中、過去の発話や他人の発話を興奮気味に再現するものもあり、この場合も同様なパターンで氣息音が生じることが観られた。

C) 驚き・意外・感心などの表現

驚き・意外・感心などの感情や態度の表現には、感動詞のように、「すごい!」、「本当!」、「そうなんですか!」、「そうなんだ!」、「なるほど!」、「うそ!」など、対話相手の発話に反応を返す場合の表現が多く出現した。

驚き・意外・感心の表現には、二つの発話スタイルが観られた。一つはパワーが強く、(B)のように話者が興奮した状態を示し、もう一つはより静かな口調 (E) の囁き声のような発話スタイルで、相手の話しに興味を示す効果があると考えられる。また、話者の興奮の度合いが増すにつれて、氣息音に伴い、声帯振動が不規則になる harsh 発声が生じる傾向が観られた。

D) 優しい口調+興味の促し

柔らかい声に伴い、リズム的に氣息音発声が見られるパターンが、母親が子供に童話を読み聞かせる場面で多く出現した。この氣息音発声を含んだ発話スタイルは（淡々と本を読み上げた場合に比べ）、優しい印象を与えると同時に、聞き手の興味や想像力などを促す効果もあると考えられる。

この発話スタイルでは、ピッチが低くなった区間に氣息音発声に伴い、ピッチが高くなると通常発声に戻るようなパターンが観られた。一方で、「興奮」の発話スタイルでは、ピッチが高くなった区間で、フレーズ全体を通して氣息音が際立つ特徴が観られた。

E) 囁き・囁き声：内緒話、戸惑い、控えめ

声帯が振動しない場合の発声を「囁き」(whisper)と呼び、声帯が振動しながら囁きの成分を有する場合も含んだ発声を総称して「囁き声」(whispery voice)と呼ぶ。

囁き声の最も知られる機能は、内緒話・ひそひそ話である。有声成分が無くなるため、音声のパワーが通常発声に比べ小さくなり、発話内の特定の単語、または全体が囁き声になり得る。考え中、戸惑い（「何やったっけな」「なんつったっけ」「…わかんない」）、または独り言や掛け声（「よいしょ」、「さて」）など自己発話にも観られた。

また、「強調」・「真の気持ち」の表現においても、氣息音発声が囁きに近づくほど、効果がある傾向が観られた。「自信の無い」、「控えめな」が付与された発話では、囁きが発話末で生じることが観られた。

F) ため息：落胆、残念、疲れ、安心

少数であったが、ため息をついた発話がデータベースに存在した。感動詞と同様に、落胆、残念、疲れ、安心などの表現で、囁き声と似た特徴を持つが、ため息を含んだ発話は低く下降するイントネーションに伴って氣息音が生じる特徴が観られた。

② 「りきみ」

氣息音は、さまざまなイントネーションに伴って生じることが示された。その一方、「りきみ」を含んだ発話では、図1の例に示すように、隣接した通常発声に対して、基本周波数(F0)が下降し、低周波数成分が弱くなり、全体的なパワーも弱くなる特徴を持つ。但し、このF0の下降は、話者がピッチを下げようと意図して生じたものではなく、喉頭を「りきませた」結果として生じたものと考えられ、聴覚印象として、イントネーションの下降型として知覚されるというよりは、声質の変化が際立って知覚される。また、「りきみ」は、声帯パルスに乱れが生じやすいフライ発声やharsh発声に伴うことが多いが、声帯パルス間で周期性の乱れが顕著に現れない発声に伴う場合もある。

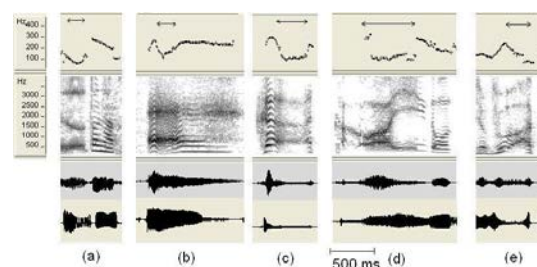


図1. 「りきみ」発声を含んだ発話の音響特徴の例：上のパネルからピッチ、酢ペクトログラム、音声波形の振幅、EGG 波形の振幅。

「りきみ」発話は共通して F0 が下降するという特徴を持つため、イントネーションに依らず、発話内のどの部分で「りきみ」が生じたかにより、表現されるパラ言語情報が変わると考えられる。以下に、感動詞を除いて、

これまで調べてきた「りきみ」発話の品詞とパラ言語的機能をまとめた。

A) 形容詞・形容動詞：真の気持ち

発話内で特定の単語で「りきみ」が起きる場合が多く見られた。形容詞や形容動詞に「りきみ」が起きる場合、強調、または話者の真の気持ちが表現される。

本データベースに出現した例として、「可愛い」、「大変」、「もったいない」、「痛い」、「きれい」、「冷たい」などが挙げられる。例えば、「りきみ」を含んだ「可愛い」という発話は、「本当に可愛い」のような表現となる。

また、「…もうわたしあれ恐怖やわ…」のように、発話全体を通して「りきみ」が生じる場合も観られた。

B) 擬音語・擬態語

擬音語や擬態語では、ほとんどの発話で、声質が変化し、その中でも「りきみ」が多く、「生き生き」感や度合いの強化が表現される。

本データでは、以下のような例が出現した：「…ギブスをガーッと切って…」(勢いよく)、「…人数がね、二人からブアーッと広がっている家系図…」(大きく広がっている状態)、「…フミヤくんが、ガーッと後ろからかかれて」(勢いよく)、「ネーて泣いてさ」(赤ちゃんの泣き声の表現)、「アー」(悲しんで興奮した泣き声の表現)、「ママは、ずーっと仲良しで…」(長く続く状態)。

C) 名詞：躊躇・戸惑い

名詞に「りきみ」が出現した例文が少なかったが、対話相手の発話に対し、強い躊躇・戸惑いを表現する機能が観られた。例文：「大津やばいっすよ…大津…、人住んでるんですか…」、「服は…やっぱり着られない…」、「プリキュア？…プリキュアは、ドリームかな」

D) 発話の再現・台詞

話者本人または第3者の過去の発話を再現する際に、「りきみ」で表現されることが観られた。形容詞を含んだ場合の再現の例：「これは心底うまいな…って思ったね…」、「なんかこう、気持ち悪い…っていうのは…」。謙遜・控えめな表現の再現の例：「やー、これはちょっとやっぱりよくないと思いますってっていうのは勇気がいると思います…」。怒りなど、興奮した発話の再現：「こら！とは言われへんよな」、「もういい加減にしー！ってまた、どうしても上に怒ってしまう…」。

興味深い結果として、実際の怒りでは、パワーが強く発せられたと考えられるが、再現の場合は、パワーが小さい「りきみ」でも、当時の興奮状態が表現されることである。

また、モデル構築には至らなかったが、「りきみ発声」と眉毛をひそめる表情との関連を見出した。

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 1 件)

Ishi, C.T., Ishiguro, H., Hagita, N. (2010). Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech. EURASIP Journal on Audio, Speech, and Music Processing 2010, ID 528193, 1-12 Jan. 2010.

[学会発表] (計 5 件)

Ishi, C., Ishiguro, H., and Hagita, N. (2010). "Acoustic, electroglottographic and paralinguistic analyses of "rikimi" in expressive speech," Proceedings of Speech Prosody 2010 (SP2010), ID 100139, 1-4.

Ishi, C.T., Liu, C., Ishiguro, H., and Hagita, N. (2010). "Head motion during dialogue speech and nod timing control in humanoid robots," Proceedings of IEEE/RSJ Human Robot Interaction (HRI 2010), 293-300.

Ishi, C.T., Ishiguro, H., and Hagita, N. (2008). "Analysis of inter- and intra-speaker variability of head motions during spoken dialogue," Proceedings of the International Conference on Auditory-Visual Speech Processing 2008 (AVSP' 2008), 37-42.

Ishi, C.T., Ishiguro, H., and Hagita, N. (2008). "The meanings of interjections in spontaneous speech," Proceedings of The 9th Annual Conference of the International Speech Communication Association (Interspeech' 2008), 1208-1211.

Ishi, C.T., Ishiguro, H., and Hagita, N. (2008). "The roles of breathy/whispery voice qualities in dialogue speech," Proceedings of Speech Prosody 2008, 45-48.

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

無し

6. 研究組織

(1) 研究代表者

石井 カルロス寿憲

(CARLOS TOSHINORI ISHI)

株式会社国際電気通信基礎技術研究所・

知能ロボティクス研究所・研究員

研究者番号：30418529

(2) 研究分担者

なし

(3) 連携研究者

なし