1—7—10

発話機能における音声の非語彙的情報の分析およびその考察*

◎田中 俊光 (奈良先端大 情報)　　△柏岡秀紀 (奈良先端大/ATR)
ニック・キャンベル (奈良先端大/ATR/CREST)

1 はじめに

近年のロボット技術の発展やコンピュータの普及を考えると近い将来の音声対話システムでは日常会話に近い発話をそのまま音声インターフェースとして使用することが期待される. そのために日常的な会話について音声の機能的側面の分析が必要である.

1.1 目的

音声の発話意図や発話機能における F_0 や発話時間長など韻律情報との関連が明らかになれば, 波形接続型音声合成 [1] における自然な韻律のターゲットとすることもでき, スムーズな対話の実現が期待できる. そこで人手付与された発話機能を示すラベルを用いて, その機能について韻律情報を説明変数とする決定木を作成し, 韻律情報と発話機能に関する知見を得ることが本稿の目的である.

1.2 音声の非語彙情報

音声情報はしばしば言語, 非言語, パラ言語に区別される [2]. 主に言語情報は内容や文構造, 非言語情報は性別などの個人的特徴, パラ言語情報はやや曖昧であるが意図, 感情などのテキストに書き起こせない情報を指すとされる. そこで本稿では境界が曖昧な非言語情報とパラ言語情報の集合を非語彙情報とする. その上で非語彙情報としての発話機能について韻律の分析を行う. 使用した発話機能はCREST/ESP プロジェクトにおいて作成された中でも出現頻度の多い発話機能に限定した. 詳細を 2 節に示す。

2 音声試料

音声資料は 1 対話あたり 30 分程度の電話対話データを 5 対話を使用する [3],[4]. 全発話数は 12715 発話である. 相手話者の違いによる使い分けの揺れを考慮し一組の女性話者同士の会話に限った. 話者は二名である. 40 代と 30 代で年齢に差があり, 対人関係による丁寧度, あるいは感情に起因する韻律の使い分けで変化する可能性を含んでいる. 発話には発話時間, 発話機能のラベルが人手で付与されている. その発話機能ラベルの出現頻度を考え, 以下の発話機能 (表 1) を使用する.

*Analysis and considerations of Non-verbal speech on speech function. By Toshimitsu Tanaka(Nara Institute of Science and Technology (NAIST)), Hideki KASHIOKA(NAIST/ATR) and Nick Campbell(NAIST/ATR/CREST)

発話機能		発話数
あいづち	(*listen*)	2478
理解	(*understand*)	556
笑い	(*laugh*)	626
フィラー	(*filler*)	364
気づき	(*notice*)	247
吸気音ノイズ	(*noise*)	996

表 1: 使用した発話機能

3 決定木分析

3.1 発話機能カテゴリの識別の可能性

各発話機能の韻律情報の分布に違いがあるのか確認するためラベル同士の識別の決定木を作成した [5]. その際の機能ラベル間の韻律による識別の精度は 72%程度であった. そのため識別は可能であると考え, ラベルを単一のカテゴリとしフィルタの作成を試みる.

3.2 説明変数

説明変数としては当該発話に関しての以下の 17 種類の韻律情報を用いる. 韻律による文脈の有無を見るためポーズを除く同様の韻律情報を直前の自発話の 16 変数と直前の相手発話についての 16 変数についても追加し計 49 変数を用いる.

○ ポーズ: 前発話とのポーズ
○ 発話時間: 人手による発話時間長 (dur)
○ 発話速度: モーラ速度
○ 基本周波数: 最大値 (F_0max), 平均 (F_0mean), 最小値 (F_0min), 最大と最小の落差 (F_0range), 時間長で正規化した最大値の位置 ($F_0maxpos$), 時間長で正規化した最小値の位置 ($F_0minpos$), F_0 の変化 $F_0range/(F_0maxpos - F_0minpos)$ (F_0vari)
○ パワー: F_0 と同様の 7 種類

3.3 各カテゴリの識別

2 節で示したカテゴリを識別するためのフィルターとして決定木を用いる. まず始めに各カテゴリとその対象とするカテゴリ以外をランダムに選びほぼ同数に調整したデータで学習を行い, 交差検定を行った際の精度を測定する. この識別精度が高いものは, 識別フィルターとしてそのまま使用できる可能性が高い. そこで自然の割合を考慮し, 識別対象のカテゴリ以外の発話に対して頑健とするため対象のカテゴリ以外の学習セットを増やし学習を行い同様に精度を測定する.

4 各カテゴリの識別結果

4.1 識別結果

カテゴリとそのカテゴリ以外がほぼ同数の学習セットの識別率の結果を表2に示す.

発話機能	function	識別率
あいづち	(*listen*)	83.5 %
理解	(*understand*)	76.3 %
笑い	(*laugh*)	80.7 %
フィラー	(*filler*)	76.4 %
気付き	(*notice*)	90.3 %
吸気ノイズ	(*noise*)	94.3 %

表 2: カテゴリの割合が同等のときの識別率

各カテゴリの決定木の特徴は以下の通りである.

あいづち: 発話時間, 発話速度, F_0mean, F_0max, F_0max_p, $Powvari$ などで構成され, 当該発話の韻律のみでも葉の数は150程度となり複雑な構造である.

理解: 発話速度と F_0mean, 直前相手発話の発話時間などの影響が大きく, 葉の数は10から30程度

笑い: 発話時間, 発話速度, F_0mean, $F_0minpos$, $Powrange$, $Powmean$, 直前の相手発話の F_0max でほぼ構成され葉の数は10から30程度である.

フィラー: 発話時間, 発話速度, F_0max, F_0mean, ポーズなどの影響が大きい. また直前の相手発話の時間も影響が大きい. 葉の数は30から50程度である

気付き: 主に発話時間, F_0min などが木の上部にみられ, その他には F_0mean などで構成されている. 葉の数は10から20程度である.

吸気音ノイズ: 今回使用したデータは片方の話者が花粉症である対話を含むため比較的多くの吸気ノイズがあった. そのため吸気ノイズのほとんどはこの一名のものである. しかし F_0mean などの分布の違いが非常に特徴的である. 葉の数は10から20程度である.

さらに識別率の高い2つのカテゴリについては実際のカテゴリのバランスに近づけ, 同時にノイズと気付きを比較するために割合の多いノイズとノイズ以外の割合に合わせ, 識別するカテゴリ以外の割合を増やし学習した. 表3に識別カテゴリ以外の識別に頑健な決定木の各識別正答率を示す.

発話機能	正答率	正答数	誤り数
気付き	80.5 %	178	43
気付き以外	97.3 %	2450	69
吸気音ノイズ	87.1 %	835	124
吸気音ノイズ以外	98.6 %	11595	161

表 3: カテゴリを自然な割合に近づけたときの正答率

ここで識別対象カテゴリとその他の割合が10倍ほど違うため各々の正答率を示めしている.

5 考察

識別に必要とされる説明変数は, 決定木の枝刈りによって単一のカテゴリの識別に必要な情報は比較的少量であることが分かる. あいづち, 理解, フィラーに関しては直前の発話の影響があり, 当該発話の前の発話の韻律情報を使用しているため, 発話機能に関して韻律的な文脈構造の様なものがあるとも考えられる. また説明変数には今回は使用していないが, 発話内で F_0 のピークが複数ある場合や笑いの発話で「ハハハ」などのように繰り返しの有無を見るような場合は標準偏差を考慮することで識別精度向上が期待できる. また今後複数話者の発話を考えると対人態度を考慮した変数を導入するべきである.

6 結論

利用した発話機能の識別はほぼ同数の学習セットのとき8割前後との結果を得た. 気付きや吸気ノイズのような書き起こしが似ている発話に関しては特に良い識別が可能であった. 吸気音ノイズでは F_0mean などの分布が2つに分かれおり特徴的であった. 気付きに吸気音ノイズの2つに関してはその他の学習セットを増やしたことでその他と判定されたものが97%以上となり非常に高い精度となった. またカテゴリ自体も良い精度を保つため実際の識別のフィルタとして期待できる.

7 課題

精度向上を目指すには説明変数の検証が必要となる. 同時に発話の機能を分析する上で今回はラベラー3名の協議後の結果を使用しているが発話機能のカテゴリがより聴覚的に妥当であるか否か検証することも研究価値が高い. また発話機能には発話内容や文構造の影響も大きいと考えられる. そこで現在, 発話内容を限定したデータについてラベラーではない学生数十名によるラベリングデータについて詳細を分析中である.

謝辞 本研究の一部は科学技術振興機構戦略的基礎研究推進事業 (JST/CREST) の援助により行われた.

参考文献

[1] ニック・キャンベル, アラン・ブラック, 信学技報, SP96-7, pp.45–52 (May.1996).
[2] 田窪, 前川ら, 「音声」『岩波講座講座言語の科学』 Vol.2, 1998.
[3] http://feast.his.atr.co.jp/
[4] 芦村, ニック・キャンベル, 人工知能学会全国大会論文集, 3C5-11, (2003)
[5] 田中俊光, 柏岡秀紀, ニック・キャンベル, 音講論, 1-8-26, pp.233–234 (Sep.2003).