

表情豊かな対話音声の感情状態に関連する F0 モデルパラメータの検討*

渡邊諒馬, 森大毅 (宇都宮大)

1 はじめに

本研究では、藤崎モデル [1] のパラメータを分析することによって、対話音声の伝達するパラ言語情報と韻律情報との関係を明らかにすることを目的とする。藤崎モデルでは、指令の生起位置が語彙・構文といった言語情報と同期している一方で、指令の大きさや継続時間などのパラメータが話者のパラ言語的メッセージや心理状態を反映して変化すると予想される。したがって、藤崎モデルのパラメータは、韻律から知覚されるパラ言語情報や心理状態を、 F_0 そのものよりも適切に説明できる可能性がある。

本稿では、表情豊かな対話音声の特徴を多く含んでいる宇都宮大学パラ言語情報研究向け音声対話データベース (UUDB)[2] の発話を分析対象とする。まず、藤崎モデルによる対話音声のモデリングの有効性と問題点を検証するために、UUDB の発話に対して藤崎モデルによる分析を行った。次に、抽出した藤崎モデルのパラメータとパラ言語情報との関係を分析した。

2 藤崎モデルによる分析

表情豊かな対話音声の特徴として、Fig. 1 に例示するような豊富な句末音調が挙げられる。これに対し、対話音声の F_0 パターンに対する藤崎モデルの適用例は少ない。したがって、句末音調を表現する適当な方法を見出す必要がある。本研究では Fig. 1(b) に示すようにアクセント指令を挿入することによって句末音調での F_0 の上昇を表現した。

また、上昇調や下降上昇調の句末音調を含む場合に、句末音調の部分をアクセント指令 1 つでは表現できない発話が数発話見られた。このような発話に対しては Fig. 1(c) や Fig. 1(d) のように連続する 2 つのアクセント指令を挿入することによって表現した。

今回、14 名の話者の 1700 発話についてモデルパラメータを分析した。分析は、言語情報を参照しながら手作業で行った。第 1 著者による聴取では、96% 以上の発話で、分析合成音声と藤崎モデルにより生成された F_0 パターンを用いた合成音声と区別できなかった。

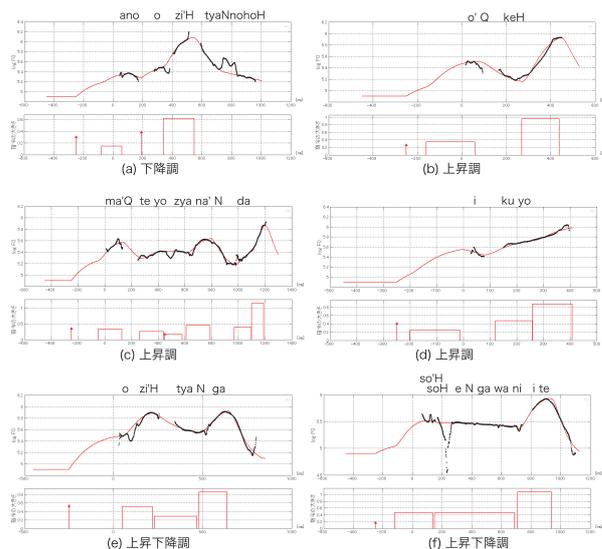


Fig. 1 各句末音調に対する藤崎モデルを用いたパラメータ付与例

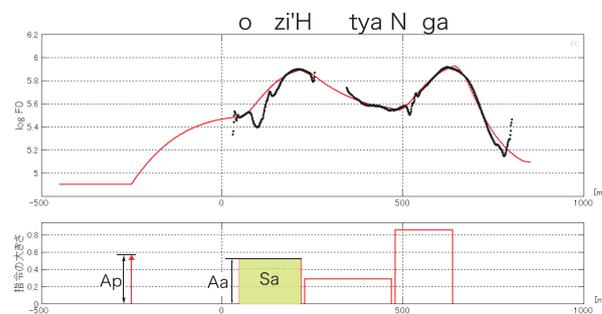


Fig. 2 分析に用いる藤崎モデルパラメータ

3 藤崎モデルパラメータとパラ言語情報との関係の分析

3.1 分析方法

各発話に対して Fig. 2 のようなパラメータを抽出し、パラ言語情報との関係について分析する。ここで、 S_a は $S_a = A_{aj} \cdot (T_{2j} - T_{1j})$ と定義するアクセント指令の面積である。パラ言語情報には UUDB に付与されているパラ言語情報のラベルを用いる。UUDB では、音声から知覚されるパラ言語情報を 6 つの抽象次元に基づいて記述している。本稿では、パラ言語情報の中でも話者個人の感情状態に注目する。

具体的には、話者の気分の良し悪しを表す「快-不快」、および発話者の心理活動の活発さを表す「覚醒-睡眠」の 2 次元を取り上げる。

* Investigation of F0 model parameters related emotional states of expressive conversational speech. by WATANABE, Ryōma, MORI, Hiroki (Utsunomiya University)

Table 1 分析したパラメータと感情状態との相関係数

	快-不快	覚醒-睡眠
A_p 最大値	0.410	0.555
A_a 最大値	0.442	0.651
S_a 最大値	0.359	0.491
A_p 平均値	0.337	0.450
A_a 平均値	0.386	0.576
S_a 平均値	0.272	0.377
フレーズ指令生起率	-0.478	-0.603
アクセント指令生起率	-0.153	-0.177
アクセント指令の幅最大値	0.241	0.278
アクセント指令の幅平均値	0.000	-0.276

3.2 話者 FTS の発話を用いた分析

対話音声に付与した藤崎モデルパラメータとパラ言語情報との関係を見るために、UADB でも特に表情豊かな話者である話者 FTS の 617 発話を分析の対象とした。分析結果を Table 1 に示す。

この結果より、話者 FTS ではフレーズ指令の大きさとアクセント指令の大きさが「快-不快および覚醒-睡眠の各次元」とやや強い正の相関があり、フレーズ指令生起率と「快-不快および覚醒-睡眠の各次元」との間にはやや強い負の相関があることがわかる。

フレーズ指令では指令の大きさ、指令生起率共に「快-不快および覚醒-睡眠の各次元」とやや強い相関があるのに対し、アクセント指令では指令の大きさのみが「快-不快および覚醒-睡眠の各次元」と強い相関がある。このことから、アクセント指令の指令生起率および指令の幅は、パラ言語情報を表現する上であまり影響がないことがわかった。

3.3 発話「うん」に対する分析

次に、「うん」という発話から伝達されるパラ言語情報と藤崎モデルパラメータとの関係を調べるために、発話内容が「うん」である発話のみを用いて分析を行う。発話「うん」の伝達するパラ言語情報は非常に多様であり、本研究の対象として適している。ここでは、14 名の話者の計 1208 発話を用いて分析を行った。分析結果を Table 2 に示す。

Table 2 は各話者で求めた相関係数の平均を示したものである。これより、アクセント指令の大きさとアクセント指令の面積が「覚醒-睡眠」との間に弱い相関を有していることが見てとれる。また、その他のパラメータについてはほとんど相関が見られなかった。

このような結果になった原因として、発話継続時間が大きく影響していると考えられる。ほとんどの発話で継続時間が 500 ms に満たない発話である「う

Table 2 発話「うん」における F_0 モデルパラメータと感情状態との相関係数

	快-不快	覚醒-睡眠
A_p	0.035	-0.092
A_a	0.275	0.358
S_a	0.215	0.313
フレーズ指令生起率	-0.040	-0.079
アクセント指令の幅	-0.033	0.215

ん」では、 F_0 の変化に対してフレーズ指令のみで対応させることができない。この時のフレーズ指令の大きさは非常に小さく、 F_0 パターンの変化をほとんどをアクセント指令によって表現している。このため、フレーズ指令の大きさとパラ言語情報との相関係数が小さくなり、アクセント指令の大きさとパラ言語情報との相関係数が大きくなったと言える。

また、今回分析に使用した発話「うん」は、話者や感情状態によらず発話継続時間がほとんど同じである。このため、フレーズ指令生起率とパラ言語情報との関係を見ることができなかった。

以上の結果から、発話内容が「うん」である発話では、アクセント指令の大きさのみがパラ言語情報の表現に寄与することが確かめられた。

4 おわりに

本稿では表情豊かな対話音声である UADB の発話に付与した藤崎モデルパラメータとパラ言語情報の関係について分析を行った。

その結果、発話「うん」では、フレーズ指令の大きさでパラ言語情報を表現しておらず、アクセント指令の大きさによってパラ言語情報を表現しているということがわかった。また、話者 FTS の発話に対する分析からは、各発話における指令の大きさの平均値よりも指令の大きさの最大値が、パラ言語情報と強い相関を有していることがわかった。今後は、使用するパラ言語情報の数を増やして、藤崎モデルパラメータとパラ言語情報との関係を分析する。

謝辞 F_0 モデルパラメータ分析ツールをご提供下さった千葉大学堀内靖雄准教授と今里圭太さんに感謝します。

参考文献

- [1] H. Fujisaki and K. Hirose, J. Acoust. Soc. Jpn. (E), 5 (4), 233-242, 1984.
- [2] Mori *et al.*, Speech Communication, 53 (1), 36-50, 2011.