

人造キャラクター(ピカチュウ)における感情音声の分析 Analysis of Emotional Speech Uttered by Artificial Characters

長谷川 真澄[†] 榎本 美香[‡] 飯田 仁[‡]
Masumi Hasegawa, Mika Enomoto, Hitoshi Iida

1. はじめに

近年感情情報のコミュニケーション上の重要性から、マン-マシンインターフェースへの応用が考えられている。そういった中で、人間の話す言語として感情音声の付加を目的としての研究は多く行われている。そこで私は人間の言葉を用いない感情音声の研究を目的とし、アニメ「ポケットモンスター」の登場キャラクターである「ピカチュウ」を研究対象として取り上げることにした。

本研究では、ピカチュウの感情音声は音声のみで人に伝わるかを調べ、感情音声を判別する際の要因として、各感情間にある特徴を探っていく。

2. 実験資料

ピカチュウの感情音声をアニメ本編から抽出し、それを桜庭他[1]の研究を参考に、各感情(怒り、悲しみ、驚き、喜び、平静)に分類する。

表-1: 各感情を分類した結果

感情	発話数
怒り	21 発話
悲しみ	20 発話
驚き	44 発話
喜び	30 発話
平静	39 発話

その中から背景 BGM が少なく、聴取実験において感情の判定に影響がでないであろう音声を各感情 8 つ用意した。このとき、平静の感情音声は今回の実験手法では複数用いないため、1 つしか用意しなかった。

また、今回は音声のみによってピカチュウの感情が人間に伝わるかを調べるために、ピカチュウの発話場面での映像を切り取り、ピカチュウの発声部分以外の音声を無音にした。このとき、背景 BGM がピカチュウの発話と被っているものは極力小さくした。

3. 聴取実験

3.1 実験内容

この聴取実験では、ピカチュウの各感情が音声のみで人間に対して伝わるかを調べる。また、実験の結果からピカチュウの各感情音声で、正解率の高い感情と悪い感情を特定し、この後おこなう音声分析のための感情音声の得点をおこなう。

そのための方法として、ピカチュウの感情(怒り、驚き、喜び、悲しみ)をアニメの映像から抽出し、各感情の種類や特徴を調べ、ピカチュウの感情音声にどのような特徴があるのかを明らかにする。

3.2 実験手法

具体的な方法として「平静」の感情を一度聞かせ、その後別の感情を聞かせて平静と比べてどのような感情だったのか評定してもらうことにした。評定する方法として用紙に項目を作り、聞いた感情がどの感情だったかを問い、該当する項目にチェックをしてもらうといったものである。

被験者の合計人数は 23 人である。被験者は全 32 問で各感情(怒り、悲しみ、驚き、喜び)8 個について評定する。これらをあらかじめ適当に並べ替えたものを音声のみで聞かせた。なお、実験をおこなった人全員がポケットモンスター並びにピカチュウを知っていたため、ピカチュウの感情音声を最低でも一度は聞いたことがあることになる。

3.3 実験結果

結果は表-1 のようになった。

表-1: 実験における各勘定の正解率の表

感情	正解率[%]
怒り	89.13
悲しみ	94.56
驚き	62.5
喜び	63.04

各感情の正解率を見てみると、全体的に 60%以上の正解率が出ており、これは有本他[2]の結果と同じであった。さらに、悲しみが 94.56%、怒りが 89.13%と高い数値を出しているが、有本他の研究と違い驚きの感情音声の正解率は低かった。

また、間違いが多い感情は「驚き」と「喜び」であった。さらに、「驚き」以外の感情で被験者全員が正解している問題が最低でも 1 問あった。このことから「驚き」には他の感情よりも編別しにくい感情だということがわかる。また、全 32 問中半数以上間違えた人はおらず、最も多く間違えた人でも 12 問であった。また大半の人の間違いが 10 問以下であり、全体で 3 分の 1 以上の正答率があった。そのため、ピカチュウの感情の判別率は高いといえる。

4. 音声分析

4.1 音声の分析方法

音声の分析は MATLAB[3]でおこない、F0 の高さや大きさを取得する。また、音声分析をおこなうにあたり、情報・システム研究機構の石本祐一さんにプログラムを作ってもらった。今回はそのプログラムを用いて音声分析をおこない、継続時間長[ms](DURATION)、音の高さ[Hz]の最大値(MAX)・最小値(MIN)・平均値(AVERAGE)・変化範囲(RANGE)、音の大きさ[dB]の最大値(MAX)・平均値

(AVERAGE)・変化範囲(RANGE)を取得した。また、これらのデータの種類は有本他の研究で用いられたものを参考としている。

4.2 音声分析の結果

各感情で聴取実験において用いた8つの音声データで音声分析をおこない、それらの音の高さの最大値、最小値、平均値、変化範囲を、音の大きさの最大値、平均値、変化範囲を、分散分析を用いて有意差があるか調べた。

その結果、有意差がみられたのは、音の高さの最大値、音の大きさの平均値と変化範囲であった。音の高さの最大値は $F(8)=9.324$ であり、0.001%水準で有意であった。音の大きさの平均値は $F(8)=4.079$ であり、0.05%水準で有意であった。音の大きさの変化範囲は $F(8)=6.803$ であり、0.005%水準で有意であった。

また、有意差がみられたものについて下位検定を行った。

その結果、音の高さの最大値では、「喜び-悲しみ」間、「喜び-怒り」間、「驚き-悲しみ」間において1%水準で有意であり、「驚き-悲しみ」間において5%水準で有意であった。音の大きさの平均値では、「驚き-悲しみ」間において5%水準で有意であった。音の大きさの変化範囲では「悲しみ-驚き」間において1%水準で有意であり、「悲しみ-怒り」間において5%水準で有意であり、「喜び-驚き」間において10%水準で有意であった。

5. 考察

聴取実験の結果、ピカチュウの感情音声は映像の有無とは関係なく伝わるということがわかった。

また、その判別率も最も低いもので喜びの63.04%と全体的に高い判別率を示した。つまり、ピカチュウの感情音声は映像の有無に関係なく、感情の判別が可能であるということが予想される。

また、その結果から分散分析をおこない、ピカチュウの感情音声を判別するにあたって、音の高さの最大値、音の大きさの平均値と変化範囲が重要となることがわかった。

さらに、各感情間の下位分析をおこなったところ、各感情間の判別の特徴として、音の高さの最大値においては、「喜び-悲しみ」間、「喜び-怒り」間、「驚き-悲しみ」間では、喜び、驚きの音声の方が、怒り、悲しみの音声よりも音が高いため、判別が可能であり、高い有意差がでたものと思われる。また、音の大きさの平均値においては、「驚き-悲しみ」間で、悲しみの音声の方が、平均値が低く、そこから判別しているものと思われる。さらに、音の大きさの変化範囲において、「悲しみ-驚き」間、「悲しみ-怒り」間において、悲しみの変化範囲が大きく、「喜び-驚き」間においては、喜びの変化範囲が大きかった。

これらの結果から、ピカチュウの感情音声を判別する際に、まず有意差が最も多く出た、音の高さの最大値である程度感情を判別し、そこで判別が難しかったものが、音の大きさの平均値と変化範囲において、判別されるものであると思われる。

6. おわりに

本論文では、人間の言葉を用いない感情音声の研究を目的として感情音声の分析をおこない、人間の言葉を用いない感情音声の実装の足がかりとして、ピカチュウの感情音声の分析をおこなった。

その結果、聴取実験において、ピカチュウの感情音声は音声のみで人間に伝わるということがわかった。

しかし、今回の聴取実験では、ピカチュウを知っている日本人のみからしか実験をおこなえなかったため、知らない人や外国人にこの実験をおこなった場合、違う結果がでる可能性がある。

また、分散分析では音の高さの最大値と音の大きさの平均値と変化範囲で有意差がみられ、それらの要因からピカチュウの感情を判別していることがわかった。

しかし、特に怒りの感情音声においてなのだが、今回使用したものに、効果音などの背景ノイズが入ってしまい、それらの影響が出てしまった可能性があるため、今後はよりクリーンな音声での分析をおこなう必要があるものと思われる。

参考文献

- [1] 桜庭 京子, 今泉 敏, 箕 一彦, “幼児・児童の感情表現における発達の研究: 聴取実験の分析——ピカチュウに込められた感性情報——”, 電子情報通信学会技術研究報告, 17-22, (1999).
- [2] 有本 奉子, 河津 宏美, 大野 澄雄, 飯田 仁, “感情音声のコーパス構築と音響的特徴の分析—MMORPGにおける音声チャットを利用した対話中に表れた感情の識別—”, 情報処理学会研究報告, 133-138, (2008).
- [3] MATLAB <http://www.mathworks.co.jp/>

† 東京工科大学大学院バイオ・情報メディア研究科 Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

‡ 東京工科大学メディア学部 School of Media Science, Tokyo University of Technology