

音声における感情パラメータ

6 C-6

小池 和仁 斎藤 博昭 中西 正和

慶應義塾大学大学院 理工学研究科 計算機科学専攻

1. はじめに

人間は、感情を伝えるだけで相手とコミュニケーションをとることができる。また、相手の感情を読み取ることができれば、言葉では逆のことを言っているにもかかわらずその真意を知ることができる。このようにコミュニケーションをとるにあたって、感情の伝達が重要な役割を担っていることは明らかである。コンピュータと、音声によってコミュニケーションをとる際にも、感情の伝達が行われればより円滑なコミュニケーションが実現されることが期待される。

そこで本研究では、平静音声(何の感情も込められていない音声)との比較により、感情を表すために必要とされる特徴を抽出した。さらに、それを用いて感情の込められた音声を再現し、合成された音声から実際にその感情を読み取ることができるかを知覚的に検証した。

2. 感情パラメータの抽出

音声において感情は、話すテンポ、声の大きさ、声の高さによって表される [1]。よって、各々に対応するパラメータとして発話速度、振幅値、基本周波数(ピッチ)を用いた。

特徴量のパターン抽出

DAT-Link[4] を用いて、サンプリング周波数を 24kHz、16bit 量子化によりデジタル化した。窓関数にはハミング窓を用い、512 個分のデータをフレーム長とし、64 個分のデータをフレーム周期とした。発話時間は、音声波形の目視により、感情ごとの長さを比較した。振幅値は、512 個分のデータの 2 乗和の平方根をその区間の振幅値とした。ピッチ抽出には、FFT ケプストラム法を用いた。また、有声音区間の判別には、以下の無音区間と無声音区間の判別法を用いた。

用いた。

無音区間の判別

ある区間の振幅値が 1000 を越えた区間から数えて、9 個目のフレームから有音区間とし、1500 を下回った区間から無音区間とした。

無声音区間の判別

ある区間の零交叉数が 150 を越える区間は、無声音区間とした。また、高ケフレンシー部における最大値を高ケフレンシー部の平均値で割った値(この値を K とする)が 8.0 を越えている場合に、それをピーク値とし、ピッチとして抽出した。

図 1 に有声音区間判別の手順を示す。ここで、1 つ前のフレームが無音区間なら $C=0$ 、有音区間のなら $C=1$ である。

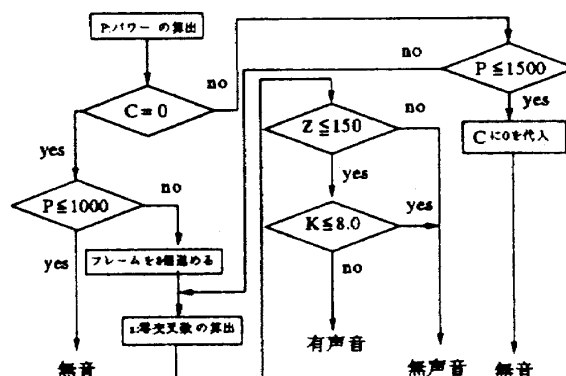


図 1: 有声音区間の判別

3. 音声合成による特徴の検証

規則合成方式の 1 つである波形合成方式による音声合成システムにより、平静音声との比較によって得られた感情固有の特徴を再現した。このシステムは、様々な感情を表現できるように、ピッチ、振幅、音節ごとの時間長を自由に設定できるものである。

4. 実験結果及び考察

実験条件

22 歳から 30 歳の男女 20 人を対象とし、合成音声から「怒り」「驚き」「悲しみ」「嫌悪」「喜び」の感情

[3]を読み取ることが出来るかを調べた。本研究では、サンプルとして人名を用いた。人名であれば、どの感情を含めても発声し易いし、語自体に感情が含まれないと判断したからである。

実験

実験は2種類行った。実験1としては、まず、平静音声を聞いてもらい、その後ランダムに選択した3つの音声について、上記の5感情のどれに感じるかを解答してもらった。実験2として、まず平静音声を聞いてもらい、次に5感情全ての音声を聞いてもらい(順番はランダム)、相対的に5つの音声がどの感情に当てはまるかを解答してもらった。

実験結果

表 1: 実験 1 の結果 (%)

	怒り	嫌悪	悲しみ	驚き	喜び
怒り	75.0	0.0	0.0	16.7	8.3
嫌悪	0.0	58.3	0.0	41.7	0.0
悲しみ	0.0	0.0	100.0	0.0	0.0
驚き	0.0	0.0	0.0	50.0	50.0
喜び	16.7	0.0	0.0	0.25	58.3

表 2: 実験 2 の結果 (%)

	怒り	嫌悪	悲しみ	驚き	喜び
怒り	95.0	5.0	0.0	0.0	0.0
嫌悪	5.0	85.0	0.0	10.0	0.0
悲しみ	0.0	5.0	95.0	0.0	0.0
驚き	0.0	0.0	0.0	50.0	50.0
喜び	0.0	5.0	5.0	40.0	50.0

表の見方としては、表1を例にとると「嫌悪」の音声を「嫌悪」と正しく知覚した確率が58.3%、「嫌悪」の音声を「驚き」と誤認識した確率が41.7%となる。

考察

「怒り」と「悲しみ」の感情については、実験1においても実験2においても正反応率が高く、合成音声の感情表現に用いたパラメータ値が各々の感情を表すのに必要かつ有用なものであることが確認できる。実験1において「嫌悪」の間違いが「驚き」に集中し、「驚き」の間違いが「喜び」に集中しているのはそれぞれに似た特徴が含まれているからであると考えられる。そのため、平静の音声との比較によるのみ感情を評価する実験1においては、共通した特徴

を持つ正解ではない感情を解答してしまうことがあると考えられる。実験1と比較して、実験2における「怒り」と「嫌悪」の感情の正反応率は大幅に上昇した。これは、他の感情の音声と比較しながら聞くことにより、その音声を持つ特徴がよりクローズアップされたことによるものと考えられる。このことから、付加した特徴は間違っただけのものではなく、それをより顕著に表現することにより正反応率がよくなることが予想される。

両実験において、「驚き」と「喜び」の正反応率は低く、それぞれの感情を表現するために有効な特徴が抽出できたとは言えない。この結果については、以下のような推測をすることが出来る。

「喜び」であるか「驚き」であるかを決めかねるなどの解答者も、2つの音声の違いは分かるのだが、それぞれがどちらの感情に当てはまるかを迷ってしまうと言っていた。このことから、感情による音声の性質の違いはあるものの、日常、人間がこの2つの感情を感じるときは話者の態度や表情とともに感じるときが多いため、音声だけによって判断することには不慣れであり、そのために間違っただけの解答をしてしまったことが1つの理由として考えられる。また、2つの感情が同居していたために、どちらか一方の感情には決めかねるということも考えられる[2]。一方で、この2つの感情を言い当てた解答者の中には、自分がその感情を含めて発声するとしたらどうなるかを考えたら、正解であったという人もいた。このことを考えると、この2つの感情を表現するために用いた特徴は間違っただけではないものの、それがあまり顕著に現れていなかったことが考えられる。

参考文献

- [1] 古井貞照: デジタル音声処理東海大学出版会, (1985)
- [2] 広瀬 啓吉, 高橋 登, 藤崎 博也, 大野 澄雄: 音声の基本周波数パターンにおける話者の意図・感情の表現, 信学技報 HC 94-41, pp.33-40(1994)
- [3] 平賀 裕, 齊藤 善行, 森島繁生, 原島 博: 音声に含まれる感情情報抽出の一検討, 信学技報, HC 93-66, pp.1-8(1994)
- [4] DAT-Link <http://www.tc.com/>