E-036

会話ロボットにおける感情を表現した音声合成
Speech Synthesis Expressing Emotions for Communication Robots

島部 貴由† 吉村 枝里子‡ 土屋 誠司‡ 渡部 広一‡
Takayoshi Shimabe Eriko Yoshimura Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年ロボットの技術は進歩し，主に産業分野において，様々な用途で利用されるようになっている．今後，求められるロボットは産業分野だけでなく，日常生活の中で人のパートナーとして活動するロボットである．そのためには，ロボットが人にとって扱いやすく操作が簡単であることが求められる．そのようなロボットの実現には会話をインターフェースとし，ロボットが円滑なコミュニケーションをできることが必要とされている．また，円滑なコミュニケーションの一つとして感情表現が挙げられる．つまり，文に感情を込めることで，人間のような発話が可能であると考える．

本研究では，ロボットに人間のような発話を行なわせるために感情を表現した音声で発話させることが目的である．ロボットが会話を行う際，聞き手に意図や感情を想起させるような韻律で発話を行うことができれば，人間が理解しやすい会話を行うことができると考える．例えば，話し手が明るく話すと話し手は喜んでいると感じたり，発話の際に語尾が上がれば，疑問文だと理解しやすい．このような韻律を用いて音声合成する手法を提案する．

2. テキスト音声合成ソフト

本研究はテキスト音声合成ソフト AITalk ver2.32 SDK[1]（以下 AITalk と呼ぶ）を用いて実験を行った．AITalk は，文章に応じて，音韻の並びや声の高さ・音韻の長さ等の条件が最も良く適合し滑らかにつながる音韻を選び出し，それらの音声波形を結合して合成音声を出力する．このため，従来の信号処理による合成音声と比べてより人間的な表現が可能である．アクセントや開始周波数，中間周波数，発話の速さや大きさのパラメータが用意されており，各パラメータはプログラミングにより変更できるため本研究では AITalk を用いた．しかし，AITalk では感情表現はできないため合成音声は感情のない平叙文しか出力できないという問題がある．そこで，上記のパラメータを変更することで，韻律を自在に変えて感情を表現した合成音声を生成する手法を提案する．また，各パラメータのことを本稿では，音声の物理的特徴と呼ぶ．

3. 提案手法

人間は会話を行う際に，感情のない平叙文で話し続けることはあまりない．会話をしているとき，会話の流れから感情が生じる場合がある．その場合，感情によって人間は声の韻律を変えて発話することで感情を表現する．従って，ロボットが発話をする際に，会話にふさわしい応答文を生成し，表現すべき感情がある場合，その感情に対応する韻律で発話をすることで人間のような発話を行うことができると考える．また，会話における音声から生じる感情を考える際，音声によって伝えられる感情には，話し手自身が表現した「話し手の感情」と，音声の聴き手が音声から受け取る「聴き手が推測する話し手の感情」（以下「聴き手の感情」と呼ぶ）の 2 つが考えられる．「話し手の感情」は，必ずしも「聴き手の感情」と一致するとは限らず，また，音声から断定することはできない．これに対し，「聴き手の感情」は，「話し手の感情」とは無関係に，音声のみから判断し推測する．従って，合成音声に含ませた感情が，聴き手に伝わるような音声を合成する規則を求めるためには，「聴き手の感情」と音声の物理的特徴の関係を定式化する必要があると考える．その物理的特徴により導出した韻律を用いて音声合成することで聴き手に「話し手の感情」と一致する感情を想起させる．

感情を想起させる音声の物理的特徴を求めるために以下の実験を行った．実験には，邦画 1 作品を用いた．大学生の被験者 15 名に映画の任意の 315 文を聞かせ，被験者に“喜び”“悲しみ”“怒り”“恐れ”“驚き”“嫌悪”“感情なし”“該当なし”の 8 個の感情の中からどのように聞こえたかを選ばせた．“感情なし”とはその文に感情はないと感じたことを示し，“該当なし”とは感情を感じたが上記の感情ではない別の感情であることを示す．映画には BGM や背景音があり機械的に韻律の周波数を取り出した場合、雑音がかなり含まれると考えられるため，今回は映画の文を聞き，その文を AITalk を用いてできる限り相違なく聞こえる様に上記の実験により人手で再現した．

映画の各文には，感情が付加されているため，各感情ごとに音声の物理的特徴を分類することができる．AITalk で再現した音声の物理的特徴を洗い出すことで感情ごとに韻律を定式化する．

実験により得られた各会話文の感情の分布を表 1 に示す．15 人全員が示した感情を含む文と 10 人以上が示した感情を含む文を抽出した．

文の数 / 感情	10 人以上	15 人	10 人未満
喜び	27	6	288
悲しみ	24	4	291
怒り	47	11	268
恐れ	2	0	313
驚き	18	1	297
嫌悪	4	0	311
感情なし	26	0	289
該当なし	0	0	315

表 1 各感情に対する文の数

† 同志社大学大学院理工学研究科
Graduate School of Science and Engineering, Doshisha University
‡ 同志社大学理工学部
Faculty of Science and Technology, Doshisha University

実験結果から“喜び”“悲しみ”“怒り”が特に顕著に感じとられやすいことがわかった．なお，“恐れ”“驚き”“嫌悪”は表 1 よりサンプル数が少ないため今回は議論しない．よって，前者の 3 つの感情の韻律を音声合成の際に用いる．その際の物理的特徴を以下の表 2 に示す．表 2 は音声における各物理的特徴が感情を含まない平叙文と比べた際にどのような相違が見られたかを示している．

表 2 の結果は一般的に心理学的見地からも正しいとされている[2]．

物理的特徴 / 感情	周波数	速度	大きさ
喜び	高い	少し速い	変化なし
怒り	高い	速い	大きい
悲しみ	低い	遅い	小さい
疑問文	語尾は高い	変化なし	変化なし

表 2　感情ごとの音声の物理的特徴

疑問文の際に語尾を上げて読むことは人間が無意識的に行っており，聞き手もその表現を経験的に理解している．そこで，それらの人間が話す際に使っている表現を用いることでより人間のような会話を実現できると考える．疑問文を表現するために物理的特徴としては語尾の周波数を上げる韻律で音声合成した（表 2）．

4. 評価

“喜び”“怒り”“悲しみ”と疑問調の物理的特徴により定式化した韻律を映画とは異なる平叙文に適応した．感情を表現する韻律を含んだ平叙文と感情を含んでいない平叙文では差があるかを比較し，実際に韻律を変えた場合に，どのように感情が伝わるかの評価実験を行った．被験者は大学生 5 名とし，評価セットとして，感情を含まない“おはよう”と“今日は寒い”という平叙文，“今日は楽しかった”という平叙文でも楽しいという快感情を含む文を用いた．

韻律を変えて音声合成した文と韻律を変えていない平叙文をワンセットとし，どのように聞こえたかという回答，疑問文の場合には疑問文または肯定文という回答を求めた．被験者に先入観を与えないために回答は自由筆記形式とした．なお，感情は 2 グループに大別できる．それは“快感情”と“不快感情”であり，前者には“喜び”“楽しみ”“満足”などが属し，後者には“悲しみ”“怒り”“恐怖”が属する[2]．そこで，その回答を「快」，「不快」，「感情なし」とまとめた．

結果としては“喜び”の場合は“快”，“怒り”の場合は“不快”，“悲しみ”の場合は“不快”と回答する人数が多いと考えられる．表 3 に“喜び”の韻律で合成した音声がどのように聞こえたかの結果，表 4 に“怒り”の結果，表 5 に“悲しみ”の結果，表 6 に語尾を上げる韻律の結果をそれぞれ示す．

表 3　感情「喜び」の韻律における結果

感情「喜び」の韻律	快	不快	感情なし
おはよう	5	0	0
今日は寒い	2	0	3
今日は楽しかった	5	0	0

表 4　感情「怒り」の韻律における結果

感情「怒り」の韻律	快	不快	感情なし
おはよう	1	4	0
今日は寒い	0	5	0
今日は楽しかった	4	1	0

表 5　感情「悲しみ」の韻律における結果

感情「悲しみ」の韻律	快	不快	感情なし
おはよう	1	4	0
今日は寒い	0	5	0
今日は楽しかった	3	2	0

表 6　語尾を上げる韻律における結果

疑問調	疑問文	肯定文
おはよう	1	4
今日は寒い	2	3
今日は楽しかった	1	4

5. 考察

表 3 では 15 人中 12 人，表 4 では 15 人中 15 人，表 5 では 15 人中 15 人が「なし」以外の回答を行った．つまり，文の音声の物理的特徴を変更し，感情を表現することで，なんらかの感情を想起したと考えられる．

また，“喜び”の場合に“快”の回答人数が多く，“怒り”と“悲しみ”の場合に“不快”の回答人数が多いことを確認した．

しかし，表 4 で「今日は楽しかった」に“怒り”の感情の韻律を付与して音声合成しても，“快”に聞こえるという回答が多かった．表 5 の“悲しみ”についても同様のことが言える．これは，感情を含む平叙文に関しては，感情がその文自体に依存してしまうため回答が逆になったと考えられる．

表 6 より，疑問調については平叙文の語尾の周波数を上げるだけでは疑問文としては伝わりにくいという結果となった．

6. 終わりに

本研究では，音声の物理的特徴を変えて音声を合成することで聴き手に感情が伝わることを確認した．しかし，合成音声に用いる韻律の種類が，“怒り”“悲しみ”“喜び”の 3 種類の感情しか検討できていない．加えて各感情における音声の物理的特徴を導出するために映画から用いる文のサンプル数が少なく，映画の文の周波数を抽出することによる具体的な数値データによる裏付けができていないという問題点がある．以上の問題点を解決することで今後さらに人間のような発話に近い音声合成ができると考える．

謝辞

本研究の一部は，科学研究費補助金（若手研究（B）24700215）の補助を受けて行った．

参考文献

[1] 株式会社エーアイ,AITalkSDK,http://www.ai-j.jp/sdk2,2012/6/21 アクセス

[2] Campbell, N. and D. Erickson ”What do People Hear? A Study of the Perception of Non-verbal Affection Information in Conversational Speech”,『音声研究』第 8 巻 1 号,pp.9-28,2004.