

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 26 日現在

機関番号：11301

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26540138

研究課題名(和文) 笑いを誘発する音声メディアの特徴分析に基づいた笑い誘発音声フィルタの試作

研究課題名(英文) Preliminary study on how to induce laughing by controlling acoustical parameters of speech signal

研究代表者

坂本 修一 (SAKAMOTO, Shuichi)

東北大学・電気通信研究所・准教授

研究者番号：60332524

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：笑いは円滑なコミュニケーションを実現する情報であるだけでなく、人間のQoLを高める重要な感性情報ともなる。本研究では、笑いを引き起こす音声メディアの特徴を工学的に明らかにし、その特徴を抽出・操作可能な合成フィルタの構築を目指した。聴衆が通常聞いたことのない予期しない音声を聞いたときに笑いが誘発されるという仮説を立て、基本周波数と話速を変化させた音声を用いて実験を行ったところ、通常音声に比べ話速が遅く基本周波数が高い音声が聴衆に対して面白さを誘発するという結果が得られた。このことから、これらのパラメータを操作可能なフィルタを構築することで、聴衆に笑いを誘発する音声を作成できる可能性が示された。

研究成果の概要(英文)：Laugh is a very important cue to communicate with each other smoothly and friendly. In this study, we focused on the acoustical characteristics of speech signals and tried to investigate how to induce laughing by controlling the acoustical parameters of speech signals. We hypothesized that people laugh when they listen to speech which has different acoustical characteristics compared to those of normal speech. As stimulus, speech signals uttered by a male and a female speakers were used. The fundamental frequency and speech rate of the signals were controlled as parameters. The results of paired comparison tests indicated that laughing occurs when people hear speech with different acoustical characteristics from original. Moreover, speech with a high fundamental frequency and slow speed induced laughing much more effectively than other types of speech. Based on this knowledge, we can generate speech signals which induce various kinds of emotional response.

研究分野：音声認知工学

キーワード：感性情報学 認知科学 ユーザインタフェース 情報システム

1. 研究開始当初の背景

笑いは、単に場を和まして円滑なコミュニケーションを実現する情報であるだけでなく、肉体的・精神的な安定を維持するといった観点から人間の QoL を高める重要な感性情報となる。このことは特に社会の高齢化が急速に進展する現在、コミュニケーション促進により地域コミュニティの活力を生み出し、延いては健康的な生活の源となる要素の一つであるともいえる。

ここ数年で、音声に関連するこのような高次感性情報の取得・生成技術は急速に進展してきた。音声を含むパラ言語情報（音声に含まれる意図や感情などの情報）と物理パラメータとの関連の解明や、様々な歌声音声の特徴抽出・合成といった技術はこの典型である。

音声の持つこのような感性情報の取得・生成技術が確立することで、発話者からの音声をその音声を持つ感性情報が正しく伝わるように物理的に精度高く伝送するだけでなく、送り手が送りたい感性情報を、使用する様々なメディア・システムの適性・制約にあわせて加工・編集して伝送・提示するようなシステム、すなわち、「感性創成型コミュニケーションシステム」の実現が見えてくる。

しかし、現状の研究を概観すると、現在多くの研究者が工学的に扱っている感性・情動情報は、Ekman が定義した基本 6 感情（喜び、嫌悪、驚き、悲しみ、怒り、恐れ）がほとんどであり、聴取者が受けるであろうより高次な感性情報は体系化がなされていない。高次感性情報を自由に操作することにより、バーチャル環境においても質の高いコミュニケーションが可能となるであろう。

2. 研究の目的

本研究では、高次感性情報として笑いに着目する。既に記述した通り、笑いはコミュニケーションを円滑に行う上で極めて有効なツールであり、QoL の向上という点で必要不可欠な要素である。様々なメディアを通して伝送される笑いの構成要素を考えると、伝達される音声を含む言語的な特徴もあるものの、例えば、同じ文章を素人が読む場合とプロが読む場合で、内容が同じにも関わらず受け手が受ける印象がまったく異なるものとなるように、パラ言語情報が極めて豊富であり、それに関連する物理的な特性にも大きな違いがあることは明白である。

本研究ではこのような視点に立ち、笑いを引き起こす音声を持つ様々な物理特徴量を定量的に分析し、その特徴を抽出・操作可能な合成フィルタの構築を目的とする。最終的には本研究で構築した合成フィルタを通常音声に適用し、聴取実験を通して、合成フィルタの有効性を検証する。

この研究が進展することにより、これまで職人や限られた人間でのみ扱うことの出来た高感性メディアに対しての理論的な裏付けを与え、日本からこれまで以上に多くの高

感性メディアの創出の実現が期待できる。

3. 研究の方法

様々なメディアを流れる笑いを誘発する音声素材を調査し、どのような要素が笑いを引き起こすのかを分析して、操作すべき物理パラメータを決定する。特に、パラ言語情報だけでなく言語情報の影響も考慮し、言語情報の取得度合いを操作した際の誘発される笑いの変化から、笑いに影響を及ぼす言語情報とパラ言語情報の影響度の差を明らかにする。その分析結果に基づき、笑いに関連する物理パラメータを精緻化し、笑いを引き起こすために必要な物理パラメータの値を算出し、笑い誘発フィルタの設計を進める。

平成 26 年度は、様々なメディアを通して素材となる笑いを誘発する音声を集める。対象として落語や漫才などの笑いを引き起こすメディアを集めるほか、日本語だけでなく外国語のメディアについても情報を収集する。それとあわせ、音声メディアだけでなく話者の映像にも着目し、話者映像の音声情報知覚に与える影響について、笑いという観点だけでなく、その基盤となる音声の聞き取りや記憶という観点からも検討を進める。

平成 27 年度は、平成 26 年度の検討に基づいて、実際に音声の持つ物理パラメータを操作した実験刺激を構築し、知覚心理実験をおして、笑いを引き起こす物理パラメータの範囲を明らかにする。その際に、パラ言語情報だけでなく言語情報の影響も考慮し、笑いに影響を及ぼす言語情報とパラ言語情報との関連を定量化する。

4. 研究成果

本研究では笑いを誘発するメディアの音響的な特徴を明らかにすることを目的としており、そのためには、文脈から発生する笑い、物理的な音響特徴量から誘発される笑いの両者を何かしらの形で分離する必要がある。そこでまず、笑いを誘発する音声素材を様々なメディアを通して収集し、それらから得られる印象を予備的に検討した。その結果、笑いを誘発する上では文脈の影響が極めて強いこと、発声される音声だけでなく、その音声に対してタイミング良く付加された笑い声など本研究が対象とするターゲットとなる音声の音響特徴量以外の物理特性の影響も大きいこと、などの結果が得られた。メディアに収録されている音声はこれらの要因が含まれてしまっていることから、本研究の目的を達成する上で、音声メディアから手がかりを得ることは必ずしも効率的ではないことが示唆された。そこで、聞いて面白いと笑う音声の有している特徴について改めて考え、例えば男性から女性のような音声を聴取した場合などのように、聴衆が通常聞いたことのない、予期しない音声を聞いたときに笑いが誘発されることを前提として研究を進めることとした。これは、単に音響特

微量の観点だけでなく、文脈についても当てはまることと思われる。

このような過程に基づき本研究では、音声の基本周波数と話速をターゲットとして扱い、これらをパラメトリックに変化させた音声を聴取した際の聴取者の印象を主観評価させ、笑いを引き起こす上で必要となる音響特徴量の範囲を分析した。

実験では、ATR 音素バランス 503 文 (NII 音声資源コンソーシアム) から、文脈から笑いを誘発されることのないニュートラルな文章 (「冷房では、冷え過ぎが、問題になる」) を男性と女性が発声したものを刺激として用いた。刺激の基本周波数と話速は、男性、女性それぞれで、76~168 Hz と 162~361 Hz, 5.8 mora/s と 5.3 mora/s であった。この音声のピッチと話速を音声分析ソフトウェア WORLD (Morise *et al.*, IEICE Trans. Info. Sci., 2016) を用いてパラメトリックに変化させた (基本周波数: 男性 -0.5, 1, 2, 4, 8 倍, 女性 -0.25, 0.5, 1.0, 2.0 倍, 話速 (男女とも): 0.25, 0.5, 1.0, 1.5, 2.0, 2.5 倍)。これらの音声を男性、女性のそれぞれについて、全ての組み合わせで一対比較させ、面白さを 5 段階で比較評価させた。

結果の例を図 1~4 に示す。各図中の $Y_{0.05}$, $Y_{0.01}$ は危険率 5%, 1% のヤードスティックで、それぞれ、男性は 0.22, 0.25, 女性は 0.22, 0.24 である。図を見るとわかるとおり、ほとんどの場合において、話速や基本周波数を変化させることで、元々の話速や基本周波数を持つ音声に比べ、聴衆に面白いという印象が与えられていることが見て取れる。このことから、実験を行う際に立てた、聴衆が通常聞いたことのない予期しない音声を聞いたときに笑いが誘発されるという仮説が裏付けられた。

さらに得られた結果を細かく見てみると、男性、女性のいずれに対しても、話速が遅く、基本周波数が高い条件でもっとも面白さが誘起されるという結果となった。メディア等で比較的耳にするものは、倍速再生などで聞くことの多い話速が早く基本周波数が高い音声や、スロー再生で聞くことになる話速が遅く基本周波数が低い音声である。これらの音声は、他の音声に比べ、聴衆にとってある程度なじみがある音声として知覚され、その結果、面白さがそれほど上がらなかった可能性が考えられる。その一方で、話速が遅く基本周波数が高い音声と、話速が早く基本周波数が低い音声は、いずれも我々にとって耳にする機会が少ないが、基本周波数が低い音声は悲嘆の感情を話者が持っている印象を与えるという研究結果もある (Murray and Arnot, 1995) ことから、今回の実験結果が示すように、話速が遅く基本周波数が高い音声は最も面白さを聴衆に印象づけたのではないかとと思われる。

その一方で、多感覚情報処理という観点から視聴覚音声情報の統合メカニズムについ

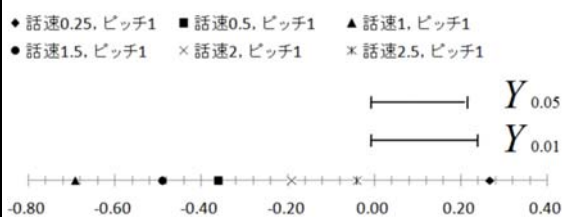


図 1: 基本周波数 1 倍の男性音声の印象評価 (パラメータ: 話速)

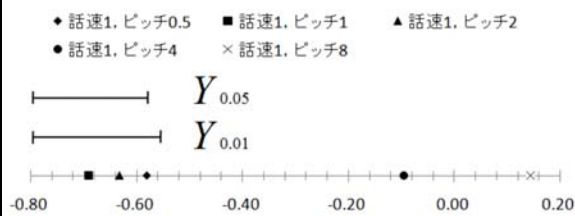


図 2: 話速 1 倍の男性音声の印象評価 (パラメータ: 基本周波数)

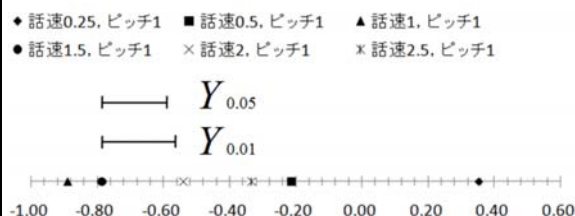


図 3: 基本周波数 1 倍の女性音声の印象評価 (パラメータ: 話速)

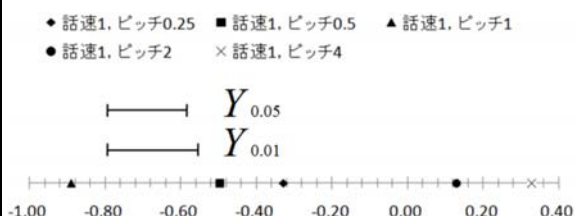


図 4: 話速 1 倍の女性音声の印象評価 (パラメータ: 基本周波数)

ても実験を行い、音韻の聴き取りやその記憶という場面において、視覚情報と聴覚情報との齟齬が影響を及ぼすという結果が得られている。この結果は、単に音韻に関連する内容にとどまらず、笑いなどの感性情報知覚にも影響を及ぼすと考えている。

以上の結果から、話速と基本周波数を操作することで、本研究の目的である聴取者に笑いを誘発する音声を作成できる可能性が示された。得られた知見は、QoL の向上といった本来の目的以外にも、例えば自動対話システムやロボットなどの音声合成技術への適用も可能である。操作がうまくいかなかったユーザの感情を何らかの形で察知して、その感情や不快感を和らげるといった目的で、笑いを誘発するように音声を変化させるといったことが考えられる。

ただし、今回実験で得られた笑いがどのようなものであるかについては、笑いというも

の自体の分類も含めてより詳細な分析が必要である。嘲笑や蔑みといった負の印象を誘起させて面白いといった反応を返した可能性も否定できず、笑いの知覚そのものをより具体的に明らかにした上でさらなる実験を行う必要であろう。

本研究は笑いという高度な感性情報の操作、制御を目指す野心的な研究であったが、今回操作した音響特徴量以外にも笑いに寄与するであろうパラメータはいくつか考えられる。例えば、しわがれ声などの声質も影響を及ぼす要因であろう。また、聴衆が予期しない音声という観点から考えると、話者の映像との印象の不整合も大きな要因と考えられる。これらの要因を体系立てて検討した上で、多感覚情報処理の観点も考慮した上で本研究を発展させていくことが、今後重要であると考えている。

5. 主な発表論文等

〔雑誌論文〕(計3件)

- [1] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “音声により提示される系列再生課題に話者映像が与える影響,” 日本音響学会講演論文集, 2-Q-24, 531-532 (2015)
(査読無)
- [2] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “話者映像が音声系列再生課題に与える影響,” 電子情報通信学会技術研究報告, HIP2014-76, 47-50 (2014)
(査読無)
- [3] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “無関連な話者映像が音声系列再生課題に与える影響,” 電子情報通信学会技術研究報告, SP2014-86, 59-63 (2014)
(査読無)

〔学会発表〕(計5件)

- [1] T. Ohtani, S. Torai, S. Sakamoto and Y. Suzuki, “Effects of a movie of speaker's irrelevant speech utterance on recall tasks,” 16th International Multisensory Research Forum (IMRF), 2015年6月13~16日, Pisa, Italy
- [2] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “音声により提示される系列再生課題に話者映像が与える影響,” 日本音響学会2015年春季研究発表会, 2015年3月16~18日, 中央大学(東京都文京区)
- [3] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “話者映像が音声系列再生課題に与える影響,” 電子情報通信学会ヒューマン情報処理(HIP)研究会, 2014年12月1~2日, 東北大学(宮城県仙台市)
- [4] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “無関連な話者映像が聴覚短期記憶に与える影響,” 多感覚研究会, 2014年11月12~13日, 広島大学(広島県広島市)
- [5] 虎井駿, 太谷智子, 坂本修一, 鈴木陽一, “無関連な話者映像が音声系列再生課題に与える影響,” 電子情報通信学会音声

研究会, 2014年10月23~24日, 南紀白浜温泉ホテルシーモア(和歌山県西牟婁郡白浜町)

6. 研究組織

(1)研究代表者

坂本 修一 (SAKAMOTO, Shuichi)
東北大学・電気通信研究所・准教授
研究者番号: 60332524

(2)研究分担者

佐藤 洋 (SATO, Hiroshi)
国立研究開発法人産業技術総合研究所・ヒューマンライフテクノロジー研究部門・グループ長
研究者番号: 10260423

大谷 智子 (OHTANI, Tomoko)
東京藝術大学・芸術情報センター・助教
研究者番号: 40422406