

要旨

本研究の目標は、非言語的な情報が表現豊かな音声の知覚においてどのような役割を演じているか、特に、文化と言語が異なる人々の場合に非言語的な情報がどのように知覚されるか、という質問への解答を見つけることである。

これを実現するために、本研究では表現豊かな音声の知覚における 2 つの仮説を提案する。

一番目の仮説は、リスナーがある音声がある感情カテゴリに属するのを決める前に、別の記述、たとえば形容詞の形を持つ記述を音声に与える、ということである。

2 番目の仮定は、文化と言語が異なる人々であっても、表現豊かな音声の知覚に対していくつかの共通の特性を持つ、ということである。

この 2 つの仮説を記述するために、人間の認知には自然な曖昧さが存在するという事実に基づいて、多層構造の知覚モデルとして仮説を表現する。このモデルでは、表現豊かな感情カテゴリ（つまり、最上位の層）の知覚は、直接に特定の音響特性（つまり、一番下の層）の変化によるものではなく、むしろ知覚的に音声を表現する記述子（つまり、中間層）を組み合わせた結果によるものであるとする。本論文では、これらの形容詞の形を持つ記述子をセマンティックプリミティブと称する。この 3 層モデルにおいて研究の焦点となるポイントは、2 つの関係、つまり最上位の層と中間層との関係と中間層と一番下の層との関係の構築とその評価である。この後、このモデルを表現豊かな音声における文化と言語が異なる人々の知覚に関して共通点と相違点を見つけることに適用する。上記した研究の目標を達成するために、提案したモデルを知覚実験によって構築し、モデルから生成したルールに基づいて音声をモーフィングすることによりモデルの検証を行い、そしてモデルにより非言語的な情報の分析を行う。

構築プロセスでは、まず表現豊かな音声を表現するための記述子（すなわち、形容詞）を見つけるためにセマンティックプリミティブの選択を行う必要がある。適切なセマンティックプリミティブが選択されるためには、セマンティックプリミティブと音響特性との関連をサポートするための解析を行う。また、音響特性と表現豊かな音声に対する知覚の関連が曖昧であるため、これを表すためにファジー推論システムを採用する。このプロセスを実行することにより、構築されたモデルから以下の現象が確認された。すなわち、表現豊かな音声の知覚プロセスでは、リスナーは、ある音声がある感情カテゴリに属すると判断する前に、その音声をいくつかの記述子の組み合わせで知覚し、さらにこの組み合わせられた記述子によって音声の感情カテゴリを判断する。

上に述べた二つの表現豊かな音声の知覚の関連を検証するために、ルールに基づく音声モーフィングのアプローチを採用する。まず構築されたモデルをルールに変換する。この変形のルールは、音響特性の解析とファジー推論システムから作成する。そして **STRAIGHT** 音声合成システムにより、ルールを用いて平静音声を各感情音声に変換した。この検証プロセスが構築プロセスをさらに一歩進め、モデルの有効性をより明らかにした。

最後に、このモデルを文化と言語が異なるリスナーの知覚特性の記述に用いた。この結果、文化と言語が異なる人々の知覚に対しても、同一の非言語情報が表現豊かな音声知覚において同様の役割を持つことが明らかとなった。

本研究の結果として、モデルの構築と検証プロセスから、一つ目の仮説が検証できた。さらに本研究の手法を適用することにより、二つ目の仮説の可能性を示唆する一例を示した。