

抑揚の変化が日本語音声の聞き取りやすさに与える影響

The Effect of Prosodic Variation on the Intelligibility of Japanese Speech

平野 達也, 中谷 裕教

Tatsuya Hirano, Hironori Nakatani

東海大学

Tokai University

4cjinm020@tokai.ac.jp

概要

本研究では、合成音声における日本語の聞き取りやすさ向上を目的に、抑揚の大きさが聞き取りやすさおよび自然性に与える影響を被験者 10 人で検討した。聴取実験とアンケート調査を通じて、抑揚の変化による聞き取りやすさの違いを確認した。また、抑揚が大きい条件では音声の好ましさが聞き取りやすさに強く関与する傾向が見られた。話し方から受ける印象を評価した好悪という因子は、自然性に対しても有意な影響を与えていた。

キーワード: 合成音声, 抑揚, 聞き取りやすさ

1. はじめに

近年、音声合成技術の飛躍的な進展により、合成音声は日常生活の中で非常に身近な存在となっている。スマートフォンの音声アシスタント、駅やバスでの自動アナウンス、読み上げソフトやカーナビゲーション、さらには視覚障がい者支援における情報提供手段としても、合成音声は多様な場面で活用されている。一方で合成音声の発話は肉声と比べた場合、聞き取りにくいと示唆されている(三谷, 2014)。日本語においては、高低アクセントを中心とする抑揚の役割が大きく、抑揚や間の取り方が意味理解や感情表現において重要な機能を果たしているためであると考えた。そこで、本研究では合成音声における「抑揚」に着目し、その変化が聞き取りやすさや自然性、ひいては聴取者の主観的評価に与える影響を検討した。

本研究では、日本語音声における抑揚変化の操作が聴取者の受ける印象にどのような影響を及ぼすかを実証的に明らかにする。特に、単なる「明瞭性」や「発音の正確さ」といった技術的指標ではなく、

「聞き取りやすさ」や「自然性」、さらには「好悪」や「流暢さ」といったより主観的な側面に対して検討を行うことにより、人間にとって快適な音声の条件とは何かを多角的に把握することを目指す。これにより、音声知覚における認知メカニズムの理解を深め、聴覚特性に即した音声モデルの構築に貢献できると考える。

2. 方法

文語的な構造を含む短編小説の冒頭部分から抜粋した、約 15 秒程度の長さのテキストを使用した。抑揚以外の外的要因が評価に与える影響を抑えるため、音声合成には VOICEVOX を用い、同一話者の声質を維持したまま抑揚のみを操作した。

話者の抑揚の値を変化させた 4 種類の音声を用意した。デフォルト (1.0 倍)、0.5 倍、1.5 倍、2.0 倍の条件を設定した。これらの音声を被験者に聴取後、各条件について七件法により主観評価を行った。アンケートは山住ら (2005) を基に作成した。評価項目は「聞き取りやすさ」、「自然性」、「好悪」、「上手さ」、「速さ感」の 5 つのカテゴリーに分類した。好悪は「心地よい—不快な」「感じの良い—感じの悪い」「好きな—嫌いな」「親しみやすい—親しみにくい」の 4 対語、上手さは「なめらかな—しどろもどろな」「話し慣れた—話し慣れていない」「上手い—下手な」「流暢な—たどたどしい」の 4 対語、速さ感は「速い—遅い」「スピード感のある—ゆつたりした」「せわしげな—のんきな」「落ち着きのない—落ち着きのある」の 4 対語で評価した。

これらの評価スコアをもとに、「聞き取りやすさ」および「自然性」を目的変数、その他の評価項目を説明変数として、重回帰分析を行った。実験には 20 代の男女 10 名が参加し、防音室内にて実施された。

表1 抑揚2倍の聞き取りやすさの重回帰分析

	標準偏回帰係数	標準誤差	p値
上手さ	0.134	0.033	0.170
好悪	0.315	0.033	0.001
速さ感	0.036	0.043	0.708

3. 結果

アンケートに対して分散分析を行ったところ、抑揚倍率における聞き取りやすさには有意差が認められた ($p<0.001$)。この結果は、合成音声の抑揚を操作することが、聞き手の聞き取りやすさの知覚に有意な影響を与えることを示している。しかし、その後の多重比較として1.0と倍率を変更した抑揚の音声間で t 検定を行ったところ、1.0倍と抑揚1.5倍の間には有意差は確認されなかった ($p>0.05$)。一方で、自然さについては、分散分析、 t 検定共に有意差が認められた。加えて、聞き取りやすさと自然さの間には相関関係は確認されなかった。

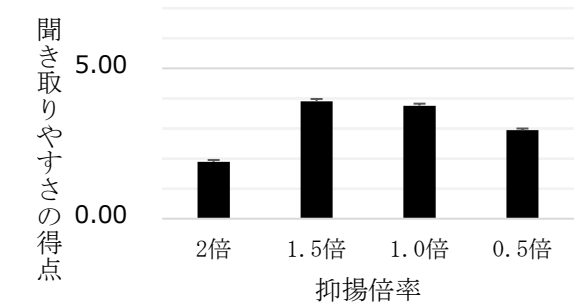


図1 倍率ごとの聞き取りやすさ

重回帰分析の結果、抑揚が2.0倍に設定された条件において、「好悪」が「聞き取りやすさ」に対して少量の影響を与えていた ($\beta = 0.315, p < 0.05$)。また、1.5倍の抑揚条件でも同様の傾向が見られ、「好悪」と「聞き取りやすさ」の間に安定した関係性が存在する可能性が高まった ($\beta = 0.321, p < 0.05$)。一方で、1.0倍条件では「好悪」は「自然さ」に対して有意な影響を示し、特に抑揚が極端に変化していない場合には印象の良さが音声の自然さとして知覚される傾向があることが示された。逆に、抑揚が0.5倍の条件ではすべての評価項目のスコアが低下しており、平坦すぎる音声聞き取りにくく、自然性や好感度を損なう原因となる可能性がある。

4. 考察

本研究を通じて、音声の設計において音響的な明瞭性だけでなく、印象評価、特に好悪といった要素が重要であることが示唆された。たとえば、同じ情報内容であっても、「感じの良い」「好きな」音声はそれだけで聴取者の注意を引き、内容理解を促進する可能性がある。

今回の結果から、合成音声においては、デフォルトの設定よりも抑揚を多少上げた方が聞き取りやすさが向上する可能性があることが示唆される。しかし、聞き取りやすさと自然さが常に強く相関するわけではないという点も考えられる。これは、合成音声という性質上、自然さという評価項目自体が、人間の肉声に対する基準と同一に適用されにくいという、本研究のアンケート設計における課題を示唆している。合成音声の「自然さ」は、必ずしも肉声にどれだけ近いのか、という一義的な尺度で測れるものではなく、聞き手が「違和感なく受け入れられるか」という側面がより重要であると考えられる。

今後としては、被験者の人数を増やすことで統計的信頼性の向上を図る必要がある。また、本研究では抑揚のみを操作したが、「話速」「間」「語尾処理」といった他のパラ言語的要素や、抑揚をより詳細な段階に分けて操作した場合の聞き取りやすさへの影響も組み合わせで調査することにより、さらなる聞き取りやすさ・快適さの向上が期待できると考える。これらの検討を通じて、合成音声における「聞き取りやすさ」と「自然さ」の複雑な関係性をさらに深く解明し、より人間中心の音声インタフェース開発に貢献していく必要がある。

文献

(1) 三谷雅純. “生涯学習施設の館内放送はどうあるべきか: 聴覚実験による肉声と人工合成音声の聞きやすさの比較.” *人と自然 Humans and Nature* 25 63-74. (2014)

(2) 西本卓也, 他. “早口合成音声の聴取実験によるテキスト音声合成の評価.” *信学技報, WIT2005-5*, (2005).

(3) 山住賢司, 他. “講演音声の印象評価尺度.” *日本音響学会誌* 61.6 303-311. (2005)