

公共放送における合成音声利用に関する基礎的検討

(指導教員 世木 秀明 准教授)

世木研究室 1131134 早野 和也

1.はじめに

公共施設のアナウンスは聞き取りやすい音声求められる。しかし、アナウンサーを常駐させるのは難しく、素人のアナウンスでは聞き取りにくくなると考えられる。このような場合の対応策としてアナウンサーの音声をもとにして開発された音素片編集型音声合成プログラムが市販されており、防災情報の発信などを目的に2014年9月時点で172の自治体や放送局に導入されている。しかし、合成品質が良いとされる合成音でも適切なアクセントやイントネーションの自動的な付与は難しく、そのまま合成しただけでは聞き取りやすい合成音声とは言い難い。

そこで本研究では、素人音声、アナウンサー音声に加え、アクセントやイントネーションを調節した合成音を用いて音声の聞き取りにくさや了解度について調査し、公共施設放送に合成音声を使用した場合、聞き取りやすく理解しやすい放送とするための検討を目的とした。

2.聴取実験

2.1.実験用刺激

第1文と第2文の内容が関連を持つ文章を8文章、聴取者に計算を要求する文章を8文章作成した。これらを女性アナウンサー1名、アナウンス素人の女性1名、音素片編集型規則合成プログラムVoice Textにより合成した音声を用意した。Voice Textにより合成した音声についてはアクセントやイントネーションの付与方法を標準設定で出力したものと、不自然さを軽減するようにアクセントやイントネーションを主観で調整したものの2種類を用意した。さらに、実験用刺激は実環境を考慮し、用意した音声にラウドネスバランスから6dB減じたマルチトーカーノイズを重畳させたものとした。

2.2.実験方法

実験は静かな部屋で行い、被験者の耳元に至適レベル(約70(dBA))でスピーカから実験用刺激を発話者ごとに提示した。被験者には、提示刺激の内容に関する質問に解答させるとともに、聞き取りにくさについての4段階評価をさせた。

被験者は健康な聴力を持つ20代の男女とし、各実験用刺激につき15名、合計60名とした。

3.実験結果

提示した実験用刺激の種類ごとの正答率の違いを図1に示す。図1に示すように素人音声、アナウンサー音声、およびアクセントやイントネーションを調整した合成音は、未調整合成音声と比べて高

い正答率を示した。各音声間の有意差検定を行ったところ、未調整合成音声は、有意水準5%で有意に正答率が低い結果となった。さらに、不自然さを軽減するようにアクセントやイントネーションを主観で調整した音声と素人音声間には有意差が見られなかったが、アナウンサー音声間には有意水準5%で有意差が認められた。また、聞き取りにくさに関しては、音声間に有意な差は見られなかったが、図2に示すように素人音声やアナウンサー音声、調整合成音声は未調整合成音声と比べて聞き取りやすいという傾向が見られた。

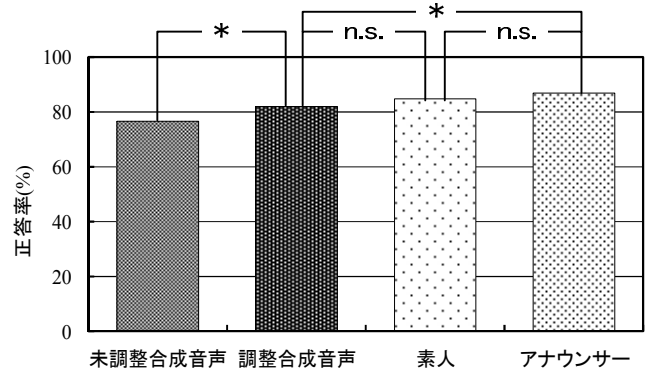


図1 提示刺激ごとの正答率の比較

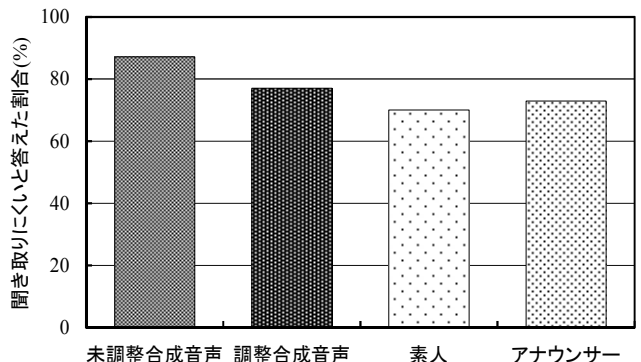


図2 聞き取りにくいと答えた割合

4.まとめ

聴取実験結果から、アナウンサー音声最も聞き取りやすい結果となったが、アクセントやイントネーションの適切な調整を行った合成音声でも80%以上の高い正答率が得られた。このことから、適切なアクセントやイントネーションの付与が聞き取りやすさや了解度の向上に貢献すると考えられた。

これらのことから、合成音声をを用いて聞き取りやすく理解しやすい公共放送を行うためには、アクセントやイントネーションに関して適切な設定を行う必要があると考えられる。