

# 公共施設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 0731130 東出 雄大

## 1 はじめに

公共施設のアナウンスにはプロのアナウンサーの音声を使用することが効果的であると考えられる。しかし、緊急時などプロのアナウンサーがアナウンスできない場合に素人がアナウンスを行うと必要な情報を伝えきれない可能性があると考えられる。

このような場合、入力した文字列を高品質な合成音声に変換できるプログラムがあれば公共施設でのアナウンスなどに役立つのではないかと考えられる。

そこで本研究では、合成音声の聞き取りにくさについて調査し、公共施設でのアナウンスに利用できるかについて検討を行うことを目的とした。

## 2 音声合成システム

本研究で使用した音声合成プログラムは、現在市販されている音素片編集型規則合成プログラムで特に、高品質に音声合成が可能であるとされている XIMERA と VoiceText の 2 種類である。

本研究で使用した 2 種類の音声合成プログラムは、入力された文字列を日本語文法に基づき、あらかじめアナウンサーが発話した音声から抽出した音素片を適切に接続し、合成するものである。このため、音素片の接続手法が合成音の品質に大きく影響を与える。

## 3 聴取実験

### 3.1 実験用刺激

刺激材料として、第 1 文と第 2 文間に意味的關係がある文章(文脈効果有り文章)、無い文章(文脈効果無し文章)をそれぞれ 24 文章作成し、これらを女性アナウンサー 1 名、関東方言の女性話者 1 名、2 種類の合成音声プログラムで読み上げた音声を用意した。また、話速は約 300 文字/分とした。さらに、聴取実験用にすべての音声に対して第 1 文と第 2 文に関する簡単な質問を作成した。

実験用刺激は、実環境を考え刺激材料にマルチトーカー・ノイズを重畳させたものと重畳させないものとした。ここで、ノイズ重畳レベルは、音声とノイズのラウドネスバランス(0dB)からノイズレベルを 12dB 減じたものとした。

### 3.2 実験方法

実験は静かな部屋で行い、被験者前方 150cm、50cm にそれぞれスピーカとモニタを置き、聴覚刺

激、問題文を提示した。被験者には、音声提示後にモニタに表示される質問と聞き取りにくさに関する調査に回答させた。ここで被験者は、聴力に異常のない 20 代男女 20 名、60 歳以上の高齢者 17 名とした。刺激音声の提示レベルは、被験者耳元で約 70dB SPL とした。

## 4 実験結果と考察

ノイズ無し条件では、20 代男女、高齢者ともに聴取実験正答率は、アナウンサー、素人、VoiceText、XIMERA の順に低下する傾向がみられ、聞き取りにくさの評価は XIMERA、VoiceText、素人、アナウンサーの順に評価値が高かった。また、文脈効果「有り」、「無し」の正答率について検討すると、両被験者群ともに文脈効果「無し」では第 1 文の正答率は第 2 文より有意に低下する傾向が見られた。しかし、高齢者では文脈効果「有り」の場合でも、第 2 文に対する第 1 文の正答率に有意な低下が観測された。これは、高齢者は 20 代男女とは異なり、文脈情報がかうまく使えないためではないかと考えられた。

一方、ノイズ付加条件では図 1 に示すように 20 代男女の正答率はノイズ無し条件よりも低下するものの音声の種類による正答率の順位は同様であった。一方、高齢者ではノイズ無し条件に比べ正答率が大きく低下し、VoiceText がアナウンサーや素人音声と同程度の正答率を示した。

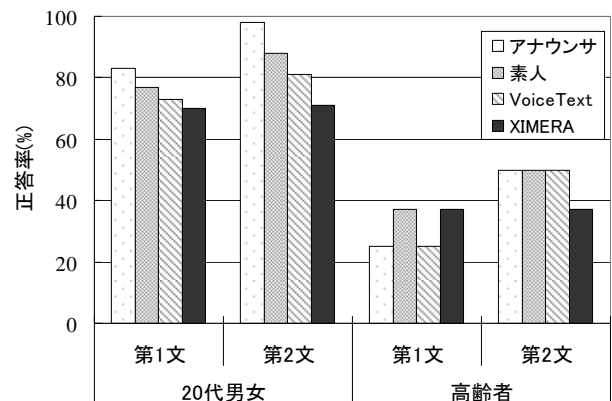


図 1 ノイズ付加条件の正答率(文脈効果あり文章)

## 5 まとめ

実験結果より、音声の品質が良い合成音声であれば、公共施設でのアナウンスに利用できるのではないかと考えられた。また、背景雑音に加えアナウンス文の文脈が聞き取りに大きく影響を与えるのではないかと考えられた。