

# 雑音除去音声の客観評価に関する研究

学籍番号：90115140 飯國研究室 牧田和馬

## 1 はじめに

現在までに提案されている種々の雑音除去システムは、その出力として得られた雑音除去音声の評価することで、システムの有効性を主張することが多い。しかし問題は、雑音除去音声の評価方式が統一されていないことである。そこで本研究では、雑音除去音声を耳で聴くという観点から、最も有効と思われる評価方式を提示することを目的とする。

## 2 雑音除去音声の評価方式

既存の雑音除去音声の評価方式は、主観評価と客観評価に大別できる。主観評価は、実際に被験者に音声を聴かせて、個人の主観によって得点を決める方式であり、最も人間の感覚を反映した評価値が得られる。ただし、個人による評価値のばらつきを抑えるためには、ある程度の人数を確保しなければならない。しかし、被験者の数が多くなると、手間や費用がかかるという問題がある。一方、客観評価は、評価式に音声データを代入することによって得られるものであり、1つの音声データに対して1回の測定で評価値が決まるという特長がある。また、手間や費用がほとんどかからず、測定者による評価値のばらつきも生じない。しかし、音声の聞き取りやすさなど、人間の感覚については、必ずしも正しい評価が得られるとは限らない。以下では、主観評価といくつかの客観評価方式を比較し、主観評価に最も近い評価値を与える客観評価方式を明らかにする。主観評価と同様の結果を与える客観評価方式が存在すれば、最小の手間と費用で、統一的に雑音除去システムを評価することができる。

## 3 客観評価と主観評価との相関

客観評価には非常に多くの方式があり、すべての方式について主観評価との類似性を調べることは難しい。そこで本論文では、雑音除去を目的とした論文の中で、引用率が高かった6つの客観評価方式についてのみ検討した。また、比較対象とする主観評価には、代表的なDMOS (Difference Mean Opinion Score)[1]を用いた。DMOSとは、雑音除去前と除去後の音声の差分に対して、主観による得点を決め、その平均をとる方法である。まず、DMOSによる評価値を得るために、被験者30名に対して主観評価を実施した。実施の際には、周囲の雑音の影響を低減するため、音声をイヤホンで聴かせた。雑音除去音声は除去前と比べて、良好になったと感じた場合は、7を最高点として、5~7の3段階で評価させた。また、変化がないと感じた場合は4、悪くなったと感じた場合は、1を最低点として、1~3の3段階で評価させた。使用した音声は男声3種類、女声3種類に、白色雑音、車内雑音、バブル雑音をそれぞれ付加したものとした。また、雑音除去音声を得るために、スペクトルサブトラクション法[2]、MAP推定法[3]、騒音再合成法[4]の3種類の雑音除去法を用いた。各手法についての説明は、本論文の目的ではないので省略する。

主観評価と客観評価の類似性を調べるために、相関係数を求めた。相関係数は次式で定義した。

$$\text{相関係数} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

表 1: 客観評価の DMOS 値との相関係数

客観評価方式	相関係数		
	白色	車内	バブル
SNR	0.53	0.92	0.84
セグメンタル SNR	0.05	0.91	0.71
スペクトル推定誤差	-0.38	0.90	0.02
板倉-斉藤距離	0.24	0.57	0.69
LPC ケプストラム距離	0.14	0.06	-0.46
改良パークスペクトル距離	0.23	0.31	0.12

ここで、 $x_i$  は DMOS 値、 $y_i$  は客観評価値である。また、 $x_i$  の平均を  $\bar{x}$ 、 $y_i$  の平均を  $\bar{y}$ 、評価値の数を  $n$  としている。表 1 に、それぞれの背景雑音に対する、6つの客観評価の結果と DMOS 値との相関係数を示す。ただし、論文での引用率が高かった順に客観評価方式を示している。表 1 より、いずれの雑音に対しても DMOS 値との相関が最も高い客観評価方法は、SNR であることがわかる。また、SNR 以外の客観評価は、雑音の種類によってばらつきがあるため、雑音の種類が複数ある場合は、有効な評価値を得ることが難しいと考えられる。よって、本研究の結果からは、雑音の種類によらず、人間の感覚を最も反映した客観評価方式は、SNR であるという結論が導き出せる。ただし、本結果は数種類の音声データに対して得られたものであるため、SNR がいかなる場合でも、最も有効な客観評価方式であるとするには、他の種類の雑音や雑音除去法を用いた、さらに詳細な検討が必要である。

## 4 結論

本研究では、雑音除去音声を耳で聴くという観点から、最も有効と思われる客観評価方式について検討した。客観評価の結果と DMOS 値との相関から、白色雑音、車内雑音、バブル雑音に対して、SNR が高い相関を示すことが明らかとなった。ただし、いかなる場合でも SNR が最も有効であることを示すにはさらに多数の音声、雑音データによる検証が必要である。

## 参考文献

- [1] 守谷健弘, 音声符号化, コロナ社, 東京, 1998.
- [2] S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-27, no.2, pp.113-120, April 1979.
- [3] T. Lotter, P. Vary, "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super Gaussian Speech Model," EURASIP Journal on Applied Signal Processing, pp.1120-1126, July 2005.
- [4] 溝川直樹, 川村新, 飯國洋二, "騒音再合成法の性能改善に関する研究," 第 50 回システム制御情報学会研究発表講演会論文集, 6F2-6, pp.445-446, May 2006.