

ACE 法による中国語音声の聴取改善技術開発 Development of Technology for Improving Listening Comprehension of Chinese Speech by ACE Method

ヨウ ケイ ガイ[†]
Jingkai Ye

阪田 治[†]
Osamu Sakata

1. はじめに

騒音環境下の健聴者の聞き取りを調べた研究より、音が大きな場合では良好に聞き取れるが、音が小さくなるにつれ急激に聞き取れなくなることが分かる。この聴覚特性は高齢者に多い感音性難聴者の聴覚特性に類似している。^{[1][2]}

本研究室では人の内耳における非線形な増幅特性に基づいて音声信号を処理する振幅帯域圧縮伸張法(ACE 法: Amplitude-bandwidth Compression/Expansion)の開発を行った。先行研究では、日本語を対象とする ACE 法の補聴処理が、感音性難聴者、及び騒音下における健聴者の聴取改善の有効性が確認されており、他言語に対しての有効性がまだ確認されていない。

本研究では ACE 法を中国語への適用方法を検討し、主に騒音下における健聴者の中国語の聴取精度改善を目的とする。

2. ACE 法による補聴処理

ACE 法は、音圧レベルと人の感じる音の大きさの関係に基づいて聴取改善を行う音声処理法である。式(1)に ACE 法の処理式を示す。

$$w = \text{sgn}(v)k_a|v|^n \quad (1)$$

w は出力音圧の瞬時値、 v は入力音圧の瞬時値、 sgn は符号関数、 k_a は線形増幅係数、 n は圧縮伸張係数である。

ACE 法の入出力特性を図 1 に示す。圧縮伸張係数 $n=1$ の無処理時に比べ、 n が小さいほど、相対的に小さい音が大きく増幅される。感音性難聴者と騒音環境下における健聴者の聴覚特性により、ACE 法の聴取改善に期待できる。また、先行研究により $n=0.6$ で聴取精度が最も高くなり、本研究も圧縮伸張係数 n を 0.6 とした。 $n=0.6$ のときの処理後音声と元音声を図 2 に示す。

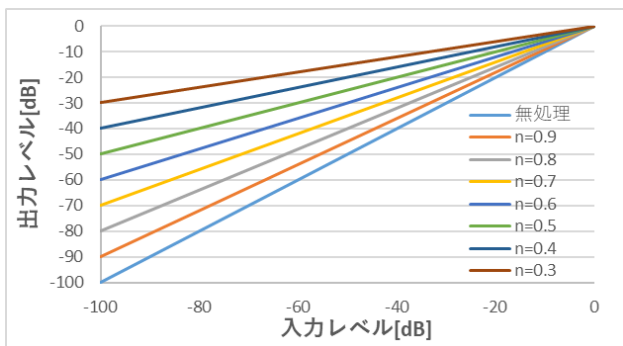


図 1 ACE 法の入出力特性

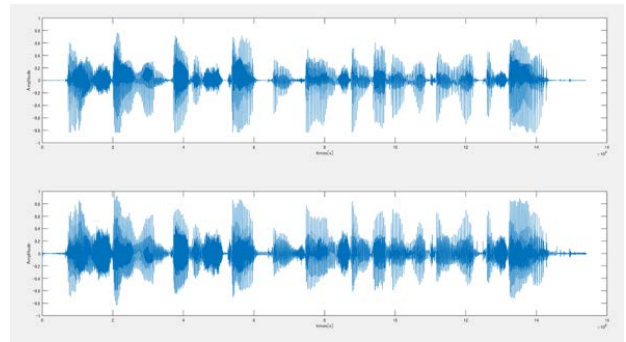


図 2 元音声(上)と ACE 処理音声(下)

3. 日本語と英語、中国語の違い

先行研究および本研究で使われた日本語、英語、中国語の 3 言語の特徴について比較する。まず母音子音の数について、日本語、英語、中国語の母音の数はそれぞれ 5, 16, 36 で、子音の数がそれぞれ 16, 25, 21 である。次に音節のパターンについて、日本語と中国語は、音節が一般的に母音を中心に、母音で終わる開音節であり、一つ一つの音節ははっきりしているに対し、英語の音節の多くは複数の子音を中心に、そして子音で終わる閉音節である。また、アクセントについて、英語は強弱アクセントを持つ言語であり、日本語と中国語は高低アクセントを持つ言語である。その中で、日本語のアクセントは高低 2 パターンしかなく、そして単語ごとのアクセントは文中の位置、イントネーションによって変化するため、意味との結びつきが弱いと考える。一方、中国語は文字ごと、つまり音節ごとに意味を持たせた 4 種類の高低アクセント(声調)が存在し、そのアクセントを弱くしたり、高くしたりすると意味が大きく変化し、アクセントが言葉の意味への結びつきが強いと考える。中国語における 4 種類の声調の音長に対しての音の高さを図 3 に示す。

4. 提案手法

従来手法では入力された音声を分割せず、全体に同様な処理を行うが、第 3 節の日本語と中国語の違いから、分割せずに処理を行うことは中国語音声のアクセントを弱くしてしまい、処理後音声の聞き取りにくくなる可能性が大きいと考えた。

したがって、日本語に有効とされる ACE 法を中国語へ応用するのに、まず入力された中国語音声の音節ごとの高低アクセントを認識する必要があると考えた。従来の ACE 法に加え、Kaldi-ASR(DNN-HMM を使った音声認識モデルの構築に使われるツール)をベースとした Montreal Forced Aligner(以下 MFA)を導入した。^{[3][4]}入力した中国語音声を MFA で強制アライメントを行い、音節

[†] 東京理科大学 Tokyo University of Science

を音素単位まで分割する。次に音節ごとのアクセントを認識し、そのアクセントの特徴に適した ACE 法をかける。アクセントを認識しそれに適した処理法をかけたことで、本来のコトバのアクセントの高低さを保ちつつ、音声全体の音量を上げることを実現した。

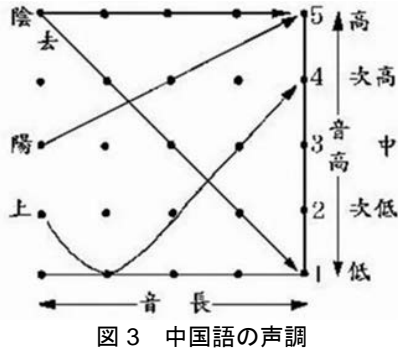


図 3 中国語の声調

5. 実験

5.1 実験方法

まず、同様かつリアルな騒音環境を再現するため、繁華街の環境音音声データを被験者に配布した。次に中国語を母国語とする健聴者の男女 7 名に、環境音を 100% の音量で再生しながら、処理前の中国語音声と提案手法で処理された処理後音声をそれぞれ 30% の音量で聞き比べてもらい、音声内容の聞き取り安さを 5 段階評価(1 が最も聞き取りにくく、5 が最も聞き取りやすい)で記録し、合計 12 個の音声についてそれぞれの評価平均点を計算した。なお、本研究で使用した音声はすべて中国語音声データセット「THCHS-30」^[5]より抜粋したものである。なお、情勢により、本実験はオンラインで実施し、各被験者の音響設備などによって聞こえ方に差異が生じる可能性があると考えられる。

5.2 実験結果

実験結果を図 4、表 1 に示す。6 つの音声データの処理前後の平均評点から、原音声の全平均評点が 1.42 であるに対し、提案した手法で処理された音声の全平均評点が 3.01 点に上がり、処理後音声は平均 1.59 点向上し、聴取精度の改善が見られた。しかし、表 1 からわかるように、処理前後の平均評点の向上がみられたが、4 点の評価が少なく、5 点の聞きやすかったという評価はなかった。

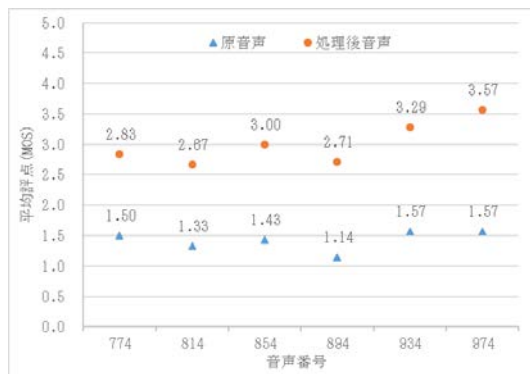


図 4 処理前後の音声の平均評価

最高点の 5 点をつける人いなかった理由として、一つは騒音に対する音声の音量が小さい、つまり SN 比がかなり小さいため、聞きやすくする処理に限界があるから。一つは ACE 処理にあたって音声の分割を細かく行ったため、音声に分裂感が生じ、快適さが足りないからと考えられる。

表 1 処理前後音声の聞き取り安さ評価

被験者番号 音声番号	被験者番号							平均評点	改善度
	A	B	C	D	E	F	G		
774_o	2	1	1	2	2	1	2	1.50	+1.33
774_ace	3	3	2	3	4	2	3	2.83	
814_o	1	1	1	2	2	1	1	1.33	+1.34
814_ace	3	2	3	3	3	2	3	2.67	
854_o	1	2	2	1	2	1	1	1.43	+1.57
854_ace	3	3	4	2	3	3	3	3.00	
894_o	1	1	1	1	2	1	1	1.14	+1.57
894_ace	3	2	3	3	3	2	3	2.71	
934_o	2	1	2	2	1	1	2	1.57	+1.72
934_ace	4	3	3	3	3	3	4	3.29	
974_o	1	1	2	2	1	2	2	1.57	+2.00
974_ace	4	4	3	4	3	4	3	3.57	

6. おわりに

本研究室で開発された ACE 法に中国語音声認識システムを加え、中国語の特徴である声調(アクセント)に合わせた ACE 処理法を提案した。音声を声調に適した補聴処理にすることによって、騒音環境下における健聴者の中国語音声の聴取精度を向上することが確認できた。今後はより中国語の特性に沿った帯域分割、圧縮伸張係数の変更などを検討し、処理された音声の違和感をなくすための処理を行っていく予定である。

謝辞

本研究を終えるにあたり、研究を進める上で終始丁寧にご指導いただいた阪田治教授に厚くお礼申し上げます。また、実験にご協力いただいたすべての皆様に心より感謝いたします。

参考文献

- [1] J.P.A. Lochner, J.F. Burger, "Form of the Loudness Function in the Presence of Masking Noise", vol.33, pp.1705-1707, (1961).
- [2] B.C.J. Moore et al, "A loudness model for impaired hearing", JASA, vol.106, p.901, (1999).
- [3] 音声認識ツール Kaldi, <<https://kaldi-asr.org/>>
- [4] McAuliffe, Michael, Michaela Socolof, Sarah Muhuc, Michael Wagner, Morgan Sonderegger, "Montreal Forced Aligner: trainable text-speech alignment using Kaldi", Proceedings of the 18th Conference of the International Speech Communication Association. (2017)
- [5] Chinese Speech Corpus by CSLT@Tsinghua University, <<http://www.openslr.org/18/>>