



# マイクスルーイヤホンを補聴器として利用するための 適応的なリアルタイム話速変換システムの構築と評価

Adaptive real-time speech rate modification for using microphone-through-earphones as hearing aids

高木周平<sup>1)</sup>, 岩田将幸<sup>1)</sup>, 坂野秀樹<sup>2)</sup>, 柳田康幸<sup>2)</sup>

Shuhei TAKAGI, Masayuki IWATA, Hideki BANNO, and Yasuyuki YANAGIDA

1) 名城大学 理工学研究科 情報工学専攻

(〒468-8502 名古屋市天白区塩釜ロー丁目 501 番地, {193426005, 183426002}@ccalumni.meijo-u.ac.jp )

2) 名城大学 理工学部 情報工学科

(〒468-8502 名古屋市天白区塩釜ロー丁目 501 番地, {banno, yanagida}@meijo-u.ac.jp )

**概要**：難聴によるコミュニケーション上の困難が、抑うつや認知機能低下などに関与することが報告されている。本研究では、外音波形を増幅・加工して提示するだけでなく、PICOLA (Pointer Interval Controlled OverLap and Add) を用いた話速変換機能を有するインテリジェント補聴器の実現を目指す。リアルタイムの会話へ適用できるように、話者音声全体の大幅な伸長を避け子音部のみを話速変換する手法により可聴性向上を図る CE (Consonant Elongation) システムを構築し、評価した。

**キーワード**：聴覚、コミュニケーション、インテリジェント補聴器、話速変換

## 1. はじめに

人間は言語を持ち、音声のやり取りとして会話を行う。会話は人の最も効率的な情報伝達・意思伝達的手段として考えることができ、人とコミュニケーションをとることに繋がる。しかし、聴覚障害者や難聴者にとって音声を聞き取ることは困難である。そのため、聴覚障害者や難聴者は、人とスムーズなコミュニケーションができず、会話することが億劫になる。難聴の存在が、抑うつ、認知機能低下、要介護または死亡等のリスク増加に関与することが報告されている [1]。

聴覚障害者や難聴者が快適に私生活を送るためのツールとして補聴器が存在する。補聴器とはすべての面で完璧に聴力を補完するものではないため、装用者の聴力レベルや使用状況など深く考慮して選択する必要がある。また、ジャパントラック 2018 [2] によると、補聴器を使わない理由として「わずらわしい」「補聴器を装用することが恥ずかしい」など、補聴器に対する世間体が悪いと多くの人が感じている点が大きな問題といえる。

これらの問題に対し、外音取り込み機能が搭載されるイヤホンやヘッドホンを補聴器として用いることで見た目からしても着用しやすい補聴器として利用できるのではないかと考えた。その上で、取り込んだ外音をそのまま流すのではなく、話速変換技術の PICOLA (Pointer

Interval Controlled OverLap and Add) [3]を応用することで、その場での会話を聞き取り易くできる可能性がある。実際に NHK では、話速変換技術を用いて高齢者や聴覚障害者にも聞きやすい番組を届けるシステムが実用化されている [4]。しかし、実際の人との会話において話速変換技術を用いた場合、音声遅延に起因して生じる話者の交替のしにくさが問題となる。

本研究では、外音波形を増幅・加工して提示するだけでなく、PICOLA を用いた話速変換機能を有するインテリジェント補聴器の実現を目指す。リアルタイムの会話へ適用できるように、話者音声全体の大幅な伸長を避け子音部のみを話速変換する手法により可聴性向上を図るシステムを構築した。また、本システムにより可聴性の向上がされたかについての被験者実験を行い、評価した。

## 2. 話速変換技術

「話速変換」とは、声の高さや質を保ちながら、任意の倍率で音声の速度を変換する技術のことである。話速を変換するだけなら、アナログテープで可変速再生をするなどにより話速を変化させることができる。しかし、同時に声の高さや質も大きく変化するため、元の高さや質を維持するには、音声波形の基本周期を保存する必要がある。

1970 年代に音声の効率的な伝送を目的に TDHS (Time Domain Harmonic Scaling) アルゴリズム[5]が開発された。この技術は周波数軸上での操作によって実現したものである。その後、品質の向上などを目指し、様々な方式が開発されている。

話速変換技術には PICOLA[6]や NHK 方式[7], STRAIGHT 分析合成[8]等が研究されてきた。PICOLA は波形をピッチ周期単位で扱えるため、処理が加えやすい。また、処理がシンプルなため、リアルタイム処理に適している。これらの理由から本研究では PICOLA のアルゴリズムを応用して音声の一部に対して伸長処理を行う。

2.1 PICOLA

PICOLA では、音声を音声の基本周期単位で分析し、自己相関関数を用いて処理を行い、ピッチを抽出する。この PICOLA は、長さ  $T_p$  の探索窓を用いる場合、探索位置から  $T_p$  シフトした箇所との差分を計算し、探索開始位置から  $T_p - 1$  地点までの差分の合計から差分平均値の計算を行う。これを窓の大きさを変化させながら、差分平均を比較し、最も差分が小さい窓の大きさが抽出ピッチとなる。

長さ  $T_p$  の探索窓を用いて音声の基本周期を抽出し、図 1 のように音声波形を (A), (B), (C) に分ける。ただし (A), (B) はそれぞれ音素の基本周期を表し、(C) はそれ以降の音声波形を表す。(A) は 0 から 1, (B) は 1 から 0 となるように重みをつけた長さ  $T_p$  のオーバーラップ波形を生成し、(A) と (B) の間に挿入することで音声を伸長することができる。

3. システム概要

実際の会話において、話速変換技術により会話速度を遅くした場合、元の音声発話時間よりも必然的に音声発話時間は長くなる。そのため、話し手と聞き手が順番に交替することで成立する会話では、音声遅延に起因して生じる話者の順番交替のしにくさが大きな問題となる。そこで本研究では、音声遅延を低減するために音声の子音部のみを話速変換するシステムを提案する。本システムを CE (Consonant Elongation) システムと名付ける。図 2 は、「アイキヤク」を、子音部のみ話速を 1.6, 2.4 倍伸長

した音声波形である。また、CE システムでは、母音部「a」「i」「u」「e」「o」以外に、「q」(小文字の『つ』) も無変換とした。

4. 音声聞き取り実験

4.1 実験方法

音声聞き取り実験は得られた単語正解率から PICOLA による子音部の話速伸長バランスの検討を行う。図 3 に示すように、実験は防音室内 (30 dB) にて行われ、被験者は配置された椅子に座り、ヘッドホン (audio-technica ATH-M30x) を装着してもらった。ヘッドホンから出力される音声を聞き、被験者は聞き取った音声を回答用紙に書き出す。その書き出した回答用紙をもとに、単語正解率から単語理解度を得る。被験者は日本語を母国語とした健聴な 20 代男性 10 名であった。

実験は図 4 のように、伸長倍率 4 種の聞き取り実験を行い、その後原音の聞き取り実験を行なった。どちらの聞き取り実験も練習セッション (5 単語) を実施してもらった後に、実験セッションを実施した。

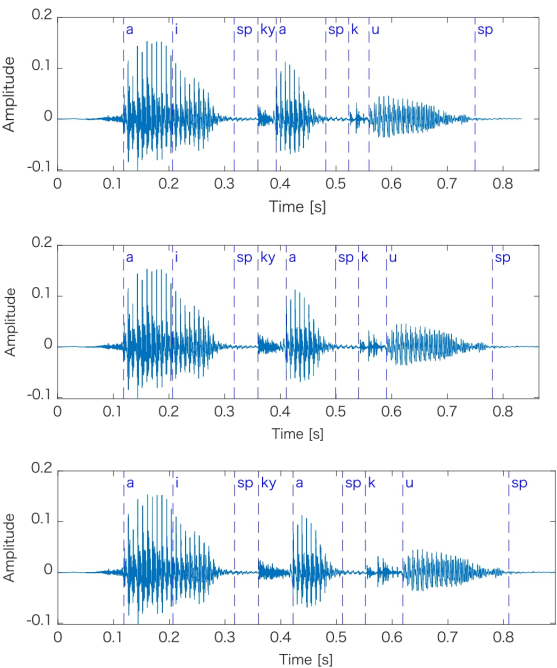


図 2: 原音声 (上), 伸長倍率 1.6 倍 (中), 2.4 倍 (下) の音声波形

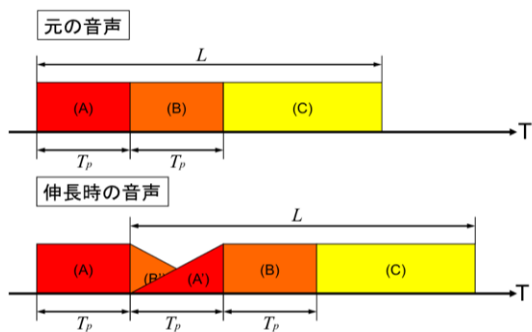


図 1: 音声伸長時の PICOLA の原理

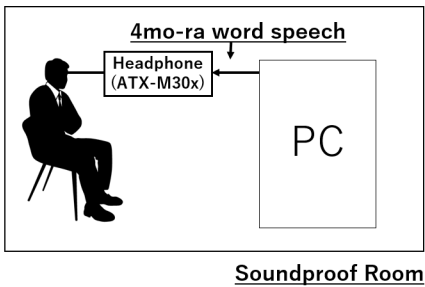


図 3: 聞き取り実験環境

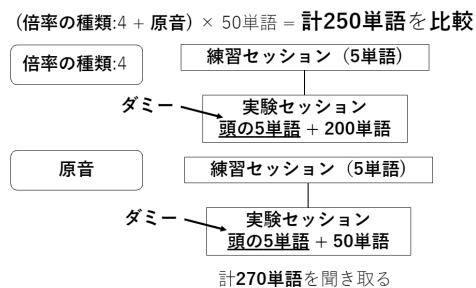


図 4: 実験フロー

4.2 実験条件

ヘッドホンに提示される音声は原音声と子音部の伸長倍率を 1.2 倍, 1.6 倍, 2.0 倍, 2.4 倍の計 4 条件で伸長したものである. 原音 (伸長倍率 1.0 倍) を含む, それぞれの倍率について 50 単語, 練習セッション 5 単語を 2 回, ダミーで 5 単語を 2 回の計 270 単語を提示した.

また, ヘッドホンから出力される音声音圧レベルは 25 dB となるように提示し, 各伸長倍率に対して, 単語正解率とモーラ毎の正解率を比較する. 使用する音声は天野ら[9]が作成した 4 モーラ単語リスト FW03 を使用した. 単語リストは 50 単語から 1 つのリストとして構成される. 各リストの単語は音韻バランスが均一となるように作成されているため, 各リストの単語了解度の比較が可能となる[10]. つまり, 1 条件につき 50 単語がセットになるため, 表 1 のように同じ単語が重複することなく比較が行える.

4.3 実験結果

聞き取り実験による CE システムの単語正解率の結果を図 5 に示す. 平均の単語正解率は, 伸長倍率が 1.2 倍のとき最も高かった.

また, 得られた単語正解率を用いて, 有意水準 5% で Shapiro-Wilk 検定により正規性を検定した. その結果,  $p$  値が 0.7 と正規性が確認された. 次に, 有意水準 5% で Bartlett 検定により等分散性を検定した結果,  $p$  値が 0.5 となり, 等分散性も確認することができた. 以上の結果をもとに, パラメトリック検定である対応のある一元配置分散分析により有意水準 5% で検定を行った. 検定した結果,  $p$  値が  $8.0 \times 10^{-3}$  となり, 5 つの伸長倍率条件間で有意差が認められた. そのため, 有意水準 5% で Holm 法により多重比較を行った. Holm 法による各伸長倍率間の対比較を行った結果を表 2 に示す. 表 2 より, 1.2 倍と 1.6 倍, 1.2 倍と 2.0 倍, 1.2 倍と 2.4 倍の間で有意差が認められた.

表 1: 単語リストと話速変換倍率の組み合わせ

伸長倍率条件	単語リスト
1.2 倍	{アイキヤク, イチハツ, ..., ワタマシ}
1.6 倍	{アイバン, イリガタ, ..., ワイリン}
2.0 倍	{アサハン, イシナギ, ..., ヒヤツソク}
2.4 倍	{アダナミ, イヤサカ, ..., ウカブセ}
1.0 倍 (原音)	{アカガシ, イドガエ, ..., ソホンサ}

各モーラの単語正解率を図 6 に示す. 各モーラの正解率は, どの伸長倍率でも第 1 モーラの正解率が著しく低いことがわかった. 第 3 モーラの正解率を比較すると, 伸長倍率 1.2 倍が最も高く, 他の伸長倍率は原音よりも低いことがわかった.

また, 得られた各モーラの正解率をモーラ毎に検定を行なった. 第 1, 第 2, 第 4 モーラでは, 対応のある一元配置分散分析及び, Friedman 検定により検定した結果, 各条件倍率間で有意差は認められなかった. しかし, 第 3 モーラにおいて, Friedman 検定により  $p$  値が  $3.3 \times 10^{-5}$  と各条件倍率間で有意差が認められた. そのため, Holm 法により多重比較を行なった. Holm 法による第 3 モーラについての各伸長倍率間の対比較を行った結果を表 3 に示す. 表 3 より, 1.0 倍と 1.6 倍, 1.0 倍と 2.4 倍, 1.2 倍と 1.6 倍, 1.2 倍と 2.4 倍, 1.6 倍と 2.0 倍の間で有意差が認められた.

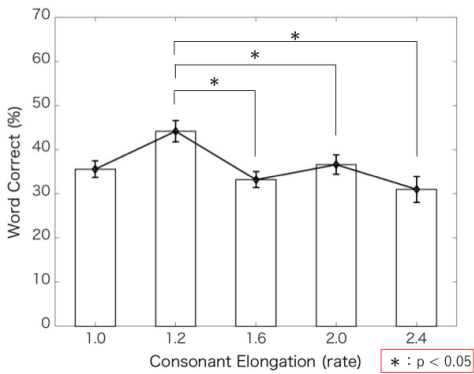


図 5: 単語正解率

表 2: Holm 法による伸長倍率間の対比較

条件	1.0 倍	1.2 倍	1.6 倍	2.0 倍	2.4 倍
1.0 倍		n. s.	n. s.	n. s.	n. s.
1.2 倍	n. s.		*	*	*
1.6 倍	n. s.	*		n. s.	n. s.
2.0 倍	n. s.	*	n. s.		n. s.
2.4 倍	n. s.	*	n. s.	n. s.	

n. s. : 非有意, \*:  $p < 0.05$

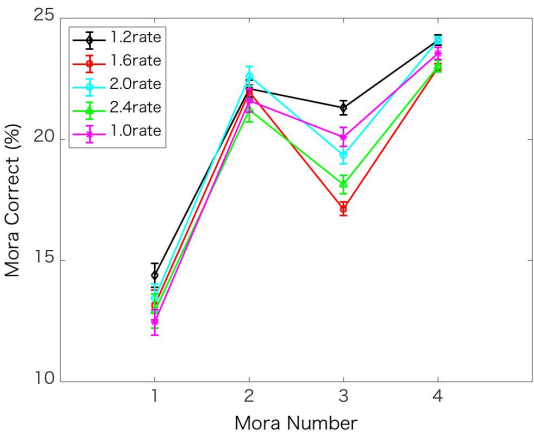


図 6: 各伸長倍率のモーラ正解率

表 3: Holm 法による第 3 モーラ条件間の対比較

条件	1.0 倍	1.2 倍	1.6 倍	2.0 倍	2.4 倍
1.0 倍		n. s.	*	n. s.	*
1.2 倍	n. s.		*	n. s.	*
1.6 倍	*	*		*	n. s.
2.0 倍	n. s.	n. s.	*		n. s.
2.4 倍	*	*	n. s.	n. s.	

n. s. : 非有意, \*:  $p < 0.05$ 

## 5. 考察

CE システムの聞き取り実験において、単語正解率は、伸長倍率 1.2 倍のときに最も高いことが確認できた。また、各モーラ正解率を比較すると、第 3 モーラのみ正解率に大きな差があり、原音よりも唯一伸長倍率 1.2 倍が高い。これは CE システムにより伸長倍率を 1.2 倍にすることにより、第 3 モーラが聞き取りやすくなったと推測できる。第 3 モーラにおいて原音と 1.2 倍間に有意差があるとは認められなかったが、CE システムの伸長倍率 1.2 倍は音声の可聴性を向上させることに有効的ではないかと考えられる。

単語正解率を比較すると、伸長倍率 2.0 倍で単語正解率が多少上がっているものの、伸長倍率 1.6 倍、2.4 倍と大差はない。そのため、全体的に CE システムで音声を過度に遅くするほど単語正解率が低くなると考えられる。特に第 3 モーラと第 4 モーラにおいて、CE システムの倍率 1.2 倍以降、音声を伸長すると正解率が低くなると考えられる。

また、PICOLA は音声波形が周期的になる有声音に対して非常に効果的な話速変換技術である。そのため、音素によっては話速変換することで別の音素に聞こえやすくなってしまったのではないかと考えられる。このことから、音素毎にどの音素に対して CE システムが効果的ではないのか検討する必要がある。

## 6. まとめ

本研究では、PICOLA を用いた話速変換機能を有するインテリジェント補聴器の実現を目指し、子音部のみを伸長する CE システムの構築、評価を行なった。CE システムでは伸長倍率が過度なものになると正解率、つまり可聴性が低くなることが考えられる。そのため、より有効的な倍率を明確にする実験を行う必要がある。また、全てのモーラにおいて伸長倍率を変えるのではなく、第 1, 2 モーラ

の子音部のみを伸長するなどの新たなシステムを構築し、より有効的なパラメータはなにかを検証していきたい。また、今回の実験は防音室内と限定的な環境で行われたものであったため、雑音、環境音のある環境でも実験を行う必要がある。

## 参考文献

- [1] 内田育恵:高齢期難聴がもたらす影響と期待される介入の可能性, 音声言語医学, Vol. 56, No. 2, pp. 143–147, 2015.
- [2] 一般社団法人日本補聴器工業会: ジャパントラック 2018 調査報告, [www.hochouki.com/files/JAPAN\\_Trak\\_2018\\_report.pdf](http://www.hochouki.com/files/JAPAN_Trak_2018_report.pdf)
- [3] 森田直孝, 板倉文忠:ポインター移動量制御による重複加算法 (PICOLA) を用いた音声の時間実での伸長圧縮とその評価, 日本音響学会講演論文集, 1–4–15, pp. 149–150, 1986.
- [4] 都木徹, 今井篤, 清山信正, 世木寛之, 田高礼子, 田澤直幸, 岩鼻幸男:話速変換技術・音声変換技術の放送および関連ビジネスへの応用, 情報処理学会研究報告, 2012–SLP–93, No. 6, pp. 1–6, 2012.
- [5] D. Malaha:Time-Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals; IEEE Trans. Acoust, Speech and Sig. Proc., ASSP–27, No. 2, 1979.
- [6] 森田直孝, 板倉文忠:ポインター移動量制御による重複加算法 (PICOLA) を用いた音声の時間実での伸長圧縮とその評価, 日本音響学会講演論文集, 1–4–15, pp. 149–150, 1986.
- [7] 都木徹, 今井篤, 清山信正, 世木寛之, 田高礼子, 田澤直幸, 岩鼻幸男:話速変換技術・音声変換技術の放送および関連ビジネスへの応用, 情報処理学会研究報告, 2012–SLP–93, No. 6, pp. 1–6, 2012.
- [8] 河原英紀:聴覚の情景分析が生み出した高品質 VOCODER: STRAIGHT, 日本音響学会誌, Vol. 45, No. 7, pp. 521–526, 1998.
- [9] 天野成昭, 近藤公久:日本語の語彙特性第 1 巻, 三省堂, 1999.
- [10] 坂本修一, 天野成昭, 鈴木陽一, 近藤公久, 曾根敏夫, 小澤賢司:単語了解度試験におけるモーラ同定に対する親密度の影響, 日本音響学会誌, Vol. 60, No. 7, pp. 351–357, 2004.