

患者の聴覚特性に応じた話し方の訓練システムの研究開発

中山功一^{†1} 谷口聖人^{†2} 町島希美絵^{†1} 大島千佳^{†1}

概要: 本稿では、それぞれ異なる患者の聴覚特性に応じて、聞き取りやすい話し方を身につける訓練を支援するシステムを提案する。本システムでは、聴力検査で用いられるオーディオメータで計測されたオーディオグラムを用いて、患者の聴覚特性を得る。得られた聴覚特性を、イコライザ機能を持つ音声変換システムで変換することで、その患者の聞こえ方を再現する。再現された音声を、音声認識ソフトウェアで文字情報に変換する。その患者に伝わりやすい話し方であったかどうかを、音声認識の正解率で発話者に提示する。その患者に伝わりやすい話し方の訓練が、提案システムで可能であるか、実験により検証する。

キーワード: 聴覚障害、難聴、話し方、訓練

Development of the speech training system using an audio gram of the hard of hearing people

KOICHI NAKAYAMA^{†1} MASATO TANIGUCHI^{†1}
KIMIE MACHISHIMA^{†1} CHIKA OSHIMA^{†1}

1. はじめに

加齢や病気などにより、人の聴覚特性は、変化する場合がある。聴覚特性の変化の内容や程度は、人それぞれである。そのような対象者（以下、患者と記す）が、日常生活において特に障害となるのが、会話が聞き取りにくくなる点である。人と話す時だけ補聴器をつける患者もいるように、会話というコミュニケーションに支障がでることで、患者のQOLが低下する。

患者と会話する際に、一般的に言われる注意点として、ゆっくり話す／大きな声で話す／低い声で話す、などがあるが、聴覚特性は一人ひとり異なるため、それぞれの患者にとって、適切な話し方は異なる。しかし、患者が、自らの聞こえ方を言葉で他者に伝えることは、必ずしも容易ではない。そのため、患者とコミュニケーションをとる場合、どのような話し方をするべきであるか、判断が難しい。それぞれの患者の聴覚特性に応じて、聞き取りやすい話し方を身につける訓練を支援するシステムが期待されている。

そこで本研究では、患者の聴覚特性に応じた話し方を訓練するためのシステムを提案する。また、提案システムを使った実験について報告する。

2. 聴覚特性に応じた聞こえ方の再現システム

聴力検査として、もっとも一般的に使われるのが、オーディオメータである（図 1）。ヘッドフォンを付けた被験者に、様々な周波数（音高）／音圧（音量）の音を聞かせ、周波数ごとに聞き取れる音圧を確認する。



図 1 オーディオメータの例：AA-74（リオン社）
Figure 1 An example of an audio meter: AA-74

検査結果は、横軸に聞かせた音の周波数(Hz)、縦軸に聞き取れた音圧に基づく聴力レベル (dB) とするオーディオグラム（図 2）で表現される。縦軸の張力レベルが高いほど、小さい音でも聞き取れることを示す。

聴覚特性が変化する例はさまざまである。例えば、老人性難聴の場合、高周波数帯が聞き取りにくくなる。そのため、右下がりのグラフとなる。また、右耳と左耳で同じようなグラフとなる場合が多い。一方、中耳炎の場合、特定の周波数ではなく、全般的に聞き取りにくくなる。このため、全体が低い位置のグラフとなる。また、炎症のある側の耳の聴力のみ低下するので、右耳と左耳で大きく異なるグラフとなる場合が多い。

^{†1} 佐賀大学大学院工学系研究科
Graduate School of Science and Engineering, Saga University

^{†2} 佐賀大学理工学部
Faculty of Science and Engineering, Saga University

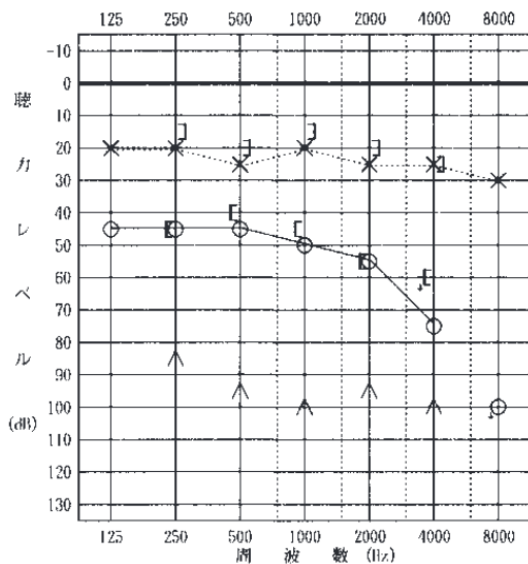


図 2 オーディオグラム例 : AA-74 (リオン社)

Figure 1 An example of an audio gram: AA-74

自分の話した音声を、これらの聴覚特性に応じて変換し、患者への聞こえ方を再現することで、自分の音声が、患者にはどのように聞こえているかが予想できると期待される。筆者らは、オーディオグラムで示されるこれらの聴覚特性をリアルタイムで再現するために、イコライザーと呼ばれる機能を持つソフトウェア[1]を用いる。イコライザーとは、周波数帯域ごとに、音圧を増加、あるいは低下させることができる。例えば、ある患者のオーディオグラムによると、高周波が聞き取りにくい聴覚特性であると判断できる場合、聞き取りにくい高周波の音圧のみ下げること、その患者が聞こえていると予想される音を、だれでも経験できると期待される。

3. 音声認識ソフトウェアによる発話内容の認識

会話を聞き取れているかどうかを、患者自身に確認することができる場合、患者に協力してもらって確認することが望ましい。しかし、本研究の目的である、患者に聞き取りやすい話し方を訓練する場合は、上達するまで何度でも繰り返して話す必要がある。そのため、聞き取れているかどうかを、毎回、患者に確認することは、患者にとって大きな負担となる。そこで、音声認識ソフトウェア[2]を用いる。



図 3 使用した音声認識ソフトウェア : ドラゴンスピーチ

Figure 3 Speech recognition software: Dragon Speech

音声認識ソフトウェアは、マイクが付いたヘッドセットに向かって話すことで、入力された音声データから、文字データへと変換するソフトウェアである。通常の日常会話であれば、話した内容は、おおむね正しく認識される。本研究では、この音声認識ソフトウェアの認識率が高い話し方は、人間にとって聞き取りやすい話し方であると仮定する。すなわち、聴覚特性の変化により会話が聞き取りにくくなった患者に対して、その患者の聴覚特性を再現した聞こえ方を音声認識ソフトウェアに入力したときに、発話内容を正しく認識できた場合に、聞き取りやすい話し方ができたと考える。

4. 実験準備

本研究では、発話者が発した音声を、2章で述べたシステムにより変換し、音声認識ソフトウェアに入力する。発話者は、変換された音声を聞くことで、自分の声が患者にはどのように聞こえているかを予測する。さらに、音声認識ソフトウェアによる認識結果を表示する。表示された認識結果および正解率を見ることで、自分の話し方が適切であったのかを確認する。正しい認識結果と高い正解率が得られる話し方を身に着けることで、同様の聴覚特性を持つ患者と円滑に会話できる話し方を身に着けられると期待する。

本稿では、老人性難聴と低音障害型感音難聴を併発した場合の聴覚特性を再現する。加齢による聴覚特性の変化である老人性難聴では、特に周波数の高い音(高音域)が聞こえにくくなる。個人により程度の差が大きいが、加齢に伴い、誰でも発症する可能性がある。老人性難聴のオーディオグラムの例を図4に示す[3]。

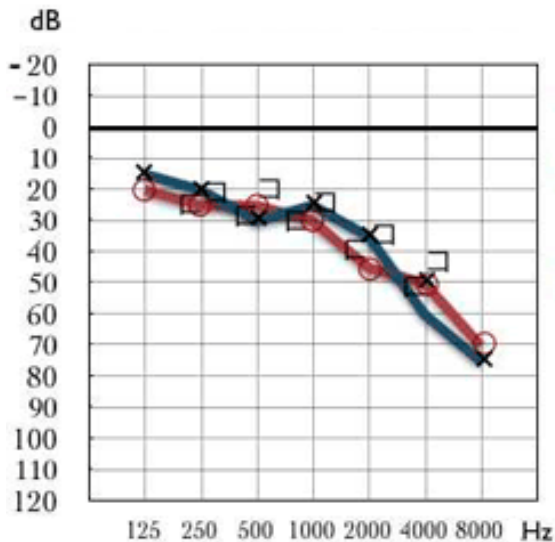


図 4 老人性難聴のオーディオグラム例[3]
Figure 4 An example of an audio gram of presbycusis

老人性難聴の場合、高音域が聞こえにくくなるという特徴はおおむね共通している。しかし、聞き取りにくくなる音の高さや、聞き取りにくくなる程度は、個人差が大きい。低音障害型感音難聴では、高齢者に限らず発症し、低音域が聞こえにくくなる。本研究では、老人性難聴の極端な例として、1660Hz以上の音域がすべて聞こえず、かつ、低音障害型感音難聴の極端な例として、450Hz以下の音域がすべて聞こえないという聴覚特性を再現する。ただし、今回、再現する聴覚特性は、あくまで一例である。本研究で想定する適切な話し方の訓練システムでは、会話したい患者ごとに、オーディオグラムで示された聴覚特性を適切に再現し、その状態で正しく音声認識される話し方を学習することを前提とする。

5. 実験

本実験では、提案システムによる訓練が有効である可能性を検証する。例文を読み、そのまま音声認識ソフトウェアで認識した場合と、老人性難聴および低音障害型感音難聴を併発した患者の聴覚特性を再現して変換した音声を、音声認識ソフトウェアで認識した場合とを比較する。例文は、タイピングの例題として用いられる文章として、「パソコン・タイピング検定[4]」および「ミシュランガイド東京2008[5]」から、末尾付録に示す3文を使用した。

それぞれの例文に対して、2種類の話し方をする。一つ目は、ゆっくり丁寧に話した場合の音声である。二つ目は、日常会話のように、普段通りに話した場合の音声である。また、2種類の話し方の音声に対して、それぞれ元音声と、患者の聴覚特性を再現して変換した音声を、それぞれ音声認識ソフトウェアに入力する(表1)。

表 1 本実験で用いた4つの条件

Table 1 4 conditions in this experiment.

	ゆっくり丁寧に話す	普段通りの速度で話す
元音声	ゆっくり丁寧に話した音声	普段通りの速度で話した音声
再現音声	ゆっくり丁寧に話した音声を、患者の聴覚特性に応じて変換した音声	普段通りの速度で話した音声を、患者の聴覚特性に応じて変換した音声

音声認識の正解率は、以下の手順で計算する。

- (1) 読み上げた文章と、音声認識ソフトウェアで認識された文章を、それぞれ全てひらがなで表記し、単語ごとに区切る。
- (2) 区切られた単語ごとに、認識された文章と、読み上げた文章を比較し、一致した(正しく認識した)単語と、一致しなかった(正しく認識しなかった)単語に分ける。
- (3) 正しく認識した単語をすべてひらがなで表した文字数の合計を、読み上げた元の文章をひらがなで表した文字数の合計で割った値を正解率とする。

例えば、以下のように計算する。

<元の文章：ひらがな 17 文字>

「私は佐賀大学の学生です」

わたし／は／さが／だいがく／の／がくせい／です

<認識された文章>

「私は皿大学の学生です」

わたし／は／さら／だいがく／の／がくせい／です

<一致した単語：6 単語 15 文字>

わたし／は／だいがく／の／がくせい／です

<一致しなかった単語：1 単語 2 文字>

さら (さが)

<正解率>

15 (文字) / 17 (文字) = 88.2%

6. 実験結果と考察

表2に、それぞれの条件において、例文ごとの音声認識の正解率と、その平均値を示す。

表 2 条件ごとの音声認識の正解率

Table 2 The correct answer rate of the speech recognition.

	ゆっくり丁寧に話す	普段通りの速度で話す
元音声	平均 : 99% 例文 1 : 99% 例文 2 : 99% 例文 3 : 100%	平均 : 90% 例文 1 : 92% 例文 2 : 85% 例文 3 : 92%
再現音声	平均 : 92% 例文 1 : 91% 例文 2 : 91% 例文 3 : 93%	平均 : 52% 例文 1 : 63% 例文 2 : 41% 例文 3 : 52%

今回、再現した患者の聴覚特性では、ゆっくり丁寧に話した場合には、元音声の正解率が3つの例文の平均で99%であったのに対し、再現音声に変換しても、正解率が平均92%までしか低下しなかった。一方、普段通りに話した場合には、元音声では平均90%の正解率であったが、再現音声では52%まで低下した。以上より、普段通りに話した場合には、今回の聴覚特性を持つ患者に対しては、半分程度しか伝わらないことを、発話者が実感できると期待される。また、様々な話し方で正解率を確認することで、どの程度ゆっくり丁寧に話せば正しく伝わるかを予測することもできる。

これらの結果から、提案手法により、患者の聴覚特性に応じた伝わりやすい話し方の訓練システムが作成できる可能性が示された。

7. おわりに

本研究では、聴覚特性が変化した患者に対して、伝わりやすい話し方を訓練するシステムを提案した。患者と会話する際に、一般的に言われる注意点の確認のみならず、患者ごとに異なる聴覚特性に応じて、伝わりやすい話し方を身に付ける訓練を支援するシステムを提案した。

実験結果から、提案システムでは、想定した老人性難聴と低音障害型感音難聴を併発した聴覚特性を持つ患者に対して、伝わりやすい話し方であるかを、判定できる可能性が示された。

今回は、1種類の聴覚特性しか想定していない。今後は、様々な聴覚特性に対応した再現システムを作成し、患者ごとに伝わりやすい話し方ができるようになるかを確認する予定である。また、一部の感音性難聴のように、特定の周波数帯が聞き取りにくくなるのではなく、音が時間軸方向にひずんで聞こえたり、断片的に聞こえたりするなど、言葉としてとらえにくくなる場合もある。このような場合に

対する有効性の検証も、今後の課題である。

謝辞 本研究は JSPS 科研費 26540068 の助成を受けたものである。

参考文献

- [1] 中山 功一, 志田玲人, 大島千佳. 音声変換フィードバックによる気分誘導システムの実装. 計測自動制御学会 システム・情報部門 学術講演会 2015, pp.SS10-9.
- [2] “音声認識・音声入力ソフト ドラゴンスピーチ 11”. <http://japan.nuance.com/dragonspeech/>, (参照 2016-07-02).
- [3] “聴力検査とは?あさひ町榊原耳鼻咽喉科医院”. <http://jibikkuma.jp/audiometry.html>, (参照 2016-07-02).
- [4] “パソコン・タイピング検定/ネット試験”. <http://www.pasotera.net/>, (参照 2016-07-02).
- [5] “ミシュランガイド東京 2008”. 日販アイ・ビー・エス, 2007

付録

本研究の実験で用いた3種類の例文を示す。

<例文 1 [4]>

「ユネスコが指定する世界遺産は自然や人類のかけがえない造型物として後世に長く保存されるべきもので自然遺産と文化遺産を合わせて現在 500 を超えている日本では 2000 年 1 月現在で白神山と屋久島が前者に法隆寺や厳島神社が後者に指定され合計でその数はやつである世界でもとりわけ森林に覆われた国土を持つ日本で自然遺産が 2 つしか指定されていないのは寂しい気がしないでもないこれに限らず手つかずの自然や歴史的価値のある建築物を後世に残そうとする世界的な取り組みが今盛んに行われている」(ひらがなで 308 文字)

<例文 2 [4]>

「情報をテレビなどで耳から得る事は簡単で早いのですが困ることもあります例えば日本語には漢字と言う独特の文化がありますこれが困ったことに同じ音で全然別の意味を持つことがあります感じはその形がいろいろな意味を持ち形から受ける印象があるので聞いただけと読んだときとは違う感じがするかもしれません逆に耳から得る情報は読み方や声の調子で印象が違ってしまいます読み手により印象が左右されます」(ひらがなで 238 文字)

<例文 3 [5]>

「ミシュランの調査員は一般の客に提供される料理の素材の質やサービスを評価するため定期的にレストランでホテルの匿名調査を行っている自費で支払いを済ませるときにはそのあとより詳しい情報を得るために名乗ることもあるガイドの読者のコメントも価値ある情報源となっておりそれに基づき調査員が再度調査を行うことができる」(ひらがなで 195 文字)