

O1-3

高齢者を対象とした合成音声の聞き取りやすさに関する研究

A Study on Ease of Comprehension of Synthesized Speech for the Elderly

○ 成田拓也 (東京大学) 二瓶美里 (東京大学) 小竹元基 (東京大学) 大中慎一 (NEC)

鎌田実 (東京大学) 井上剛伸 (国立障害者リハビリテーションセンター研究所)

Takuya NARITA, The University of Tokyo

Misato NIHEI, The University of Tokyo

Motoki SHINO, The University of Tokyo

Shinichi OHNAKA, NEC Corporation

Minoru KAMATA, The University of Tokyo

Takenobu INOUE, Research Institute, National Rehabilitation Center for Persons with Disabilities

Abstract: The purpose of this study is to clarify the characteristics of synthetic speech easy to hear the elderly. The evaluation experiment was focused on three elements, (a) speaker, (b) the fundamental frequency of speech, (c) the rate of speech. The participants were 20 elderly. As a result, the male speech with low frequency made hearing easier than others. In addition, speech rate 4 and 5 mora/s made hearing easier than speech rate 6 mora/s. On the other hand, it was clarified the relationship between hearing reduction (average hearing level) and difficult to hear the speech of female and child. Among the 15 different speech conducted in this study, male speech (120Hz fundamental frequency) with 5 mora/s was the most easy to hear by elderly.

Key Words: Elderly, Auditory Function, Synthesized Speech

1 緒言

近年、日本では人口の高齢化が進んでいる。それに伴い、加齢によって身体機能や認知機能が低下した高齢者を考慮し、建物の段差を失くす、手すりを設置するなどの設備のバリアフリーデザインやユニバーサルデザイン、必要最低限の機能を持つシンプルな家電製品の開発が行われている。また、認知機能の低下を考慮し、機器の操作方法を音声案内する銀行 ATM や券売機が製品化されている。最近では家電などの一般製品の操作にも音声ガイドが用いられている⁽¹⁾。

本研究では、もの忘れのある高齢者や認知症高齢者を対象とし、スケジュール情報などの自立した日常生活を送るために必要な情報を支援するシステムの開発を行っている⁽²⁾。スケジュール情報は個別性が高いため、合成音声による音声ガイドを用いる。先行研究では、音声ガイドの聞き取りやすさは、音量や帯域などが関係しており、高齢者と若年者では主観評価が異なるという結果が報告されている⁽¹⁾。また、高齢難聴者に適した話速変換方式の研究では、速度を下げ、明瞭な言葉を聴かせると言語理解力が上がることを示している⁽³⁾。話者については視覚障害者を対象としたスクリーンリーダーの利用調査では、男性音声を使用している割合が8割以上を占めていることが報告されている⁽⁴⁾。しかし、高齢者が聞き取りやすい合成音声の仕様や特性は明らかになっていない。そこで、本報では、合成音声の聞き取りやすさに影響する要素を、①高齢者入居施設での介助者と高齢者の対話によるやりとりを観察し、影響が大きいと考えられる要素を抽出し、②それらの要素ごとに設定した音声を用いて、聞き取り評価と主観評価を実施した。

2 事前調査

2-1 事前調査の方法

調査は高齢者入居施設で行い、著しい聴覚機能低下のな

い高齢者5名(85.8±7.29歳)について、対象高齢者の生活支援を担当する介助者に、高齢者に話しかけをする際に配慮している点に関する聞き取り調査を行った。音声刺激はNEC社製PaPeRoに搭載された合成音声を用い、各人に聞き取りやすい音声とするための試行錯誤的な調整を行った。

2-2 事前調査の結果

各対象高齢者に共通する項目と、関連する項目を次にまとめる。

[介助者による声掛け]

- ・ 介助者が話しかけをする際には、ゆっくりとはっきりとした声掛けを行う。
- ・ 高齢者への情報伝達は、わかりやすい言葉で文章の長さにも配慮する。
- ・ 高い声ではなく、低めの声で話しかけを行う。

[合成音声の調整]

- ・ 子ども音声(合成音声)は、「何か言っているみたいけど、聞こえない」など聞き返すことが多く、聞き取りにくく、内容を認識しにくい
- ・ 成人男性の肉声の録音声をもとに、話速を遅く、基本周波数を低く調整した際の音声聞き取りやすい。また、成人男性の合成音声聞き取りやすい。
- ・ 聞き取ることができる音量は、個人の聴力によって異なる。

2-3 評価項目

合成音声の話者は、個人の好みによって設定されることが多く、聞き取りやすさに個人差があることがわかっていくが⁽⁴⁾、事前調査の結果から、子ども音声が聞き取りにくく、男性音声が聞き取りやすい傾向にあったことから、高齢者が聞き取りやすい話者を調べる必要があると判断した。また、同じ話者内においても周波数の違いによる聞き取りやすさが異なると考えたため、同一話者に対して数種類の周波数を比較することとした。話速については、視覚障害

者の場合は早い速度を好む傾向があるが⁽⁵⁾、高齢難聴者の場合は、「ゆっくり」話すことが情報伝達を行う上で重要であると言われている⁽³⁾。本研究の対象者は、前期高齢者(65歳~74歳)から後期高齢者(75歳以上)であることから、難聴者も含む場合があるため、「ゆっくり」とした速度で正確に情報を伝えるための適切な速度を調べることにした。

以上の事前調査の結果をもとに、本研究で検討する高齢者の聞き取りやすさに関係すると考えられる合成音声の要素として、次の3つ(a) 話者、(b) 音声の基本周波数、(c) 発話速度(話速)を設定した。

3 合成音声の聞き取り評価実験方法

3-1 対象者と評価実験条件

実験対象者は65歳以上の男女20名とし、シルバーサービスセンターに登録された聴覚障害のない健聴な高齢者とした。実験はテレビやラジオの音声の入らない一般的な日常生活音の範囲内である暗騒音約30dBの静かな屋内で行った。音声刺激はスピーカ(YAMAHA-NX-U02)から呈示され、音源の距離は対象者から約30cmの位置に設置した。音源のサンプリング周波数は22kHzである。

3-2 評価実験方法

評価実験は、対象者の基本的な聴力レベルを測定する聴力測定と15種類(男性5種類、女性5種類、子ども5種類)の合成音声を聴取し、再生回答する聞き取り評価実験を行った。

(1) 聴力測定

対象者の基本的な聴力レベルを測定するために、オーディオメータ(リオン社製(AA-73A))を用いて各対象者の平均聴力レベルを算出した。平均聴力レベルの算出は一般的な聴力検査で用いられる4分法を用いた。4分法による平均聴力レベルAは、式(1)で表わされる。

$$A = \frac{B + (C \times 2) + D}{4} [dB] \quad (1)$$

ここで、Bは500Hzでの聴力レベル、Cは1000Hzでの聴力レベル、Dは2000Hzでの聴力レベルである。

(2) 合成音声の聞き取り評価実験

聞き取りやすさの評価実験は、合成音声を読み上げた文章を対象者が口頭で再生する形式で行った。正答率は、文章が合致しているかどうかで判断した。また、聞き取りやすさに関する主観評価(5段階)を行った。

音声刺激は話者、基本周波数、話速の異なる計15種類の音声を作成した。合成音声のパラメータを表1にまとめる。ここで、基本周波数とは、各話者の最低周波数と最高周波数を設定し、スペクトル中の最も低いピーク値を基本周波数とした。また、実験に用いる音量は、普通の会話と同程度の音量である60dBに設定した。

合成音声を読み上げる文章は、日常的に用いられる文章とし、補聴器適合評価用CD(TY-89)に収録されている日常生活文リストを参考に作成した。読み上げる文章の長さの違いが正答率に影響を与えると考えたため、文章の長さは約10モーラを基準とした。また、対象者の言語能力の違いが正答率に与える影響を考慮し、単語の「なじみ」の程度を表わす単語親密度を基準とし、多くの人になじみのある単語(親密度:5.5以上7.0未満)で構成した文章を作成した。実験に使用した文章を表2に示す。

Table 1 Parameter of Synthesized Speech

Speaker	Male	Female	Child	
Fundamental frequency [Hz]	Low	Middle	High	
	M	101	120	143
	F	189	225	268
C	252	300	357	
Speaking rate [mora/s]	4	5	6	

Table 2 Example of Speech Text

風邪で学校を休む	ゆっくり風呂に入る
忘れ物を確かめる	今日はとても暑い
電話でタクシーを呼ぶ	朝食にパンを食べる
食後に歯を磨く	電気のスイッチを入れる
テーブルの上を拭く	ちゃんとガスを消す
ワイシャツを毎日洗う	おいしいケーキがほしい
はがきをポストに入れる	ちゃんと布団を片付ける
部屋の電気を消す	頭がとても痛い
パンにバターを塗る	テレビをいつまでも見る
友だちと電話で話す	石鹸で顔を洗う
スーツをクリーニングに出す	駅に傘を忘れる
豆腐にしょうゆをかける	明日は起きるのが早い

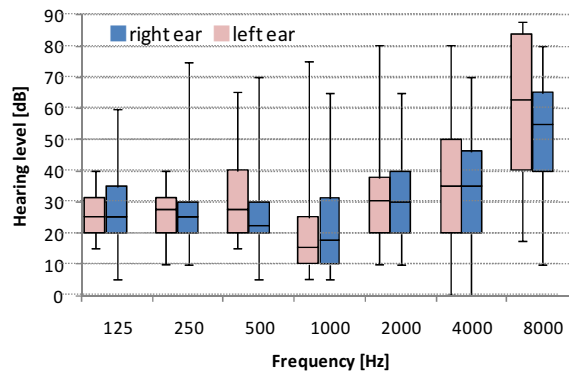


Fig. 1 Hearing Level of Participants

4 合成音声の聞き取り評価実験の結果

4-1 評価実験結果

(1) 実験参加者と平均聴力レベル

実験参加者は66歳から81歳の高齢者20名(男性10名、女性10名、平均年齢71.35歳)である。図1の箱ひげグラフに各周波数の左右聴力測定結果を示す。ここで、箱は4分位、ひげは最大値、最小値、箱中の線は中央値を示す。平均聴力レベルは平均 24.7 ± 14.4 dB(最小値7.5dB、最大値71.3dB)であった。実験参加者に聴覚障害の判定基準(両耳の平均聴力レベルが70dB以上もしくは一側耳の聴力レベルが90dB以上で他側耳の聴力レベルが50dB以上)に該当する者は含まれていなかった。

(2) 各条件における評価実験結果

(a) 話者の性別

各話者について、話速や周波数の条件を変えた各5種類の平均正答率は、男性音声(92.5%)が最も高く、次いで女性音声(90%)、最も低いのは子どもの音声(80%)であった。

(b) 音声の基本周波数

図2に基本周波数ごとの正答率を示す。子ども音声(Low, 252Hz)の基本周波数は、女性音声(High, 268Hz)の基本周波数より高い設定になっている。全体的に周波数が高い

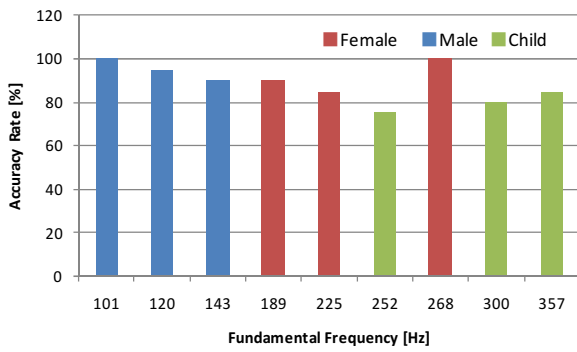


Fig. 2 Fundamental Frequency

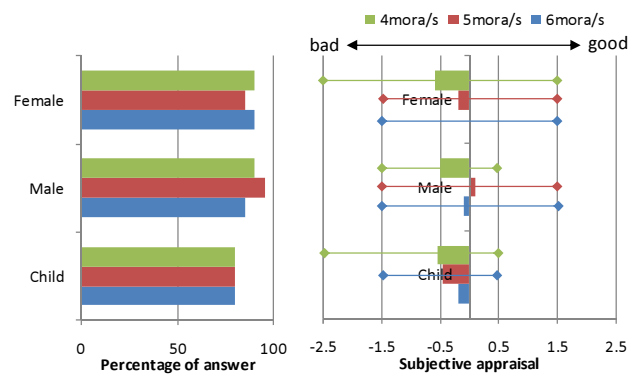


Fig. 4 Accuracy Rate of Difference of the Rate of Speech

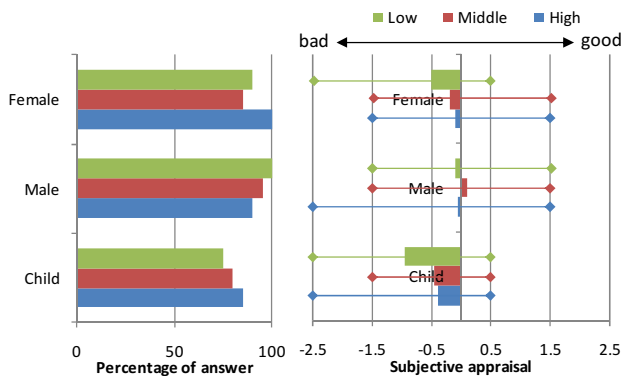


Fig. 3 Accuracy Rate of Difference of Fundamental Frequency

ほど正答率は低くなる傾向があるが、子ども音声 (Low, 252Hz) と女性音声(High, 268Hz) では、女性音声の方が聞き取りやすく、子ども音声は周波数が高いほど正答率が高くなる傾向があることがわかった。

(a)-(b)話者と基本周波数

図3に話者(子ども, 男性, 女性)と周波数ごとの正答率および主観評価の結果を示す。周波数は表1に示した各話者の3種類の基本周波数である。主観評価は5段階(1:とても聞き取りにくい, 2:聞き取りにくい, 3:ふつう, 4:聞き取りやすい, 5:とても聞き取りやすい)を, 0を中心として, 負の値が聞き取りにくく, 正の値が聞き取りやすいスケールに変換した。その結果, 男性音声は, 低い周波数(Low, 101Hz)の正答率が高く, 女性音声については, 高い周波数(High, 268Hz)の正答率が高く, 子どもの音声については, 高い周波数の方が, 正答率が高い傾向にあることがわかった。一方, 主観評価の結果は, 男性音声聞き取りやすく, 特に男性音声(Middle, 120Hz)が最も聞き取りやすいことがわかった。

(c)話速

話速については, 4 mora/s, 5 mora/sが正答率86.7%, 6 mora/sが正答率85%であり, 6 mora/s以上になると聞き取りにくくなる可能性があることがわかった。

(a)-(c)話者と話速

図4に話者と話速ごとの正答率および主観評価の結果を示す。主観評価は0を中心として, 負の値が聞き取りにくく, 正の値が聞き取りやすいことを表す。男性音声は, 5 mora/sが最も正答率が高く(95%), 女性音声では4 mora/s, 6 mora/sの正答率が高い(90%)。子ども音声では, 各速度とも正答率は80%であった。また, 主観評価についても5 mora/sが最も聞き取りやすいと感じていることがわかった。

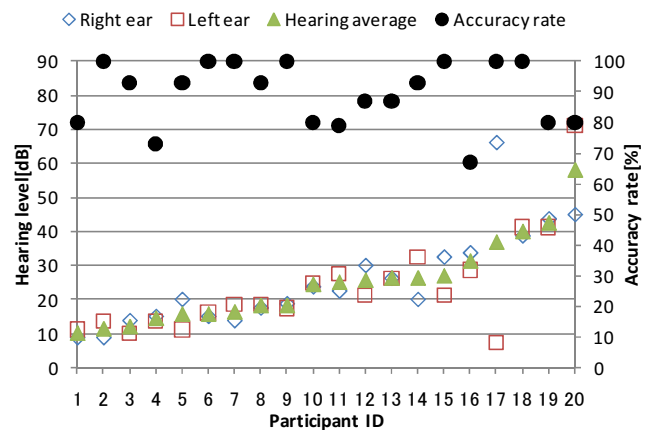


Fig. 5 Average of Hearing Level and Accuracy Rate

Table 3 The Number of Correct and Incorrect (Male)

	M5M	M5H	M5L	M6L	M4M
Correct	19	18	20	17	18
Incorrect	1	2	0	3	2

総合的には, 話者は男性音声, 基本周波数120Hz, 話速は5 mora/sが最も聞き取りやすいことがわかった。

4-2 聴覚特性と聞き取りやすさの評価実験結果

(1)平均聴力レベルと正答率

実験参加者の平均聴力レベルと合成音声の聞き取りやすさを比較する。実験参加者の左右平均聴力レベルと15種類の音声の聞き取り評価実験結果を図5に示す。ここで, 正答率とは, 15種類の文章聞き取りに正答した割合(正答率)を表す。その結果, 平均聴力レベルと正答率には関連性が認められなかった。

(2)平均聴力レベルと各条件の正答率

平均聴力レベル(左)と話者5種類の聞き取りやすさとの関係を調べるために, 各条件で正解(correct), 不正解(incorrect)の平均値を比較した。それらの結果を図6~8に示す。ここで, 項目は話者・話速・周波数を表わし, 例えばM5MはMale, 5 mora/s, Middleを意味する。

(a)男性音声

各項目の正解/不正解数は表3に示す。また, 図6に男性音声刺激に対する正解/不正解と平均聴力レベル(左)の結果を示す。その結果, 男性音声については, 実験参加者の聴力レベルと正解/不正解は関連がないことがわかった ($p>0.05$)。

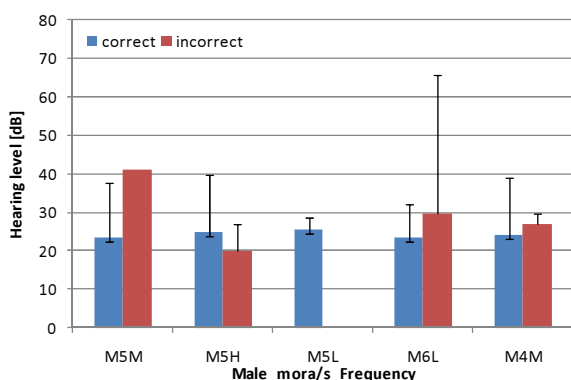


Fig. 6 Hearing Level and Accuracy Rate of Male Speech

Table 4 The Number of Correct and Incorrect (Female)

	F5M	F5H	F5L	F6L	F4M
Correct	17	20	18	18	18
Incorrect	3	0	2	2	2

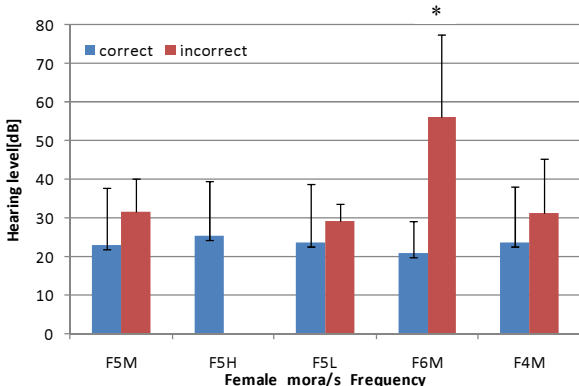


Fig. 7 Hearing Level and Accuracy Rate of Female Speech

Table 5 The Number of Correct and Incorrect (Child)

	C5M	C5H	C5L	C6L	C4M
Correct	16	17	15	16	16
Incorrect	4	3	5	4	4

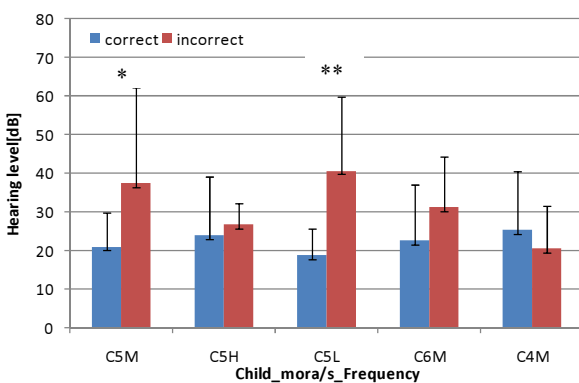


Fig. 8 Hearing Level and Accuracy Rate of Child Speech

(b)女性音声

各項目の正解／不正解数は表 4 に示す。また、図 7 に女性音声刺激に対する正解／不正解と平均聴力レベル(左)の結果を示す。その結果、女性音声については F6M (6 mora/s, 225Hz) の条件のみにおいて、不正解であった実験参加者の聴力レベルが低いことがわかった ($p<0.01$)。

(c)子ども音声

各項目の正解／不正解数は表 5 に示す。また、図 8 に子ども音声刺激に対する正解／不正解と平均聴力レベル(左)

の結果を示す。その結果、子ども音声については C5M (5 mora/s, 300Hz) ($p<0.05$), C5L (5 mora/s, 252Hz) ($p<0.01$) の条件のみにおいて、不正解であった実験参加者の聴力レベルが低いことがわかった。

以上より、合成音声の女性音声や子ども音声は、聴力レベルが低下した高齢者に対して聞き取りにくいことがわかった

5 結言

本研究では、高齢者に聞き取りやすい合成音声のパラメータを明らかにすることを目的として、高齢者を対象とした合成音声の聞き取り評価実験を行った。まず、合成音声の聞き取りやすさに影響する要素を、高齢者入居施設での介助者と高齢者 5 名の対話によるやりとりを調査することによって(a) 話者, (b) 音声の基本周波数, (c) 発話速度(話速)の 3 つの要素を抽出した。次に、それらの要素について 15 種類の合成音声を高齢者 20 名が聴取し、再生回答する評価実験を行った。

その結果、話者では男性音声、基本周波数は話者によって異なるものの、低い周波数の方が、聞き取りやすかった。また、話速は、4, 5 mora/s が聞き取りやすかった。一方、実験参加者の平均聴力レベルと聞き取りやすさの関係を調べたところ、聴力レベルが低下すると女性音声や子ども音声の聞き取りにくくなることがわかった。本研究で実施した 15 種類の音声の中では、男性音声(基本周波数 120Hz)、話速 5 mora/s が最も聞き取りやすかった。今後は、認知症者を対象とした合成音声の聞き取りやすさを明らかにし、高齢者と認知症者での合成音声の聞き取りやすさの違いについて明らかにしていく予定である。

謝辞

本研究に協力いただいた公益社団法人文京区シルバー人材センターおよび実験参加者のお皆様に感謝の意を表す。また、本研究は科研費(21300213)の支援を受けて実施した。

参考文献

- (1) 渡辺英樹, 有川泰史, 小川哲史, 濱野博司他, ユニバーサルデザインに配慮した音声ガイドの「聞き取りやすさ」評価法, 松下電工技報 Vol. 53, No.4, pp.28-33.
- (2) Takenobu Inoue, Misato Nihei, Takuya Narita, Minoru Onoda, Rina Ishiwata, Ikuko Mamiya, Motoki Shino, Hiroaki Kojima, Shinichi Ohnaka, Yoshihiro Fujita, Minoru Kamata, Field-based Development of an Information Support Robot for Persons with Dementia, Assistive Technology Research Series, Volume 29, pp.534-541, 2009.
- (3) 伊福部達, 高齢難聴者のための音声情報変換方式, 電子情報通信学会誌, vol.84, No.5, pp. 325-328, 2001.
- (4) 渡辺哲也, 視覚障害者向け音声インタフェースに関する研究—合成音声の話速, ピッチ, 性別の設定値に関する調査—, 電子情報通信学会, 信学技報WIT2004-6, pp. 13-18.
- (5) 林実, 合成音声の話速に対する高齢者および弱視者の聴覚特性, 電子情報通信学会, 信学技法WIT2007-3, pp.13-18.