

高齢者支援のための聴覚補正技術におけるリップシンク効果の確認

鈴木雅寿^{*1}, 菅原庸平^{*1}, 佐藤美恵^{*1}, 森俊文^{*1}, 杉原敏昭^{*2}, 春日正男^{*1}

Experimental Verification of the Lip-Synchronization Effect in the Auditory Compensation System for the Elderly People

Masatoshi Suzuki^{*1}, Youhei Sugawara^{*1}, Mie Sato^{*1}, Toshifumi Mori^{*1}, Toshiaki Sugihara^{*2} and Masao Kasuga^{*1}

Abstract : This paper describes an experimental verification of the lip-synchronization effect in the auditory compensation system for the elderly people. Because the aging effect to the human sensory mechanism, the functionality of the elderly people's sensation is declined. The auditory sensation is also declined. Many of the elderly people have difficulty with a listening in the radio or TV program. Recent digital media technology can compensate such declined auditory functionality. The compensation mechanism regulates a speech speed. However, in the television system, there is well known problem, that is called lip-synchronization problem. Our previous related work already evaluated the efficiency of the speech speed controlled compensation method for the broadcasting equipments, but we also confirmed unnatural feelings with the lip-synchronization skew. So we intend to verify the lip-synchronization effect in the speech speed controlled compensation method, by using a subjective assessment. In this assessment, we used psychological index for evaluating the unnatural feelings corresponding to lip-synchronization skew. As a result of this assessment, we can evaluate the unnatural feelings quantitatively, also discuss the relationship between the lip-synchronization skew and the subjective impression.

Keywords : auditory sensation, compensation technology for the elderly people, lip-synchronization

1. はじめに

現在、聴覚特性の衰えた高齢者支援のために、放送の音声の話速を変換して聞きやすくする技術、話速変換技術が開発され、ラジオについては実用化、製品化もなされている。しかし、この技術をテレビ放送に応用する場合には、所謂リップシンクのずれによる違和感の問題が生じ、テレビや他の映像機器にこの機能を実装する上での問題点となっている。本研究では、聴覚補正のための話速変換技術に映像のリップシンク機能を取り入れた際の効果を、被験者実験手法によって確認した事例を報告する。

2. 研究の背景と目的

高齢者においては、各種の感覚特性の衰えが生ずることが広く知られており、聴覚特性においては、

- 1) 可聴域の狭まり、特に高音域での感度の低下
- 2) 音の分離能力の低下
- 3) 認識可能な話速の低下
- 4) リクルートメント現象として知られている微小音の聞き取りの困難さと大きな音に対する不快感

が生ずるものとされている。これら加齢に伴う聴覚特性の衰えに対する高齢者支援技術として、近年、デジタル信号処理技術を応用し、放送される音声の話速を高齢者にとって聞き取りやすい速度に調整することで、上記の2)および3)の現象を補償、支援する話速変換技術が開発され、ラジオ受信機については実用化、製品化も行われている[1]。この聴覚補償技術の基本的な仕組みは、受信機に内蔵されたバッファメモリと信号処理回路を用い、バッファリングした放送の音声をリアルタイムに信号処理して、高齢者にとって聞きやすい話速となるように音間を引き伸ばすものである。音間を引き伸ばすことで生ずる音声全体での遅延への対応は、無音区間を間引くことや、音声の開始では遅めに話速変換を行い、音声の終止に向けて徐々に話速を早めて行くことが行われている[2]。これらによって、文脈的、意味的に、ある一まとまりとなっている音声の終止時には概ね遅延が無い程度の、つまりは、実際の放送番組において、その放送時間枠をはみ出してしまうような遅延という放送受信機としては致命的な弊害となる点を避けるための方策、工夫がなされている。

我々は、この話速変換技術が、高齢者支援技術としてどのように有効、有益であるのかを検証するために、実際

*1: 宇都宮大学大学院

*2: 株式会社リコー研究開発本部

*1: Graduate School of Engineering, Utsunomiya University

*2: R&D Center, RICOH Co., LTD.

に高齢者を被験者として用いた心理学的評価実験を行った[3,4]。この評価実験により、前述の2)や3)に示す聴覚特性の劣化を補償する話速変換技術の効果を定量的に測ることができ、どのような点において有益な効果をもたらしているのかを確認した。しかしながら、この話速変換技術をテレビ放送に応用することを考慮すると、視覚情報を伴うものであるために、所謂リップシンクのずれによる違和感が生じていることも実験を通じて確認している。一般的に、人間の感覚特性では、腹話術効果[5]に代表される視覚優位の特徴がある。前述の評価実験では、現実的な応用を想定し、かつ、準備が容易で条件が統制しやすいという点より、テレビ放送の録画を刺激として用い、その音声に対して聴覚補償を適用した場合の有無による聞き取り易さ、聞き取り難さの差異を評価することを行った。この刺激内容としては、ニュース番組、気象予報、スポーツ中継、ドラマ等の複数の録画を用いている。この中で、気象予報士の上半身の映像を含む気象予報の場合において、顕著な違和感が生ずるとの報告が被験者より得られている。この場合、気象予報士の発話に伴う口唇の動きが視覚情報として明確に被験者に与えられていることが、違和感の原因であるものと推測できる。また、ドラマの場合においても、聴覚補償による音声の遅延によって、カメラアングルの変更によるシーンチェンジ後にもシーンチェンジ前の配役の台詞が残ってしまうと、同様に違和感や不自然さを被験者が感ずることが確認できている。

以上の点より、テレビ放送に対して聴覚補償を適用する際に、発話や音声の切れ目あるいは転換点であることを明示的に与える視覚的な手がかり(キュー)が伴う場合には、違和感や不自然さの問題が生じることが明らかとなった。

本研究では、この前報にて明らかになった問題点、聴覚補償による音声遅延と映像間のずれ、すなわち、所謂リップシンクのずれによる違和感が、どのように、かつ、どの程度生じているのかを、定量的に把握し確認することを目的とした。

本研究では、被験者実験の手法を用い、リップシンクのずれ量と主観的に感ずる違和感、不自然さの関係を定量的に評価することを行った。ここでは、前報にて刺激として用いたものと同じ録画を、映像編集ツールを使い、手

業で確認を行いながら、映像のフレーム間を間引く、あるいは、引き伸ばすなどの加工を施して、リップシンクのずれが殆ど無いものから、大きくずれているものまでの複数のものを異なる刺激として用意した。これらの異なる内容の刺激における、ずれ量の過多と被験者の受ける違和感や不自然さの印象の差異を、主観評価の手法で計測し評価することを行った。次節に、実験の具体的な手法とその実施について述べる。

3. 被験者実験

3.1 実験手法

本実験では、主観評価手法によって、被験者が感ずる違和感や不自然さに関わる印象を定量的に評価するものである。ここでは、話速変換による聴覚補償を施した音声を含むテレビ放送の録画映像を刺激として用い、ここでのリップシンクのずれ量と被験者のうける違和感、不自然さの印象との関係を、7点法の主観評価指標で被験者に判定させることを行った。

3.2 刺激映像

既に述べた通り、今回の実験では、前報[3,4]で刺激として用いたものと同じ、以下の5種類のテレビ放送録画を刺激として用いることにした。

刺激a: ニュース

刺激b: ドラマ(時代劇)

刺激c: 天気予報

刺激d: 野球中継

刺激e: ドキュメンタリー

ここで、被験者にとって出来るだけ短時間で印象の判定が行え、かつ、話速の変化も判りやすいものであることを、前報での経験的知見も加味して考慮し、各刺激の時間は30秒前後と設定した。

次に、リップシンクのずれ量を刺激の提示条件とすべく、映像編集ツール(Adobe Premiere™)を用いて映像を編集加工し、以下の2条件を用意した。

A: 録画映像をそのまま用いたもの

C: 話速変換機を用いて、話速をゆっくりにした映像に対してリップシンクのズレを緩和したもの

Cの条件では、手作業で画面を確認しながら、話速変換を

用いた音声に対して映像が音声に同期するように映像を編集、特に口が動いているときに音声同期することに注意を払った。

3.3 実験環境

本実験における実験環境を図1に示す。実験は外部の音を遮断するために防音室で行った。音の提示は、テレビに内蔵されているモノラルスピーカを使用する。音量は刺激の種類に関わらず一定とし、被験者にとって適当な、大き過ぎずかつ物足りないと感じない程度の、音量となるように、実験手続きに関する教示の際に調整を行った。画面のサイズは29インチであり、テレビまでの距離は家庭での平均的な視聴距離である約200cmに設定した。

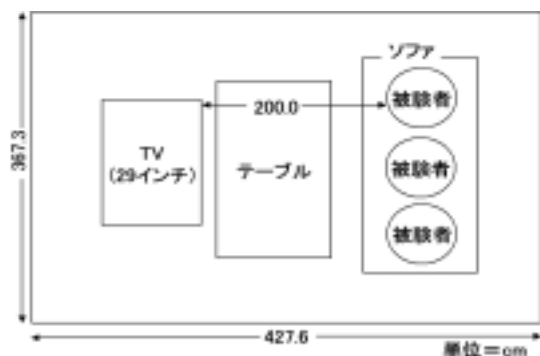


図1：実験環境

3.4 実験手順

実験の流れを以下に示す。

問診：被験者に、プロフィール、実験日当日の健康状態等を問うアンケート（問診表）への記入を行わせる。

刺激提示：前述の刺激条件(A)と(C)の映像・音声を提示する。

主観評定：刺激提示後、刺激の与える印象についての7点法の主観評定を、質問紙への回答で行わせる。

全員が主観評定の質問紙への回答を済ませたことを確認した後に、次の刺激を提示する。

～ の手順を全5種類の映像に対して、同様に繰り返す。

3.5 評価語

今回は、リップシンクのずれに対する違和感、不自然さ

を評価するために、以下の6つの語彙を主観評定の指標として用いた。

- ・自然である
- ・見やすい
- ・聞きやすい
- ・心地よい
- ・分かりやすい
- ・好ましい

今回は、図2に示す7点法で質問紙を作成し、記入は被験者自身に行わせることとした。

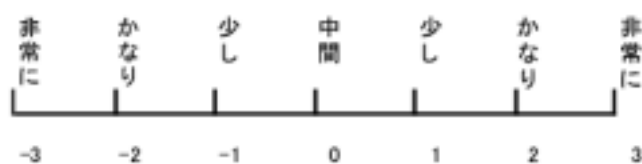


図2：質問紙

3.6 被験者

今回、被験者は視聴覚ともに正常な20代の男子大学生12名を用いた。図1に示すソファに被験者を着座させ、1回の試行につき被験者1～3人を用いた。

3.7 実験結果

以下、図3～図7に主観評定の結果を示す。ここで、主観的印象上で良好な評価結果となった提示条件を比較結果の項に示す。ここに示される通り、自然さについては、全ての刺激においてリップシンクのずれを少なくした条件で印象が好転し、心地よさと好悪については、5つの内の4つで好転している。また、見やすさや分かりやすさについては、大差が無いものと言える。

4. 考察とまとめ

本研究の背景と目的で述べた通り、我々は前報において、聴覚補償技術自体の利点、すなわち、実際の高齢者被験者にとっては、適切な聴覚補償を施すことで、放送の音声聞き取りやすくなり、聞き取り易さが内容の理解のしやすさにつながることを確認している。しかしながら、一方で、前報での刺激として用いた視覚情報を含むテレビ放送においては、顔画像を含む場合のような直接的にリップシンクのずれが分かる場合や、カメラアングルの切り替えなどで間接的に話者の交代が視覚的な情報として与えられる

ような場合には、違和感や不自然さが生じていることが、特徴的に被験者の印象より得られている。このことは、テレビ放送に音声補償技術を適用する際には、聞き取りやすさ、理解のしやすさという利点と、違和感、不自然さという欠点が共存してしまうことである。本研究では、リップシンクのずれによって与えられる違和感や不自然さを定量的に確認するための被験者実験を行った。主観評価の結果、リップシンクのずれが少なくなることで、不自然さが改善されることが確認できた。前述の通り前報で確認した聴覚補償による聞き取りのしやすさが理解のしやすさにつながる点と、前節に述べた今回の実験結果ではわかりやすさに大差が生じなかった点とを合わせて考えるならば、リップシンクの効果は、主として好悪などの情動的な面に寄与するものと考えることができ、このことは、前節の結果に示した、心地よさと好悪の印象が好転していることと良く合致するものである。既に述べた通り、放送機器として致命的な音声全体での遅延をおさえるために、聴覚補償の手法では無音区間の間引きや速度変化のある話速変換を行っており、映像の同期を自動的に行うことが困難な点がある。今回の実験でも、各刺激映像を手作業で加工しているのは、この事由に因る。聴覚補償に合わせて映像の同期を自動的に行う手法の開発は今後の課題と考えている。

		5%有意	1%有意	
ニュース	分散比	4.695	10.044	比較結果
自然である	7.206			A
見やすい	0.464			-
聞きやすい	9.698			-
心地よい	12.250			A
わかりやすい	1.106			-
好き	8.459			A

図3：評価結果(1)刺激(a)ニュース

		5%有意	1%有意	
時代劇	分散比	4.695	10.044	比較結果
自然である	8.721			A
見やすい	4.549			-
聞きやすい	0.536			-
心地よい	3.854			-
わかりやすい	0.891			-
好き	6.102			A

図4：評価結果(2)刺激(b)ドラマ

		5%有意	1%有意	
天気予報	分散比	4.695	10.044	比較結果
自然である	38.684			A
見やすい	18.766			A
聞きやすい	2.581			-
心地よい	7.656			A
わかりやすい	3.265			-
好き	9.423			A

図5：評価結果(3)刺激(c)天気予報

		5%有意	1%有意	
野球	分散比	4.695	10.044	比較結果
自然である	43.214			A
見やすい	9.423			A
聞きやすい	0.800			-
心地よい	15.652			A
わかりやすい	12.500			A
好き	18.443			A

図6：評価結果(4)刺激(d)野球中継

		5%有意	1%有意	
ドキュメンタリー	分散比	4.695	10.044	比較結果
自然である	20.610			A
見やすい	3.902			-
聞きやすい	3.270			-
心地よい	10.714			A
わかりやすい	0.978			-
好き	4.286			-

図7：評価結果(5)刺激(e)ドキュメンタリー

5. 参考文献

- [1] 井上, 岩崎, 大西, 安藤, 小倉: 人にやさしいゆっくり通話技術, SANYO Technical Review, Vol.33, No.3, pp.46 - 54, (2001).
- [2] 今井, 池沢, 清山, 中村, 都木, 宮坂, 中林: ニュース音声を対象とした時間遅れを蓄積しない適応型話速変換方式, 電子情報通信学会論文誌, Vol.J38 - A, No.8, pp.935 - 945, (2000).
- [3] 菅原, 生田目, 佐藤, 杉原, 春日: 高齢者の加齢現象に着目した視聴覚にやさしい映像コンテンツの検討, 映像情報メディア学会技術報告, Vol.28, No.27, pp.25 - 28, (2004).
- [4] Sugawara, Y., Sato, M., Sugihara, T., Kasuga, M., : Experimental Approach to User Friendly Audio Visual Contents for the Elderly People, Proc. of IEEE TENCON2004, pp. 391-394, (2004).
- [5] 八木昭宏著: 知覚と認知, 培風館, (1997).