

合成視覚情報を付与した英語音声聴解学習法の検討

Investigation of English Speech Listening Comprehension Learning Method with Synthetic Visual Information

佐竹 望愛^{*1}, 蜂須賀 知理^{*2}, 栗田 佳代子^{*3}, 割澤 伸一^{*1},
Mie SATAKE^{*1}, Satori HACHISUKA^{*2}, Kayoko KURITA^{*3}, Shin'ichi WARISAWA^{*4},

^{*1} 東京大学大学院新領域創成科学研究科

^{*1} The University of Tokyo, Graduate School of Frontier Sciences

^{*2} 東京大学大学院情報学環

^{*2} The University of Tokyo, Interfaculty Initiative in Information Studies

^{*3} 東京大学大学院教育学研究科

^{*3} The University of Tokyo, Graduate School of Education

Email: msatake@s.h.k.u-tokyo.ac.jp

あらまし：英語は日本語に比較して複雑な音韻を持ち、話者の口元を見て発音を判別する必要性の高い言語であると言われている。しかしながら、日本で実施されている英語音声指導においては視覚情報が提示されることはまれで、聴覚情報のみを与えて聴取（リスニング）させるケースが一般的である。本研究では日本語母語話者が特に混同しやすいとされる R 音と L 音との効果的な聞き分け手法の提案を目指し、発音を強調するための合成視覚情報を付与する学習法の有効性について検討した。

キーワード：マガーク効果、錯聴、視聴覚、英語音声聴解学習、日本語母語話者

1. はじめに

1.1 マガーク効果

イギリスの心理学者である McGurk らは 1976 年、人間の聴覚認識は視覚刺激に影響を受けるという実験結果を報告した⁽¹⁾。例えば、/ga/と発音する話者の口元の映像に、/ba/と発音された音声を合成した動画を提示した場合、被験者の 98 %が/da/と発音していると答えた。この「錯聴」現象はマガーク効果 (McGurk Effect) と呼ばれている。

1.2 マガーク効果の言語・文化差

マガーク効果について、日本語母語話者と英語母語話者とでは異なった現象が確認されている。1994 年、英語母語者と比較して日本語母語者にマガーク効果が生じにくいとする実験結果が示された⁽²⁾。この結果には日本語における母音と子音の数が英語と比較して圧倒的に少ないことが関係する可能性がある⁽³⁾と指摘されている。これは日本語母語話者が聴覚情報に依存して音声を判別する傾向がある一方で、英語母語話者は比較的複雑な英語の音韻を区別する必要があるために話者の口元の視覚情報を活用しているという説である⁽³⁾。

2. 目的

1.2 で述べたことから、日本語母語話者は話者の口元を注視しない傾向があることが考えられる。実際、日本の英語音声教育において話者の口元に注目させるような教授方法が取られることは少ない。例えば、聴解学習（リスニング）については聴覚情報のみを頼りに英語音声を聴取させる形式が一般的である。

しかしながら、現行の聴解学習の効果には限界がある。例えば日本語母語話者は R 音と L 音を聴取する際にどちらも日本語におけるラ行の音として認知

する傾向にある⁽⁴⁾。ここで、話者の口元の視覚情報を付与し、マガーク効果を活用することで音声を判別しやすくできるのではないかと考えた。

マガーク効果は、ある音声に対し、調音位置の異なる映像が合成されている場合に発生する⁽¹⁾。ここで、日本語母語話者が一般に認識できる調音位置は両唇音 (/p/, /b/, /m/), すばめ唇音 (/w/), 非唇音 (/t/, /d/, /n/, /k/, /g/, /r/, /z/等大部分を占める子音) の 3 つしか存在せず⁽⁵⁾、日本語母語話者に対してマガーク効果を検証する際は非唇音を発音する音声に両唇音を発音する映像を合成することがほとんどである。そこで本稿では新たな試みとして、すばめ唇音を含む“White”と発音する映像を“Right”と発音する音声に合成することで、聞き手が R の発音の特徴をより強く感じることができると可能性を検証する。

3. 実験

3.1 用いる素材

実験に用いた動画・音声素材を以下に示す。なお、発音と編集は著者が実施した。



図 1 WR 合成動画のキャプチャ

“Right”と発音する話者の口元を映した動画（以下 R 動画），“White”と発音する話者の口元を映した映像と、R 動画から抽出した“Right”と発音する音声を合成した動画（以下 WR 合成動画），“Light”と発音する話者の口元を映した動画（以下 L 動画）、R 動画

から抽出した“Right”と発音する音声(以下R音声), L 動画から抽出した“Light”と発音する音声, 以上の5通りの素材を用い実験を実施した。

3.2 手順

本実験は東京大学倫理審査専門委員会の承認を得て実施した(承認番号: 21-123)。実験は Google フォームを用い, リモートで実施した。実験参加者には実験手順を口頭および文面で説明し, フォームにも同様の説明を記載した。参加者は20代から30代の健康な男女10名であり, ランダムに5名ずつをA群, B群に振り分けた。参加者が体験した実験の流れは以下の通りである。

- ① 動画・音声ファイルをそれぞれ2回ずつ再生し聴取する。再生順は各群で以下のように振り分けた。A群: R音声, L音声, R動画(1回目), L動画, WR合成動画, R動画(2回目)の順。B群: R動画(1回目), WR合成動画, L動画, R動画(2回目), L音声, R音声の順。
- ② 聴取した単語を以下の5段階で評価する。1. 確実にLightだと思われる 2. おそらくLightだと思われる 3. 判別できない 4. おそらくRightだと思われる 5. 確実にRightだと思われる

実験結果の集計のため, 以下のような点数計算を実施した。ある参加者がR音声について「確実にR」を選んだ場合は2点, 「おそらくR」では1点, 「判別不能」では0点, 「おそらくL」ではマイナス1点, 「確実にL」ではマイナス2点とした。

3.3 結果

以下, 表に実験結果を示す。なお, A群1名に実験内容の認知齟齬があったことが実験後に判明したため当該データは結果の分析から除外した。よってA群については4名分の結果を示している。

表1 A群の点数

A群	確実にL	おそらくL	判別不能	おそらくR	確実にR	合計
(1)R音声	0	0	0	1	6	7
(2)L音声	8	0	0	0	0	8
(3)R動画	0	0	0	0	8	8
(4)L動画	8	0	0	0	0	8
(5)WR合成動画	0	0	0	0	8	8
(6)R動画	0	0	0	1	6	7

表2 B群の点数

B群	確実にL	おそらくL	判別不能	おそらくR	確実にR	合計
(1)R動画	0	0	0(1)	1	6	7
(2)WR合成動画	0	0	0(2)	1	4	5
(3)L動画	8	0	0	-1	0	7
(4)R動画	0	-2	0	1	4	3
(5)L音声	4	1	0	-2	0	3
(6)R音声	0	-1	0(1)	1	4	4

表3 A群とB群を合わせた点数

A群+B群	確実にL	おそらくL	判別不能	おそらくR	確実にR	合計
R音声(A1, B6)	0	-1	0(1)	2	10	11
L音声(A2, B5)	12	1	0	-2	0	11
R動画(A3, B1)	0	0	0(1)	1	14	15
L動画(A4, B3)	16	0	0	-1	0	15
WR合成動画(A5, B2)	0	0	0(2)	1	12	13
R動画(A6, B4)	0	-2	0	2	10	10

4. 考察

今回の実験において, 聴覚情報のみを用いて聴取する場合よりも視覚情報を付与して聴取する場合で全体的に正答率が高かった。このことから, 視覚情報を付与した聴解学習の有効性が示唆された。

一方で, マガーク効果を活用することができたとは言いがたい結果となった。表3内の合計網掛け部に示したように, 実験参加者が1回目にR動画で聴取した場合は合計15点, WR合成動画で聴取した場合は合計13点, 2回目にR動画で聴取した場合は合計10点であった。この結果から, 1回目に見たR動画と次に見たWR合成動画, そして2回目に見たR動画とをそれぞれ無意識のうちに比較して混乱が生じてしまったことが考えられる。このように, 情報の提示順が実験結果に影響する可能性が判明した。特にマガーク効果の活用を狙ったWR合成動画の視聴のタイミングは, 平常通りの発音動画に対する聴取正答率を下げる可能性が示唆されたことより, 慎重な検討が必要である。また, R動画を見せずにWR合成動画のみを見せた場合の結果は今回検証できていない。よって今後は, 視聴する素材の提示順序を考慮して最初にR動画を提示することなく合成動画を提示する場合の効果を検証したい。

5. まとめ

本研究では, マガーク効果(錯聴)を活用した効果的な英語音声聴取学習法の開発を目指して実験を実施した。結果として, 聴覚情報のみで聴取する場合と, 視覚情報を付与して聴取する場合とでは後者で聴取正答率が向上することは確認できた。

しかしながら, マガーク効果の活用を目指して作成した合成動画で聴取する場合の正答率向上は認められなかった。この原因としては, 情報の提示順により参加者の混乱を招いてしまったことが考えられる。次回以降の実験では純粋にマガーク効果の有効性を検証し, 効果的な教育システムへの応用可能性を明確にする。

参考文献

- (1) McGurk, H., & MacDonald, J: "Hearinglips and seeing voices" Nature, 264 (5588), pp.746-748 (1976)
- (2) Sekiyama, K: "Differences in auditoryvisual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility" Journal of the Acoustical Society of Japan, 15 (3), pp.143-158 (1994)
- (3) 積山薫: "視覚と聴覚による音声知覚—言語/文化による差とその発達—", Cognitive Studies, 18(3), pp.387-401, (2011)
- (4) 岡本真砂夫: "音響音声学に基づく“Left”と“Right”の分析—児童の混乱要因—", 小学校英語教育学会誌, 19(1), pp.86-100, (2019)
- (5) 積山薫, 城和貴, 梅田三千雄: "単音節の読唇による混同行列の分析: 多次元尺度法による知覚属性の検討", 電子情報通信学会技術研究報告, 87(417), pp.29-36, (1988)