

## 促音聞き取りにおける視覚情報の効果

本橋美樹

### 要旨

近年の IT テクノロジーの向上により、第二言語習得においてピッチ曲線を見せながら聞き取りトレーニングを行うというような聴覚と視覚の二つのモダリティーを用いたトレーニングの有用性が報告されている。本研究は学習者の促音の聞き取り向上に、音声波形という視覚情報が効果があるのかを検証した。視覚情報が与えられた実験群と音声情報のみの統制群にプリテスト、ポストテストを行い結果を比較したところ、実験群に聞き取り能力の向上が見られた。学習者の苦手な音素の聞き取り向上に、視覚情報も合わせたトレーニングを積極的に取り入れることを提言したい。

【キーワード】 促音、モダリティー、視覚情報、音声波形

### 1. はじめに

私達は実際のコミュニケーションにおいて音声を知覚する際に、聴覚という単一のモダリティーだけに頼るのではなく、相手の発話時の口、唇の動き等の視覚という他のモダリティーからの情報も用いて情報を処理している。このような複数のモダリティーによる情報処理における研究は、成人の第一言語や聴覚障害者の言語療法において主になされてきている。代表的な例でいえば、発声された音声と口唇の動きが一致しない場合、口唇の動きに惑わされるという所謂マガーク効果をあげることができる (McGurk & McDonald 1976)。また、乳児の第一言語習得においても、口唇の動きが音声理解に重要な役割を持つことが様々な研究によって明らかにされている (Melzoff & Kuhl 1994 ; Patterson & Werker 1999 ; 麦谷ら 2004, 2006 など)。さらに、聴覚と視覚情報を統合させることにより音声処理能力が向上することが報告されている (Massaro 1998 ; Grant & Seitz 2000 など)。

第二言語習得の分野では主に単一のモダリティー、つまり聴覚のみに焦点を置いた研究が主であったが、近年では以下に示すように複数のモダリティーによるインプットが

第二言語の学習を促すという報告が増えてきている。実際に教室で促音のような特殊拍を初めて導入する際に、拍に合わせて手を叩いたり、黒板に書かれた文字をたたいたりして指導したことがある教師も多いであろう。教師は経験上、発声という聴覚からの情報だけでなく、手の動きや黒板の文字という視覚情報も統合させた方が指導に効果的であるということを知っているのである。そのような第二言語習得における複数のモダリティの働きに関する研究は、口唇 (Hardison 2005 ; Sekiyama 1997)、手振りなどのジェスチャー (Sueyoshi & Hardison 2005) などが学習者の音声知覚を促すという報告がされている。また、第二言語習得における従来の聴覚トレーニングは LL 教室などでヘッドホンから流れる音声テープを聞く、というのが主流であったが、近年の IT 技術の発展によりコンピューターによる視覚情報を用いたトレーニングの効果も報告されている。早くからテクノロジーを用いた第二言語学習への応用が示唆されていたものの実験による実証までは至らなかったが、de Bot (1983) において初めて、イントネーションのディスプレイを用いたオランダ語話者による英語学習への効果が実験により示され、Hardison (2004) ではさらにフランス語を学習する英語母語話者を対象に、従来の音声のみのトレーニングに加えてコンピューターのディスプレイ上に発話と同時にピッチ曲線を見せるトレーニングを行い、その効果を報告している。その他ピッチ曲線と共に知覚と生成のトレーニングを行う効果を検証した研究に Anderson-Hsieh (1994)、Chun (2002)、Hardison (2008)、Hincks & Edlund (2009)、Levis & Pickering (2004) などがある。

上記のような研究において、日本語学習者を対象としたものはまだ見られない。そこで本研究では日本語学習者にとって習得が難しいとされる促音の習得に、そのような視覚情報が役に立つかを検証する。日本語母語話者にとって促音を知覚する上での重要な音声上の手がかりは促音部の持続時間であり、破裂音の場合は (例 : /p/t/k/) 無音区間、摩擦音であれば (例 : /s/) 摩擦区間の持続時間である。さらにその他の要因、たとえばピッチアクセント (大深 2003) や先行母音の長さ (平田 1990) も手がかりとして用いていることが報告されている。対して学習者は母語話者と違った手がかりを用いて促音を認識しているという研究報告も多い。例えば、Hayes (2002) では 2 拍語 (非促音語) とそのミニマルペアである 3 拍語 (非促音語) の同定テストを行ったが、促音と非促音の持続時間の差が大きいほどその区別は簡単であり、小さいほど難しいという結果であった。つまり、破裂音/t/は促音と非促音の閉鎖持続時間の差は大きいので区別は容易で、逆に摩擦音/s/は、促音と非促音の時間差が小さいので区別は困難であるという、促音種による知覚の違いがあると結論付けられている。さらに、Yamagata & Preston (1999) は英語からの外来語を聞かせカタカナで書かせるという調査をしたところ、日本語母語話者が促音とする音を長音として聞く場合が多いという結果になった。学習者は促音の拍

数、つまりモーラウェイト (mora weight) の認識は正しくできるものの、そのウェイトをどこに置くかが難しいことが伺える。このモーラウェイトのミスポジションについては Motohashi (2007)、Motohashi-Saigo & Hardison (2010) でも報告されている。これらの実験では、大学で日本語を学習中のアメリカ人を対象に /t/、/k/、/s/ の促音種別に聞き取りテストを行ったところ、摩擦音 /ss/ が一番結果が悪く、破裂音 /k/ が一番良かった。興味深いのは Yamagata & Preston (1999) の結果と同じように摩擦音の促音の摩擦持続時間を、長音として聴き間違える学生が多かったことである。やはりモーラウェイトの認識はできるもの摩擦の持続に惑わされ、聞き取ったモーラの「割り当て」かがわからない学生が多いのであろう、と結論づけている。

本研究では、その摩擦の持続が正しい促音の知覚を妨げていると仮定し、音声教育のトレーニングツールとしての音声波形の可能性を検証する。音声波形は、以下のように発話の流れの中でターゲットとなる促音の長さを示す視覚情報として提示する役割をする。本研究で用いられた音声波形の例を以下に示す。図 1 は促音部が破裂音 /k/ である語 /akku/ の音声波形であるが、促音の閉鎖部分が無音として示されている。対して図 2 は促音部が摩擦音 /s/ である語 /assu/ の音声波形であるが、持続時間にも全くの無音とはならず摩擦が続いている。

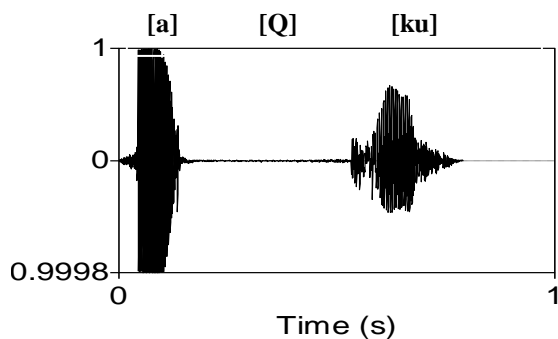


図 1 /akku/ の音声波形

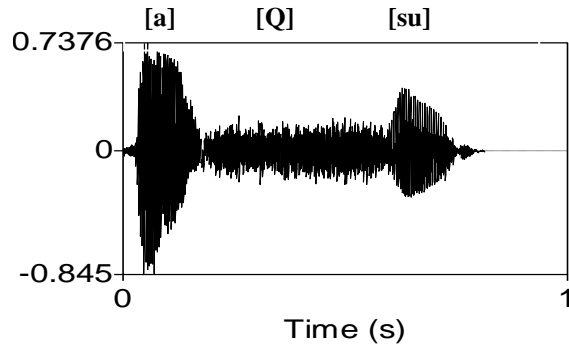


図2 /assu/の音声波形

このように、音声波形を示しながら音声を聞くと促音の閉鎖時間と隣接する音との時間的割合がはっきりする。また、摩擦音の促音は無音区間ではなく摩擦が続いていることがわかる。この視覚情報が、学習者の促音の知覚、特に Motohashi-Saigo & Hardison (2010) で報告されている正解率が一番低く、長音と間違える傾向にあった摩擦音/ss/の促音の、音声処理を向上させるか検証する。

## 2. 実験

### 2.1 概要

本実験は音声波形という視覚情報を与えられた実験群 (AV-group) と音声のみの情報しか与えられない統制群 (A-only group) に促音聞き取りのプリテストとポストテストを行った。実験群と統制群のスコアを比較することにより促音の学習を促すツールとしての音声波形の効果を検証した。

### 2.2 被験者

アメリカの大学で日本語を学習中である英語母語話者 64 名が実験に参加した。参加者を実験群 (AV-group) 33 名、統制群 (A-only group) 31 名の二つのグループに分けた。参加者の変数をできるだけ少なくするため、実験以前に学習経験のない初級学習者に制限した。参加者は全員、初級日本語コースに在籍中で実験の時点で学習歴 3 ヶ月であった。日本語のクラスは週に 5 日、50 分間ずつで英語による文法のレクチャーと日本語の口頭練習からなる。日本語の発音について大まかな説明はあるが、特に促音について特化した練習はされていない。

## 2.3 刺激音

刺激音は促音と非促音のミニマルペアで、日本語の実在語（例：/assi/）と無意味語（例：/asse/）から成った。全 20 語のうち 10 語は促音の後続子音が摩擦音/s/の三音節語（例：/assu/）と促音を含まないミニマルペアの二音節語（例：/asu/）で、それぞれの後続母音は/i/ /e/ /a/ /o/ /u/から成った。他の 10 語は破裂音（/k/と/t/）と長音を含む語から成りダミーの刺激語として用いられた。先行母音長などの影響も考えられるため、第一拍は全て/a/で統一した。また、全て頭高アクセントに統一して発音された。音声はすべて東京方言の日本語母語話者の女性によって、ミニディスク（SONY MD MZ-RH10-S）とマイクロフォン（SONY ECM-CS10）を用いて録音された。

## 2.4 手順

実験は AV グループと A-only グループに対しそれぞれ別の日程で、プリテスト、レクチャー、ポストテストの順で別々に行われた。

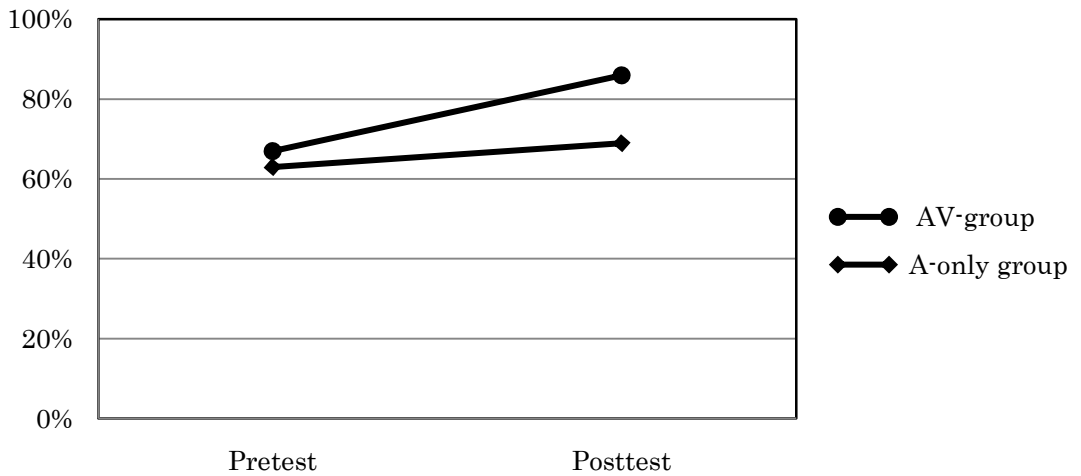
まず初めにプリテストが行われた。通常日本語のクラスが行われている教室でスピーカーを通し刺激音が一度ずつ再生された。それぞれの刺激語に対し、ミニマルトリプレットである三つの選択肢、すなわち促音（例：/assu/）、非促音（例：/asu/）、それから長音（例：/aasu/）を与える三肢強制選択法を採用し、被験者は刺激音を聴き、聞こえたと思ったものに回答用紙に丸をつけるように指示された（テストの例については添付資料を参考されたい）。

プリテストに続いて AV-group の被験者は音声波形を用いて摩擦音の促音について著者（当時の担当教師）によりレクチャーを受けた。どのように摩擦音の促音部が音声波形で視覚化されて見えるか、また、破裂音の無音区間の音声波形との違いも視覚情報と共に説明された（例：/akku/ vs. /assu/）。また、著者が促音を含む語をマイクロフォンを通して発音し、音声分析ソフトの Praat<sup>1</sup>を用いてリアルタイムでスクリーンに音声波形が作られる様子を見せた。このレクチャーの目的は被験者が、様々なタイプの摩擦音を含む促音とスクリーン上の音声波形の関係を理解できるようにすることであった。このレクチャーは5分間で、その後ひき続きポストテストが行われた。プリテストと同様の手順で同じ刺激音を用いて行われたが、違う順序で提示された。A-only group にも全く同じ手順で摩擦音と破裂音を含む促音の違いについてのレクチャーが行われたが、口頭による説明のみで音声波形による視覚情報は与えられなかった。どちらのグループのレクチャーとも、予め用意した同じ原稿に沿って行われた。

## 3. 結果

AV-group と A-only group のプリテスト、ポストテストの結果を以下に示す。

表1 プリテスト、ポストテストの正解率



	AV-group Group (M, SD)	A-only Group (M, SD)
Pretest	67% (6.7, 1.25)	63% (6.3, 1.01)
Posttest	86% (8.6, 0.95)	69% (6.9, 1.19)

(Test items N=10)

両グループのプリテストの結果は、t 検定の結果、有意な差は見られなかった ( $t(18) = -0.726, p = .477$ )。両グループの聞き取り能力はレクチャー前では同等と言える。また、プリテストにおける解答の詳細を見ると、促音を長音とした解答が全誤答の 80%にも上った。そこで、視覚情報の効果をみるために、被験者グループ (AV と A-only) を被験者間変数、テスト (プリテストとポストテスト) を被験者内変数として分散分析を行った結果、グループおよびテストに主効果があり ( $F_t(1, 18) = 15.918, p = 0.001, F_g(1, 18) = 7.694, p = 0.013$ )、被験者グループとテストに交互作用 ( $F(1, 18) = 4.818, p = 0.042$ ) が確認された。つまり、プリテストからポストテストの比較により、レクチャーを通して視覚情報を得たグループの方が摩擦音を含む促音の聞き取りにおいて正答率が高くなるという結果が認められた。

#### 4. 考察

本研究では、摩擦音の促音の聞き取りを困難にしているのは持続する摩擦のためであるという仮定から、摩擦音の音声波形という視覚的情報を用いたトレーニングを行い、その効果を検証した。他の研究 (Motohashi-Saigo & Hardison 2010) において、学習者は正しくモーラウェイトを認識できても、そのモーラの「割り当て」に困難を見せていた

という報告があったが、本調査においては摩擦音の促音についての説明が、コンピューターを用いた音声波形をリアルタイムな提示とともになされたため、モーラウェイトをどこに割り当てるべきか、注意が正しい位置に向けられたと思われる。ポストテストでは視覚情報は与えられなかったが、音声波形の提示がなくとも被験者はモーラを割り当てる位置を正しく認識できたということができていたと考えられる。この結果は音声指導において、従来聴覚という一つのモダリティーに頼ってきた音声教育に、視覚という他のモダリティーを用いることの有用性を示唆する。

本実験では視覚情報をレクチャーを通して一度与えただけでありその効果の持続性は検証できなかった。また知覚と生成の関係は様々な報告で示唆されているが (Bradlow et al. 1997 など)、本研究では生成への影響は検証できなかった。トレーニングとして長期間にわたって視覚と聴覚の二つのモダリティーへの情報を与えた場合の知覚と生成への効果については、Motohashi-Saigo & Hardison (2009) を参照されたい。

## 注

(1) Praat はアムステルダム大学の Paul Boersma 氏と David Weenink 氏によって開発された音声分析ソフトウェアである。 <http://www.fon.hum.uva.nl/praat/>

## 参考文献

- 大深悦子 (2003) 「促音/t/の知覚：アクセント型と促音・非促音の音響的特徴による違い」『音声研究』 7(1), 70-76.
- 平田由香里 (1990) 「単語レベル・分レベルにおける日本人の促音の聴き取り」『音声学会会報』 194, 23-28
- 麦谷綾子・小林哲生・開一夫 (2004) 「日本人幼児における母音の視聴覚音声口形マッチングの検討」『音声研究』 8(1), 85-95
- 麦谷綾子・小林哲生・石塚健太郎・天野成昭・開一夫 (2006) 「日本語学習乳児の音声口形マッチングの発達に関する母音/i/を用いた検討」『音声研究』 10(1), 96-108
- Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, 11(4), 5-22.
- Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101 (4), 2299-2310.
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.

- de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, 26(4), 331-350.
- Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of Acoustical Society of America*, 108(3), 1197-1208.
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language*, 8(1), 34-52.
- Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, 26(4), 579-596.
- Hardison, D. M. (2008). The visual element in phonological perception and learning. In M. C. Pennington (Ed.), *Phonology in context* (pp. 135-158). Basingstoke: Palgrave Mcmillan.
- Hayes, R. L. (2002). The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. *Coyote Papers*, 12, 28-41.
- Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology*, 13(3), 32-50.
- Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32, 505-524.
- Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Meltzoff, A. N. & P. K. Kuhl (1994). Faces and speech: Intermodal processing of biologically relevant signals in infants and adults. In D. J. Lewkowitz & R. Lickliter (eds.), *The development of intersensory perception: Comparative perspectives*. Hillsdale, NJ: Lawrence Erlbaum, 335-369.
- Motohashi, M. (2007). Acquisition of geminate consonants in Japanese by American English speakers (Unpublished doctoral dissertation, Michigan State University, 2007). *Dissertation Abstracts International*, 68, 1912
- Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. With M. Motohashi Saigo. *Language Learning & Technology*, 13, 29-47.
- Motohashi-Saigo, M., & Hardison, D. M. (2010). Development of perception of second language Japanese geminates: role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, 31(1), 81-99.
- Patterson, M. L., & J. F. Werker (1999). "Matching phonetic information in lips and voice is



robust in 4.5-month-old infants." *Infant Behavior and Development*, 22(2), 237-47.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, 59(1), 73-80.

Sueyoshi, A. & D. M. Hardison (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661-699.

Yamagata, A., & Preston, D. (1999). English learners' acquisition of the Katakana spelling of English loan-words in Japanese. In M. Wysocka (Ed.), *On language theory and practice*. In honor of Janusz Arabski on the occasion of his 60th birthday (pp. 276-292). Katowice, Poland: University of Silesia.

(motohash@kansai.ac.jp)

添付資料（プリテスト、ポストテストの例）

You will listen to some Japanese words. Each question has 3 options. Circle the word you're hearing. You will hear each word only once. If you are not sure which word you heard, make your best guess. Please do not leave any numbers unanswered. Don't be concerned if you don't understand the words.

Examples:

You will hear: kitte

Circle: a. kite  
⑥ kiite  
c. kitte

---

- 1) a. asa  
b. aasa  
c. assa
- 2) a. asi  
b. aasi  
c. assi
- 3) a. ata  
b. atta  
c. aata
- 4) a. ase  
b. aase  
c. asse