

## 見真似発音練習システムにおける口唇動作の誤り指摘

小田 まり子<sup>†</sup> 一ノ瀬 峻<sup>†</sup> 小田 誠雄<sup>‡</sup>

<sup>†</sup>久留米工業大学工学部 <sup>‡</sup>福岡工業大学短期大学部

E-mail: <sup>†</sup> {mari, g07m202} @cc.kurume-it.ac.jp, <sup>‡</sup> oda@fjct.fit.ac.jp

人間の音声認識における双方特性化や読唇術に着目した聴覚障害児用発音練習システム "Lip Reading あい"を開発している。本システムでは、学習者が発音時にネットワークカメラから取り込んだ学習者自身の口唇動作をパソコン画面上に表示し、モデル話者の口唇動作や口唇動作CGアニメーションと比較しながら発音練習が行える。本研究では、見真似発音練習をより効果的にするための口唇動作誤り指摘機能について報告する。

## An Error Detection Function for a Pronunciation Practice CAI system "Lip Reading あい"

Mariko ODA<sup>†</sup> Shun Ichinose<sup>†</sup> Seio ODA<sup>‡</sup>

<sup>†</sup> Kurume Institute of Technology

<sup>‡</sup> Fukuoka Institute of Technology

E-mail: <sup>†</sup> {mari, g07m202} @cc.kurume-it.ac.jp, <sup>‡</sup> oda@fjct.fit.ac.jp

We developed a pronunciation practice CAI system called "Lip Reading あい" for deaf children which is based on the bimodality of human speech and lip reading techniques. This system enables a learner to practice pronunciation by observing the model speaker's mouth movement and allows to compare his/her own mouth movement. It simultaneously allows to compare his/her own mouth movement with model speaker's mouth and a computer graphics animated mouth on a computer screen connected to a video camera. This system also provides an error detection function instructing users how to change their mouth movements and correct pronunciation problem.

## 1. はじめに

人間は、音声言語を確認する際、聴覚からの情報だけでなく無意識に視覚からも多くの情報を取り込んで認識している。すなわち、聴覚と視覚という2つの異なる感覚を統合して音声を知覚している[1]。騒音などにより音声が聴き取り難い状況下においても、人間は発話者の舌、歯、口唇の形などの視覚情報を利用して話された内容を理解できる[2]。なまりがある話者や外国人による音声言語の場合にも、話者の顔を見ることが理解への大きな助けになる[3]。また、幼児は大人が話す際の顔の表情をじっと見ながら聞いた言葉を何度も真似をして繰り返す行為をする。幼児が唇を使って出される音(唇音)を初期に習得しやすいのは、唇音が容易に「目に見える」音で、音そのものだけでなく音を出す時の顔の動きが模倣しやすいためだと考えられている[4]。そこで、我々は、人間の持つ読唇術のテクニックや幼児の音声言語習得過程に着目した見真似による /l-/r/ 発音練習 CAI システムを開発した[5]。評価実験により、モデル話者の口唇動作を真似して発音練習することにより自分自身の口唇動作が矯正され、発音も改善できることを確認した[5]。

我々は、現在、小学生以下の聴覚障害児のための見真似発音練習システム“Lip Reading あい”を開発している[6]。聴覚障害者の場合、口唇動作を読み取る読唇の能力を身につけることによりコミュニケーションがとりやすくなる。また、わかりやすい口唇動作が発音を補い、相手に正しく伝わりやすくなるという意味でも口唇動作は重要である。

実際の聾学校の「口話」の授業においても、先生が最初に口唇動作を強調しながら発音をしてみせ、子供たちがそれを真似る形式で発音練習をしている。また、子供たちが自分自身の口唇を鏡で見ながら行う発音練習もある。従って、

“Lip Reading あい”による見真似発音練習は聾学校の授業内容とも合致しており、児童にもわかりやすい発音練習方法であると言える。見真似発音練習の際には、口唇動作の特徴が掴みやすいモデル話者の口唇映像だけでなく、モデル話者の口唇動作に基づいて作成した口唇動作CGアニメーションも見ることができる[7]。

“Lip Reading あい”による見真似発音練習では、学習者自身がモデルとなる口唇動作と自己のものを比較し、良否を判定するため、一旦、学習者が口唇動作を正しいと判断した場合、それ以降の口唇動作に改善が見られないという問題が生じる。また、口唇動作が改善してはいても、まだ十分に矯正できていない場合もある。

そこで、我々は“Lip Reading あい”に口唇動作の誤り指摘機能を付けた。この機能は、学習者の発音時における口唇動作(口唇の縦幅、横幅の変化)を分析し、その誤りを指摘する。学習者は自己良否判定による口唇動作誤り矯正(以後、自己矯正と呼ぶ)の後、システムによる誤り指摘を受けることができるので、自己矯正では十分な効果が得られなかった学習者も、システムによる誤り指摘により自己の口唇動作の誤りを矯正(以後、システム矯正と呼ぶ)できる。

本稿では、口唇動作誤り指摘機能を用いた見真似発音練習の方法と、学習効果評価実験について述べる。

## 2. “Lip Reading あい”による自己矯正とシステム矯正

### 2.1 “Lip Reading あい”の概要

“Lip Reading あい”はWWW上で実現されており、WEBブラウザを用いて簡単に利用できる。発音時における口唇動作をネットワークカメラより取り込み、WEBブラウザ上に表示し、画面上でモデル話者の口唇動作と比較しながら発

音練習できる。

ネットワークカメラを用いることにより、生徒がブラウザ上で学習している様子を教師も教師用パソコンから確認することができ、生徒に対してアドバイスをすることができる。

また、ネットワークカメラから取り込んだ口唇動画像をサーバにアップロードすることにより、口唇画像ファイルをサーバに保存できる。保存したファイルは、生徒も教師もブラウザ上で再生し、口唇動作を確認できる。学習者は保存していた自分自身の口唇動作をモデル話者のものと比較しながら見ることができ、口唇動作を矯正しながら発音練習ができる。また、教師は保存していた学習者の発音時口唇動画像を見ることにより、その上達度を把握できる。

## 2.2 “Lip Reading あい” による発音練習

“Lip Reading あい” を用いて発音練習をする際、通常、図1の順序で発音練習することを想定している。

### 2.2.1 口唇動作の理解・確認部

学習者は、まず発音時のモデルとなる口唇画像を見たり、モデル話者の音声や口唇動画像を再生して、発音方法を理解、確認する。学習の際、モデル話者の口唇形状を基にして開発した口唇動作CGも見ることができる。図2.に母音発音時における口唇形状のCG例を示す。

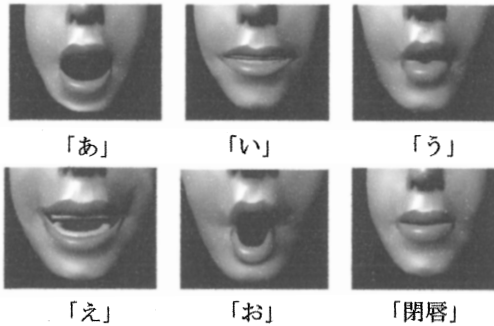


図2. モデル話者の口唇形状CG例

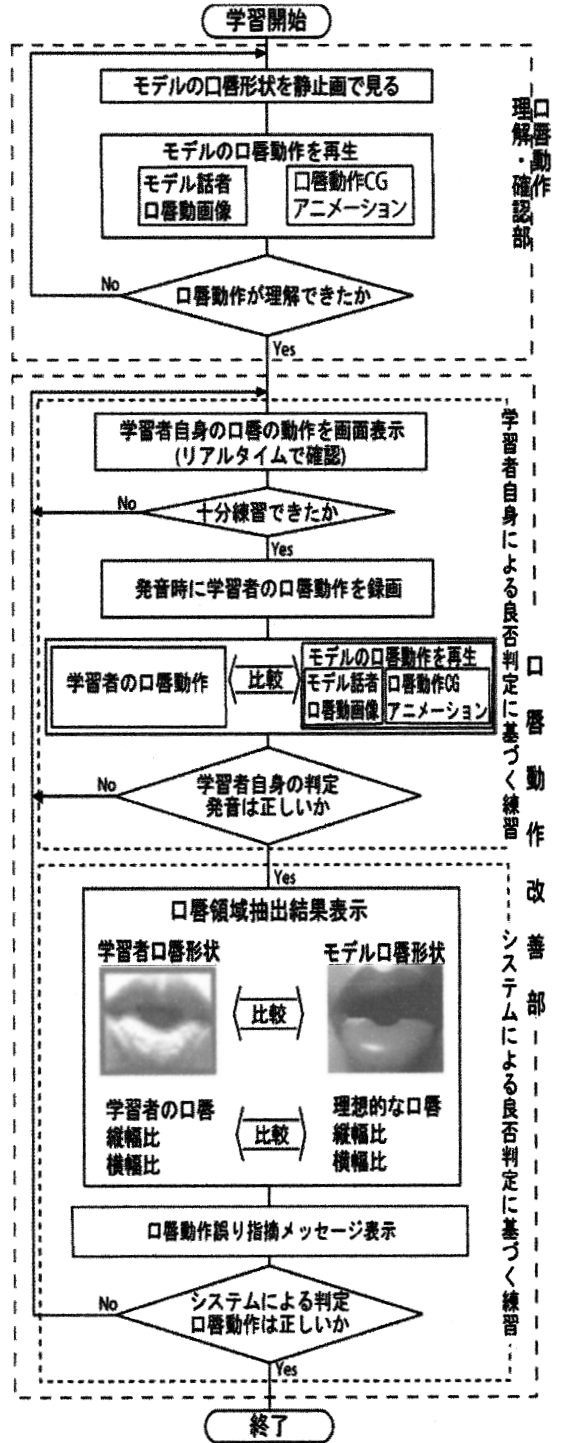


図1. “Lip Reading あい” の発音練習モデル

## 2.2.2 見真似による自己矯正型発音練習

口唇動作改善部では、自分の口唇をビデオカメラで撮影し、画面上に表示しながら自己矯正する。学習者は、画面上で自分の口唇動作と、モデル話者の口唇動作や口唇動作CGアニメーションと比較し、発音時の口唇動作が正しいかどうかを確認する。図3、図4に、見真似による自己矯正型発音練習画面例を示す。

口唇動作を比較しながら発音練習する際の音の組み合わせは文献[8]を参考に選んだ。



図3. モデル話者による自己矯正型発音練習画面例

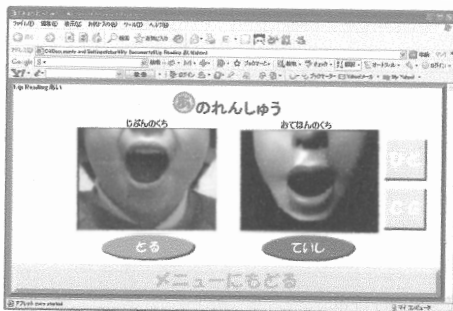


図4. 口唇動作CGアニメーションによる自己矯正型発音練習画面例

## 2.2.3 口唇動作誤り指摘機能によるシステム矯正型発音練習

学習者がメニューより「あやまりしてき」を選ぶと、口唇動作誤り指摘が行われる。システムは取り込まれた学習者の口唇動作を解析し、発音

時における口唇領域画像を判定結果画面上に提示する。学習者は自分自身の発音時の口唇形状と口唇形状モデルを比較できる。また、自分の口唇画像上に理想的な口唇のサイズが示されるので、どの程度口唇形状を変更すべきかが具体的にわかる。さらに、口唇動作の良否判定結果および学習者へ誤り指摘メッセージも提示する。図5に、口唇動作誤り指摘画面を示す。

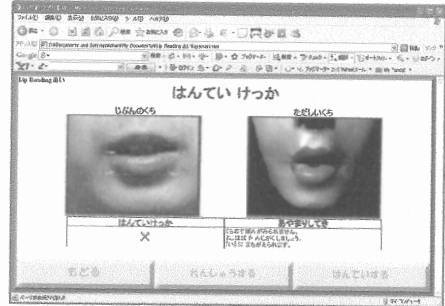


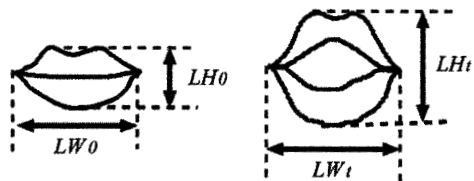
図5. 口唇動作誤り指摘画面例

誤りが指摘された場合には、再度、見真似による発音練習に戻り、口唇動作の自己矯正を行う。最終的に、システム矯正により正しいと判断されるまで練習を継続する。

## 3. 口唇動作誤り指摘機能の実現

### 3.1 口唇の特徴パラメータ

口唇動作は口唇の横幅、口唇の縦幅、上歯の幅、下歯の幅、歯間幅、上唇の幅、下唇の幅の7つのパラメータによって表現される[9]。



(左)基準：閉唇 (右)tフレーム後の口唇

図6：口唇特徴パラメータ

図6のように  $LH$  は口唇縦幅（上唇の上端と下唇の下端の間の距離）を  $LW$  は口唇の横幅（口唇



の左端と右端の間の距離を表す。この  $LH$  と  $LW$  を用いて、口唇縦幅比  $R(S, H, t)$ 、横幅比  $R(S, W, t)$  を求め、これを発音時の口唇動作を表す特徴パラメータとして定義する。ただし、縦幅、横幅の比率は、閉唇（口を閉じた状態）を基準の 1 とした場合の比率であり、(1)、(2) 式で表せる。

$$R(S, H, t) = LH_t / LH_0 \quad (1)$$

$$R(S, W, t) = LW_t / LW_0 \quad (2)$$

### 3.2 母音における口唇形状の違い

日本人話者 20 人が母音 ( $/a/$ 、 $/i/$ 、 $/u/$ 、 $/e/$ 、 $/o/$ ) を発音した時の、声が出た瞬間における口唇形状（縦幅比  $R(S, H, t)$ 、横幅比  $R(S, W, t)$ ）を分析した。

図 7 は、 $/a/$ 、 $/e/$  発音時における口唇形状を示している。 $/a/$  における口唇縦幅比は平均 2.13（分散 0.157）、口唇横幅比は平均 0.995（分散 0.00434）で、 $/e/$  における口唇縦幅比は平均 1.85（分散 0.108）、口唇横幅比は平均 1.08（分散 0.0110）である。 $/a/$  と  $/e/$  はどちらも大開き母音であるが、口唇形状が異なることがわかる。T 検定の結果からも、 $/a/$  と  $/e/$  の縦幅比、横幅比には各々 0.01% 水準の有意差が認められた。

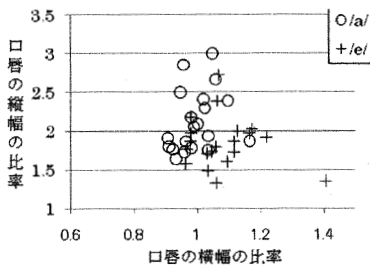


図 7 :  $/a/$ 、 $/e/$  発音時の口唇形状の違い

図 8 は、日本人話者の  $/i/$ 、 $/e/$  発音時における口唇形状を示している。 $/i/$  における口唇縦幅比は平均 1.56（分散 0.0677）、口唇横幅比は平均

1.16（分散 0.0100）で、 $/i/$  と  $/e/$  の口唇形状も異なる。T 検定の結果からも  $/a/$  と  $/e/$  の縦幅比には 0.000001%、横幅比には 0.1% 水準の有意差が認められており、縦幅の違いが大きい。

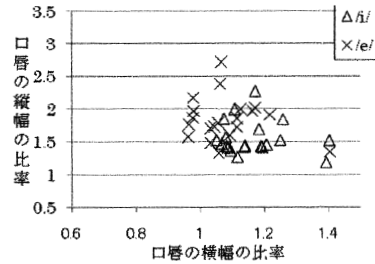


図 8 :  $/i/$ 、 $/e/$  発音時の口唇形状

図 9 は、日本人話者の  $/u/$ 、 $/o/$  発音時における口唇形状を示している。 $/u/$  における口唇縦幅比は平均 1.45（分散 0.0713）、口唇横幅比は平均 0.831（分散 0.0106）で、 $/o/$  における口唇縦幅比は平均 1.80（分散 0.214）、口唇横幅比は平均 0.810（分散 0.00505）である。 $/u/$  と  $/o/$  はどちらも小開き母音であるが、口唇の縦方向の開きが  $/o/$  の方が大きく、口唇形状が異なることがわかる。T 検定の結果からも、横幅比には有意差が認められないが、縦幅比は 0.01% 水準の有意差が認められた。

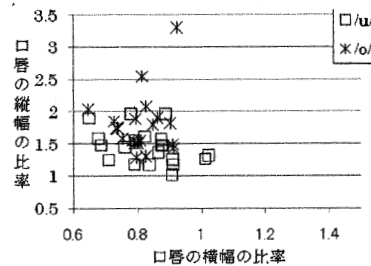


図 9 :  $/u/$ 、 $/o/$  発音時の口唇形状

### 3.3 口唇動作分析結果のフィードバック

発音時における平均的な口唇形状や理想的な口唇形状（音声なしでも相手に正しく伝わるよう

な特徴の強く表れた口唇形状)を、学習者に提示することは有効である。さらに、学習者自身の口唇動作を分析し、理想的な口唇形状との類似度を求めることにより、口唇動作の良否判定(誤り指摘)ができれば、有効なフィードバックになる。母音発音時の口唇形状には有意差が認められたので、“Lip Reading あい”による母音の発音練習において口唇動作誤り指摘機能を実現した。



図 10. 口唇動作の誤り指摘処理手順

図 10 は、口唇動作の誤り指摘処理手順を示している。発話時にカメラから入力された口唇動作画像は、1 秒間に 30 フレームの動画像として保存する。システムは、フレーム毎に画像をスキャンし、口唇の領域分割を行う。口唇領域の抽出に

は、Haar-Like 特徴と AdaBoost アルゴリズム [10][11] による Open CV のオブジェクト検出プログラムを用いた。次に、領域分割された口唇部分の縦幅、横幅を測定する。この時、学習者  $S$  の  $t$  枚目のフレームにおける口唇幅を  $(S, D, t)$  と表す。ただし  $D$  は、縦  $H$  もしくは横  $W$  である。また発声前の口唇の閉じた状態の口唇幅を基準として、口唇幅比  $R(S, D, t)$  を以下の式により計算する。

$$R(S, D, t) = L(S, D, t) / L(S, D, 0) \quad (3)$$

全てのフレームにおける口唇領域分割、口唇縦幅比  $R(S, H, t)$ 、口唇横幅比  $R(S, W, t)$  の計算処理が終了すると、学習者の口唇動作を表す口唇幅比の時系列データが得られる。この学習者口唇動作時系列データとあらかじめ用意したモデル話者による口唇モデルパターン、口唇動作誤りパターンとのパターンマッチングにより、口唇動作の良否判定と誤り指摘を行なう。

現在のシステムでは、母音発音時における以下の誤り指摘が行える。

- ・口唇縦方向への開きが小さすぎる場合
- ・口唇縦方向への開きが大きすぎる場合
- ・口唇横方向への開きが小さすぎる場合
- ・口唇横方向への開きが大きすぎる場合
- ・口唇のすばみがみられない場合
- ・口の動きが縦横ともに小さすぎる場合
- ・/a/と/e/, /u/と/o/などの区別ができない場合

#### 4. 発音練習における誤り指摘の効果

誤り指摘機能を用いた口唇動作矯正の有効性を調べる実験を実施した。

##### 4.1 評価実験の目的

学習前後における口唇形状の違いを分析し、誤り指摘機能の有効性を調べる。また、正同定率(被

験者に口唇動作のみで正しく伝わる率)による有効性についても調べる。

#### 4.2 実験方法

- (1) 日本人学習者 8 人 (L1~L8) の発音練習前の音声および口唇動作をビデオ撮影する。
- (2) “Lip Reading あい” を用いて、誤り指摘機能を用いた見真似発音練習を行なう。
- (3) 発音練習後の学習者 8 人の音声および口唇動作をビデオ撮影する。
- (4) 12 人の被験者が、学習者 8 人の学習前後における口唇動作を見て発音を同定し、その正同定率を調べる。
- (5) 学習前後における口唇形状(縦幅比、横幅比)の違いを調べる。

#### 4.3 実験結果と考察

表 1 : 学習前後における平均正同定率の比較

学習者	学習前	学習後
L1	83%	96%
L2	85%	97%
L3	92%	95%
L4	86%	94%
L5	73%	93%
L6	87%	96%
L7	86%	95%
L8	69%	88%
平均	83%	94%

表 1 より、誤り指摘機能を用いた口唇動作の矯正により、学習後の平均正同定率が 11% 高くなっている。口唇動作の誤りを効率良く学習し、口唇動作を改善できたため、口唇動作のみで正しく同定される割合が向上したと言える。

また、図 11 から図 15 に、学習前後における口唇形状変化図(図 11 は/a/、図 12 は/i/、図 13 は/u/、図 14 は/e/、図 15 は/o/)を示す。誤り

指摘により効率的な練習ができるため、学習後には、全ての母音において口唇形状がより理想的な値に近づいており、分散値が小さくなっている。

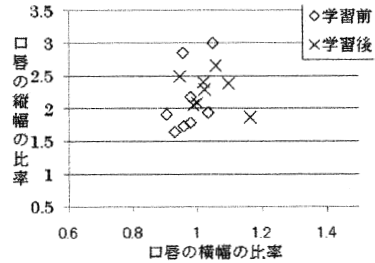


図 11 : 学習前後における/a/の口唇形状変化

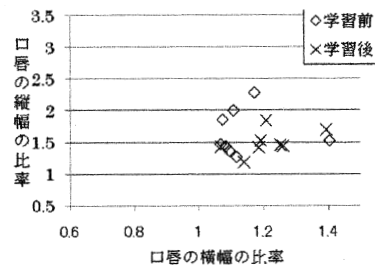


図 12 : 学習前後における/i/の口唇形状変化

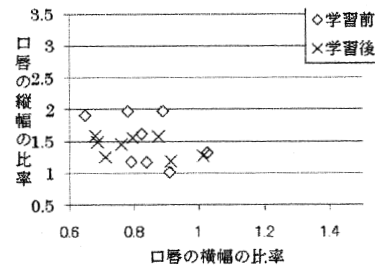


図 13 : 学習前後における/u/の口唇形状変化

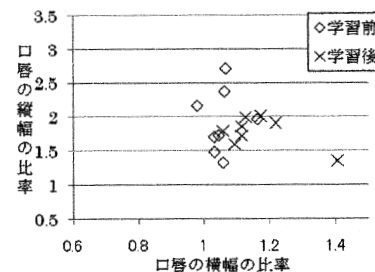


図 14 : 学習前後における/e/の口唇形状変化

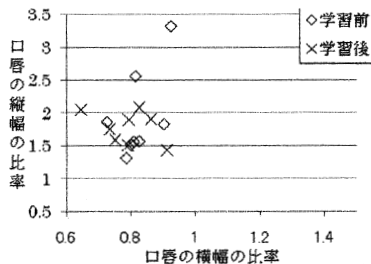


図 15：学習前後における/o/の口唇形状変化

## 5. おわりに

“Lip Reading あい”を用いた発音練習において口唇動作誤り指摘機能を実現した。日本語母音の場合は、全ての組み合わせにおいて口唇形状に有意差が見られたため、口唇動作の誤りを指摘することができた。また、誤り指摘機能を用いた学習前後で、口唇形状が理想的な値に近づくことも確認できた。発音練習する際、/a/と/e/、/i/と/e/、/u/と/o/などを組み合わせて学習し、口唇形状の違いを確認することにより、効果的に口唇の特徴を掴むことができる。

今後は、母音以外の口唇形状も分析し、誤り指摘できる発音の種類を増やしていく。

本研究を進めるにあたり、久留米工業大学情報ネットワーク工学科の学生諸氏の御協力を賜りましたことを感謝致します。また、本研究の一部は文部科学省科研費基盤研究 (C) (NO. 20500600、小田まり子)の補助を受けた。

## 文 献

- [1] H. McGurk and J. Macdonald, J. (1976)  
Hearing lips and seeing voices, Nature, pp.746-748
- [2] Sumbly, W and Pollack, I (1954)  
Visual contributions to speech intelligibility in noise. J.Acoust.Soc. Amer.,26(2), pp .212-215
- [3] Reiseberg, D., Mclean, J. and Goldfield, A. (1987)  
Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli, in hearing by eye
- [4] Susan. H. Foster-Cohen (1999)  
An Introduction to child language development. Addison Wesley Longman Limited(今井邦彦(訳)(2001) 子供は言語をどう獲得するか. 岩波書店
- [5] 小田まり子, 小田誠雄, 新井康平,  
“見真似による/l-/r/発音練習システムの効果”, 日本教育工学会論文誌 26(2), pp65-75(2002)
- [6] 小田まり子, 一ノ瀬 峻, 小田誠雄,  
“聴覚障害児のための見真似による発音練習システム” 電子情報通信学会技術研究報告, 信学技報 vol.107 No.179 WIT2007-25
- [7] 小田まり子, 一ノ瀬 峻, 小田誠雄,  
“見真似による発音練習システムのための口唇動作CGアニメーション”  
電子情報通信学会技術研究報告, 信学技報 vol.107 No.555 WIT2007-107
- [8] 社団法人, 全日本難聴者・中途失聴者団体連合会, “豊かなコミュニケーションに向けて—読話のためのビデオテキスト—”
- [9] 藤井智子, 黒川隆夫 (1995)  
単語発話時の口形変化の分析と口話画像合成のための規則化, 11th Symposium on Human Interface, pp.191-196
- [10] Paul Viola and Michael J. Jones (2001)  
Rapid Object Detection using a Boosted Cascade of Simple Features, IEEE CVPR
- [11] Rainer Lienhart and Jochen Maydt(2002), An Extended Set of Haar-like Features for Rapid Object Detection, IEEE ICIP 2002, Vol.1, pp.900-903, Sep. 2002.