幼児の言語獲得に寄与するディジタル絵本の試作

川合 康央† 池辺 正典† 佐久間 拓也†

概要

本研究は，幼児の言語獲得に着目し，音声認識による幼児のためのディジタル絵本を提案するものである．このシステムは自然言語処理に基づいて動作し，音声入力された単語によって絵本のキャラクターを作成することができる．また，幼児の音声入力データの分析から本システム独自の辞書を作成する．さらに，幼児にとって使いやすいインタフェースと幼児が好む色や形による親しみやすいキャラクターを設計した．本システムは実際の幼児による操作によって評価され，いくつかの問題点と改善手法を明らかにした．

Development of digital picture books for language acquisition in infant

Yasuo Kawai† Masanori Ikebe† Takuya Sakuma†

Abstract

This study focused on language acquisition in infant, we propose a digital picture book with speech recognition for infant. This system operates based on the natural language processing, and infant can create some characters on the picture book based on words input by voice. In this system, we create a dictionary of infant voice from analysis of speech input data by infants. Moreover, we designed an easy to use interface for infants and adorable characters used infant's favorite colors and shapes. This system was evaluated by the actual operation by infants, and these results showed some problems and improvement methods.

1. 研究目的

1.1 はじめに

人間にとって言語の習得は，人間社会においてコミュニケーションを行う上で重要なものである．言語を獲得することによって，様々な概念を音声や文字で表現し，他者へ情報を伝達することが可能となる．世界には様々な言語が存在するが，どのような言語であっても，人間は言語によるコミュニケーション手法を，幼児期における環境からの経験によって，習得することができる．本稿では幼児期の言語獲得に着目し，その初期段階における学習支援システムの試作について報告する．

1.2 研究の目的

本研究は，音声認識によって画像情報のインタラクションが発生するディジタル絵本の開発である．このシステムは，入力インタフェースであるマイクを通じて，音声入力された発話データを，音声認識によって分析し，本システムを動作させるために独自に作成された単語辞書を参照して，画像を描画するものである．

本研究では，幼児の言語獲得期における発話音声とその意味する概念に対して，インタラクションが発生するディジタル絵本というメディアを用いて，その理解を深めるための手助けとなるシステムの試作を目標とする．

1.3 研究の背景

幼児の言語獲得に関しては，先行して主に英語圏での研究[1]が進んでいたが，日本語においても言語学や心理学，情報学などからの言語発達研究[2][3][4]が盛んに行われている．音声認識の分野においても，幼児の日本語音声の認識に関する研究 [5][6][7]が進んでいる．

幼児を対象としたデジタルコンテンツの研究[8][9][10]は，さまざまなシステムの開発とその評価についての研究が行われてきた．ま

†文教大学
Bunkyo University

た，そのインタフェースについても，多くの研究成果[11][12]の蓄積がある．一方，デジタルコンテンツなどのメディアが幼児とその周囲の社会に与える影響についても，様々な研究が行われている[13][14][15][16].

本研究では，これら先行事例を踏まえた上で，幼児の音声認識によるディジタル絵本システムを開発する．

2. 音声認識によるディジタル絵本

2.1 幼児の言語獲得

言語の獲得は，人間の幼児期における重要な発達の一つである．環境における様々な概念を認知し，その概念を言語と結びつけて理解する．

発話は，0 歳児の意味のない声である喃語から始まる．やがて，周囲の言葉から単語を切り出し，意味と結びつけを行うことで単語を獲得し，「マー」「マッ」など，意味のある単語としての発話が開始される．その後，1 歳から 2 歳にかけて多くの単語を覚え始め，急激に語彙の数が増加し，3 歳頃までにおよそ 1000 語の単語を獲得する．ここでは，「コレナァ（ニ）？」などと周囲の人間に質問することによって，環境に存在する種々の概念とそれを示す単語の結びつきを確認する作業が繰り返されることで，語彙の蓄積と分類が進んでいく．

幼児は，1 歳半から 2 歳頃に二語発話を獲得し，「オオキイバス」「シロイネコ」など，単語を修飾する発話も見られる．3 歳児程度になると，3~5 語分の連続した関連のある単語の集合による多語文が見られる．そこでは，他者との会話が自然と成立する複雑な文章を組み立てられるようになり，社会における活動範囲が拡張する．

これらの音声による言語獲得は，周囲の人間が話す言葉や，映像，音楽などのメディアを通じて，段階的に獲得される．

また，音声による言語獲得とともに重要なものとして，文字や記号による言語獲得がある．1~2 歳児頃から，絵本などを通じて，写真やイラストなどの記号化された図像と概念の結びつきの理解が始まり，3 歳児前後から，音声言語を表す記号である文字を徐々に理解し始めていく．そこでは，絵本などのメディアを通じて，概念の図像や文字による抽象化と，その抽象化された記号が表す概念との結びつけが行われる．

2.2 絵本というメディア

絵本は，写真やイラストレーションによる色と形で構成された「図像」とひらがなやカタカナ，漢字，或いはアルファベット，数字など記号化された「文字」の，大きく分けて二つの要素で構成されている．幼児はこの 2 次元平面上に印刷された図像と文字から，対象となる物体とその性質や状態，行為，物語，心情などの概念を，段階的に理解する．

絵本の読み聞かせによって，幼児はまず音声言語と図像の関連に興味を示す．対象の写真や抽象化された線と面による 2 次元のイラストレーションが，身の回りにある現実の 3 次元空間上に存在する物体と関連付いており，それを指し示していることを理解する．また，日常的に接することのない非日常的な対象，たとえば珍しい動物などに対しても，図像を通じてそれが何を示している概念であるかを理解している．

次に図像とともに記されている文字に対して，それらが読み聞かせている発話者の音声を記号化していることに気付く．ひらがなやカタカナ，漢字，アルファベット，数字などが，音声による言語を表す記号であることを理解する．

幼児は文字を理解することによって，物語のある絵本の読解が可能となる．物語性のある絵本を読むことによって，日常生活による経験とは異なる非日常的な世界観からの物語経験を得る．

幼児にとって絵本の理解は，それぞれの形や色から始まり，その形や色によって抽象的に表現される記号の意味，さらにその意味のある記号の集合による場面構成，連続継起的な場面の展開による時間軸のある物語の理解，その物語に登場するキャラクターの心情表現の理解へと発展していく．

本研究では，幼児の言語獲得初期段階に着目し，図像と言語の結びつきを，話し言葉による音声での理解に資するディジタル絵本を作成した．

図 1 絵本を読む幼児

2.3 対象とするユーザ

本システムの対象となるユーザを，小学校就学前の幼児，特に言語獲得の初期段階である1~3歳児として，システムの設計を行った．語彙の蓄積段階の幼児に対して，対象となる概念の名詞，形容詞などと，その発話について理解を支援するシステムとする．

2.4 システムの概要

本システムは，ユーザが発話した言葉を認識し，それに対応したキャラクターオブジェクトをシーン上に配置するものである．「ゾウ」と発話すれば「象」のキャラクター画像が画面上に表示される．また「オオキイゾウ」と発話されれば，画像サイズが一回り大きい象のキャラクターオブジェクトを描画する．

また，本システムで扱う名詞の種類として，対象となるシステム利用者が特に関心を示すと考えられる概念である「生き物（陸，海）」「乗り物」「食べ物」を，キャラクターオブジェクトとして登録した．

さらに，キャラクターオブジェクトの性質を表す修飾語を登録した．本システムでは，拡大縮小に関する「大きい」，「小さい」とその類義語，移動速度に関する「速い」，「遅い」とその類義語を扱うこととする．

表1 キャラクターオブジェクト

陸の生き物	海の生き物	乗り物	食べ物
アヒル	イカ	機関車	かぼちゃ
ウサギ	イルカ	救急車	きのこ
馬	エビ	自動車	さつまいも
カバ	貝	消防車	じゃがいも
キリン	カニ	新幹線	大根
熊	カメ	電車	とうもろこし
象	クジラ	トラック	トマト
パンダ	クラゲ	バス	にんじん
羊	魚	パトカー	バナナ
豚	サメ	飛行機	ぶどう
ヘビ	タコ	船	みかん
ワニ	ヒトデ	郵便車	りんご

表2 修飾語表現

	1.5倍	0.5倍
拡大縮小	「大きい」と その類義語	「小さい」と その類義語
移動速度	「速い」と その類義語	「遅い」と その類義語

3. システム

3.1 システムの基本構成

本システムは，Webブラウザ上で動作するWebアプリケーションであり，対応するブラウザはGoogle Chromeである．また，本システムは，複数のクライアント間で認識されたテキストデータの共有を行うために，サーバーを介した送受信機能を備えている．

クライアントの機能としては，音声認識機能，映像合成機能，単語分割機能，描画機能から構成される．

まず，ユーザの音声認識により取得したデータによって背景となるシーン画像の指定を行う．背景として設定可能な画像として，3種類のシーン画像を用意したが，この背景画像はJavaScriptで作成された辞書から追加が可能である．また，背景としてカメラが選択された場合には，カメラのストリーミングを取得し，videoタグとの関連付けを行うことで，カメラからのストリーミング映像を背景データとして取り扱うこととした．

シーン選択後，背景に合成する各種キャラクターのオブジェクト画像の指定を音声認識によって行う．音声認識はWebブラウザGoogle Chromeで音声入力APIとして提供されているモジュールを利用した．音声認識の後にテキスト化されたデータは，辞書との比較の際に事前に分かち書き処理が行われる．本システムにおける，分かち書き表現への分割は，JavaScriptによる単語分割が可能なライブラリであるTinySegmenterを採用した．また，入力データとの比較を行う辞書としては，名詞辞書と修飾語辞書の2種類を作成した．単語分割の後に各辞書とのマッチングにより，対象となる画像データの特定とオブジェクト動作時のパラメータの設定を行う．今回，システムのプロトタイプで登録したオブジェクト数は48個であり，これに対応する辞書としては，類義語や判定精度向上のための誤認識率の高い類似表現を併せて登録することで103件の表現からオブジェクトを取得させることとした．さらに，修飾語表現として，オブジェクトの拡大縮小および移動速度に関する修飾語とその類義語を22件登録した．画面上にオブジェクトの描画を行う際に，入力時に修飾表現があわせて取得された場合，対応するパラメータを描画機能に渡す構成とした．

入力音声が辞書と一致した場合には，実行中の他のクライアントとデータ共有を行うために，サーバーに対してWebsocketを用いた通信を行い，現在表示中の全てのクライアントに一致単語を一斉送信する．このデータ通

信によって，単語を認識したクライアント以外にも同期して表示させることを実現した．一致した単語もしくは他のクライアントから受信した単語は，描画機能により，表示中の背景もしくはストリーミング映像とあわせて，画面上に描画される．

画面への描画は，JavaScript から canvas 要素に対して，対象となる画像データを描画し，JavaScript でアニメーション制御を行うことでオブジェクトの移動等の描画を行う．また，この画面への描画は 30fps でのアニメーションで表示することとした．

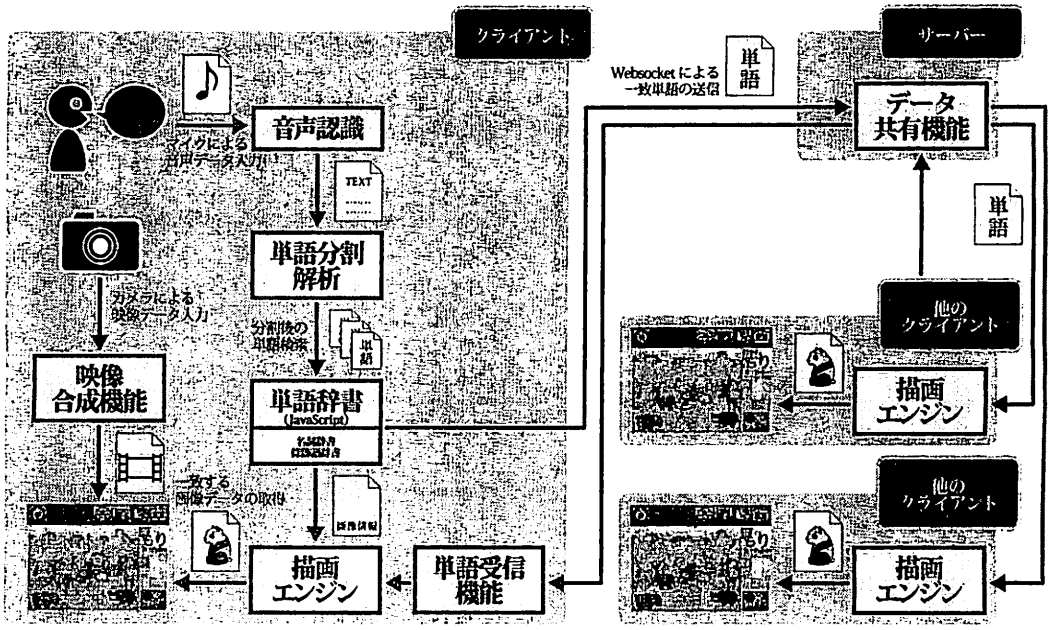

図 2 システム基本構成

3.2 カメラ機能による AR

本システムにおける AR 機能では，PC のカメラデバイスを用いることで，動画像の取り込みを行なっている．データの取得には，W3C で規格が策定中である WebRTC の Web API を用いて Video タグとの関連付けを行なった．また，取得したストリーミングデータは，30fps で Canvas タグに描画を行うことで，画面に動画を表示している．

表示されている動画像から，顔の位置を検出することで，音声認識による描画オブジェクトの表示を行なった．顔の検出には，face.js および ccv.js という JavaScript による顔検出および画像処理用ライブラリを利用した．face.js は，対象画像から領域の切り出しを行い，その明度差のパターンから顔部分を認識するという方式を取っている．また，ccv.js は画像処理用の汎用ライブラリである OpenCV を JavaScript に移植したものである．

今回の実験システムにおけるシーン表示画像の解像度は，800pixel×600pixel である．このサイズの動画像をそのまま認識対象とした場合，一度の認識処理において 2 秒程度の処理時間を要した．そこで，非表示領域にて，200pixel×150pixel の同一画像を生成し，この画像に対して認識処理を行なったものから，画像比率によって元画像の顔位置検出を行い，その座標に描画オブジェクトを表示することで，追跡処理を行なっている．顔位置の追跡処理は 0.5 秒毎に上記手順で行なっている．処理は，非同期で行われているために，若干の遅延が生じるが，動画像に大きな影響を与えることはなく，映像をストリーミングとして描画することを可能としている．

カメラの認識精度としては，認識可能な最短距離が 20cm 程度であり，最も認識の良い距離は 60cm 程度であると考えられる．この場合においては，秒間で約 2 回の顔認識の結果，対象者をほぼ完全に追跡可能であると考えられる．また，顔認識は原則正面画像を対象として行うが，本システムでは，正面画像から左右 45 度程度が認識限界であった．

3.3 環境構成

本システムは，ネットワークに接続されたPC と，音声入力インタフェースとしてのマイク，画像出力インタフェースであるモニター或いはプロジェクターによって構成される．また，コンテンツの背景となるシーン画像として実世界を取り入れることが可能であり，そのため，画像入力インタフェースとしてwebカメラも用いることとする．

ネットワークを介して，入力された単語データを共有することによって，複数のユーザが同一のシステムに描画することが可能である．プロジェクターを使ったお話会や，遠隔地にいるユーザ間でのコラボレーションなどの利用が考えられる．

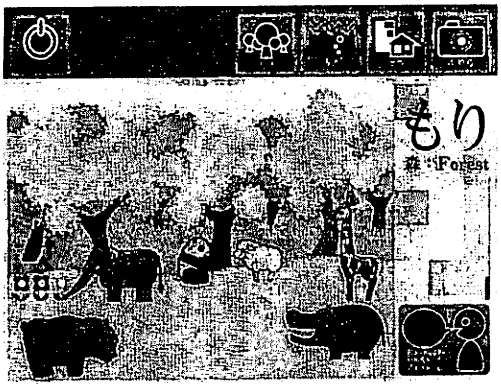

図3 森のシーン画像を表示したインタフェース

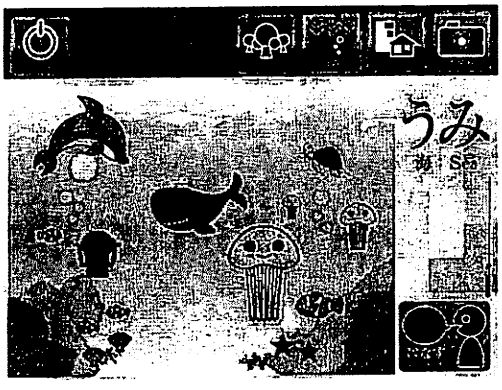

図4 海のシーン画像を表示したインタフェース

4. インタフェース

4.1 シーンとキャラクター

本システムでは，4種類のシーン（森，海，町，カメラ）と48種類のキャラクターを登録した．本システムのユーザ層を考慮した際，検索による画像抽出は不適切なものも多いため，今回はあらかじめ用意した画像データを用いることとする．

キャラクターオブジェクトは，画像解像度縦横200pixelのアルファチャンネルを持ったpng ファイルを用いた．本システムでは，通常時この画像ファイルをそのままの画像解像度である200pixel四方で表示する．また，修飾語「大きい」とその類義語が文中に含まれる場合，画像のスケールを 1.5 倍にした300pixel四方で表示させ，修飾語「小さい」とその類義語が文中に含まれる場合には，スケールを0.5倍にした100pixel四方で表示するものとした．さらに，修飾語「速い」とその類義語が文中に含まれる場合，キャラクターオブジェクトのアニメーションによる移動量を 1.5 倍とし，修飾語「遅い」とその類義語が含まれる場合は移動量を0.5倍とした．

シーン画像は，あらかじめ画像として登録してある「森」「海」「町」の3つの場面とともに，Webカメラにより撮影された実空間の映像上にキャラクターオブジェクトを生成する「カメラ」というシーンも用意した．システム全体の画像解像度は 1024pixel × 768pixel であり，その中にシーン画像が800pixel×600pixel で構成されている．このシーン画像の範囲内にキャラクター画像を配置し，アニメーションを描画する．

絵本の図像が幼児の対象イメージに与える影響を考慮し，キャラクター造形は単純な色と形で構成したものを用い，レジビリティの高い記号として扱うこととした．背景となるシーン画像は，塗りのみの面で構成されたものを使用し，他方，キャラクター画像には輪郭線のある画像を用いることで，キャラクターオブジェクトが前面に浮き出て知覚されるよう構成した．

また，キャラクター生成時の効果音を付し，音声認識の障害にならないよう配慮した上で，インタラクションに聴覚情報による効果を持たせることとした．

4.2 インタラクション

ユーザは，画面上にある「はなす」ボタンを押下して，話し言葉で音声入力を行う．入力された文節は形態素解析され，画面上にどのように認識されたのかを表示する．

音声入力によって生成されたキャラクターオブジェクトは，X 軸，Y 軸に対して，それぞれ一定速度で移動するアニメーションが描画される．このキャラクターオブジェクトは，シーン範囲の境界に接する値まで移動すると，移動量が-1 倍され，移動方向が反転する．生成されたキャラクターオブジェクトは反転を 10 回繰り返すと，そのままシーン範囲外へと移動し，描画を終了させる．

また，キャラクターオブジェクトの生成とともに，シーンの切り替えも音声で操作できるものとした．現在表示されているシーンは，システム画面上部にピクトグラムで表示する．

5. ユーザビリティ

5.1 ユーザビリティテストの概要

本システムのプロトタイプを幼児に対して提示し，ユーザビリティテストを実施した．対象となった被験者は 3 歳児の男子 2 名である．言語理解レベルとして，ひらがなと一部のカタカナ，漢字の読解が可能であり，発話内容が単語から文章へと移行した後の段階である．

テストは，まず実験者が先行してシステムを動作させ，実際に発話して画像を発生させることで，システムの使用方法について見本を提示した．その後，被験者には実験者が適宜アドバイスしながら，実験者と共に自由に操作してもらい，その際の行動と発話の内容を映像と音声データで収集し分析を行った．

表 3 ユーザビリティテストの概要

ID	回	年齢	日時	時	環境	助言
A	1	3歳 5か月	2012 年 5月 28 日	15 分	ノート型 PC	実験者と共に操作
A	2	3 歳 5か月	2012 年 6月 2 日	10 分	タブレット型 PC	実験者の助言で操作
A	3	3歳 7か月	2012 年 7月 27 日	15 分	ノート型 PC	被験者のみで操作
B	1	3 歳 3か月	2012 年 7 月 27 日	10 分	ノート型 PC	実験者が操作

5.2 ユーザビリティテストの結果と考察

被験者は，事前に提示した実験者の行動によって，本システムのルールを類推し，音声による発話と表示される図像と文字が関連していることを容易に理解し，自律的に動作させることが可能であった．

実験の結果，発話における幼児特有のアクセントによって，音声認識の精度が成人に比べ低いことが明らかとなった．また「トウモロコシ」を「トッコロモシ」と発話するなどの，幼児独特の言い回しも多く見られた．そこで，幼児の発話に対して音声認識された語のうち頻出するものを，キャラクターオブジェクトに対応させ辞書に登録して行くことで，認識率の向上を図った．一方で，被験者は先行して提示されたシステムのインタラクションを理解していることから，任意のキャラクターオブジェクトを発生させるため，様々なアクセントでの音声入力を何度も試み，発話を修正しようとする行為も見られた．このことから，本システムは言語獲得にある程度資する教材に成り得ると考えられる．

一方で，未登録の語を入力しようとする行動も多く見られた．本システムは，キャラクターオブジェクトを適宜追加登録することが可能である．被験者の興味に応じていくつかのキャラクターオブジェクトの追加登録を行った結果，被験者が興味を持つことが観察された．キャラクターの追加登録機能を利用することで，ユーザやその家族が作成した画像を用い，システムをより身近なものとして扱うことが可能である．

また，本システムでは「大きい｜象」や「速い｜車」など，いくつかの修飾語に応じて，異なるインタラクションを準備していたが，「オイシイ｜ニンジン」や「クラゲ｜ワラッ｜テ｜イル｜ヨ」など，さまざまな修飾語が発話要素として見られた．未登録の修飾語に対してインタラクションは発生しないが，そのことによって幼児の創造力を規制しないよう，何らかのインタラクションを付加するなどの改善を行う必要がある．また，登録されている修飾語である大きさと速度の変化についても，インタラクションを確認した後に，「モット｜オオキイ｜バス」などといった発話要素も見られた．

図 5 プロトタイプによるシステムのユーザビリティテスト

ノート型 PC で操作を行った際，キーボードなどの本システムが使用しないインタフェースが，操作の障害となっていることが確認された．インタフェースの評価実験として，リモートデスクトップ接続によって，タブレット型デバイス上に画面表示を行った結果，操作性の向上が確認された．本システムでは，インタラクションに不要なインタフェースは，ユーザに極力提示しないような形での利用が適していると考えられる．

図 6 タブレット型デバイスによるインタフェースのユーザビリティテスト

6. まとめ

6.1 今後の課題と応用可能性

本研究で開発した音声認識によるディジタル絵本を，実際の幼児によって操作してもらった結果，本システムに興味を持たせることが可能であることを確認した．また，結果として，いくつかの課題と応用可能性が明らかとなった．

音声認識において，幼児の発話データは正しく認識されない事例が多く見られたため，幼児による発話データをより多く収集し，専用の単語辞書を作成する必要がある．

また，インタフェースデザインでは，あらかじめ登録する画像データを，早期表出語彙を中心に拡充していく必要があるとともに，ユーザによる画像登録やカスタマイズが容易に可能となるよう，より自由度の高いキャラクター生成システムを用意する必要があると考えられる．ユーザやその周囲の人間が描いた画像をインタラクション可能なものとすることで，システムの使い方に広がりを持たせるとともに，本システムをより身近な親しみのあるものとして利用されることが可能であると考えられる．

さらに，本システムによって幼児が作成した物語を，ネットワークを通じて記録・蓄積することで，単語辞書の改良に資するデータを収集することが可能である．

本システムでは日本語辞書を用いたが，日本語以外のさまざまな言語辞書を用いることで，語学学習教材としての応用が考えられる．辞書の多言語化と精度向上によって，正確な発音を学習できるシステムへと発展させることも，今後の重要な課題である．

参考文献

[1] Steven Pinker:The Language Instinct: How the Mind Creates Language, New York: Harper Perennial Modern Classics, 1994.

[2] 小椋たみ子: 日本の子どもの初期の語彙発達, 言語研究(132), pp.29-53, 2007.

[3] 小林春美, 佐々木正人(編): 新・子どもたちの言語獲得, 大修館書店, 2008.

[4] 麦谷綾子: 乳幼児の音声言語獲得, 電子情報通信学会技術研究報告. TL, 思考と言語 104(316), pp.13-18, 2004.

[5] Janet F. Werker, Ferran Pons, Christiane Dietrich, Sachiyo Kajikawa, Laurel Fais, Shigeaki Amano: Infant-directed speech supports phonetic category learning in English and Japanese, Cognition, Volume 103, Issue 1, pp.147-162, 2007.

[6] 桐山伸也, 竹林洋一, 堀内裕晃, 石川翔吾, 笠見朋彦, 北澤茂良: マルチモーダル幼児行動コーパスを用いた発話分析, 情報処理学会研究報告. SLP, 音声言語情報処理 2007(75), pp.25-30, 2007.

[7] 小川厚徳, 山口義和, 松永昭一: 小学生音声データベースを用いた子供音声認識の検討, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J87-D-II(8), pp.1572-1580, 2004.

[8] 大即洋子, 坂東宏和, 大島浩太, 小野和: 絵本を題材とした活動的な保育を支援する PC 利用の一例, 情報処理学会研究報告. コンピュータと教育研究会報告 2010-CE-103(17), pp.1-8, 2010.

[9] 角薫, 長田瑞恵, 田中克己: アニメーションメディア変換システム Interactive e-Hon における親子エージェント情報提示モデル, 知能と情報: 日本知能情報ファジィ学会誌 : journal of Japan Society for Fuzzy Theory and Intelligent Informatics 18(2),

pp.240-250, 2006.
[10] 朱文昌, 小宮山美緒, 古井陽之助, 速水治夫: 小学生向けデジタル絵本教材システムを用いた学習効果の検証, 情報処理学会研究報告. GN, [グループウェアとネットワークサービス] 2007(32), pp.103-108, 2007.
[11] Donald Arthur Norman: Turn signals are the facial expressions of automobiles, Basic Books, 1993.
[12] 米澤朋子, Brian Clarkson, 安村通晃, 間瀬健二: ぬいぐるみインタフェースによる音楽コミュニケーション, 情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告 2001(3), pp.17-24, 2001.
[13] 森田健宏: 保育所におけるパソコン利用に対する保育士の抱く問題点の検討, 日本教育工学雑誌 26(2), pp.87-94, 2002.
[14] 堀田博史, 金城洋子, 新田恵子: コンピュータ遊びに対する保護者の考え方, 日本保育学会大会研究論文集 (53), pp.692-693, 2000.
[15] Dimitri A Christakis: The effects of infant media usage: what do we know and what should we learn? , Acta Pediatrica,vol.98, pp.8-16, 2009.
[16] 沢井佳子, 藤永保, 竹林圭子: テレビ幼児教育番組に対する 2 歳児の視聴反応, 日本教育心理学会総会発表論文集 (29), pp.394-395, 1987.