

合成音声によるインタラクティブな教示音声 CAI システムへ導入するための検討

篠崎 翼 阿部 匡伸

NTT ヒューマンインタフェース研究所
〒239 神奈川県横須賀市光の丘 1-1

電子メール: tsubasa@nttspch.hil.ntt.co.jp, ave@nttspch.hil.ntt.co.jp

あらまし

我々は規則合成音声インタフェースを利用して、システムを使う気になるとかシステムに親しみを持てるといったユーザフレンドリなシステムを実現したいと考えている。本稿ではユーザフレンドリ化を検討するための具体的なタスクとして、合成音声によって指示を行う CAI ペン習字システムを設定し、インタラクティブな教示音声を CAI システムへ導入することを検討した。その結果、誤りの指示では 200 ms ないし 500 ms の遅れで指示音声を出すことが好まれること、ユーザの動作をインタラプトするための間投詞が付いた指示音声好まれること等が明らかになった。更に、誤り指示の繰り返し回数に応じた指示内容のバリエーションの付け方についても検討を行った。

Experiments to introduce interactive instructions by synthesized speech for CAI system.

Tsubasa SHINOZAKI, Masanobu ABE

NTT Human Interface Laboratories.
1-1 Hikarinooka Yokosuka-Shi Kanagawa 239

E-mail: tsubasa@nttspch.hil.ntt.co.jp, ave@nttspch.hil.ntt.co.jp

Abstract

In this paper, we examined methods to use synthesized speech as interactive instructions for Computer Assisted Instruction system that teaches how to write calligraphy by hand. The CAI system has a display-pen tablet interface. In the experiments, a user draws a character using pen tablet then the system instructs by synthesized speech. The experiment results show users prefer 200 ms or 500 ms delayed voice responses for the error indications, and show a number of users prefer an interruption message such as "et," or "this," before the system outputs instructions. In addition, it is necessary to change the variation of instructions according to the number of error occurrence frequency.

1. はじめに

システムとユーザとのインタフェースとして合成音声による出力を考える場合、大きく分けて二つの捉え方がある。一つは、電話を通しての電話番号案内や電子メールの読み上げ、情報検索の結果取得といったものであり、他の出力手段にはない特徴を生かした情報提供手段としての合成音声である。もう一つは、テレビゲームや CAI (Computer Assisted Instruction) システムの出力手段として用い、システムに対して親しみやすさや使いたいという気持ち、楽しさや面白さといったものを与えるユーザフレンドリなインタフェースとしての捉え方である[1][2][3][4]。我々は、後者の立場に立って、ユーザフレンドリなシステムを構築するうえで合成音声を用いることを検討している[5][6][7]。

これまで、合成音声によるシステムのリアクション音声について、合成音声の韻律特徴とユーザが受ける印象との関係を検討し、合成音声に意図した印象を与えることが可能であることを示した [6]。一方、合成音声によるマンマシンインタフェースのユーザフレンドリ化の検討を進める上では、具体的なタスクを考える必要がある。本報告では CAI ペン習字システムを具体的なタスクとしてとりあげ、合成音声によるインタラクティブな教示音声を導入するための検討を行う。

2 節では、CAI ペン習字システムについて説明する。3 節では、ペン習字における指示発話の収録について述べる。4 節では、3 節で得られた知見をもとにインタラクティブな教示音声をペン習字システムへ導入するための検討について述べる。

2. CAI ペン習字システム

音声によってインタラクティブに指示を行う CAI ペン習字システム「CALTURE (CALligraphy Training system Using voice REsponse)」を図 1 に示す。このシステムは液晶画面付きペンタブレットを用いて、画面に表示されたお手本をペンでなぞることによりペン習字の練習を行うものである。ユーザが表示されたお手本をなぞると、書きはじめの位置やストローク(お手本をなぞるペンの軌跡)のずれについてイン

タラクティブに音声で指示を行う。また、正しくなぞれている場合には褒めるといったことを行う。

以下にタスクとして CAI ペン習字システムを選択した理由を挙げる。

- なぞるといふ単調な動作の繰り返しにおいて、すぐにあきつぱくになりがちなものを音声によるリアクションを返すことによりゲーム感覚で楽しく練習を行えると期待できる。
- 字をなぞるペンの動きに対してリアクションを返すことにより、ユーザとシステムの間にインタラクションが生じる点で人同士のコミュニケーションに似ている。
- 手と目を使う作業をしているユーザにとって、音声による誤りのフィードバックは注意を引き、ユーザにすぐに反応を起こさせるのに適していると考えられる。
- 「ペン習字によって字がうまくなる」というように、ある目的に向かってユーザとシステムが協調して作業を行うといった、世の中によくみることのできる協調作業である。
- お手本をなぞるといったように文脈があらかじめわかっているタスクであり、履歴情報が有効に利用できる。
- 入力にペンタブレットを使っているのでユーザの動作が確実にわかりインタラクションの制御がしやすい。
- 音声出力に加え画面表示もあり、音声による位置や動きに対する指示不足をマルチモダリティで補うことができる。

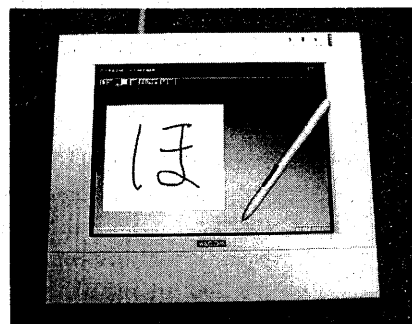


図 1. CAI ペン習字システム「CALTURE」
液晶画面付きペンタブレットを用い、音声指示を用いてインタラクティブにペン習字を行うシステム。

3. お手本と指示発話の収録

今回の検討では、ペン習字の題材として「ひらがな」を選んだ。そこでまず、ペン習字の指示発話における基本的な知見を得るために、「ひ」「お」「ほ」の3つの「ひらがな」についてお手本の収集と指示発話の収録を行った。指示発話の収録では、ペン習字の方法を「お手本をなぞる練習」に限定し、なぞり書きにおいてインタラクティブにお手本とのずれを指摘、書き直させることを想定した。

3.1. お手本の収集

お手本の収集は、図1のペンタブレットを用い、図2に示すように補助線によって16マスに区切られた正方形の文字枠の中に、一文字ずつお手本の字を先生役の被験者に書かせ時系列データとして収集した。また別途、お手本の文字を書くうえで注意すべき点についても文字を書きながら口頭で説明してもらい、その様子をVTRで収録した。先生役の被験者はペン習字経験のある成人女性一名である。

3.2. 添削時の書き直し指示の収録

ペン習字の経験のない成人男女10名にペンタブレットを用いて「ひ」「お」「ほ」の文字を書かせた。そして、紙に出力したそれらの文字を先生役の被験者に赤ペンで直させ、赤ペンによる指示の様子及び指示発話をVTRで収録した。このとき、先生役の被験者には文字を書いた本人を目の前にして「ここはこう書くんだよ」といった感じで説明をしながら文字を添削していることを想定するように指示した。

3.3. 指示発話に関して得られた知見

添削時の書き直し指示の収録を通して得られた知見のうち、誤り指示に関する知見を以下に示す。

- 指示は「書きはじめ」の位置や「跳ね」、「止め」といった各ストローク(一筆)のはじめと終わり及び、「曲げ」や「結び」といった特徴的な点について主に行われる。
- 指示の内容は「跳ね」や「止め」、特徴的な点について各ストロークごと細かく指示するものと、バランスやつながりといった全体に関する指示とに大きく分けられる。

- 指示の内容は「右」や「左」といったように共通して用いられる指示と、「ここ(「ほ」の第一筆の書きはじめの位置)よりちょっと下から書き始めて」といったように他の部位との相対的な関係を使った指示とがある。
- 「ここは」や「こういうふう」のようにペンによる指示と指示発話を同期させて用いることがよく見られる。

3.4. リアルタイムな書き直し指示の収録

文字の「書きはじめ」や「跳ね」、「止め」、ストロークの「曲がり」といった各部分について、間違ったその場で指示を行う場合の発話の収録を行った。収録のための実験システムを図3に示す。実験は、先生役側にディスプレイを置き、お手本を表示しておく。操作側では先生の指示の様子をVTRカメラによってモニタできる。操作者がペンタブレットの液晶画面に表示されたお手本をペンでなぞると、先生側のディスプレイになぞり書きの様子が表示される。先生には予め、操作側ではお手本が見えない状態で文字を書いていると伝えておく。そして、操作者は、3.2

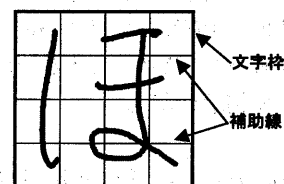


図2.ペン習字のための文字枠

正方形の文字枠の中を縦横それぞれ3本の補助線で均等に区切る。

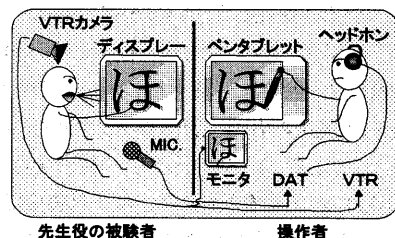


図3.指示発話収録のための実験システム

被験者はディスプレイに映った操作者がお手本をなぞる様子を見て指示を行う。その様子をDAT及びVTRで記録する。操作者は被験者の様子をモニターとヘッドホンを通して知ることができる。

節での収録で見られたお手本をなぞるときの誤りのパターンと同じ書き誤りをわざと行う。先生にはディスプレイに映るなぞり書きの様子を見て音声で指示させた。このとき、画面への指差しによる指示についても許した。この指示発話と指差しの様子をDAT及びVTRで収録した。

3.5. 得られた知見

3.3 節で述べた知見に加えて、リアルタイムな書き直しの収録を通して得られた誤り指示に関する知見を以下に示す。

- 「書きはじめ」の誤りは直ぐに指示する。
- 「書きはじめ」以外は、1 ストローク書き終わってから指示を出すことが多い。
- 指示を出す際、ユーザの動作にインタラプトをかけるために「あっ」や「えっ」、「そこっ」といった間投詞や指示語が付くことが多い。
- 誤りを指摘するとき、「そこ」とか「ここは」といった指示語に指差しを添えた指示がよくみられる。
- 誤りがない場合には「そうそう」や「そう」といった指示発話がなされる。
- 添削時の指示に比べると、相対的な位置関係を使つての指示より、大きさや向き、長さを使った指示が多い。

4. インタラクティブな教示音声導入の検討

3 節で得られた知見からペン習字には文字全体の大まかな形やバランスを練習するモードと各ストロークごとの「跳ね」や「曲げ」といった部分を細かく練習するモードがあることがわかる。後者の各ストロークごとの練習モードでは、「書きはじめ」や「止め」、「跳ね」といった指示を出す点でインタラクティブに指示が行える。そこで、1 ストロークずつ練習を行う場合の「書きはじめ」や「書き終わり」といった部分に、インタラクティブな教示音声を導入することを検討する。

4.1. 検討のためのペン習字システム

1 ストロークずつの練習は、以下のような方法で行

う。ペン習字システムは図 2 に示す文字枠の中に文字を書くものとする。はじめにお手本が表示され、これから 1 ストロークずつの練習を行うことを知らせる合成音声流れ、およそ 5 秒後にお手本は消える。ユーザは文字枠と補助線を参考に、お手本と同じ文字を書く練習を行う。このとき、合成音声によって誤りの指摘や指示といったリアクションを行い、必要なら書き直しをする。指示を効果的に行うために、指示音声と同期して画面上に練習中のストロークを表示する。

以上のような場合において、ユーザへの教示をインタラクティブに行うためには、指示音声の出力タイミング[8]や動作への割り込みといったものが重要である。また、誤り指示の内容については、指示が単調なものにならないようにするためのバリエーションの付け方について検討する必要がある。そこで、本節では、(1)指示音声の出力タイミング、(2)指示のバリエーションの付け方について検討を行った。

4.2. 指示音声の出力タイミングについて

ペン習字の各ストロークの「書きはじめ」と「書き終わり」で指示を出すことを想定し、このときユーザに対してインタラクティブに教示を行うための指示音声の出力タイミングについて検討を行った。

4.2.1. 実験方法

実験にはひらがな「ほ」の練習を用いた。お手本を表示した後、文字の練習を行うと、「書きはじめ」と「書き終わり」の位置のずれに対して合成音声によって指示を出す。指示音声は合成音声の韻律を自由に操作できる *Speed97* [9]を用いて、韻律パターンを適切に調整した合成音声を用いた。指示の出力タイミングは「書きはじめ」についてはペンをタブレットに下ろしてストロークを書き始めたとき、「書き終わり」についてはペンをタブレットから上げて書き終わったときを基準とした。このとき、指示の出力タイミングを表 1 のように 0 ms から 1000 ms の間で変化させたものをランダムに被験者に提示し、指示音声の出力タイミングの好ましさについて 5 段階で評価させた。また、評価と共に評価の理由について簡単に記述(アンケート)させた。指示音声は間投詞(表 2)に

「もっと右から」や「上に向かって跳ねて」といった指示を組み合わせたものを用いた。実験は成人男女5名の被験者で行った。

(実験の実施例、ひらがな「ほ」)

- 動作 : ペンを下ろす(第1ストロークの書き始め)
- 指示 : 「えっ、もっと右から」(書き直し)
- 動作 : ペンを下ろす(第1ストロークの書き始め)
- 指示 : 「そうそう」
- 動作 : ペンを上げる(第1ストロークの書き終わり)
- 指示 : 「もう少し大きく跳ねて」(書き直し)
- 動作 : ペンを下ろす(第1ストロークの書き始め)
- 指示 : 「そう」
- 動作 : ペンを上げる(第1ストロークの書き終わり)
- 指示 : 「うーん」(了解・感心した感じ)
- 動作 : ペンを下ろす(第2ストロークの書き始め)
- 指示 : 「もっと左から」(書き直し)
- 動作 : ペンを下ろす(第2ストロークの書き始め)
- ...
- 動作 : ペンを上げる(第4ストロークの書き終わり)
- 指示 : 「うーん、もう一度練習しましょう。」

4.2.2. 実験結果

図4に間投詞別の評価の平均と分散を示す。評価は指示音声の出力タイミングについて「好ましくない」を1点、「好ましい」を5点として行った。この結果から500ms以内に指示が出されることが好まれることがわかる。また、指示を出すうえで遅れの全くない0msに関しては、その評価の分散が大きいことから、その好みは被験者に大きく依存することが分かる。アンケート結果によれば、この原因は「早いほうが説得力がある」といったように指示が早く出されることを好む被験者(2名)と「早いとせかさされている感じがする」といったように指示が早く出されることを好まない被験者(3名)とに分かれたためと考えられる。さらに図4の200msの部分に注目すると、間投詞が付く場合には200msが最も平均が高く分散も小さいが、間投詞が無い場合には分散が大きく、平均が下がっている。このことは、3節の知見で述べたような「えっ」といった間投詞を付けることにより、指示を出す際に注意を引き、ユーザの動作をインタラプト

表1.指示音声の出力タイミング

指示音声出力までの遅延時間 (ms)			
0	200	500	1000

表2.指示の前に付けた間投詞

「えっ」	「えー」	(無し)
------	------	------

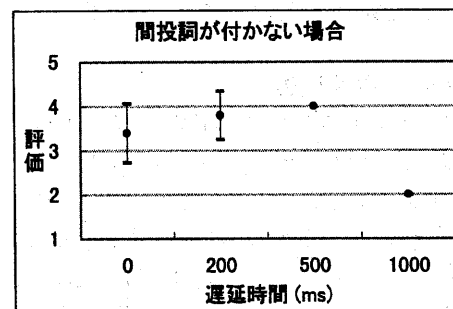
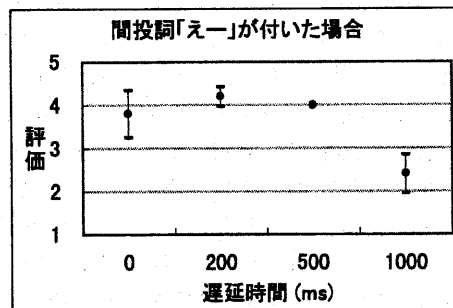
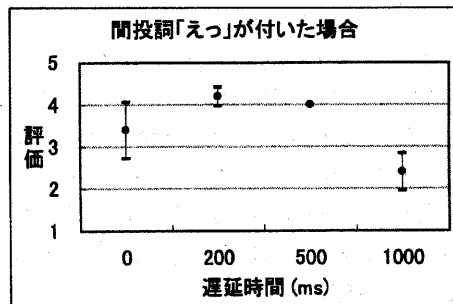


図4.指示音声の出力タイミングについての評価

間投詞別に「好ましくない」を1点、「好ましい」を5点としたときの評価の平均と分散を示す。間投詞「えっ」が付いた場合(上)、間投詞「えー」が付いた場合(中)、間投詞が付かない場合(下)。

する場合に比べて、間投詞が無い場合は指示のタイミングが早く感じられたなどで好みが分かれたためと考えられる。被験者のアンケートからも「指示を出す場合、間投詞があったほうが良い」という答えが8割を占めた。

指示音声の出力タイミングとして 200 ms ないし 500 ms 遅れることが好まれることは、文献[6]に見られるユーザの入力に対するリアクション音声の出力タイミングの実験結果とも合致する。

4.3. バリエーションの付け方について

誤り指示に用いる指示音声について検討を行う。誤りの指示を行う場合、指示が単調にならないようにするためにバリエーションを付ける必要がある。そこで、繰り返して指示を行う場合の、指示音声のバリエーションと冗長性との関係について実験を行った。

4.3.1. 実験方法

実験はひらがな「ほ」の1ストローク目の「書きはじめ」に対する指示を用いて行った。指示の前に修飾する言葉を付けた「長い指示」(表3)と指示語を言い換えた「短い指示」(表4)とを、表5のように組み合わせる4つの指示セットを用意し被験者に聞かせた。このとき、被験者には、各指示セットを被験者が納得いくまで何度でも聞かせた。「書きはじめ」の位置を6回繰り返して誤るように実験を仕組んでおき、各指示セットについて評価させた。評価は、各指示セットを好ましい順番に理由を添えて順位付けするとともに、それぞれの順位間の距離を測るために7段階評価させた。被験者は成人男女7名で行った。

4.3.2. 実験結果

図5に被験者別の7段階評価の結果を示す。この図を見ると「長い指示」を好むもの(被験者4~7)や、「長短」「短長」のようにバリエーションの付け方に変化があることを好むもの(被験者1)、また、逆に変化がないことを好むもの(被験者7)、とにかく「短い指示」を好むもの(被験者2)等の傾向が見られる。このように被験者の反応は様々であるが、「長い指示」から「短い指示」へと指示のバリエーションの付け方を変えることは、比較的多くの被験者に受け入れられていることが分かる。また、大まかな傾向を調べるために、ある指示セットについて被験者を大きく2

表3.「書きはじめ」の指示に用いた「長い指示」

えっ、もっと 右から	もう少し 右から	えー、もうちょっと 右から
えっ、もっと 左から	もう少し 左から	えー、もうちょっと 左から

表4.「書きはじめ」の指示に用いた「短い指示」

右	右右	右から
左	左左	左から

表5.提示した指示セット

セット	誤り	1~3回目	4~6回目
「長長」		長い指示	長い指示
「長短」		長い指示	短い指示
「短長」		短い指示	長い指示
「短短」		短い指示	短い指示

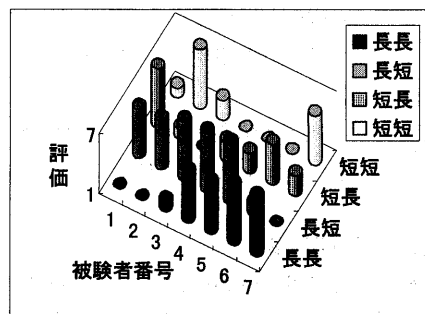


図5.指示音声の指示セットについての評価1

表5に示す指示セットに対しての好ましさについて、「好ましくない」を1点、「好ましい」を7点としたときの被験者別の評価を示す。

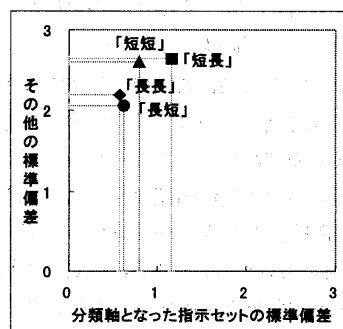


図6.被験者を2つのグループに分けた場合の分散

表5の各指示セットを軸として、好むグループと好まないグループに被験者を分けた場合の分散を表す。横軸は分類の基準となった指示セットの標準偏差の平均、縦軸は分類の基準以外の指示セットの標準偏差の平均である。

つのグループに分け、その指示セットに対する評価の分散と、その指示セット以外に対する評価値の分散を求めた。結果を図6に示す。「長短」「長長」の分散が小さいことから、「長短」「長長」を好むグループと好まないグループでは、比較的意見が一致する傾向が見られるといえる。

以下にアンケートの結果をもとにまとめる。「長い指示」の繰り返しを好まない人は「しつこい感じがする」という理由で、そして、好む人は「親切な感じ、丁寧な感じがする」という理由で判断している。「長短」の評価では、途中で「短い指示」に変えることによりしつこい感じが消えるため好まれている。また、「長い指示」の繰り返しを好む人では「短い指示」の割合が増えるほど評価が下がっている。これは、「短い指示」は不親切であるとか冷たい感じがすると判断しているためである。

4.4. 短い指示音声について

ここでは更に、4.3節の「短い指示」の繰り返しにおける指示のバリエーションの付け方について実験を行った。

4.4.1. 実験方法

実験は 4.3 節の実験と同様にひらがな「ほ」の「書きはじめ」について、「短い指示」に表 6 に示す方法でバリエーションを付けた指示音声を用いて行った。それぞれの指示音声を 6 回ランダムに提示することを 1 試行として、被験者には 4 試行の練習試行を含めた 12 試行を提示し、好ましさについて 5 段階で評価させた。被験者は 4.3 節の実験と同じ被験者で行った。

4.4.2. 実験結果

図 8 に好ましさの評価として「好ましくない」を 1 点、「好ましい」を 5 点としたときの平均と分散を 4.3 節の「短い指示」の繰り返しを好むグループと好まないグループ別に示す。図 8 より、「短い指示」では言葉の言い換えによるバリエーションの付け方がもっとも好まれるといえる。この結果から短い指示については、指示音声にバリエーションを付ける方法として言い換えを行うといった言葉を変化させることは有効であ

表 6. 「短い指示」のバリエーションの付け方「右」について

方法	指示音声のバリエーション		
言葉の言換え	「右」	「右右」	「右から」
F ₀ 平均	-40 Hz	±0 Hz	+40 Hz
韻律パターン	図 7 に示す		
(無し)	「右」		

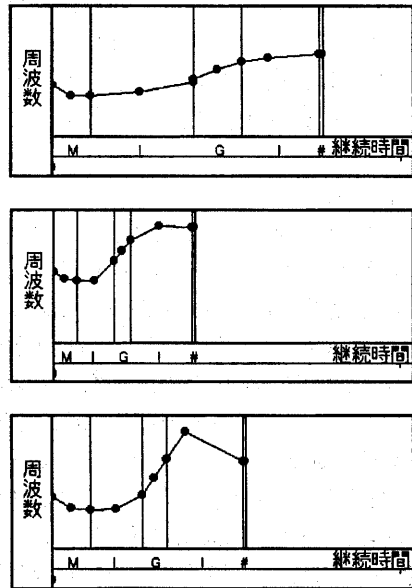


図 7. 指示音声の「右」の韻律パターン

指示音声のバリエーションを付ける方法として韻律パターンを用いた場合の指示音声「右」の韻律パターンのうち、F₀ パターンと継続時間を表す。短く区切って(上)、F₀ 高めの早口で(中)、強めの口調で(下)。

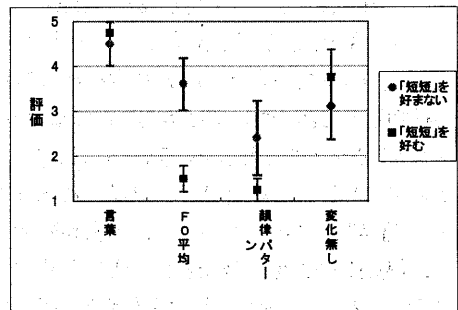


図 8. 指示音声のバリエーションの付け方についての評価
被験者を「短い指示」の繰り返しを好むグループと好まないグループに分け、バリエーションの付け方について「好まない」を 1 点、「好ましい」を 5 点としたときの評価の平均と分散を示す。

ると考えられる。また、韻律パターンを変えてバリエーションを付けた場合、両グループとも最も低く評価された。アンケートによれば、これらの音声は「怒った感じ」や「突き放した感じ」を与え、「こわい」というイメージを持たれたため好ましくないと評価されていた。これは、バリエーションを付けるとき、 F_0 を上げたりパワーを上げるといった操作を行ったためであると考えられる。この結果は、短い指示について韻律によってバリエーションを付ける際には、単に変化をつけることに注意をするのではなく、どのような印象を与えるかも考慮する必要があることを示唆している。

5. まとめ

本報告では、CAI ペン習字システムへインタラクティブな教示音声を導入する検討を行った。ペン習字における誤り指示について収録を行い、得られた知見をもとに、CAI ペン習字システムにおける誤り指示の方法として、指示の出力タイミング、指示のバリエーションについて実験を行った。その結果、指示を出す場合、200 ms ないし 500 ms の遅れをともなう指示が好まれることが明らかとなった。また、指示の前に「えっ」といったような間投詞が付くことはユーザの注意を引くこととなり、指示の前に間投詞が付くことが好まれることを明らかとなった。また、間投詞や程度を表す言葉、履歴を利用した言葉で指示を修飾しバリエーションを付ける場合、誤りの繰り返しの回数に応じて「右」や「右右」といったような短い指示に指示の方法を変えることが必要であることを示した。

今回、音声による指示出力中のユーザからの割り込みやそれにもなって出力音声へ次の指示音声が入り込むことについては検討を行わなかったが、インタラクティブな教示を行うためにはこのような指示への割り込みも重要であると考えられる。また、今後は音声によるインタラクティブな教示を行うと同時に、「やさしい先生」であるとか「厳しい先生」といったようにシステムに対して個性を与えることによって、ユーザがシステムに対して人間っぽさや親しみを感じさせることも検討していく予定である。

参考文献

- [1] 大石, 岸本, 武田, 岡本, 上谷: “子どもたちによるコミュニケーションと、HI のデザイン,” 人工知能学会研究会資料 SIG-HIDSN-9603-6(3/7), pp.28-33, 1996.
- [2] 畑岡, 菊池: “音声を利用したマルチモーダルインタフェース,” 電子情報通信学会誌 Vol80 No. 10, pp.1031-1035, 1997.
- [3] 尾崎, 足達, 石井, 小谷津: “感性情報を考慮したファジー手書き文字学習システム,” 電子情報通信学会論文誌 D2 Vol.J79-D2 No. 9, pp.1554-1561, 1996.
- [4] 片桐: “社会的エージェントとメタ・コミュニケーション,” ATR 第 10 回研究発表会資料集, pp.44-45, 1997.
- [5] 阿部: “誤り指示音声の特徴分析と音声出力への適用,” 電子情報通信学会論文誌 D-2 Vol. J79-D-2 No. 12 pp. 2191-2198.
- [6] 篠崎, 阿部: “規則合成音声で躍動感を実現する方略について,” 情報処理学会研究報告, 97-SLP-17, pp. 81-88, 1997.
- [7] 篠崎, 阿部: “合成音声によるリアクションがユーザに与える印象について,” 日本音響学会講演論文集秋, pp.229-230, 1997.
- [8] Carol A. Simpson and Douglas H. Williams, “Response Time Effects of Alerting Tone and Semantic Context for Synthesized Voice Cockpit Warnings,” Human Factors, 22(3), pp.319-330, 1980.
- [9] 阿部, 水野, 中嶋: “様々な音声表現を実現できる音声作成ツール - Speed97 -,” 情報処理学会研究報告, 97-SLP-17, pp. 67-72, 1997.