

音声ペン: "WRITING" on Speech

栗原 一貴	後藤 真孝	緒方 淳	五十嵐 健夫
東京大学情報理工学系 研究科コンピュータ科 学専攻	産業技術総合研究所	産業技術総合研究所	東京大学情報理工学系 研究科コンピュータ科 学専攻 / JST PRESTO
qurihara@nifty.com	m.goto@aist.go.jp	jun.ogata@aist.go.jp	takeo@acm.org

本論文では音声認識と手書き文字認識を組み合わせ、講義・プレゼンテーションなどにおける手書き文字入力を効率化する「音声ペン」システムについて述べる。講演者が普段通り自由に発言し電子白板に板書を行うと、システムがバックグラウンドでその音声と手書き文字を認識し、その結果に基づき次に書かれるであろう内容が予測され、入力候補として提示される。適切な予測が得られた場合講演者はそれを選択することで文章に挿入でき、その分の手書き作業が不要となる。適切な予測が得られなかった場合、もしくは予測システムに関心がない場合には単純に提示される予測を無視すればよく、再入力や訂正を強制されることはない点が特長である。さらに音声ペンシステムは講演者の音声認識結果を一人一人の聴衆のPCに配信し、同様の予測入力に活用することで講演者と聴衆の間で予測入力用のコンテキスト情報を共有することができ、参加者全員の「書く」作業を促進することができる。簡単なユーザスタディによりシステムの有効性が示され、さらなる改善点が得られた。

Speech Pen: "WRITING" on Speech

<i>Kazutaka Kurihara</i>	<i>Masataka Goto</i>	<i>Jun Ogata</i>	<i>Takeo Igarashi</i>
Department of Computer Science	National Institute of Advanced Industrial Science and Technology (AIST)	National Institute of Advanced Industrial Science and Technology (AIST)	Department of Computer Science
The University of Tokyo	The University of Tokyo	The University of Tokyo	The University of Tokyo / JST PRESTO
qurihara@nifty.com	m.goto@aist.go.jp	jun.ogata@aist.go.jp	takeo@acm.org

This paper introduces a multimodal input system, called *speech pen* that assists digital writing during lectures or presentations with background speech and handwriting recognition. The instructor basically freely speaks to the audience and writes on an electronic whiteboard as usual. The system recognizes those speech and handwriting in the background and provides the instructor with predictions for the further writing by using the recognized results. The instructor can accept a prediction and paste it in the board to save manual writing. If all predictions are wrong or useless, the instructor can simply ignore them. The speech-pen system also allows the sharing of context information for predictions among the instructor and the audience; the result of the instructor's speech recognition is sent to the audience to support their own note taking. A preliminary study shows the effectiveness of this system and the implications for further improvements.

1 はじめに

講義においては講演者・講師が黒板に情報を書き込みながら話を進め、聴衆・学生がそれを見聞きしながら各自のノートにまとめていく。これは今日においても日常的に繰り返されるコミュニケーション様式のひとつであるが、音声やジェスチャー、文書などの複数のモダリティが自然かつ効果的に活用されている好例であるといえる。近年の技術進歩により講義において講演者や聴衆をサポートするシステムの開発が盛んになってきた。事前にスライド資料を用意し電子的に順に提示していく手法は一般的になりつつあるが、一方で進行中に資料を作成・提示するという講義本来の

スタイルを電子的に支援する手法の研究も進められている。[1]によると、「書く」という作業は単に準備されたスライド資料を見せるだけの講義・プレゼンテーションにくらべてより柔軟であり、聴衆が集中して取り組めるために優れているという。「書く」ことをテーマにした研究として、事前に作ったスライド資料に対して発表時に追加的に電子ペンによる書き込みを行えるようにしたもの[2][3]、電子ペンによる資料作成作業自体を扱ったもの[4][5][6]などを例として挙げる事ができる。

しかし書くという作業には大きな問題がある。キー

ボードを用いたテキスト入力に比べて、長いテキストを入力するのに労力と時間を必要とするという点である。岩田ら[7]は調査の中で講義時間全体の18%の時間が板書に費やされていたことを報告している。

それでは、別の可能性として音声認識を用いてリアルタイムに資料作成を行うことは現実的だろうか。講義においては講演者の発言内容と板書内容は密接にかかわるため、ディクテーションの利用は妥当性があり、また長い文を一度に入力可能な点も魅力的である。しかし講演者は発言した内容をそのまま板書するわけではなく、またディクテーションを行うと必ず誤認識が発生するために修正作業が必要となる。一般に修正作業は煩雑なので、そこで話が途切れ聴衆の理解は妨げられてしまう問題が生じる。

本論文では、ディクテーション以外の音声認識の新たな利用法として、講義中における講演者および聴衆の文字入力作業を支援する「音声ペン」システムを提案する。本システムは、上記のように講義中の音声を認識して板書の労力を削減することを目指す。認識結果をそのまま板書内容として表示せず、電子ペンによる手書き文字入力用の予測候補の生成にのみ用いる特長を持つ。講演者は普段通り話しながら手書き文字入力で行っていき、うまく認識されて適切な候補が表示されたら、選択して入力できる。たとえ誤認識があっても、入力予測候補の質が落ちるだけで板書を妨げられることはない。また自然に発話された内容を入力予測として副次的に利用する形態のため、特別に音声認識器向けに話す必要はなく普段どおりの講義を行うことが可能である。本システムではさらに、聴衆も講演者の音声認識結果に基づいて各自の手書き文字入力の予測候補として見ることができるようになることで、ノートを取る作業も省力化する。

音声認識のインターフェースとしての新たな可能性を論じた関連研究として、後藤ら[8][9]の非言語情報を活用した「音声補完シリーズ」があげられる。本論文もそのような試みのひとつであると位置づけられる。また、複数のモダリティの認識技術を相補的に組み合わせることで全体の認識率を向上させるOviatt[10]提唱のmutual disambiguationを、本論文では音声認識と手書き文字認識を組み合わせた文字入力に適用している。中川ら[11]はそのための統合的な認識エンジンを開発しテキスト入力インターフェースの一例を示したのに対し、本論文では音声認識器用ではない日常の自然な発話の活用を取り扱う。このテーマについてはHidusら[12]、Lyonsら[13]が取り組んでいるが、文字入力へと活用する我々の目的とは異なる。Mankoffら[14]のように認識技術の誤認識とそれに伴うあいまい性の存在を前提としたツールキットの開発を研究した例もある。

2 音声ペンシステム

音声ペンシステムは音声認識と手書き文字認識を用いて、講義・プレゼンテーション時の板書およびノート取り作業を支援するものである。通常良く行われる「活字」テキストの入力でなく、手書き文字をそのまま記載するという従来の伝統的な講義・ノート取りのスタイルを維持しながら、次に書く語の予測を提示することで、すべて手書き入力する場合に比べて労力を軽減することができる。

2.1 システム構成

図1に現在の音声ペンシステムのシステム構成を示す。講演者は音声認識用のマイクに向かって話し、プロジェクタに接続されたTablet PCもしくは電子白板にペンで書き込みを行うことで講義を進める。このとき聴衆もそれぞれ独立して各自のノートをTablet PCで取る。講演者の音声は音声認識サーバで処理され、認識結果が講演者、聴衆を含むすべてのユーザにネットワーク経由で配信される。

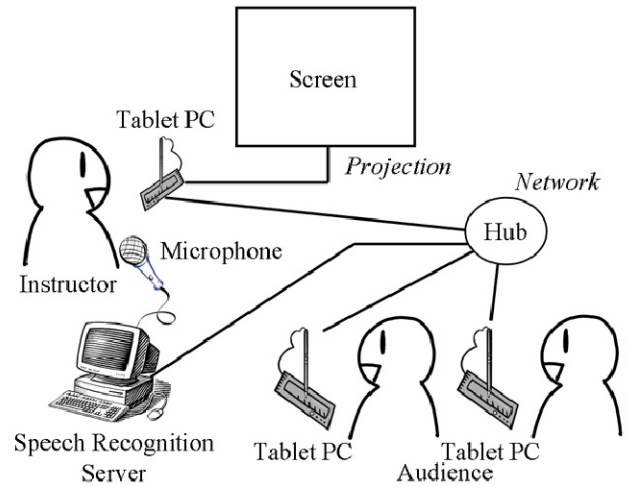


図1：システム構成

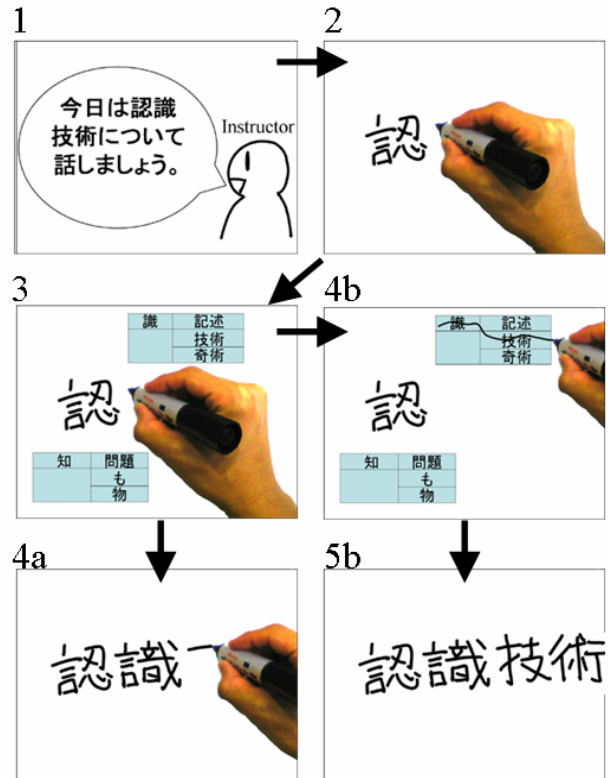


図2：ユーザインターフェースの概要

2.2 ユーザインターフェースの概要

図2はユーザの視点から音声ペンシステムがどのよ

うに動作するかを示した図である。講演者は自由に発話しながら電子白板に板書を行う(図2の1および2)。書くのを少し静止すると、システムは音声認識と手書き文字認識結果に基づき次に書く可能性が高い文字、語、文を提示する。これらの予測候補は書く作業の邪魔にならないように手の周りに表示される(図2の3)¹。予測候補は過去の発言履歴(音声認識結果)や予め設定しておいた辞書から生成される。もし講演者が予測を利用したくない時や正しい予測候補が得られない場合はそのまま手書き作業を継続できる(図2の4a)。入力したい候補を発見できた場合は候補をなぞるジェスチャーによりそれを白板上に挿入できる(図2の4b 5b)。この文字は、講演者の筆跡に似せて作られたフォントで表示される。

ユーザインタフェースの観点から重要である点は、ユーザが予測候補を使いたいと思う時以外は音声ペンシステムの存在を無視できることである(ディクテーションでは無視できない)。これは本システムが基本的にバックグラウンドで働くものであり、使用を強制するものではないからである。初心者ユーザは初め普段のように手書きを行えばよく、段階的にシステムのサポートを受けるように慣れてゆけばよい。

次の節以降ではユーザインタフェースデザインの詳細を示す。

2.3 予測の表示

本システムでは複数の予測候補がユーザの最新の書き込み位置の周辺に表示される(図2の3)。これらの予測結果は過去の発話の音声認識結果に対応しており、直前の手書き文字認識結果に基づき音声認識結果データベースから検索された「最後に書かれた文字や語から始まるような過去の発言」である。得られた発言は音声認識結果の複数の可能性とともに尤度の高い順に並べられて表示される(図3)。この表示は、音声認識の誤り訂正用インタフェース「音声訂正」[15][16]の競合候補表示を応用したものである。

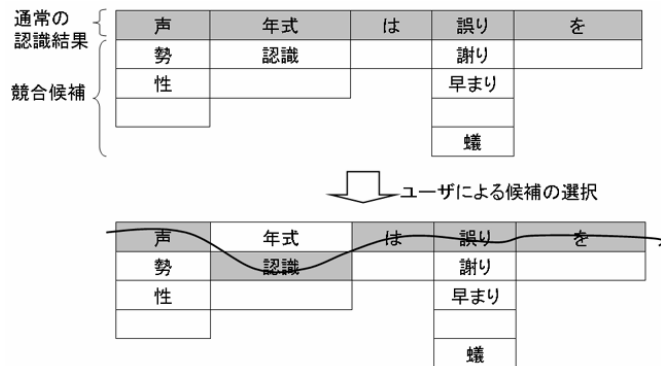


図3: 予測候補の提示・選択方法

2.4 予測の選択と無視

入力予測候補が表示されたとき、ユーザはそれらを選択して挿入するか、無視して手書き作業を続行する

¹講演者の書き込み用画面と表示用画面を分けられる場合(例えばTablet PCとプロジェクタ)、これらの予測候補は書き込み用画面のみ表示するとよい。

か任意に決定できる。予測の選択は一筆書きでリスト中の候補をなぞっていくcrossing interfaceで行われ(図3)、選択された文字列は手書き中の白板領域にユーザの筆跡を模したフォントを用いて挿入される。そうしたフォントは商用サービス[17]を利用して用意でき、フォントの表示サイズは直前の手書き文字を分析し自動的に決定される。予測候補が役に立たないときは、再び手書きを始めるだけで予測候補は消えるので、他の余分な操作なくユーザは予測候補を無視できる。また、最後に手書きを行ってから一定の時間が経過すると予測候補はやはり消える。以上のように本システムは[18][19]などの手書き認識技術を用いた従来のテキスト入力システムとは異なり、認識誤りを訂正する作業が必要ない。これは雑音下で誤りを起こしやすい音声認識技術および文字認識技術を有効に活用できる、特記すべき特長の一つである。

2.5 Ambient Contextの共有

講演者の音声認識結果は逐次聴衆にも分配され、共有される。共有されている音声認識結果は各ユーザ(講演者、聴衆)がこれまでに述べたような予測つき手書き入力を行う際にデータベースとして用いられる。システムは各ユーザの手書き文字認識を行い、最近書かれた文字や語から始まるテキストをデータベースの中から検索する。講演者の場合と同様、聴衆も提示された予測を適用するか無視するかの選択権を常に持つ。図6は後述のユーザスタディにおいて、講演者と聴衆が同時に各自作成した板書もしくはノートの例である。講演者の発言というある種のコンテキスト情報が、入力支援というアンビエントな形で共有されるため、各ユーザの主体性を反映した資料作成が可能であることが示されている。われわれはこれをambient contextの共有と呼んでいる。現在のプロトタイプシステムでは共有する対象が音声認識結果に限定されているが、今後手書き認識情報もambient contextとして共有したり、各ユーザがどのような認識候補を採用して挿入したか、といった情報も共有したりすることを検討している。

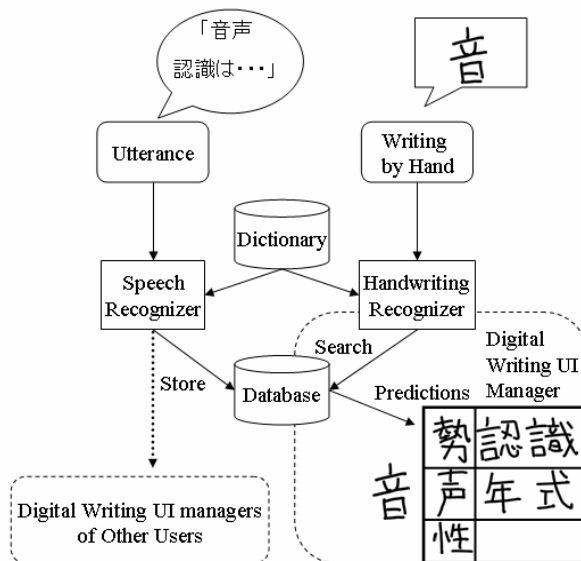


図4: システムのアーキテクチャ

3 実装

3.1 アーキテクチャ

「音声ペン」を実現するシステムは、図4のように、主に音声認識部と予測つき手書き文字入力管理部で構成される。音声認識部は、ユーザの発話を常時認識しており、手書き入力予測候補の元となる confusion network (ambient contextの実体となるデータ、次節参照)を生成して出力する。それと平行して、手書き文字入力管理部では、ユーザの手書き文字を認識し、その先の予測候補を画面表示する。通常の使用では、講演者のみが音声認識部、手書き文字入力管理部を両方使い、聴衆は後者のみを用いる。これらの構成要素は別々のプロセスとして実装され、ネットワーク(LAN)上の複数の計算機で負荷分散して実行することが可能である(前者をワークステーション(Xeon 3.06 GHz CPU, Linux 2.4)、後者をタブレットPC(Pentium M 1.4GHz CPU, Windows XP Tablet Edition)上で実行した)。プロセス間の通信には、音声言語情報をネットワーク上で効率よく共有することを可能にするネットワークプロトコルRVCP (Remote Voice Control Protocol)[20]を用いた。

音声認識部では、効率的なN-best探索手法を用いることにより、confusion networkの生成・送信をリアルタイムで実行可能となっている[15][16]。音響モデル、言語モデルには、CSRCソフトウェア2000年度版[21]から、PTM triphoneモデル、新聞記事テキストより学習された20000語のbigramをそれぞれ用いた。手書き文字入力管理部は、Microsoft Tablet PC Platform SDKを用いて実装した。

3.2 音声認識と入力予測候補の生成方法

提案する音声ペンシステムを実現するためには、逐次入力される講演者の発話に対して認識を行い、図3に示されるようなシンプルで入力予測候補をリアルタイムで生成する必要がある。講演者により発音が開始されると、発音中のパワー情報をもとに発話区間が検出され、得られた区間ごとに認識処理が実行される。ここで音声認識器は、入力発話に対してN-best探索を行い、何らかの中間結果(複数候補)を生成する。通常、大語彙連続音声認識における中間結果(例えば単語グラフ、N-bestリスト)は非常に膨大な数の候補を含んでおり、図3のようなシンプルで入力予測候補を表現することは不可能である。そこで、本システムでは、大規模な単語グラフを効率よく圧縮した形式である confusion network を、ユーザ側に提示する入力予測候補として利用する[15][16]。confusion networkを利用することにより、図3で示されるように、各単語候補間の競合関係が明確化し、ユーザは効率よくペン等による候補の選択が可能になる。

一般的に、ディクテーションを目的とした音声認識を考えた場合、少しでも高い認識率を得るためには、システムに用いる語彙や言語モデルの精度が重要なファクターとなる。特に講演や講義を認識対象にした場合、話題が多岐に渡ることや、話し言葉の影響などから、どのような音声に対しても高精度に認識可能な言語モデルを作成することは不可能である。それに対して、音声ペンでは、ユーザは誤りを含めた全ての音声認識結果を利用することは想定しておらず、認識誤り

を避けながらユーザの欲しい結果だけを積極的に利用するインタフェースとなっている。したがって、ディクテーション目的の音声認識システムのように、言語モデルや語彙の不備による認識誤りが、システム全体に大きく影響することはないと考えられる。実際に本システムでは、講演者の音声認識のための言語モデルとしては、より多くの話題をカバーし、比較的学習テキストも利用しやすい新聞記事から学習された N-gramを用いている。

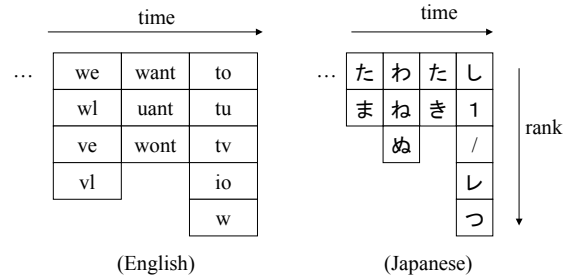


図5：手書き文字認識結果
認識結果は複数の区画からなり、その一つ一つが最大5個までの認識候補からなる。

3.3 手書き文字認識方法

音声ペンシステムでは、ユーザは電子白板上の任意の場所に任意の大きさで手書きを行うことができる。即ち[19]などの多くの従来の手書き文字認識によるテキスト入力システムとは異なり、文字入力用のセル(長方形領域)への書き込みを強制されない。その反面、文字認識に先立ちシステムはまずストロークのセグメンテーション(ストロークを文字単位にグループ化する作業)を行う必要がある。図5にセグメンテーションと手書き文字認識の結果の例を示す。日本語では文字が意味的な単位となるので文字単位のセグメンテーションを行うが、例えばヨーロッパ言語の場合には、語を単位とするセグメンテーションを行う必要がある。手書き文字認識の結果はN-bestリストの系列として次の処理段階に送られる。現在の実装では、Microsoft Tablet PC Platform SDKの文字認識エンジンを用いており、セグメンテーション結果に複数の可能性が考えられる場合については考慮していない。

3.4 入力予測候補の決定方法

システムは手書き文字認識の結果をクエリとして confusion network のデータベースを検索する。その際、まず一番最近書かれた文字(もしくは語)を取り出してクエリとし、対応するデータベース上の confusion network を検索する。もしもたくさんの候補がマッチした場合は、クエリに最近書かれた文字の一つ前の文字を加える。つまり confusion network の中から2文字の文字列と同じものを探す。この作業により、一般的にマッチした候補の数は1文字クエリの場合よりも減少する。この作業を繰り返し、マッチする候補がなくなるまでクエリの文字数を多くしていく。図5右の例では、「し」、「たし」、「わたし」、「たわたし」の順で検索を行っていく。最終的にシステムはもっとも長い

クエリにマッチした候補を出力とする。

この方法でははじめから検索候補が見つからない場合がある。つまり confusion network データベースの中に最近書いた文字が存在しない場合である。このような場合、システムは代わりに次に尤度の高い文字認識結果を用いる。そしてマッチする検索結果が多かった場合は、先述のように最近の文字からさかのぼってマッチしなくなるまで検索を進めていく。図5右の例では、もしも「し」が見つからなかった場合、「1」、「た1」のように進める。

システムは今まで述べてきたような作業を、予め設定してある数（現在の実装では3つ）のマッチする検索結果が得られるまで行う。得られた検索結果は尤度の高い順にソートされ、ユーザに提示される。予備実験ではこの単純なアルゴリズムでも比較的うまく機能していたが、今後の改善の余地は多い。例えば現在は confusion network 上の文字に付随している音声認識時の尤度を単に最終的なソートの目的のみに利用しているが、検索候補を得る過程でもこれを考慮することで性能向上が期待できる。

4 ユーザスタディ

提案システムの有効性を確認するとともにさらなる改善へ向けての知見を得るため、簡単なユーザスタディを行った。ユーザインタフェース研究分野とは無縁な3人のテストユーザがボランティアで参加した。

4.1 手順

用意したタスクは、講演者と聴衆に扮して模擬的な講義を行い、板書およびノート取り作業を行うというものである。それぞれのテストユーザは、講演者と聴衆を一人の実験監督者と交代で演じる。トータルでは、各テストユーザが講演者と聴衆を1回ずつ、実験監督者は各3回ずつ演じることになる。実験監督者は著者の一人であり、音声ペンシステムについて熟知しているので、事前にテストユーザにシステムの説明を行うが、タスク遂行中は一切手助けを行わない。

本実験では模擬的な講義のテーマとして「たこ焼きの作り方」を選んだが、音声認識エンジンの語彙や言語モデルには特に変更は加えていない。これは本システムが、認識誤りを起こしやすき環境にあっても有用であることを示すためである。音響モデルも、インフォーマルな会話用のものではなく、話者適応もしていない。

4.2 結果

図6に得られた板書・ノートの例を示す。予測により挿入されたテキストが区別できるように下線を図の作成時に引いた。[1]などの従来のコンテキスト共有システムでは見られない、表現の多様性が観察された。これは本システムの自由度の高さを示すものである。また、手書き風フォントの採用により板書の仕上がりがとても自然になっていることが観察された(図6左)。手書き文字と予測の挿入によってシステムが生成した文字の移り変わりはほぼ見えない。一方、図6右は敢えてフォントとしてMSゴシックを採用した例だが、手書き文字との移り変わりに違和感があることがわか

る。

4.3 サポート率

音声ペンシステムがどの程度ユーザを支援できるかを分析するため、以下の「サポート率」という評価尺度を提案する：

$$S = \frac{N_{Sup}}{N_{All}}$$

ここで N_{Sup} はシステムによって生成された文字のストローク数²、 N_{All} は全ストローク数である。サポート率はすべてのストロークが手書きによって書かれた場合に最小値0となり、すべてのストロークがシステムによって生成された場合、理論値として最大値1を取る。本システムではユーザは基本的に手書きを行っており、予測は必要などきのみ用いるという方針のため、サポート率を上げて最大値1を目指すことが目標ではない。また、箇条書きやアノテーション記号などの非文字列ストロークもサポート率を低下させる要因となる。

図7に全テストユーザのサポート率を示す。監督者の値はユーザスタディの観点からはあまり意味がないので参考値として掲載する。訓練時間はわずかであったが、ほとんどのテストユーザはシステムのサポートを得ることができた(平均0.31)。さらに重要な知見は、すべてのテストユーザが認識誤りからの復旧失敗によるタスク中止に追い込まれることなくタスクを完遂できたことである。また、講演者と聴衆の間でサポート率に明確な差が無かったことも観察された。我々は話しながら書かなければならない講演者の方が作業負荷は大きいとサポート率が低下すると予想していたが、これは確認されなかった。

図7に示される結果によると、テストユーザBは講演者を演じた際、システムの支援を一切受けなかったことがわかる。彼はその理由について「私はそもそも書きながら話すこと自体に慣れていないので、システムの支援を利用するゆとりがなかった」と述べている。このような場合でも、単純に提示される予測をすべて無視して書き続けることでタスクを問題なく完了できる点は注目に値する。また、そのときも聴衆に扮する実験監督者がテストユーザBの音声認識情報を効果的に活用することができた事実は ambient context 共有の有効性を示唆するものである。

4.4 テストユーザからのフィードバック

タスク終了後にテストユーザに対しインタビューを行った。まず、本システムに対する一般的な印象を尋ねたところ、3人全員がポジティブな印象を持っており、特に「使用を強制されるのではなく、活用したいときだけ存在を意識すればよい」という点が魅力的だと指摘されていた。次に、本システムの更なる改善に向けてのコメントおよび提案を尋ねたところ、(1)入力予測候補の表示場所には改善の余地がある (2)各予測候補の表示が小さすぎる (3)すべての文字を手で書く時間

²厳密に言えばシステムによって挿入される文字は活字であり、ストローク情報は無い。ここでは活字をもし手で書くとしたときに必要な手書きストローク数を数えている。

とそう変わらないようであれば、予測候補を選ぶメリットは小さいだろう などが得られた。

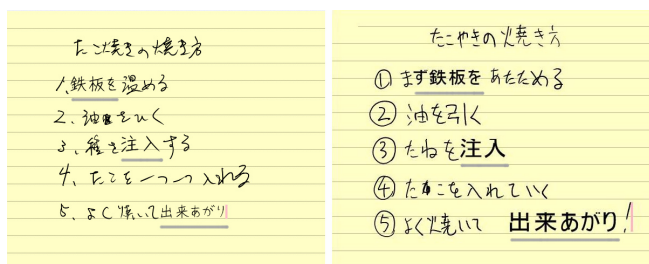


図6：ユーザスタディで得られた板書・ノートの例
(左) 講演者の板書，手書き風フォントを利用
(右) 聴衆のノート，MSゴシックフォントを利用

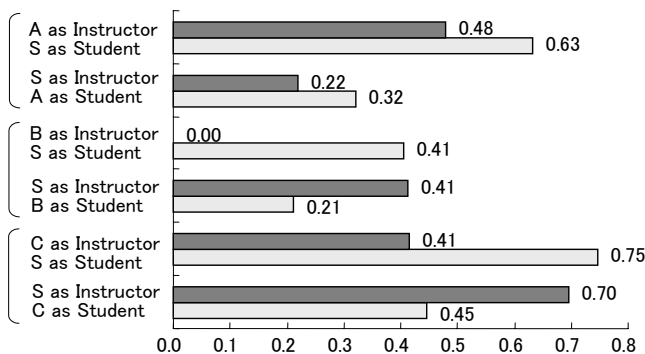


図7：サポート率

A,B,Cはテストユーザのサポート率，Sは実験監督者のサポート率（参考値）を示す．A~Cはシステム利用経験が乏しいにもかかわらずサポートが得られていたことがわかる。

5 まとめ

本研究では、講義やプレゼンテーションなどにおいて音声認識と手書き文字認識を用いつつユーザが手書き文字を書く作業を支援する「音声ペン」システムを開発した。本システムでは、講演者および聴衆が予測入力用の情報をambient contextという形で共有することができる。簡単な実験により本システムの有効性が示され、更なる改善に向けてのユーザからの意見が得られた。

参考文献

- Schilit et al.: Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations. *CHI'98*, pp.249-256, 1998.
- Kam et al.: A System for Cooperative and Augmented Note-Taking in Lectures. *CHI'05*, pp.531-540, 2005.

- Anderson et al.: A Study of Digital Ink in Lecture Presentation. *CHI'04*, pp.567-574, 2004.
- Elrod et al.: Liveboard: A Large Interactive Display Supporting Group meetings, Presentations, and Remote Collaboration. *CHI'92*, pp.599-607, 1992.
- Denoue et al.: Shared Freeform Input for Note Taking across Devices. *CHI'03*, pp.170-171, 2003.
- Stifelman et al.: The Audio Notebook. *CHI'01*, pp.182-189, 2001.
- 岩田 陽子, 加藤 直樹, 中川 正樹: "対話型電子白板を用いた電子化授業への遠隔受講者参加方式の試作", 情処研報 2002-CE-67, pp.33-40, 2002.
- 後藤 真孝: 解説 "音声補完: 言い淀むと助けてくれる音声インタフェース", 情報処理(情報処理学会誌), Vol.43, No.11, pp.1210-1216, 2002.
- 後藤 真孝: "非言語情報を活用した音声インタフェース", 情処研報 2004-SLP-52-7, pp.41-46, 2004.
- Oviatt: Mutual Disambiguation of Recognition Errors in a Multimodal Architecture. *CHI'99*, pp.576-583, 1999.
- 中川 竜太, 小林 唯, 小林 隆二, 篠田 浩一, 古井 貞照: "音声と手書き文字の同時入力インタフェース", 情処研報 2005-SLP-56, pp.29-34, 2005.
- Hindus et al.: Ubiquitous Audio: Capturing Spontaneous Collaboration. *CSCW'92*, pp.210-217, 1992.
- Lyons et al.: Augmenting Conversations Using Dual-Purpose Speech. *UIST'04*, pp.237-246, 2004.
- Mankoff et al.: Interaction techniques for ambiguity resolution in recognition-based interfaces. *UIST'00*, pp.11-20, 2000.
- 緒方 淳, 後藤 真孝: "音声訂正: ``CHOICE" on Speech", 情処研報 2004-SLP-54-54, pp.319-324, 2004.
- 緒方 淳, 後藤 真孝: "音声訂正: 認識誤りを選択操作だけで訂正ができる新たな音声入力インタフェース", WISS 2004論文集, pp.47-52, 2004.
- マイフォントサービス, TechnoAdvance, Co Ltd., <http://www.techno-advance.co.jp/product/myfont/>
- Masui: An Efficient Text Input Method for Pen-based Computers. *CHI'98*, pp.328-335, 1998.
- 福島 俊一, 山田 洋志: "予測ペン入力インタフェースとその手書き操作削減効果", 情処学論, Vol. 37, No. 1, pp. 23-30, 1996.
- 後藤 真孝, 伊藤 克亘, 秋葉 友良, 速水 悟: "音声補完: 音声入力インタフェースへの新しいモダリティの導入", コンピュータソフトウェア, Vol.19, No.4, pp.10-21, 2002.
- Kawahara et al.: Recent Progress of Open-source LVCSR Engine Julius and Japanese Model Repository. *ICSLP*, pp.3069-3072, 2004.