音声インタフェースシステムの
効果的設計と評価に関する研究

Research on Effective Designs and Evaluation
for Speech Interface Systems

2011 年 2 月

早稲田大学大学院 基幹理工学研究科

西本 卓也

論文要旨

本論文は，音声合成および音声認識を用いて構成される情報通信システムを，誰にでも使いやすいものにするための体系的な方法論を提案する．また，提案する方法論に基づいて行われた応用システムの設計，実装，評価について述べる．

本論文の構成は以下のとおりである．

第 1 章では，ヒューマンインタフェースに関する基礎理論を踏まえて，独自の視点を加えた「基本原則」「構成原則」「導入原則」の 3 つのインタフェース原則を提示する．これらの原則の音声応用における重要性が，本論文におけるもっとも大きな仮説である．

第 2 章以降では，音声応用システムの発見，設計，評価において，本研究で提案した原則論が有効であることを示した具体的な成果について述べる．その主要な部分は，音声技術の応用分野を発見し，その有効性を検証する 2 つの研究と，特に心的負荷（メンタルワークロード）に着目した音声インタフェース評価に関する 2 つの研究である．前者の具体的なシステムは，音声認識・マウス・キーボードを併用した作図アプリケーションと，インターネットを介した非同期の音声コミュニケーションを実現した音声会議システムである．後者は具体的には，車載情報システムを想定した音声対話において二重課題法を適用し，比較的高い時間分解能でワークロードの増加する箇所を特定できることを示した研究と，視覚障害者のためのスクリーンリーダを想定した超早口音声の聞き取りにおいて NASA-TLX 法を適用し，被験者への教示によって被験者の知識や慣れの効果が制御されることを示した研究である．

第 6 章では，さまざまな音声インタフェースの研究として，インタフェースの理論および原則について再考しつつ，本研究から今後期待される展開を論じる．情報通信分野で近年注目されているバリアフリー（アクセシビリティ），エンタテインメント，User Generated Media (UGM)，ソーシャルメディア，などの概念についても，本研究における位置づけが議論される．

第 7 章では，本研究の主たる貢献を検証する．すなわち，本研究の特徴は，ヒューマンインタフェース技術の幅広い要素を網羅した原則の考察と，音声認識および音声合成の利用に特化した具体的な検討が，明確に分離されて議論されている点である．既存のガイドラインは従来のインタフェースデバイスに特化していたり音声技術に特化したりしたのに対して，本研究はインタフェース技術の普遍的なガイドラインから自然に導かれる音声応用システムの設計や評価

を論じている．本研究は新たなインタフェース技術や応用分野に対して柔軟に適用できる枠組みを提供している．

Abstract

This paper describes a systematic way of enabling of developers and designers to build information-communication systems successfully with speech technologies, such as speech synthesis and speech recognition. As the results of this work, application systems of speech technologies can be used easily for everyone. This work also describes four research projects including the development of speech applications and the evaluations of speech interfaces, which are performed based on the proposed methodology.

This paper consists of following chapters.

The first chapter proposes "the principles of interfaces," consists of (1) basic principles, (2) organization principles, and (3) adoption principles, which are based on the fundamental theories of human-machine interactions. The most important hypothesis of this work is that these principles are indispensable for accomplishing the projects to investigate on various human interface systems with speech technologies.

The following four chapters show the effectiveness of proposed principles, which comes along with the effectiveness of speech technologies and importance of designs and evaluations of speech interfaces.

One of the application systems proposed here is the S-tgif multimodal drawing system, which uses isolated speech command recognition, mouse and keyboard. Using speech recognition with the application, average operation time or the number of command inputs can be reduced.

Another proposed system is the AVM asynchronous voice messaging system. The system displays the voice message as the threaded written words. The user can manipulate voice messages just as if they are text messages. If the participant wants to quote or annotate to a message, the user has only to play the sound and barge into the message while it is playing. Such user interface increases the usefulness of the voice-mail system.

Other two projects are related to the evaluations of speech interface systems.

One of the projects proposed an improvement of the dual-task method to measure the

workload of spoken dialog tasks. In this method, subjects play a game using visual display and keyboard input. We evaluated the effectiveness of our method with a word shadowing task and a spoken dialog application. Our method can measure relatively small workload differences, such as in word shadowing tasks, which are difficult to measure with the previous works. Also, our method can identify positions in the dialogs which cause some users significant difficulty.

Another project was related to the ultra fast speech for the computer application of persons with visual disability. For the evaluation of such speech, consideration of learning effect of the listener is important. In this research, the learning effects of listening to ultra fast speech with the control of word familiarity were investigated with the considerations of (1) the changes of the familiarity condition during the experiments, and (2) existence or nonexistence of instructions of the familiarity. The experiments to observe the intelligibility and mental workload were performed, using the speech with the speed of approximately 21 mora per second. The results supported the hypothesis that the intelligibility increases and mental workload decreases if the listener is aware of high word familiarity because the access to mental lexicon is promoted.

The chapter 6 describes the various research projects related to the interface principles, which give the perspectives of human-machine interactions and applications of speech technology.

The most important contribution of this work is discussed in last chapter, which is the separation of general human interface principles and utilization of speech technologies. The well-organized principles naturally give the guidelines for designing and evaluating interface systems with various applications, modalities and devices. Speech technology is one of the applications of the principles.

目次

図目次

表目次

第 1 章

序論

1.1 はじめに

本論文は，音声合成および音声認識を用いて構成される情報通信システムを，誰にでも使いやすいものにするための体系的な方法論を提案する．また，提案する方法論に基づいて行われた応用システムの設計，実装，評価について述べる．

本研究の背景は以下の通りである．

近年，音声合成および音声認識は，ヒューマンインタフェースの新たな構成要素として，広く用いられつつある．過去数十年に渡って，多くの研究システムや製品が開発されてきた．

しかし現在においても音声技術は完璧ではない．さらに，完璧でない技術は日常的な利用に値しない，という理由で，音声技術の利用は広まっていない．

この状況を打破するためには，インタフェースシステムの設計において音声技術をどのように活用すべきか，指針を示せるための体系的な方法が必要である．

例えば，入力手段としてマウスとキーボードがすでに広く受け入れられている応用システムにおいて，音声入力がどのような役割を果たすべきか，音声入力によるインタフェースシステムをどのように実装すべきであるか，できるだけ試行錯誤に頼らずに設計できることには大きな意義がある．

さらに，音声技術は，バリアフリーやエンタテインメントなど，新たな応用システムにおいては重要な構成要素である．これらをどのように設計・実装・評価すべきであるか，といった知見は，それぞれの応用分野の中でのみ議論されており，分野をまたいで適用できる方法論は検討されて来なかった．

例えば，あらゆる情報通信技術を視覚に頼らないで利用できるようにすべき，という主張が，現実的に考慮されはじめている．これは，障害の有無や加齢の影響に関わらず情報通信技術が利用できることは，国際的な要求として認識されており，これらがユニバーサルに実現可能な技術基盤が整備されつつあるからである．バリアフリーの概念を広く捉えるならば，カーナビ

ゲーションシステムの操作も，運転中という制約された状態の人間の支援技術であり，音声の重要な応用分野と言えよう．

さらに，歌声合成システムの流行，ユーザが作る動画や音楽コンテンツの流通などのいわゆるVOCALOID [*1] ブームによって，コンピュータの中のバーチャルキャラクタがヒューマンインタフェースの手段として認知されつつある．

1.2 インタフェースの理論

本節では文献 [1] に基づいてヒューマンインタフェースに関する重要な基礎理論を概観する．

ここでまとめる理論は，ヒューマンインタフェースの問題を人間工学，認知工学の立場から捉えた提案の草分けであり，現在も多くの分野に影響を与えている．

1.2.1 行為の 3 階層モデル

Rasmussen[2] は，人間の行為を以下の 3 つの認知的階層に支配されているものとして捉える観点を導入し，ユーザとインタフェースとの関わりを考えていく基盤とした．

技能ベースの行為 技能ベースの行為は，一度その行為を始動させる刺激が存在すると，意識的なコントロールのないまま，自動的に最終ゴールまで進行する（signal-driven である）．熟達化した日常的な行為の多くがこれに該当する．

規則ベースの行為 規則ベースの行為は，特定の目的を指向している（goal-driven である）．新しいインタフェースの操作を習得しようとして，マニュアルから規則を読み取って逐次操作をしているような状態が該当する．

知識ベースの行為 知識ベースの行為は，直面する事態に対して積極的にモデルを立てて関わっていくものである（model-driven である）．知識ベースの行為が必要とされるのは，事態が曖昧であったり，複雑過ぎたり，馴染みがなかったり，といった場合である．このような事態においては，まず事態の解釈（同定）が行われ，情報の意味づけがなされる．

1.2.2 情報処理特性のモデル

Card ら [5] は，エディタや Desk-top Publishing (DTP) システムなど情報機器の使い勝手を分析・評価するために，GOMS および KLM などのモデルを提案し，合わせてユーザである人間の情報処理過程について，特にその時間特性や容量特性に着目した認知的モデルを提案した．この情報処理特性モデルは，知覚系，認知系，運動系の 3 つの下位システムから構成される．

[*1] ヤマハ（株）が開発した音声合成技術．

Card らのモデルは，各過程の特性を数値的に明確化することにより，システムの使いやすさの定量的な分析や評価を可能にした．例えば，何らかの操作を n 回目に実行するときの所要時間が $n-1$ 回目の実行時間に対し，べき関数的に短くなっていくという習熟性に関する「べき法則」が提案された．また，画面の上でマウスを使って大きさ S 距離 D にある目標に手を移動する所要時間を

$$T_{POS} = I_M \log_2(D/S + 0.5)$$

$$I_M = 100[70 \sim 120]msec/bits$$

で表した「フィッツの法則」(Fitts's law) [7] もこのモデルの中で再評価された．

1.2.3 行為の 7 段階モデル

Norman[6, 13] は，インタフェースシステム使用時の認知過程を以下の 7 段階でモデル化している．

1. ゴール（目標）を立てる
2. 意図を形成する
3. 行為系列を特定化する
4. 行為を実行する
5. システムの状態を知覚する
6. 状態を解釈する
7. システムの状態を目的や意図と比較して評価する

すなわち，人がシステムを使う過程は，何かをしたいという目標を持つところから始まる．ユーザは心理的な世界の目標を実行するために，物理的な世界に働きかけなくてはならない．

なお，(1) ～ (4) の過程におけるユーザの困難を「実行の淵」，(4) ～ (7) における困難を「評価の淵」と呼ぶ．物理的世界と心理的世界を越えるこの 2 つの淵こそがヒューマンインタフェースの問題の源であり，この淵の橋渡しがインタフェースシステムを設計する際の主要な問題であると Norman は主張した．

1.2.4 道具型と秘書型

インタフェースシステムは「道具型システム」「秘書型システム」に分類できるという主張がある．これについて文献 [1, 3] に基づいて概説する．

ユーザはシステムを媒介として自分が望む世界となるように物理世界を変化させると考えられる．ユーザと物理世界の間にシステムがあり，そのシステムの両面が第一接面（人の側），第二接面（物理世界の側）であるとする．このときインタフェースシステムは以下のように分類で

きる．

秘書型システム　第一の接面に注目し，その接面より後ろすべてを外的世界として扱うインタフェースシステム．ユーザはシステムに向かっているという意識を持っており，対象とする物理世界には直接には向かっていない．操作の結果はシステムの報告によって知ることになる．理想的には，ユーザが自分の要求を自分の言葉で述べることで，システムがユーザの要求を解釈し，ユーザの望む結果を導くことが求められる．

道具型システム　第二の接面に注目し，ユーザ自身を行為者として自覚させ（直接操作），システムをユーザの一部分のように感じさせるインタフェースシステム．システムから外的世界への入出力がユーザ自身からの入出力であるかのように受け取られる．ユーザとシステムの情報の流れは完全に予測可能であり，システム自体はユーザというシステムの一部分となりユーザの意識から消えてしまう．すなわちシステムは透明性 [4] を獲得する．

1.2.5　インタフェースの理論と原則

インタフェースにおいて新しいモダリティを有効に活用し，その応用システムを発見していくためには，本節で挙げたような普遍的な理論に基づきつつ，システムの開発や改良に役立つような「インタフェースの原則」が有用である．

優れたインタフェースシステムを実現するための方法論は数多く提案されてきた．例えば Apple や Microsoft などのベンダーは，アプリケーション開発者に向けたインタフェース設計ガイドラインを作成している．

一方で，パーソナルコンピュータやモバイル機器の操作に限定しても，デバイス技術やソフトウェア技術は日々進歩している．音声以外の技術に限ってみても，インタフェース技術に影響を与える多くの変化が日々おきている．例えば，グラフィックスの表現力の向上や高速化，高品質ビデオ，画像認識，三次元モデル，立体視デバイスなどの進化，マウスからタッチパネル，加速度センサ，マルチタッチ操作への発展，二次元バーコードや IC タグなど実環境指向の自動認識技術の普及，高速インターネットを介して高性能のサーバと常時接続できる環境の整備，これらを実現するデバイスの小型化と低価格化，などである．

そして，ユーザも日々新たなツールの操作を学習しており，インタフェースシステムが前提とするユーザのスキルも変化している．

インタフェース技術の個別の技術をどのように使うべきか，という具体的な方法論やガイドラインは，実状に合わせて改訂されつづけなくてはならない．例えば Shneiderman の “Designing the User Interface” [14] は 2009 年に第 5 版が発行され，版を重ねながら最新のデバイスやアプリケーションに関する言及が追加されている．しかし，個別の技術に合わせたガイドラインだけでは，インタフェース技術を革新することは難しい．

1.3 インタフェースの原則

1.3.1 インタフェース原則の必要性

音声認識技術の進歩は目ざましく，種々の優れた応用システムが実現されている．しかし，代替の入力手段を考慮してなお音声入力の必然性を感じさせるアプリケーションは少なく，音声認識の応用がヒューマンマシンインタフェースの改善につながることの主張には必ずしも成功していない．

過去の音声研究は，要素技術の個々の改良の積み重ねが洗練されたトータルシステムを作るという，いわば還元主義的な思想で進められた．しかし，望まれるヒューマンマシンインタフェースは何かといった全体像を描いた上で要素技術たる音声入力に求めるものを検討する，グローバルな意識が必要である．

そこで本研究では，優れたインタフェースが満たすべき用件を整理した上で，他の手段の特質と音声特有の性質との対比に基づいて音声が担うべき役割を考え，さらにその機能をどのように実現するべきかを検討することで，有用な音声インタフェースを実現することを試みる．また，このような方法によって作成したシステムの評価を通じて，音声の利用がシステムの利便性の改善にどのように貢献したかを総合的に検討する．

インタフェースの原則としてはさまざまな提案がなされている [1, 13, 14, 15]．本研究では，先行の提案を参考にしながら，音声の特質の検討の指標として利用しうる新たなインタフェースの原則を提案する．

まず，望まれるインタフェース像を，インタフェースが持つべき機能に関する基本的要求と，それらの機能を組み上げてシステムを構成する際に考慮すべき要求との 2 つの側面から見直し，インタフェースの基本原則および構成原則として整理する [16]．さらに，応用システムの開発と評価のプロセス全体に関するインタフェースの導入原則を整理する．

1.3.2 インタフェースの基本原則

人間は道具を使って作業を行なうにあたっては，楽に仕事を進めることを何より望んでいるものと考える．ここでは，この観点から問題を整理する．

「楽である」とは，労少なく操作を実行できること，労少なく操作法について知ることができることの 2 つに分類できる．また，これらの 2 つにかかわる重要な問題として，システムの頑健性があげられる．

この 3 つの観点を満たすためにシステムが従うべき原則を「インタフェースの基本原則」とよぶ．また，前述の 3 つの観点からの要求に応える原則を，それぞれ操作労力に関する原則，透過性に関する原則，頑健性に関する原則，とよぶ．なお，透過性および頑健性については

Norman[13] による同様の主張などを踏まえている．

以下にこれらの原則について述べる．

I 操作労力に関する原則

a．位置移動最少の原則: コマンドを入力するために必要となる，手・マウス等の位置移動は少ないほど良い．

b．指定操作回数最少の原則: 一つのコマンドを指定するための操作回数は少ないほど良い．

c．指定操作容易性の原則: 一つのコマンド操作は単純で容易なほど良い．

II システムの透過性に関する原則

a．理解容易性の原則: システムの提示する情報，システムの状態とそこで可能な命令は，容易に知覚，把握，記憶できることが好ましい[*2]．

b．手順連想容易性の原則: ある命令を実行するための操作は容易に連想できることが好ましい．このため，操作の手順はアプリケーションの内外で一貫性を持ち，コマンド名なども連想しやすいことが好ましい．また，操作がもたらす結果は，常に予測可能であることが好ましい[*3]．

c．フィードバックの原則: 操作には，常に適切なフィードバックを得られることが好ましい．

III 頑健性に関する原則

a．誤入力防止の原則: システムに対する誤入力はなるべく防止できることが好ましい．

b．修復容易性の原則: 操作はできるかぎり可逆にし，誤操作が致命的な影響を及ぼさないことが好ましい．誤り易い操作に対する修復の操作は特に容易であることが望ましい．

1.3.3 インタフェースの構成原則

前述した基本原則の間にはトレードオフが存在する．例えば，キーボードでコマンドを入力する場合，労力最少化の立場からは，コマンド名は短いほどよく，1 回程度の打鍵で一つのコマンドが実行できることが望ましい．しかし，その手順の連想を容易にする立場からは，求める機能に対応する名称をフルネームでコマンド名とする方が望ましい．このように，どれかの基本原則を立てれば，どれかは立たなくなるといった関係が随所にある．

したがって，前述の基本原則個々を同時に最大化するシステムは存在し得ず，実際には目的に応じてどれかの原則を重視しながらバランスのとれたシステム構成を実現しなければならない．

[*2] これまで「状態理解容易性の原則」と呼んできた項目であるが，「システムの提示する情報」を追加し，項目名を改めた．また「知覚」「記憶」を新たに加えた．

[*3] 「予測可能」の主張は「フィードバックの原則」から「手順連想容易性の原則」に移した．結果が予測可能であることは手順連想を容易にするからである．

幅広いユーザに受け入れられる入力形態を実現するためには，ユーザの経験や使用頻度，好みなどに応じた最適形態の違いが入力システムに考慮されていなければならない．

どのような割合でどの基本原則を重視してシステムを構成すべきかについての指針をインタフェースの構成原則と呼ぶ．ここでは，構成原則として次の 3 つを挙げる．

a. 初心者保護の原則 [13]**:** 入力システムには，当該アプリケーションならびに計算機自体に不慣れなユーザのための，透過性の高い入力手段を用意すべきである．

b. 熟練者優遇の原則 [15]**:** 入力システムには，精通したユーザのための，アプリケーションに特化した効率的な入力手段を用意すべきである．

c. 上級利用移行支援の原則 [1]**:** 入力システムは汎用手段を用いてシステムを利用しているユーザに対して，特化手段の存在を知らせ，この利用を促す自然な枠組みを用意すべきである．

これらは初心者に対する閾を下げながらも，使い込んだときの利便性を重視し，利用者にいち早く熟練者モードへの移行を促すのが良いという主張である．

1.3.4 インタフェースの導入原則

インタフェースシステムを成功させるためには，アプリケーションそのものの選択や設計により深く関与し，システムをどのような状況に適合させ，どのように評価や改良を行っていくか，というプロセスが重要になっている．本節では，このような観点から，以下の「有用性」「適合性」「妥当性」の 3 項目から構成される「インタフェースシステムの導入原則」を提案する．

有用性の原則

インタフェースの有用性の原則とは，以下の主張である．

- 使用される場における必然性を考慮して設計と導入を行う．
- ユーザに動機付けを与える．
- ユーザや開発者による有用性の発見を支援する．

有用性の発見について以下に補足する．ある技術が有用であることを発見するためには，特殊な現場や立場からの要求，日常の些細な不便，などを幅広く検討する必要がある．また，ちょっとした思いつきを開発者が気軽に実装して試せる，ということが重要である．そのために必要なのはツールキット，API，Web サービス等の整備と戦略的な無償化，フリーソフトウェアの整備である．ノウハウを簡便化して共有するための標準化，そして国際化も必要である．

適合性の原則

インタフェースの適合性の原則とは，以下の主張である．

- あらゆる年齢や能力の人々に対して使いやすさを提供する（ユニバーサルデザイン）.
- 使われる状況・環境を考慮する.
- ユーザ以外の人に悪影響を与えない.
- ユーザが行っている他のタスクに悪影響を与えない.

ユニバーサルデザインについて以下に補足する．障害の有無に関わらず誰でも幅広く利用できるシステムを実現するためには，「聴覚だけでの利用（まったく見えない状況での利用）」と「視覚だけでの利用（まったく聞こえない状況での利用）」を考慮すべきであると，浅川は主張している[*4].

妥当性の原則

インタフェースの妥当性の原則とは，以下の主張である.

- 適切なタイミングで，妥当な手法と尺度で評価を行う.
- 結果を生かして反復的な開発・改良を行う.

反復的な開発プロセスについて以下に補足する．評価結果がシステム改良に生かされるために，反復的な開発プロセスが有効である．例えばクーパー [8] はインタフェース・システムの開発プロセスに起因する問題を取り上げ，「まずデザインを作り，それから機能を実装せよ」と主張している.

1.4 インタフェース原則に基づく音声技術の検討

1.4.1 本研究における仮説

本研究の大きな目標は，提案するインタフェース原則が，音声応用システムの開発者や評価者に役立つものであるかを，実践を通じて検証することである.

いわゆる「原則」の有効性を評価する方法として，本研究では「幅広い視点に基づく複数の問題を，この原則を用いることで適切に扱える」という主張そのものを仮説と考えて，この仮説を検証していく.

1.4.2 検討すべき視点

音声技術の応用について幅広く検討を行うためには，以下のような複数の視点で，バランスを考慮することが望ましい.

[*4] 浅川智恵子のアクセシビリティ論 http://itpro.nikkeibp.co.jp/article/COLUMN/20060920/248534/

1. 音声認識に着目した検討と，音声合成に着目した検討
2. 視覚要素と音声技術を組み合わせたマルチモーダル技術と，視覚に依存せず音声のみによって構成されるインタフェース技術
3. 音声認識技術について，音声コマンド入力とディクテーション入力
4.「道具型」システムと「秘書型」システム
5. 時間や労力などの効率化の観点，理解容易性の観点，および頑健性の観点
6. 客観的評価と主観的評価
7. システムの提案と，インタフェース技術の評価

1.4.3 インタフェース導入原則と音声技術

本節では 1.3.4 で述べたインタフェースの導入原則を音声技術に当てはめたときに，音声技術に期待される役割や，求められる研究のアプローチについて論じる．

有用性の原則と音声技術

主張：使用される場における必然性を考慮して設計と導入を行う．

ユーザに動機付けを与える．

ユーザや開発者による有用性の発見を支援する．

必然性の考慮

音声認識技術の利用が必然となる「現場」を探すことが重要である．初期の実験システムでは「電話応答システムにおいて数字を入力する代わりに，音声で読み上げた数字を認識できる」といった提案が行われた．これに対して「数字は電話機の数字ボタンで入力すればよいので，音声入力が優位性につながらない」という意見があった．一方で「数字は任意のリソースを示すユニバーサルな手段」として重宝される場面もある．

音声技術をマルチモーダル・インタフェースの要素技術として捉えなくてはならない．「音声と非音声の手段を適切に組み合わせる」ことが有効な場合もある．さまざまな立場の設計者との協力も必要である．

動機付け

音声認識に熟練しているユーザは少ない．音声を使ってみたいと思わせる「動機付け」が重要である．「楽しさ」をもたらすもの，特に「一人ひとりの人間が主役になるメディア環境時代」の新たなアプリケーションも，音声技術を使いこなすための視点である．

適合性の原則と音声技術

主張：あらゆる年齢や能力の人々に対して使いやすさを提供する．
使われる状況・環境を考慮する．
ユーザ以外の人に悪影響を与えない．
ユーザが行っている他のタスクに悪影響を与えない．

ユニバーサルデザイン

音声認識は聴覚障害の支援に有効なメディアである．具体例は，放送の字幕作成支援技術などである．また音声合成は（健聴者については）視覚障害の支援に有効なメディアである．具体例はスクリーンリーダ（画面読み上げソフトウェア）である．

状況と環境の考慮

カーナビゲーションシステムの目的地設定においては音声入力が有効である．これは自動車の運転において「ハンズフリー」「アイズフリー」が要求されるからである．これに加えて「マインドフリー」の考慮も重要である．例えば右左折の直前など「運転そのものが大きな負荷」という状況では，音声対話システムがユーザに話しかけたり，操作を求めたりすべきではない．

妥当性の原則と音声技術

主張：適切なタイミングで，妥当な手法と尺度で評価を行う．
結果を生かして反復的な開発・改良を行う．

インタフェースの妥当性評価

妥当性を保証するためには評価手法が重要である．音声対話システム・音声インタフェースがどれだけユーザにとって「楽であるか」を評価できる手法が重要となる．

所要時間に関する評価　同じタスクを与えて複数のインタフェースシステムを被験者に利用させ，所要時間を比較する．

身体的・物理的な負荷に関する評価　音声入力とポインティングデバイスの比較のために「手や指を動かす距離の累積」「コマンドを入力した回数」を使用できる．

心的要因に関する評価　二重課題法は心的要因を客観的な尺度で測ることができる一手法である．また，客観評価では比較できない要因も，主観評価では観測できる可能性がある．

1.4.4 本研究で取り上げる対象

本研究では，以下の 4 つの具体的なシステム設計およびインタフェース評価を通じて，本研究が提案するインタフェース原則そのものの有効性を検討していく．ただし，具体的な研究内容は次章以降で詳述する．

研究 1 音声作図システムの設計と評価
研究 2 非同期型音声会議システムの設計と評価
研究 3 音声対話による情報検索システムのワークロード評価
研究 4 スクリーンリーダを想定した超早口音声聴取の了解度およびワークロード

それぞれの研究が前述した要素の何を検討しうるかを，表 1.1 に示す．ただし○で示された項目は，個々の研究が左列の要素を直接的に検討していることを示す．空欄の場合も間接的に検討が行われる可能性がある．

この 4 つの研究は，いくつかの視点において，相互に補完しあう関係にある．例えば (2a) および (2b) に着目すると，マルチモーダル技術と音声のみの技術はどれか 1 つだけの研究では両方を扱えない．音声コマンド入力とディクテーション（(3a) および (3b)），道具型と秘書型（(4a) および (4b)）も同様である．

このように，4 つの研究を取り上げることによって，さまざまな音声技術，音声応用システムを扱う場合のインタフェース原則の有用性を幅広く検証できる．

表 1.1　検討すべき要素と研究対象.

記号	検討要素	研究 1	研究 2	研究 3	研究 4
(1a)	音声認識	○	○	○	
(1b)	音声合成			○	○
(2a)	マルチモーダル技術	○	○		
(2b)	音声のみの技術			○	○
(3a)	音声コマンド入力	○		○	
(3b)	ディクテーション入力		○		
(4a)	道具型インタフェース	○	○		○
(4b)	秘書型インタフェース			○	
(5a)	労力最少性	○	○		
(5b)	理解容易性	○	○	○	○
(5c)	頑健性	○	○		
(6a)	客観的評価	○	○		○
(6b)	主観的評価			○	○
(7a)	システムの提案	○	○		
(7b)	インタフェースの評価			○	○

第 2 章

音声利用作図システムの設計と評価

2.1 はじめに

本章では，タスクとしてコンピュータによる作図を取り上げて，インタフェースの原則に基づいて，音声をどのように利用すべきか，あるいは予想される弱点をどのように補うべきかを検討する．

作図タスクでは描画・修正などのモード切換と，キャンバス上の座標指定が混在して頻繁に行われる．このため，マウス・キーボードのみでは操作が煩雑になりがちであり，効果的な音声の併用により，インタフェースが格段に改善される可能性を持つ．

このようなアプリケーションを前提として，本研究では，音声の利用を単なるモダリティの追加ととらえるだけでなく，その中で音声のどのような特徴，あるいはどのような音声と他手段の協調のあり方が有用となりうるかを，インタフェースの原則論に基づいて考える．

音声と他の手段とを比較し，その特質を明らかにしようとする研究は古くから試みられている [9, 10, 11]．しかしそれらの多くは，キーボード，マウスなど他の入力手段との比較において，音声入力が，速さ，正確さの点で優れるか否かを論じている．結果はさまざまで音声の有効性を主張する共通の結論を得ていない．ここでの問題は，音声を他の入力手段と競合する手段としてとらえていることにある．

この中にあって Martin は，音声を他の入力手段と協調して利用する手段としてとらえ，こうすることでインタフェースの利用効率をいかに改善できるかという問題を扱っている [12]．彼は VLSI CAD において，音声という新たなモダリティの追加によってシステムの利用効率が向上することを確認している．本報告においても Martin [12] と同様にマルチモーダルな入力環境における音声の利用を考える．

2.2 音声作図システム設計における配慮

2.2.1 操作労力に関する配慮

位置移動を最少化する観点からは，各入力手段の役割分担を明確にすることが重要になる．そこで，モード切換を音声で行うことを可能にし，マウスが座標指定に専念できるようにする．このことによりマウス操作における手の移動を大幅に減少できる．

指定操作回数を最少化する立場からは，ショートカット（深い階層メニューの中の一つのコマンドをキー入力などで直接指定する枠組）の役割を音声で担わせる．従来，多くのコマンドを持つシステムでは，階層メニューからいくつかの手順を経てコマンドを選んでいる．音声認識を用いることにより，メニューを開くことなく直接コマンドを選択することが可能になる．これにより操作回数の軽減が期待できる．一般にはキー入力がこの目的に用いられるが，この操作では実質的に，押しやすい位置にある 10 個程度のキーに利用が制限される．これに対して音声の利用により，より多くのコマンドに対するショートカットが提供可能となる．

2.2.2 透過性に関する配慮

理解容易性という観点からは，音声の利用は負の効果を生じる可能性を持つ．マウスによる操作時には操作対象であるメニューなどに注目しているため，その操作に伴うモード切換等の情報をメニュー項目に記せばその変化を容易に確認できる．これに対して音声入力では，一般に視点は描画対象にあるので，モードの情報が画面隅に表示されてもこれを確認するのが難しい．そこで作業モードの表示をマウスポインタの形状で知らせることにより視点とモード表示位置を一致させることをねらう．

手順連想容易性に関しては音声の利用は有利に働く．例えば，一般にキーボードによるショートカット操作は一文字コマンドが多く，忘れやすく混同しやすい．音声ではコマンドの機能から連想される語彙をできるだけ多く登録することで，コマンド名を覚えやすく忘れにくくできる．

フィードバックに関しては，音声入力が受け付けられたかどうかをユーザに知らせるために，認識終了時にビープ音を返すこととする．また，認識結果も文字で副画面に表示する．

2.2.3 頑健性に関する配慮

誤入力を防止する観点から最も重要なのは，コマンド以外の発話（独り言，隣人との雑談など）の棄却である．この問題を回避するために，コマンド発話の開始をスイッチを押してシステムに知らせることも考えられるが，これではマウスなどとの協調的作業が妨げられる．ここではフロアマウントのマイクを採用することで，顔の向きが音量に与える影響を大きくし，非コマ

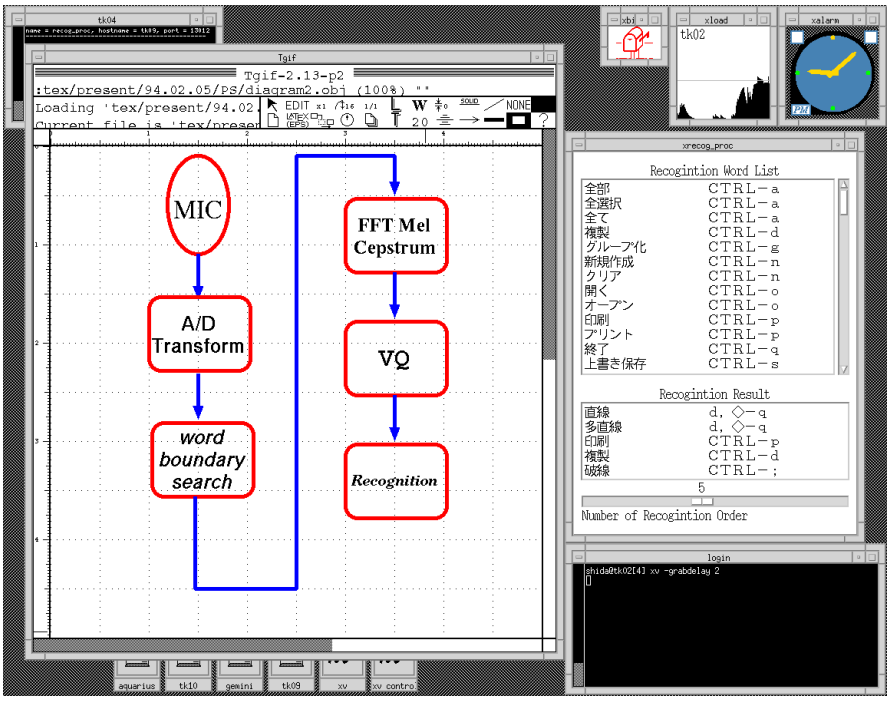

図 2.1　S-tgif の実行画面例．左側のウィンドウは作図部，右側は音声認識部で，上部は認識対象語を，下部は認識結果を示す．

ンド発話の棄却を容易にする [17].

修復容易性については取消操作を最も利用しやすいキーに配置することで対処する．

2.2.4　構成原則に関する配慮

幅広いユーザに受け入れられるインタフェースを目指す立場からは，比較的初心者向けと思われる音声・マウスによるコマンド入力と，熟練者向けと思われるキーボード操作を切換なしに同時に提供することで，初心者を保護しつつ熟練者を優遇できるインタフェースを目指す．

マウス・音声・キーボードの効率的な複合操作を提示するのが上級利用移行支援であると考えられる．今回のシステムでは，音声によるコマンド入力時に同等なキーボード操作を画面隅に提示することとする．

2.3　音声利用作図システム S-tgif の構成

本節ではシステムの構成を述べる．システムは S-tgif と呼ばれる．S-tgif は音声認識部とパブリックドメインの作図ソフト Tgif[18] の組合せによって構成され，X のプロトコルで交信する．S-tgif の実行画面を図 2.1 に示す．以下にそれぞれの特徴を述べる．

2.3.1 音声認識部

音声認識部は単語単位の離散出力分布型 HMM を用いて約 80 語の不特定話者音声認識を行う．高速化のため処理は複数の汎用ワークステーションで分散して行うことができる．認識処理は常時行われており，適当な音声認識結果が得られた段階で X のイベントを Tgif に渡す．

2.3.2 作図部

ベースとなる作図アプリケーションとして Tgif を用いた．tgif の基本的な操作は，パネル・ウィンドウ（画面上部に常時表示）またはポップアップメニュー（作図領域の任意の位置に表示できる）によって必要な機能や属性（線の太さ・塗りつぶしパターン・色など）を選んで実行する方式である．

2.3.3 操作ログモニタ

評価に用いるために，マウスカーソルの移動量，所要時間を記録するモニタープログラムを並列に動作させる．なお，これとは別に，ビデオによって操作の様子を記録し，認識誤りの個数などについてはこれを参照しながら手作業で数えることとする．

2.4 音声利用作図システムの評価実験

2.4.1 課題

実験時には課題としてこれから作る図の概略を与え，被験者にその作成を求めた．図には文字の大きさ・色・線種・塗りつぶしパターンなどの条件を文章で併記した．図形の大きさ・配置などの厳密さは問わないこととした．

課題は，練習用も含めて 7 つの異なる課題を用意した．課題はいずれも同程度の複雑さとなるよう配慮した．

2.4.2 被験者

被験者として 16 人（実験によっては 8 人）理工系の学生を選んだ．被験者は計算機操作にはある程度親しんでおり，マウス，キーボードなどについては一通りの利用経験がある．一方，S-tgif に関しては全く予備知識がない．

2.4.3　1 セッションの構成

a. 実験配置

1 回の実験をここではセッションと呼ぶ.

1 セッションにおいて被験者は 2 つの課題について作図を行った．一方の図は音声を用いて (S-tgif を用いて) 作図を行い，もう一方の図は音声を用いずに (Tgif を用いて) 作図を行った.

それぞれの被験者が先に音声を用いて実験を行うか後に用いるかはクロスオーバー法の実験配置に従った．すなわち，被験者は 2 つのグループ G_1, G_2 に分けられ，2 つの課題 A,B に対し，グループ G_1 に属す被験者は最初に音声を使って課題 A に取り組んだ後，音声を用いずに課題 B に取り組んだ．グループ G_2 に属す被験者は，まず音声を使わずに課題 A に取り組んだ後，音声を使って課題 B に取り組んだ．この配置によって，課題の違い，慣れの効果，被験者の能力の違いなどの影響を補正した上で，音声利用の効果だけを分離して抽出することができる.

b. 客観的評価

実験においては，課題図を完成させるまでの操作時間，マウスカーソルの移動量，各コマンドの操作回数，各コマンドの入力手段，音声認識エラー，音声認識エラーの訂正に用いられた取消操作 (undo) の回数を計測した.

課題図を完成させるまでの操作時間，マウスカーソルの移動量，各コマンドの操作回数，の 3 項目については分散分析によって音声の利用の効果を測定した．上記 3 項目を目的変数 (x_{ijkl}) とし，音声の利用の有無による効果 (α_i)，なれ・課題差の影響 (β_j)，グループの影響 (γ_k)，被験者による影響 (δ_l) を説明変数として式 (2.1) に示すモデルを立て，Scheffe 法により音声利用効果の 95% 信頼区間を求めた．なお μ は平均，ε_{ijkl} は誤差項である.

$$x_{ijkl} = \mu \cdot \alpha_i \cdot \beta_j \cdot \gamma_k \cdot \delta_l \cdot \varepsilon_{ijkl} \tag{2.1}$$

c. 主観的評価

セッション終了後にアンケートを行った．アンケートの内容は表 2.1 の通りである．項目 1 から 5 については，音声ありのシステムの使用感を，音声なしのシステムと比べ 5 段階（5：良い，3：同等，1：悪い）で評価させた．項目 6 においては，項目 1 から 5 の中で最も重要視する項目を答えさせた.

表 2.1　S-tgif アンケート項目.

項目番号	質問
1	操作の軽快さ
2	やりたい操作を実現する方法の分かりやすさ
3	ある状況で使える操作と使えない操作の区別
4	図形の変形や移動などの操作の簡単さ
5	疲労感の改善
6	上記 5 項目のうち最も重要なものはどれか

2.4.4　実験構成

全体の実験は，練習，セッション，中断などを以下のように配して行った.

a. ガイダンスと導入練習: まず，被験者に対してシステムの機能についての簡単な説明を行った後，15 分程度自由な作図をさせた.

b. 第 1 セッション: 最初の練習が終った直後，第 1 セッションを行った．被験者は 16 人とした.

c. 練習: S-tgif にさらに慣れさせるために，第 1 セッションの終了後の 1 週間のうち 3 日間に約 30 分程度のシステムの練習を課した．実習は自由な作図とした.

d. 第 2 セッション: 1 週間の練習を経て，第 2 セッションを行った．被験者は第 1 セッションに参加したもののうち 8 名を選んだ．実験方法は用いた課題が異なることを除いて第 1 セッションと同じである.

e. 中断: 利用中断の影響を調べるため，2 カ月の間被験者が S-tgif を利用することを禁じた.

f. 第 3 セッション: 2 カ月の中断を経て，第 3 セッションを行った．被験者は第 2 セッションの参加者 8 名とした．実験方法は用いた課題が異なることを除いて第 2 セッションと同じとした.

2.5　評価実験の結果

課題図を完成させるまでの操作時間，マウスカーソルの移動量，各コマンドの操作回数の 3 つの評価項目における音声利用の効果と，その 95% 信頼区間を図 2.2 から図 2.3 に示す.

第 1・第 3 セッションにおける操作時間と全セッションにおけるマウスカーソルの移動量に対する音声利用効果は，危険率 5% の F 検定において有意に認められた．コマンドの操作回数においては有意差は認められなかった.

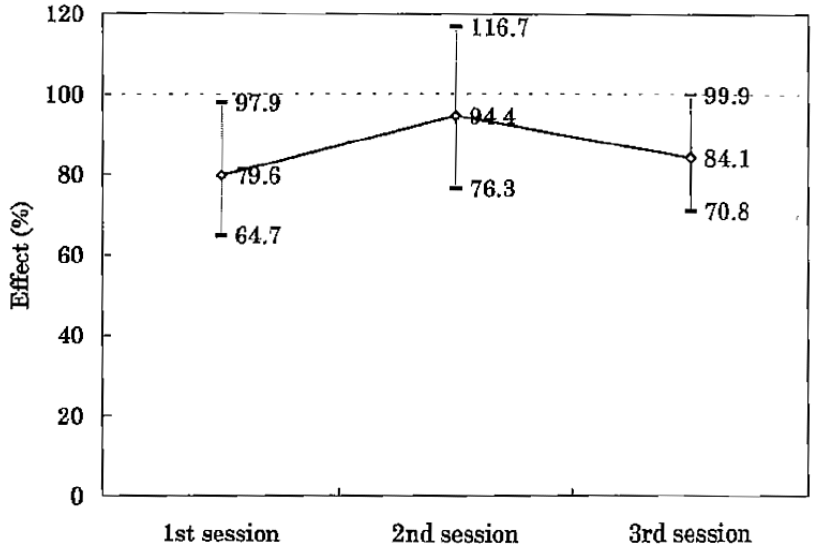

図 2.2　操作時間に対する音声利用効果の 95% 信頼区間.

それぞれのセッションにおいて，音声，マウス，キーボードの入力手段をどの程度の比率で用いたかを図 2.4 に示す．図中，a. は描画モードの変更関連のコマンドを集めて集計したものであり，b. は図の複写や削除といった図形操作，c. は線の太さなどの図形の属性変更，d. は色変更を集めたものである．f. は a. から d. およびそれ以外も含めたコマンド全体での集計結果である．

また，3 セッションを通じての音声認識率は 86%，それを undo で取り消した割合は，全誤りの 14% であった.

主観評価としてのアンケートの結果は表 2.2，および図 2.5 の通りである.

2.6　検討

2.6.1　操作労力に対する音声利用の影響

操作の労力あるいは効率に関する評価のために最も総合的な指標となりうるものは，作図完了までの所要時間と考える．この作図完了時間を第 1 から第 3 セッションでそれぞれ 20.4%，5.4%，15.9% 減じることができた（図 2.2 参照）．アンケートの結果では，疲労感の軽減，作業の軽快感の改善に音声が高い評価を得た．これらのことから，音声の協調的利用が操作労力の軽減に有効に機能したことがわかる.

実際のマウスの動きはどのセッションにおいても著しく減少しており，手の動きに関する労力を軽減できていた．このことは，右手が描画に専念し，複数の作業を同時並行的に行いやすくなったことを示唆する．アンケートの結果で，操作の軽快感に関する評価が高くなったのは，このようなことが影響したものと考える.

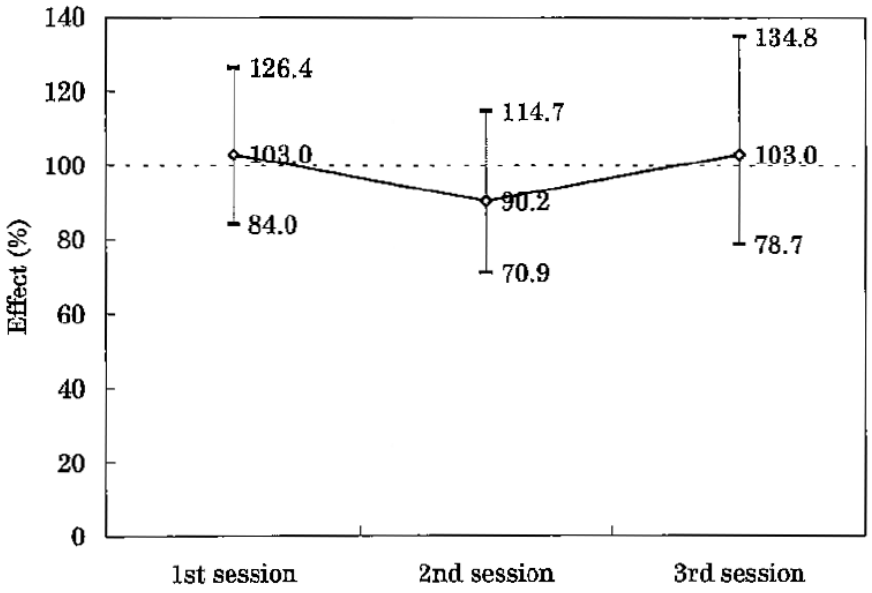

図 2.3　コマンド数に対する音声利用効果の 95% 信頼区間.

表 2.2　S-tgif アンケート結果.

項目	評価値 (分散)			
	第 1 回	第 2 回	第 3 回	全体
1　操作の軽快さ	3.88	4.25	4.50	4.21
	(0.70)	(0.79)	(0.29)	(0.61)
2　わかりやすさ	4.25	4.25	4.13	4.21
	(0.50)	(0.50)	(0.41)	(0.43)
3　使える操作の区別	3.13	3.50	3.50	3.38
	(0.98)	(0.57)	(0.57)	(0.68)
4　図形操作の簡単さ	3.75	3.75	3.88	3.79
	(1.07)	(0.21)	0.98)	(0.69)
5　疲労感の改善	4.00	4.25	3.88	4.04
	(1.14)	(0.79)	(1.55)	(1.09)

操作回数に関しては音声利用の効果は見られなかった（図 2.3 参照）．当初，音声の利用によって階層的なメニューからの選択操作が減少することで操作回数も減ることを期待したが，Tgif ではこのような操作は複合的な 1 回の操作として実現されることが多かった（マウスボタンを押したままメニュー階層の中を探索し，所望の項目が見つかったところでボタンを離す，など）．また，音声の利用によって階層メニューの利用が減った分を音声認識誤りによる再入力の

増加分が相殺してしまった側面もある．

2.6.2 透過性に対する音声利用の影響

アンケートでは音声の利用は「やりたい操作を実現する方法の分かりやすさ（項目 2)」で高い評価を得た（表 2.2 参照)．音声を利用可能にすることで，手順連想を容易にさせるという当初の目的は達成された．

また，「ある状況で使える操作と使えない操作の区別（項目 2)」では，当初音声の評価が低くなることを予想したが，アンケート結果では，平均的な評価を得た．モードの表示がユーザの視点に入るよう工夫をしたことと，音声の利用によってコマンドの階層性を気にする必要性が薄くなったことなどが影響したものと考える．

2.6.3 頑健性に関する検討

音声の利用は頑健性に対し負の影響を与える可能性が高い．

本システムでは認識率は 86% であった．しかし，認識誤りに対し取消操作を行ったのはわずか 14% であった．これは，色変更時に認識誤りによって他の色になってしまった場合などは，単に色を指定し直すだけで取消操作を必要としないことなど，入力エラーが作業遂行上致命的な状況を引き起こさないことによる．本システムにおける音声の利用は，このような意味で修復容易性の原則を満たしている．このことが，86% という認識率にも関わらず，音声の利用の評価が良好だったことにつながっているものと考える．

2.6.4 初心者における音声の利用の効果

作業完了時間に関する音声利用効果を時間の経過を追って調べると，1 週間の練習を積んだあとの第 2 セッションでは音声の利用による減少は 5.6% にとどまっており，有意差もなくなった．

これは，図 2.4 からわかるように，練習によってより効率的なショートカットであるキー入力の利用頻度が増えたことによる．キー入力によるショートカットは覚えにくいため，第 1 セッションのような利用し始めてまもない状態では使いこなせなかったが，練習とともに利用が増えた．

しかし 2 カ月の休止期間を置いた後の第 3 セッションでは，音声利用の効果，利用入力手段の比率ともにほぼ第 1 セッションと同じ傾向に戻った．これは，キーによるショートカットはその利用法を覚えていることが難しいためである．

アンケートにおいて，どの項目を最も重要と考えるかという問いに対し，第 1，第 3 のセッションでは約 50% の被験者が「やりたい操作を実現する方法の分かりやすさ（項目 2)」を選ん

だ．使い始めて間もない場合や休止期間が入った場合に音声利用の効果が特に大きいのは，このような傾向のある非熟練者に対し，音声が実現する手順連想の容易性が受け入れられたためと考える．

2.6.5　熟練者における音声の利用の効果

熟練に伴って効率の側面における音声利用の効果は目立たなくなることを前述したが，このことは，熟練者にとって音声は価値がないということを示すものではない．図 2.4 が示すように，熟練が進んだ段階においても，色の変更に代表される選択肢が多く言葉で表現しやすい属性の指定などにはほとんど音声が利用された．このように，作業効率の向上や軽快感の改善に結び付く音声の利用は，ユーザの熟練が進んだ後も強く支持された．第 2 セッションにおいては，アンケート項目の中で一番重要なものを「疲労感の改善」とした被験者が 50% おり，これに結び付く音声の利用は熟練者にも支持された．

2.7　評価手法に関する考察

本研究で用いたインタフェース評価手法については，以下のような課題が残されている．

本研究は，音声入力とポインティングデバイスの比較のために「手や指を動かす距離の累積」「コマンドを入力した回数」を使用した．具体的には，モード切替のコマンドを音声入力で行うことにより，すべてをマウス操作で行った場合と比較して，マウスポインタの移動量を削減できたことが，本研究の主張である．

しかしこの議論にはいくつか検討の余地がある．まず「画面上のマウスポインタの移動距離」と「マウスを持つ手の移動距離」が同じであるという仮定を検証しなくてはならない．これは，本実験で用いられたワークステーションのマウス環境においてはほぼ問題なく成立していたが，常に成り立っているとは限らない．特に，マウスではなくタブレット操作やタッチ操作を対象にする場合は，身体的な負荷の評価のために，画面上のポインタの移動距離に依存しない方法が必要となる．

また，マルチモーダルインタフェースにおいて，マウスの移動量だけを負荷と見なすのは妥当ではなく，「キーボードでショートカット操作を行う場合の指の動き」も検討する必要がある．さらに，「音声コマンドを発話することの身体的・物理的な負荷」も無視できない．声の大きさや発声方法も考慮すべきである．

2.8 まとめ

優れたインタフェースが満たすべき原則に基づいて，音声入力のマウス，キーボードとの協調のあり方を考えた．その設計指針に基づいて音声作図システム S-tgif を作成し評価を行った．

一般に，音声の利用が操作の軽快感の改善，操作のわかりやすさ，操作に伴う疲労感などの改善に役立つことを検証できた．

特に利用し始めて間もない時期においては，音声を利用することで音声を利用しない場合に比べ，課題完了までの操作時間を 79.6% に減じることができ，作業効率の改善に貢献することがわかった．

一方システムの利用に熟練するとともに，ユーザは他のより効率的な入力手段の利用に慣れ，音声利用の効果は薄れることがわかった．しかしながら，特定のコマンドでは音声の利用率が常に 90% を越え，また主観評価でも高い得点を得るなど，音声の利用はユーザから支持された．

また，システムの利用を中断すると，手順連想容易性が劣る他の入力手段の利用は減り，再び音声利用の効果が上がることを確認した．

これらの結果は，音声の持つ操作性と手順連想容易性とが効果的に機能した結果と考えることができる．他の音声応用を考えるにあたっても，ここで行ったように，インタフェースの原則を参考にしながら音声の特性を生かし，また欠点を補う機構を併用することによって，優れたインタフェースシステムを実現できる．

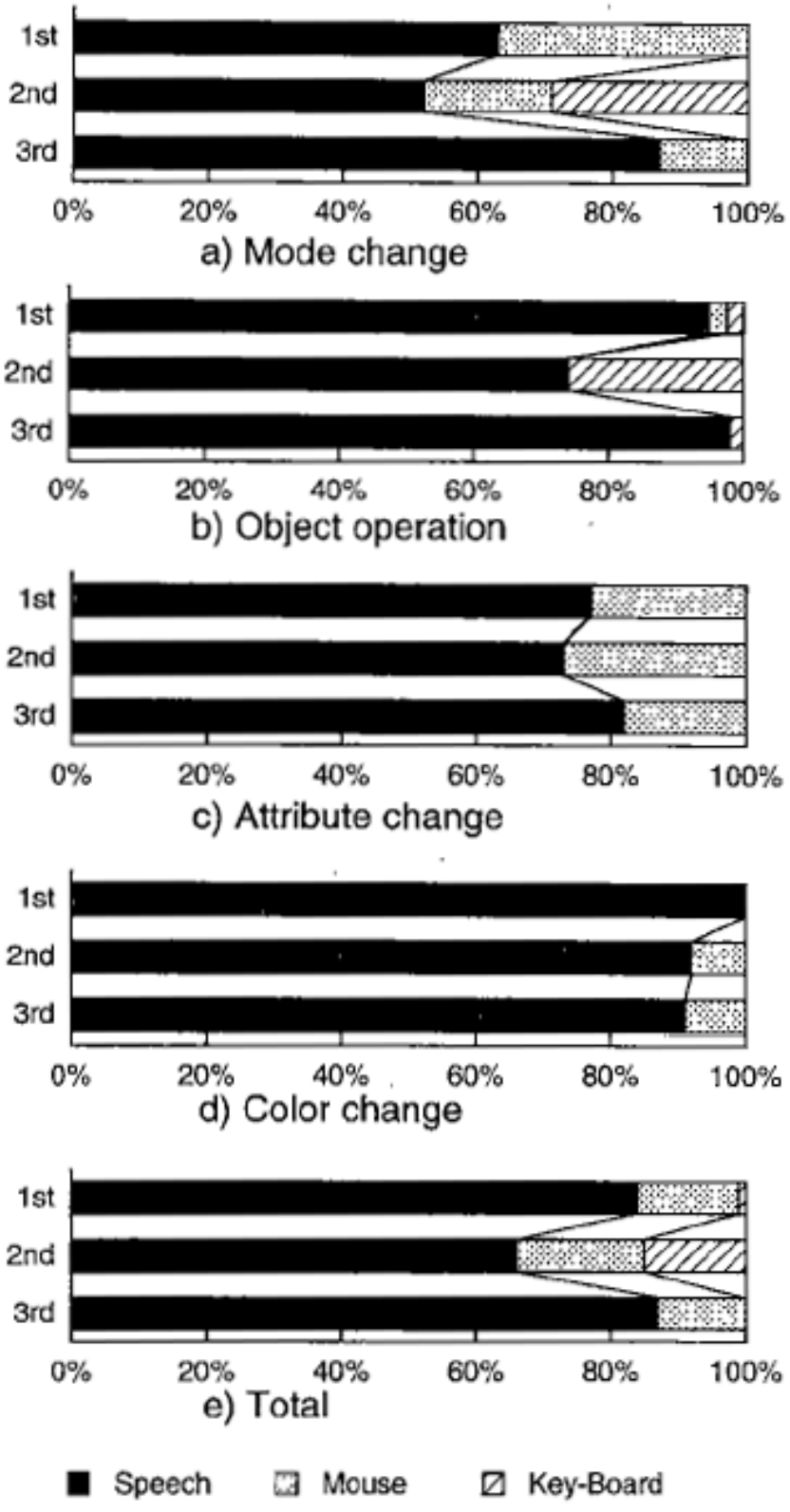

図 2.4　コマンド入力における各入力手段の利用比率.

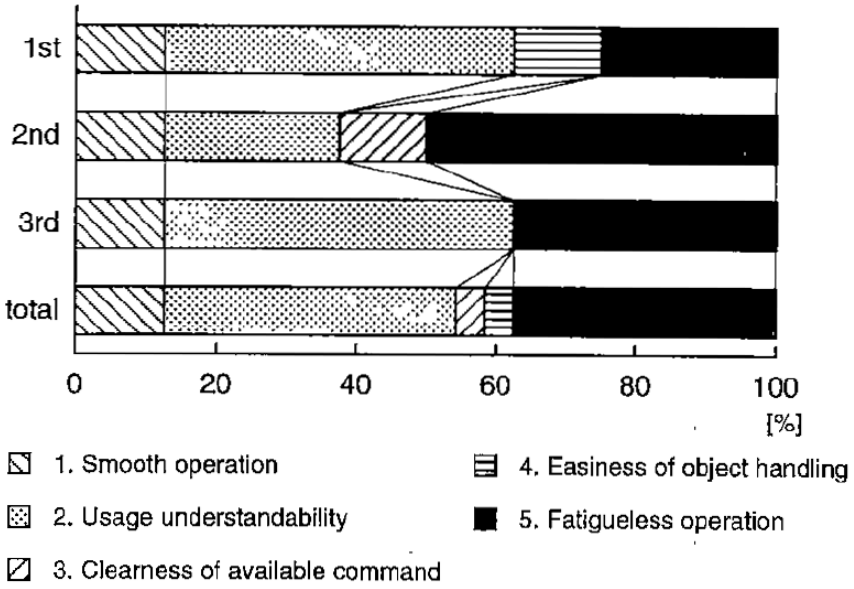

図 2.5　S-tgif アンケート「最も重要な項目はどれか」結果.

第 3 章

非同期型音声会議システムの設計と評価

3.1 はじめに

時間を同じくしなくても，複数の人間が相互に発言し，コミュニケーションを行う手段として，電子メールや電子掲示板などの非同期・蓄積型メディアがある．この種のメディアには，時間的拘束がない，メッセージが常に保存される，時間をかけて返答を作成できる，複数の相手への同報が可能である，などの利点がある．これらの利点は多様な生活習慣を持つ多数のメンバーがコミュニティを形成することを支援する．また，蓄積された情報を検索することによって新たな仲間を探すことも容易になる．インターネットの普及は，このようなコミュニティ形成機能によってもたらされたとさえ言える．

一方で，コンピュータの高性能化やマルチメディア符号化技術の進歩によって，ディジタル化された音声や画像による通信が可能となっている．特にインターネットにおける音声メディアの応用を表 3.1 のように，リアルタイム型と非同期・蓄積型，片方向通信と双方向通信，という観点で分類すると，コミュニティ形成機能を実現するためには，非同期・蓄積型の双方向通信が必要とされる．しかしこれまでに，リアルタイム型の片方向通信（インターネットラジオなど），リアルタイム型の双方向通信（音声チャットやインターネット電話など），非同期・蓄積型の片方向通信（音楽配信など），といった応用は提案されているが，非同期・蓄積型の双方向通信に関する応用はほとんど進んでいない．

非同期・蓄積型の音声メディアを対話的に用いるための提案としては Hyper-Audio[20] などがある．また，音声や動画などの配置や同期，リンク情報を表すインターネット標準規格として SMIL [*1]が広く利用されている．しかしこれらはコンテンツの選択を対話的にしているが，発言

[*1] Synchronized Multimedia, http://www.w3.org/AudioVideo/ なお SMIL は読書に障害がある人のための電子書籍フォーマット DAISY (Digital Accessible Information SYstem) の要素技術としても活用されてい

表 3.1　インターネットにおける音声メディアの応用.

	Realtime	Asynchronous
1way	Internet Radio	Radio Archive Music Archive
2way	Voice Chat Internet Phone	(not available)

そのものの双方向化を実現するものではない.

コミュニティ形成機能を音声によって実現できれば，使いやすくモバイル環境に適したサービスが実現できる．幅広いユーザによる利用が期待できるため，高齢者や視覚障害者の社会参加の支援や，社会のディジタル格差（インターネットの接続手段や利用技術の有無による格差）の解消などに貢献できる．また，肉声を使うことにより本人の発言であることを確認しやすくなるため，発言内容の信頼性も高まる．このため，文字メッセージと比較してセキュリティ面でも有利となる.

そこで本研究では，非同期的な音声会議を効率的に実現するためのインタフェースを新たに提案し，クライアント・サーバ型の会議システム AVM の作成・評価を行う．特に音声メッセージの蓄積方法，録音および再生方法，文字による発言の視覚化などに関して詳細に検討し，評価実験によってその有効性を検討する.

3.2　音声会話の漸次性と相槌

メールやウェブ情報の音声リーダーソフトや，音声認識機能を統合した電子メールソフトはすでに製品化されている．これらによって，コンピュータに不慣れなユーザや視覚障害者などが容易に通信を行える環境が整備されつつある．しかし，文字による通信手段に音声インタフェースを付加したシステムでは，入力された音声が含んでいた声質や感情などの非言語情報が欠落して伝わってしまう．これは，音声による豊かなコミュニケーションの可能性を切り捨てていると言える．そこで本研究では，例えば音声認識機能は発言の閲覧や検索のための補助的手段として用い，発言そのものは肉声で録音・再生することを前提とする.

また，書き言葉と話し言葉の違いにも考慮する必要がある．日常の会話や電話などで用いられる話し言葉には漸次性があり，相手の知らない情報だけを伝えたり，発話の断片を思いつくまま次々に伝えたりすることが多い．この漸次性が，話し手と聞き手で共有されている知識や情報を省略したり，統語構造が簡単で認知的な負荷の小さい表現を可能にしている [21]．メッセー

る.

ジシステムにおいて音声が有効に利用されるためには，このような漸次的発話を許容する設計が重要となる．

非同期・蓄積型でありながら擬似的にリアルタイムの会話を実現するためには，漸次的な会話を促すための配慮が必要である．そこで本研究では，発話にオーバーラップする他の話者の発話や相槌などの現象に注目する．

過去の対話研究では，一人の話者が話し終わる前に次の話者が話し出す，という現象は稀有であるとして例外視されてきた．しかし，我々はこれまでに，RWC プロジェクトで収録された対話について検討を行い，応答発話の過半数でオーバラップ現象が見られ，これにより 2 話者の発話時間を合計した時間の 13% が節約されていたことを確認した [22]．また榎本ら [23] も，地図課題対話コーパス中にオーバラップ発話が全発話中の 45% にも昇っていると報告している．川口ら [24] は，Grice の会話の含みの理論に基づいて，「相手が何を言おうとしたかを推論し理解した」ことを示すものとしてオーバラップ現象の一部を説明している．

オーバラップ発話の中で，特に相槌の生成に注目した研究もある．最も簡単な相槌の生成方法は，声の出されていない無音時間が一定の長さ以上続いた場合に相槌を発する，というものである [25, 26]． また，ピッチパターンに基づいた相槌生成モデル [27] も提案されている．音声対話における相槌の役割という観点からは，話者交替におよぼす影響の検討 [28, 29] や確認の役割に関する検討 [30, 31] などがなされている．また，実時間性の高い対話制御によって相槌を生成する音声対話システム [32] が構築されている．

これらの先行研究を踏まえ，本研究では，オーバラップ発話を例外ではなく一般的な現象としてモデル化する．また，相槌が何らかの役割を持つ，という観点からのインタフェース設計を行う．

3.3 音声メッセージの相互編集機能

文字メッセージとの比較において，音声メッセージは，高い現実感や豊富なパラ言語情報を持ち，また，キーボード操作などを用いず容易に入力することができる，という利点がある．その一方で音声は，多くの情報の中から欲しい項目を流し読みして探すことや，必要な部分を引用したり加工したりする編集行為が困難である．ここでは，音声のこのような問題点を補う方法として二つの提案を行う．

第一の提案は，音声メッセージの検索や流し読みを実現するために音声認識を用いる，というものである．システムの見かけの機能は，音声で録音されたメッセージを音声で再生する，というものに限定する一方で，音声メッセージがあたかも文字情報であるかのように扱えるようにする．これにより，音声メッセージの持つさまざまな利点と，文字メッセージの扱いやすさを両立させることを目指す．

さらに，我々はこの提案を通じて，たとえ音声認識の性能が完璧でなくても，何らかの実用的

な応用が可能である，ということを実証したいと考えている．本提案では音声認識の結果を最終目的としてユーザに与えるのではなく，聞きたい音声を選ぶための手段として利用できればよい．元の音声を聞けば内容は理解できるため，誤認識があっても実用性が大きく損なわれることはない．既存の技術によって「いますぐ使える」と感じさせる応用システムを提供することによって，音声応用システムのユーザに対する啓蒙や市場開拓が進むだろう．

第二の提案は，非同期・蓄積型メディアによる双方向的な議論は発言の相互編集行為なしには不可能である，という立場から，テキスト編集とは異なる発想によって同等の機能を提供する，というものである．例えば，電子メールや電子掲示板では，メッセージを読み，その一部を引用し，それにコメントをつける，といった操作を容易に行うことができる．ここで行われている操作の機能は，(a) どのメッセージに対する返答であるかを示すこと，(b) そのメッセージ内でどの部分に注目してるかを示すこと，に大別されると考えられる．ツリー表示によるメッセージの操作は (a) を実現するためのインタフェースであり，発言に「⟩⟩」などの文字を付与して部分引用するのは (b) を実現するためのインタフェースである．これらの機能はいわば発言の相互編集機能である．この機能こそが，非同期型メディアによる双方向的な議論や雑談を可能にしている．

文字による発言を編集するもっとも簡単な手段は，カット＆ペースト機能などを含む，コンピュータのテキスト編集機能である．しかし，音声会話において同等の機能は，簡便なインタフェースでは実現できない．そこで本研究では，前章での検討を踏まえ，オーバラップ発話とそのタイミング情報に積極的な意味を持たせる，新たな操作体系を提案する．つまり，音声メッセージの再生中に，自由なタイミングでの割り込み (barge-in) を許すこととし，その割り込みが，(a) どの発言に対して行われたか，(b) 発言のどの部分の再生中に行われたか，という情報を相互編集機能の代替として使用する．この操作体系は，我々が日常行っている自然会話から類推しやすいインタフェースであると考えられる．また，メッセージの関連付け操作をユーザに委ねているため，システム側がメッセージ内容を理解する必要がない．基本的に音声入出力以外のデバイスを必要としないため，電話やモバイル環境に適したインタフェースとなることも期待できる．

3.4 AVM システムの設計

前述の議論をふまえて設計された非同期型音声会議システム AVM（Asynchronous Voice Meeting）は以下の通りである．

3.4.1 サーバ・クライアント構成

AVM システムはメッセージサーバとクライアントから構成される．ユーザはクライアントを用いてメッセージの録音を行い，録音されたメッセージはサーバに集中的に蓄積される．またサーバは，クライアントからの要求に応じて，複数のメッセージを一つの連続した音声ファイルに編集してクライアントに送信する．この編集は動的に実行されるものであり，サーバには常に，クライアントで録音された個々のメッセージがそのままの形で蓄積される．

3.4.2 利用方法の流れ

AVM システムのユーザから見た操作の流れは次のようになる．

1. 参加したいグループを選んで，過去に発言されたメッセージの一覧を取得する．
2. メッセージの一覧から特定のメッセージを選択して，音声を取得する操作を行う．
3. 取得された音声を再生しながら，それにオーバラップするように返答を発声し，録音する．
4. 返答音声を聞き返し，録音を取り消すならば（3）に戻る．
5. 録音された返答音声をサーバに登録する．

3.4.3 通信プロトコルとデータ構造

AVM システムでは XML[*2] 準拠の情報ファイルである AVML ファイルと音声ファイルとがサーバ・クライアント間で転送される．音声ファイルは非圧縮のフォーマット（11.025KHz サンプリング，16bit 量子化，モノラル）を用いているが，効率的な転送のために圧縮を施すことも可能である．

（1）通信プロトコル

ファイル送受信に使用するプロトコルとしては WWW で用いられる HTTP[*3] の GET および PUT メソッドを流用した．URL の path 部分を用いてグループ名とデータを表現する．例えばグループ名 room1 の AVML 情報を指定する場合には /room1/text/avml が path として用いられる．また，メッセージ ID を指定するために path の末尾に ?index=1,2,3 といった形式の文字列が付加される．

（2）サーバが生成する AVML 情報

サーバが生成する AVML 情報は，グループ内のメッセージ一覧として単独で使用される．また，音声ファイルを編集してクライアントに送信する際には，その音声に付随する属性情報とし

[*2] Extensible Markup Language, http://www.w3.org/XML/

[*3] Hypertext Transfer Protocol HTTP/1.1, RFC2068.

```
<?xml version="1.0" ?>
<avml>
  <segment mesid="1" sender="nishi" playtime="0"
     mestime="0" length="1.5" indent="0">
    <text mesid="1" begin="0.0" end="1.01">
      Good morning!
    </text>
  </segment>
</avml>
```

図 3.1　サーバが生成する AVML 情報の例.

ても使用される.

1 つのファイルは複数の segment エンティティによって構成される．segment は 1 つの音声ファイルを時系列上のいくつかの音声区間に分割したものであり，segment エンティティは，その音声区間がどのメッセージのどの部分から生成されたかを示す．segment エンティティの属性には，mesid（元メッセージの ID），sender（元メッセージの発言者），mestime（segment 先頭に対応する元メッセージ上の位置），playtime（segment 先頭に対応する音声ファイル上の位置），length（segment の長さ），indent（ツリー表示時の階層の深さ），がある．また，segment エンティティはメッセージ自身の内容や他の話者の相槌などを表す text エンティティを含むことができる．サーバが生成する AVML 情報の例を図 3.1 に示す.

（3）クライアントが生成する AVML 情報

クライアントが生成しサーバに送信する AVML 情報は，発言された音声をサーバに登録する際に用いられるものであり，音声区間として切り出された 1 つの範囲が 1 つの message エンティティに対応する．その音声が録音されたときに再生されていたメッセージ（親メッセージ）に関する情報を音声に付与するのが目的である．message エンティティの属性には，parent（親メッセージの ID），reltime（新規メッセージ先頭に対応する親メッセージ上の位置），length（新規メッセージの長さ），overlap（新規メッセージが相槌であるか否か），がある．overlap 属性については 4.5 節で詳しく述べる．クライアントが生成する AVML 情報の例を図 3.2 に示す.

3.4.4　メッセージの録音と関連付け

AVM クライアントにおいては，既存のメッセージを再生しながら，全二重的に新規メッセージが録音される．録音された音声は始終端検出によって無音部分が除去される．また，再生中のメッセージの segment 情報に基づいて，既存メッセージとの相対的な時間関係が新規メッセー

```
<?xml version="1.0" ?>
<avml sender="canny">
  <message parent="2" reltime="0.4" length="0.3"
     overlap="1">
    <text begin="0" end="0.3">
      Yes.
    </text>
  </message>
</avml>
```

図 3.2　クライアントが生成する AVML 情報の例.

表 3.2　メッセージの関連付けデータの構造.

フィールド名	内容
mesid	メッセージ ID
length	メッセージの長さ（秒）
parent	親メッセージ ID
offset	親メッセージとの 開始時間の相対位置（秒）
memberid	発言者 ID
wavefile	音声ファイル名
overlap	オーバラップ属性
date	メッセージ登録日時

ジに付与される．segment 情報は，サーバで編集された音声の特定の区間が元のメッセージのどの位置に相当するかを示しているため，録音されたメッセージがサーバに登録される際には，サーバに登録されている元メッセージと，新規にサーバに登録されたメッセージとの相対的な時間関係を保存できる．サーバ上でのメッセージの関連付けデータの構造を表 3.2 に示す.

3.4.5　BISP 機能

本システムの予備的な実装による実験 [33] を行ったところ，音声メッセージを再生しながら新規メッセージを録音する際に，再生されている音声を一時停止するか否かを適切に制御する必要があることが明らかになった．つまり，既存メッセージの再生を行いながら長い発話の録

音を行うと，録音中にシステムが再生してしまった既存メッセージの内容をユーザが把握することができず，再度既存メッセージを聞き返さなくてはならなくなる．しかし，録音すべき音声区間において常に既存メッセージの再生を止めてしまうと，たとえ相槌であっても必ず再生が止まってしまうため，相槌を打たない方がメッセージを聞きやすい，という事態が生じる．

また，サーバでのメッセージ編集においては，発話時間の長いメッセージ同士が重なって再生されると内容を聞き取ることが難しくなる．このため，長いメッセージに関しては，親メッセージにオーバラップせずに挿入するような編集が望ましい．しかし相槌などは親メッセージにオーバラップするような編集をしたほうが，再現される会話の自然性が高まる．

そこで，ある発話区間が単なる相槌（相槌メッセージ）であるか，内容的に意味を持つ発言（非相槌メッセージ）であるかをオーバラップ属性によって区別し，サーバに登録することとする．発言内容を再現する場合には，オーバラップ属性に応じて，相槌であれば親メッセージと重ね合わせ，非相槌であれば親メッセージに割り込ませる形でメッセージを編集し，仮想的な対話音声を再現する．このようなサーバ側の処理は次のアルゴリズムで実現される．

1. メッセージ ID のリストをクライアントから受け取り，そのリスト中からルート（根）となるメッセージを検索する．
2. ルートメッセージの子となるメッセージを検索し，非相槌メッセージのみを親メッセージの対応する位置に再帰的に挿入する（オーバラップさせない）．このとき，対応する segment 情報も同時に作成する．
3. 全ての非相槌メッセージの挿入を繰り返して作られた音声に対して，相槌メッセージの重ね合わせ（オーバラップ）を行う．相槌メッセージはそれぞれ親メッセージとの相対時間によって管理されているので，segment 情報を用いて重ね合わせる場所を決定する．

また，クライアント側での新規発言の録音においては，始終端検出と同時に，発話長によって相槌であるか非相槌であるかを簡易的に検出するようにした．新規発話の発話長が短ければ相槌とみなし，再生中の音声は途切れることなく，単に録音だけが行われる．しかし，新規発話が一定の長さ（現在の実装では 1.0 秒）を超えると，新規発話の終端を検出するまで既存発話の再生を中断し，この新規発話を非相槌として保存する．この機能を BISP（Barge-in to Stop Playing）と呼ぶ．これにより，任意のタイミングで自由に発話した相槌を再現できると同時に，再生中に新たに長い発話を行っても，再生中の既存発話の内容を聞き漏らすことがなくなった．この機能を実現したクライアントの状態遷移図を図 3.3 に示す．

3.4.6 既読管理機能

1 つのグループに多くのメッセージが蓄積されていくと，ユーザがすでに聞いた発言がどれであるかを把握することが困難になる．これを解消するために，サーバ側でユーザごとに既読

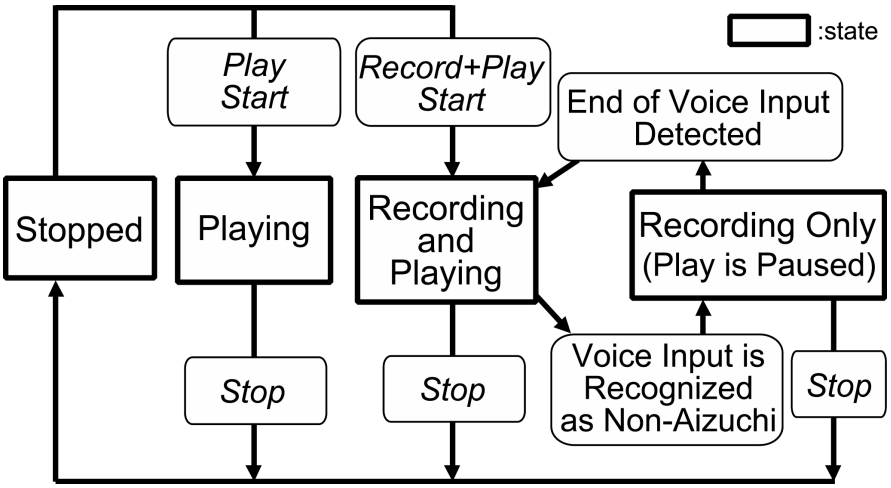

図 3.3　クライアントの状態遷移図.

メッセージの ID を管理し，クライアントでの一覧表示時に既読メッセージをマークで示す機能を設けた.

3.5　評価システムの構成

前述した設計に基づいて，以下のような評価システムを構築した.

3.5.1　サーバ

AVM サーバ Voxer は移植性を考慮して Perl および C 言語によって実装されており，HTTP によるクライアントからの要求を処理する．音声ファイルとメッセージ間の関連付けなどの情報を蓄積するデータベース機能と，再生用音声の編集機能を備える．Linux 上でも動作するが，後述する実験では Windows 上で運用した．音声認識機能の実装も行っているが，本実験では使用していない.

3.5.2　クライアント

AVM クライアント Voyager は Microsoft Visual C++ を用いて実装され，全二重で音声入出力が可能な Windows98 システム上で動作する．HTTP によるメッセージの送受信機能と，BISP 機能を含む音声の録音・再生機能，音声メッセージのツリー表示機能などを備える．再生中の音声がそのまま全二重で録音されることを防ぐために，メッセージの録音と再生にはヘッドセットを用いる．Voyager の画面表示を図 3.4 に示す．左側はツリー表示によるメッセージの選択ウィンドウで，右側は音声認識によってつけられた text エンティティ（4.3 節参照）に基

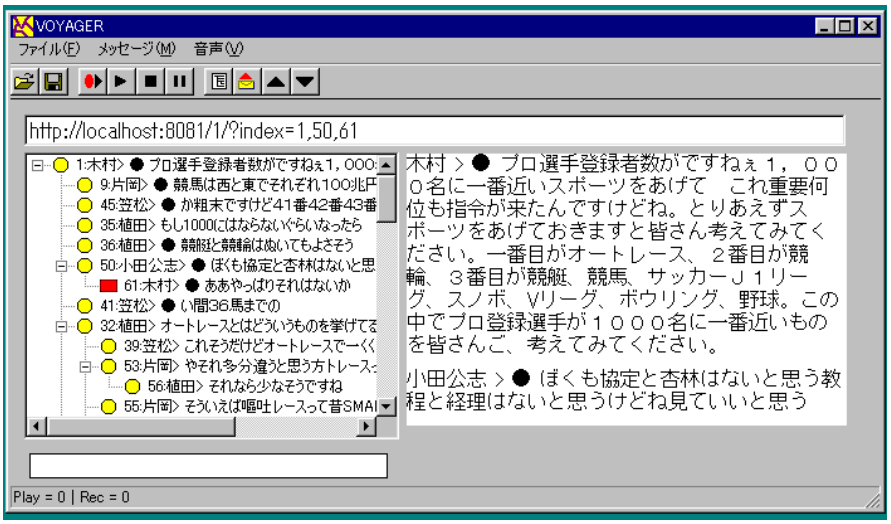

図 3.4　クライアントの画面表示.

づいてメッセージの内容を表示する部分である．上部ツールバーには「再生＋録音」「再生のみ」「停止」などのボタンがあり，下部にはマイク音量が表示される．

3.6　実験

3.6.1　実験方法

議題として「次にあげるスポーツの中で，プロ登録者数が 1000 人に近いスポーツを挙げて下さい．オートレース，競艇，競輪，騎手（中央競馬会），サッカー J1 リーグ，スノーボード，V リーグ，ボウリング，野球.」というクイズを与えて，AVM システムまたは電子掲示板（BBS）を用いて議論をさせた．

このような設問においては，すべての項目について深い知識を持っている参加者はいないが，参加者がそれぞれの知識を相互補完的に提供し合うことが可能であり，双方向的な議論によって正解に近づくことが期待できる．議論の結果が正解に近いかどうかによって活発な議論が行えたかどうかを判断できると同時に，議論が収束するまでのシステム利用回数や発言回数などを定量的に評価することもできる．このような観点から本実験を計画した．

被験者たちが真剣に議論することを促すために，1000 人により近いスポーツ名を回答したチームには報酬を多く支払うことを予告した．

AVM システムとしてはクライアントとサーバを 1 台の Windows98 搭載 PC で実行させた．BBS 用のソフトウェアとしては WWW サーバ上で動作し，発言をツリー構造で管理できる

表 3.3　AVM および BBS におけるメッセージの分析結果.

	AVM	BBS
システムののべ使用回数	20	24
メッセージ数	71	24
総発言文字数（引用を除く）	2303	5657
メッセージの平均文字数	32.4	235.7
総発言時間（秒）	501.3	–
平均発言時間（秒）	7.1	–
オーバラップ発話数	12	–
非オーバラップ発話数	59	–

WebForum[*4]を用いた.

被験者は理工系の大学研究室に所属する 20 代男性の学生（音声の研究に従事する学生を含む）10 名であり，全員がキーボード操作に熟練している．この 10 名が 5 名ずつ 2 チームに分かれて，AVM と BBS を各 1 チームが用いて実験を行なった．各チームから議長を 1 名ずつ選出し，まず議長からチームのメンバーに対して，各システムを用いてクイズの問題を通知させた．以後は乱数により 1 名ずつユーザを選び，各システムを交替で使用させた.

AVM の音声認識機能に関しては，実用的な音声認識性能を保ちつつ，ある程度の誤認識を含んだテキストを得るために，ユーザが使い終わるたびに，新たに発言された音声をオペレータが ViaVoice98（日本 IBM 社製の音声認識ソフト）に対して再度読み上げて，認識結果を AVM サーバに登録することとした.

議論によってチーム全員による結論が得られたら，議長がオペレータに口頭で回答することとした．最後に各被験者に対してアンケートを行った.

3.6.2　実験結果

議論の結果，どちらのチームも正解または正解に近い回答が得られた（勝者は AVM チームであった）．発言の定量的分析を表 3.3 に，メッセージあたりの文字数の比較を図 3.5 に示す．AVM におけるメッセージの文字数は，実験終了後に音声を再度手作業で書き起こしたテキストを用いて求めた．また，BBS の発言においては引用部分を除いて文字数を求めた.

ViaVoice98 によるメッセージの認識性能は，単語認識率 85.2%，認識精度 82.5% であった.

各チームのアンケート結果（1〜5 の 5 段階評価，1 が「まったくそう思わない」，5 が「とて

[*4] http://www.kent-web.com

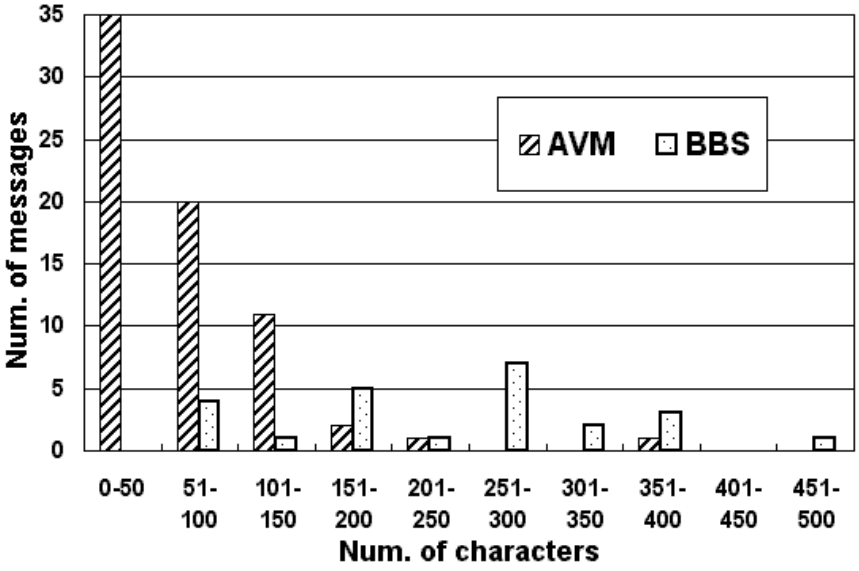

図 3.5　メッセージの文字数の比較.

もそう思う」に対応）の平均値を表 3.4 に示す．ただし項目（k）（l）は AVM のみで行った質問である.

3.6.3　検討

ここでは,「AVM によって話し言葉的な音声会話を非同期的に実現できた（仮説 1)」「AVM において話し言葉の漸次性が生かされた（仮説 2)」「BISP が有効に機能し，発言しやすさに貢献した（仮説 3)」「誤認識を含んだ認識結果が有効に活用できた（仮説 4)」「AVM は BBS の代替として充分に機能した（仮説 5)」という仮説を挙げ，実験結果がこれらを支持したかという観点から検討を行う.

まず仮説 1 について検討する．AVM では，例えば「サッカー言うたら今ー，20 チームぐらい．で 1000 人やとしたら 1 チーム 50 人．そんなにおらんでしょう.」といった口語的でくだけた表現が多用された．これに対して BBS では,「クイズの僕なりの解釈ですが，みんなの言われているように野球は 1,2 軍あわせると 1000 ÷ 12=84 人以上いるような気がします．しかし，1 球団で選手登録人数の上限が決められていますので……」といった書き言葉による表現が多用された．被験者はすでに互いに親しい間柄であるため，グループによる差異は考えにくい．従ってこの実験結果は仮説を支持している.

仮説 2 については，これが成立していれば，第 2 章で述べたように，情報が省略されたり，簡潔で断片的な表現が多用されたはずである．AVM によって総発言文字数が 48% になったこと（表 3.3）や，各メッセージの文字数が少ない値に集中したこと（図 3.5),「あー近いといえば近

表 3.4　AVM アンケート結果の平均.

質問内容	AVM	BBS
(a) 納得できる答えを得られた	4.4	3.6
(b) 自分の意見を十分に言えた	4.0	4.4
(c) 活発な議論が行えた	3.8	3.4
(d) 議論の流れに不自然さがなかった	4.0	4.0
(e) 実際に集合して交わす議論と雰囲気が近い	2.2	2.8
(f) 反論・同意といった意見を出しやすい	4.4	3.8
(g) メッセージを入力しやすい	3.4	4.2
(h) メッセージ内容の把握がしやすい	3.4	4.0
(i) 会議全体の流れを把握しやすい	3.4	3.8
(j) ユーザインタフェースがわかりやすい	3.8	4.0
(k) 音声再生よりも文字表示をよく利用した	4.0	–
(l) 誤認識を含むテキストの利用価値があった	4.6	–

いですよねー」といった文脈に依存した発言が多くみられたことなどが，この仮説の正しさを裏付けている．

仮説 3 については，発言の 17% を占めるオーバラップ発話を聴取したところ，発話区間の検出に失敗して断片化されたものが多かった．これは表 3.4 の項目 (g) の評価の低さとも関連しており，オーバラップ発話の処理が不充分だったことを示す．しかし，表 3.4 の項目 (a) 〜 (d) の評価は高く，本実験の用途での発言しやすさは実現されていたと言える．

仮説 4 については，表 3.4 の項目 (k) (l) の評価が高得点であることと，アンケートの自由記述における「大まかな内容をテキストで確認し，詳しい内容は音声を再生し確認した」「(文字情報があれば) いろんな箇所の声を聞くより手間が少なくなる」などの回答によって支持されたと考えられる．

仮説 5 については，表 3.4 で BBS と比較して AVM が著しく劣っている項目がなかったこと，議論の結果どちらも適切な答えが得られたこと，などにより支持されたと考えられる．ただし表 3.4 の項目 (e) の評価が BBS と同様に低かったことから，対面での議論の代替として AVM を位置付けることは難しい．

以上より，本実験によって仮説 1,2,4,5 は全面的に，仮説 3 は部分的に支持されたと考えられる．

3.7 まとめ

非同期・蓄積型で双方向通信が可能な音声会議システム AVM を提案し，その設計と評価について述べた．電子メールや電子掲示板などが持つ非同期型通信の利点を損なわずに，表現力が高く発言しやすい，という音声メッセージの利点を生かせたことが確認できた．

今後の課題としては，サーバに実装中の音声認識について，特に話し言葉での性能向上を目指す必要がある．また，音声のサーバ側での編集方法に関して，既読発言の再生を省略したり，メッセージが単語の途中で分割されることを防ぐなど，より会話しやすくするための改良が必要である．また，キーボード操作に不慣れなユーザなど，多様なユーザによる大規模な運用実験を行う必要がある．さらに，電話回線や携帯情報端末による AVM の実現方法についても検討し，本システムが幅広いユーザに利用されるようにしていくことが望まれる．

第 4 章

音声インタフェースの負荷測定法の検討

4.1 はじめに

近年，音声合成や音声認識の技術が成熟したことにより音声インタフェース（音声による機械と人間のインタフェース）が実現され，さまざまな音声対話システムも実用化されている．音声による情報取得や機器操作の利点には，キー操作やポインティング操作が不要であり非デスクトップ環境で利用しやすいこと（ハンズフリー），ディスプレイに視覚的な注意を払わずとも使用できること（アイズフリー），などがある．

しかし，音声のみを用いたインタフェースでは，長時間喋ることに起因する疲労に加えて，「何を喋ったらよいのか」「発話が正しく認識されたか」といったことがわかりにくく，心理的にも疲れやすい場合がある．これらは音声合成性能（正聴率など）や音声認識性能（認識率や応答速度など）の低さにも起因するが，インタフェース設計（対話パターンや入力語彙・応答文などの設計）の不適切さも大きな原因である．

本研究では，音声対話におけるユーザのさまざまな負荷を総称して対話負荷と呼ぶ．対話負荷には，発話動作などによる物理的負荷と，覚える，探す，推論する，注意する，などの心的な行為による認知的負荷 (cognitive load) の両方が含まれると考える．対話負荷を測定するためには，タスク達成に要したユーザ発話回数などの物理的負荷を比較するだけでは十分でなく，ユーザがシステムにどれだけ注意をしていたか，といった認知的負荷も考慮した尺度が必要となる．対話負荷の尺度は，ユーザにとっての使いやすさや，別のクロスモーダルな作業を同時に行なうための余裕度を反映すると考えられ，例えば歩行中や自動車の運転中などに使用する音声インタフェースとしての適切さの指標ともなる．

音声対話システムを使いやすくするためには，インタフェース設計はシステム開発の初期段階で行ない，実際のユーザによる評価を早い段階から繰り返しながら改良することが重要だと

される（例えば [34, 35]）．その際に，ユーザの内観的な報告に頼らずに対話負荷を計測し，対話負荷の大きい箇所を特定できる手法があれば，効率的に評価や改良を行なえる．本研究ではこのような用途を想定して，音声インタフェースにおける対話負荷の測定方法を提案し，その有効性を検討する．

ある対象にどれだけ注意をしていたかを測定するために，二重課題法が広く用いられている [36, 37, 38] ．人が一つの課題を行っている間に別の課題が加わると，心理的な折り合いをつける必要が生じる．人は心的負荷を複数の課題に配分することができても，費やすエネルギーを一定以上に増やすことができない，とされる（「キャパシティ一定の法則」）．心的エネルギーの限界を超えた場合は，課題に対する応答が遅れたり誤りが増加したりする．

4.2 関連研究

音声インタフェースの負荷測定において，二重課題法を用いる場合には，例えば，音声インタフェースの利用を第一課題とし，ゲームなどの客観的に負荷が測定可能な第二課題を被験者に並行して行わせ，第二課題の成績が低い状況において音声インタフェースの負荷が大きい，と判断すればよい（図 4.1）．

この目的のためには，「キャパシティ一定の法則」が成立すること，被験者が意図的に第一課題を第二課題に対して優先できること，第二課題の成績差が比較しやすく信頼性の高い値として得られること，が必要である．

複数の同時課題を用いて音声インタフェースを評価する試みには，自動車の運転中を想定したいくつかの研究がある．

小島ら [39, 40] は，実車の運転，LED 刺激への反応，単語の記憶復唱の三重課題を被験者に与えて，記憶課題の有無が刺激反応タスクの成績に関連することを示した（平均反応時間の有意差は確認されていないが，反応時間が長い試行の割合は有意に変化した）．この実験は音声対話を用いることの影響は測定しているが，どのような対話がより大きな影響を与えるか，といった測定はなされていない．

清水らの関連研究 [41] では，実車の運転，LED 刺激への反応，音声対話の三重課題を被験者に与えて，車載用音声対話システムの安全性を評価している．しかし，刺激反応タスクの遅れ割合について，対話パターンの違いによる有意差は得られていない．この実験では，巡回コースを運転しながら刺激反応タスクと音声対話（交通情報検索）タスクを同時に繰り返し行っているが，第二課題に相当する課題が運転と刺激反応の 2 つであり，被験者が課題の優先度を意識すること（運転をおろそかにして刺激反応に集中するのか，刺激反応をおろそかにして運転に集中するのか）や，2 つの課題を加味した成績評価を行なうことが難しい．

Strayer ら [42] は二重課題法を用いて携帯電話での会話が運転に与える影響を調べている．被験者は，模擬運転課題（3 種類の正弦波を加算したコースにおけるハンドル操作）を行いながら

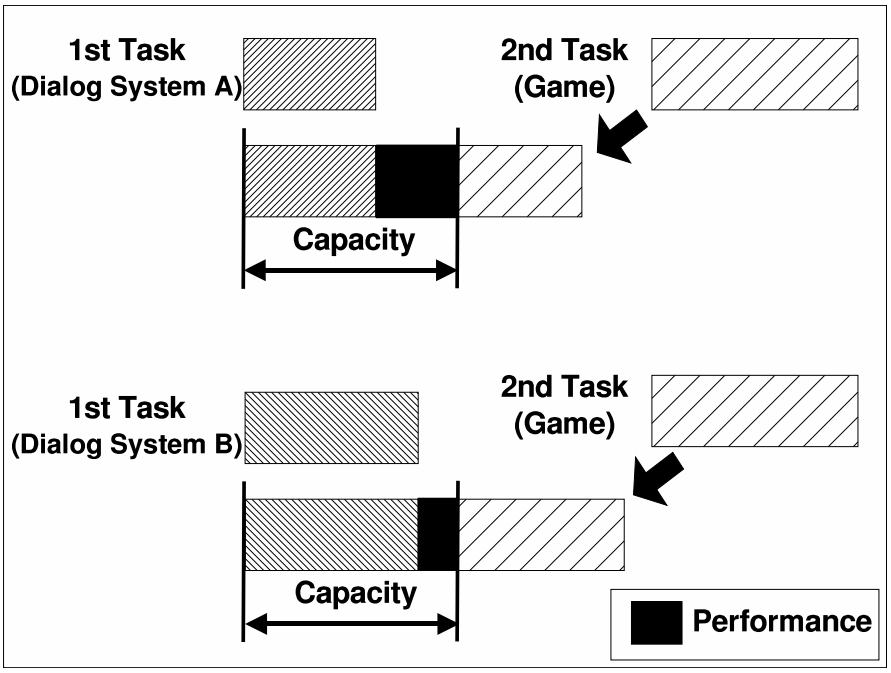

図 4.1　音声インタフェースのための対話負荷測定.

信号反応課題（赤信号が点灯したらできるだけ速くブレーキを踏むこと）を行い，これらと並行して音声による課題を行わせている．結果として，携帯電話での会話は模擬運転課題や信号反応課題に悪影響を与えること，携帯電話の形状（ハンズフリー型，ハンドヘルド型）は課題への影響において違いをもたらさないこと，単語復唱課題は運転に影響を与えないが単語生成課題（聴取した単語に基づいて新たな単語を考える課題）は影響を与えること，などが確認されている．

この研究は，視覚と手の操作を伴う第一課題と音声による第二課題の間で二重課題法が利用できることを示しているが，一部の実験はハンドル操作と信号反応を含む三重課題になっている．また，単語復唱課題の負荷は確認されていないが，その原因としては，測定対象の分解能が不十分であった可能性がある．

本研究と目的は異なるが，自動車の運転状況から運転者の余裕度を推定し，運転余裕のあるときに音声対話を用いた情報提示を行う運転状況適応型音声インタフェースの提案があり，内山ら [43] は二重課題法を用いた運転余裕度の定量化を試みている．

4.3　提案手法

第二課題の選択においては，対話負荷の比較しやすさを優先する立場から，運転など実際の作業に内容が類似していることは重視せず，周期的に一定の条件で結果を測定可能であること，課題に対する慣れの効果が少ないこと，第一課題である音声対話を被験者が優先できやすいこと，

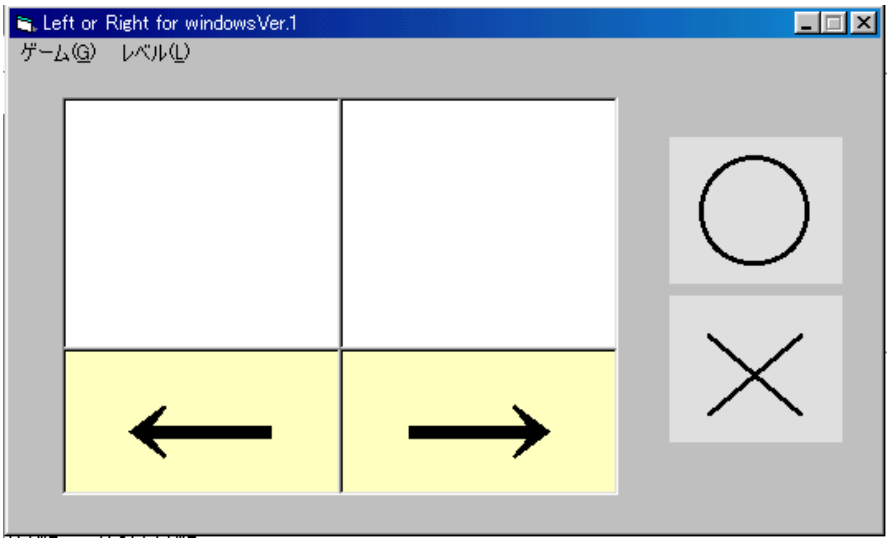

図 4.2　早押しゲームの画面構成.

などを重視する必要がある.

すなわち，ある実験によって複数の第一課題の負荷の違いが確認できたとき，第一課題の負荷の大小関係は状況によって変化しないと仮定できる．この仮定が成り立つ範囲においては，負荷の大きな箇所を特定する目的であれば第二課題の違いは問題にならない.

以下の早押しゲームは，音声対話システムと併用できる第二課題として適切なものの一つである [44, 45] .

早押しゲームの画面構成を図 4.2 に示す．画面左には 2 つの空欄および 2 つの矢印が並んでおり，ゲームが始まると 2 つの空欄の一方に左右いずれかの矢印（←，→）がランダムに現れる．左の空欄に左向きの矢印が表示された場合と，右の空欄に右向きの矢印が表示された場合には，○に対応する上矢印（↑）キーを押すこととする．それ以外の場合には，×に対応する下矢印（↓）キーを押すこととする．キーを押すと，正解不正解に関わらず 1 秒後に次の矢印が現れる．ゲーム実行中は出題や応答の内容および 1 ミリ秒単位でのユーザの応答時間を記録している．実装には Microsoft Visual Basic 6.0 を用いており，動作環境は Microsoft Windows 2000 である.

実験手順としては，まず，被験者が十分慣れるまで早押しゲームを練習させる．続いて，被験者に第二課題として早押しゲームをさせながら，同時に第一課題である音声課題を行わせる．被験者には事前に，二つの課題を同時に行う場合には音声課題を優先すること，早押しゲームについては正確さを優先しつつ，できるだけ速く応答すること，を指示しておく.

被験者実験の後，同時に行われた音声課題の内容と対応付けて各試行の結果を集計する．ある音声課題における早押しゲームの応答速度が，他の課題と比較して有意に遅れていれば，その音声課題は有意に対話負荷が大きいとみなすことができる.

表 4.1　2 単語復唱課題の例.

プロンプト	応答
「ろぼっと」(3.5 秒のポーズ)	
「たれまく」(3.5 秒のポーズ)	
ビープ音	
(6 秒のポーズ)	「たれまく」
ビープ音	
(2 秒のポーズ)	「ろぼっと」

4.4　予備実験

提案手法の妥当性を検証するために予備実験を行った．目的は，音声対話がもたらす負荷の違いが自明であるような課題を被験者に与え，有意差を確認することである．

提案手法である早押しゲームの応答速度に影響をもたらす要因は，システム発話の聴取，内容の記憶，発話タイミングの判断，発話に伴う音声器官の運動，などであると考えられる．これらは認知的負荷と物理的負荷の両方を含んでいる．このような対話負荷を模擬的に与える課題として，ここでは，単語を聴取して復唱する課題（単語復唱課題）を用いる．同時に記憶する単語数の大小によって，対話負荷の大小差がもたらされることが自明であると考える．この実験は内山らの関連研究 [43] を参考にしているが，内山らが単語復唱課題を第二課題として用いるのに対し，我々は第一課題として用いる．

4.4.1　実験手順

単語復唱課題の例を表 4.1 および表 4.2 に示す．プロンプトとして 2 つまたは 4 つの単語が提示され，ビープ音に続いて，被験者は記憶した単語を 1 つずつ答える．復唱する単語の順番は自由とする．これを単語復唱課題の 1 試行（約 20 秒）とし，各試行は 10〜20 秒のランダムな間隔で繰り返される．1 試行の時間をほぼ同じにするように各課題の単語提示や応答の間隔を設定する．

単語復唱課題の語彙サイズは 166 単語である．単語の頭文字が 50 音（濁音を除く）としてバランスよく出現するように 2〜4 モーラの普通名詞を無作為に選び，聞き取りにくいものなどを除く．読み上げには合成音声（東芝 LaLaVoice2001）を使用してランダムな順序で提示する．

被験者は 20〜25 歳の男性 9 名女性 1 名である．被験者を 5 名ずつ A, B の 2 群に分け，A 群は 4 単語課題の後に 2 単語課題を行うこととし，B 群は 2 単語課題の後に 4 単語課題を行うこととする．A 群の実験手順は以下の通りである（B 群の場合は 3-4 と 6-7 が入れ替わる）.

表 4.2　4 単語復唱課題の例.

プロンプト	応答
「ろぼっと」(1 秒のポーズ)	
「たれまく」(1 秒のポーズ)	
「けしごむ」(1 秒のポーズ)	
「すきやき」(1 秒のポーズ)	
ビープ音	
(2 秒のポーズ)	「たれまく」
ビープ音	
(2 秒のポーズ)	「すきやき」
ビープ音	
(2 秒のポーズ)	「ろぼっと」
ビープ音	
(2 秒のポーズ)	「けしごむ」

1. ゲームの説明と練習
2. ゲームだけを実行 (8 分)
3. 4 単語課題を十分慣れるまで練習
4. ゲームと 4 単語課題を同時に実行 (8 分)
5. 休憩
6. 2 単語課題を十分慣れるまで練習
7. ゲームと 2 単語課題を同時に実行 (8 分)

4.4.2　結果と考察

実験に先立って，20〜25 歳の男性 5 名が前述の手順 1-4 のみを行なったところ，4 単語復唱課題を実行している間（二重課題時）のゲームの応答時間は，課題の待機中（単一課題時）よりも有意に増加した（5 人全員が 1% 水準で有意）．これによって，本実験においてキャパシティ一定の法則が成り立つこと，また，音声課題を行いながら最速でゲームに応答することは一般にキャパシティを越えていること（すなわち第二課題の負荷量として適切な範囲であること）が確認できた．

また，2 単語課題と 4 単語課題を比較する実験において，各被験者の応答時間の平均は表 4.3（**は 1% 水準での有意差）のようになり，被験者 10 人中 9 人において 1% 水準で有意差が現れた．

これらの結果より，提案手法によって 2 つの音声課題の負荷の大小を比較できることが確認された．

表 4.3　予備実験における応答時間の平均 (msec).

被験者	2 単語課題	4 単語課題	有意差
A1	542	1191	**
A2	458	589	**
A3	486	675	**
A4	496	683	**
A5	426	562	**
B1	695	1060	**
B2	401	418	-
B3	523	868	**
B4	712	941	**
B5	478	519	**

4.5　音声対話システムの評価

本節では，被験者に音声対話システムを使用させ，提案手法を用いて対話負荷の測定を試みた実験について述べる．

4.5.1　システムの概要

自動車運転中にレストランを検索することを想定したアプリケーションを作成した．音声対話システムの実行環境として，Nuance Voice Web Server (VWS) の日本語版[*1]を使用する．被験者はノート型コンピュータに接続されたヘッドセットを装着し，インターネット電話クライアント Pingtel Instant Xpressa [*2] によって VWS に接続して対話を行う．

音声対話アプリケーションは VoiceXML 1.0 [*3] によって記述し，静的な状態遷移と変数管理のみによって対話を制御する．システムからの応答には全て合成音声を使用する．音声認識および音声合成は VWS が標準で備えるものを用いる．バージイン（システム発話中の割り込み）は全ての状態において可能とし，ボタン（数字）入力を併用せず音声入力のみを用いる．音声認識には主に孤立単語を用い，24 種類の音声認識文法（語彙サイズの平均 8.3 個，最大 30 個）を対話状態に応じて切り替える．

[*1] オムロン（株）社製.
[*2] http://www.pingtel.com/
[*3] http://www.w3.org/Voice/

4.5.2 対話の流れ

被験者には「現在地から 10 分以内で到着できるレストランを検索し，その中から一番予算の安い中華料理店を予約する」という課題を提示する．この課題を達成するための対話に含まれる状態 (S1〜S6) は次のようになる．

S1: メインメニュー いくつかのサービス項目の中から「周辺情報検索」を選ぶ.
S2: 周辺情報検索 ジャンルおよび所要時間から周辺情報の検索対象を絞り込む．ここではジャンルとして「レストラン情報」を，所要時間として「10 分」を選ぶ．ジャンルと所要時間の入力は任意の順序で可能である（この箇所のみ「レストランで 10 分」のように，1 発話で同時に 2 つのスロットに値を埋めることを許可している）.
S3: レストラン絞込み「レストラン情報」の中から,「中華料理」を選択する.
S4: 予算絞込み 一人あたりの予算を「1500 円以上」「1500 円以下」から選択する.
S5: 候補選択と詳細情報 条件を満たすレストランの件数と，レストラン名が提示される．ユーザがレストラン名を発話すると，選択されたレストランの予算を含む詳細情報が提示され，最後に「このレストランを予約しますか」という質問が行われる．ユーザが「いいえ」と答えると，S5 の最初に戻る.「はい」と答えると S6 に進む.
S6: 予約とサービス終了 決定したレストランの予約を行い，サービスを終了する.

4.5.3 実験方法

被験者は 20〜25 歳の男性 4 名および女性 1 名の計 5 名である．全員，音声対話システムの使用経験はあるが，本アプリケーションの使用経験はない．各被験者に，早押しゲームを十分慣れさせておき，対話システムを用いて 1 回だけ課題を実行させる．予備実験と同様，音声課題を優先するように教示する.

被験者実験の後，早押しゲームのログと対話の書き起こしの対応付けを行う．S1〜S6 の状態は，VoiceXML で記述された対話パターンと対応しており，各状態ごとの第二課題応答時間を集計する.

4.5.4 結果

被験者 5 名全員が与えられたタスクを達成した．平均タスク達成時間は 343 秒，ユーザ発話数は 16〜22 回，ユーザ発話がリジェクトされた回数は平均 1.2 回，最大 5 回である.

早押しゲームの応答時間は，ゲーム誤答も欠損値とせず，全データを使用する．また，個々の応答時間はばらつきの大きい値であるため，実測値 R に対して前後 5 点の移動平均値 R' を

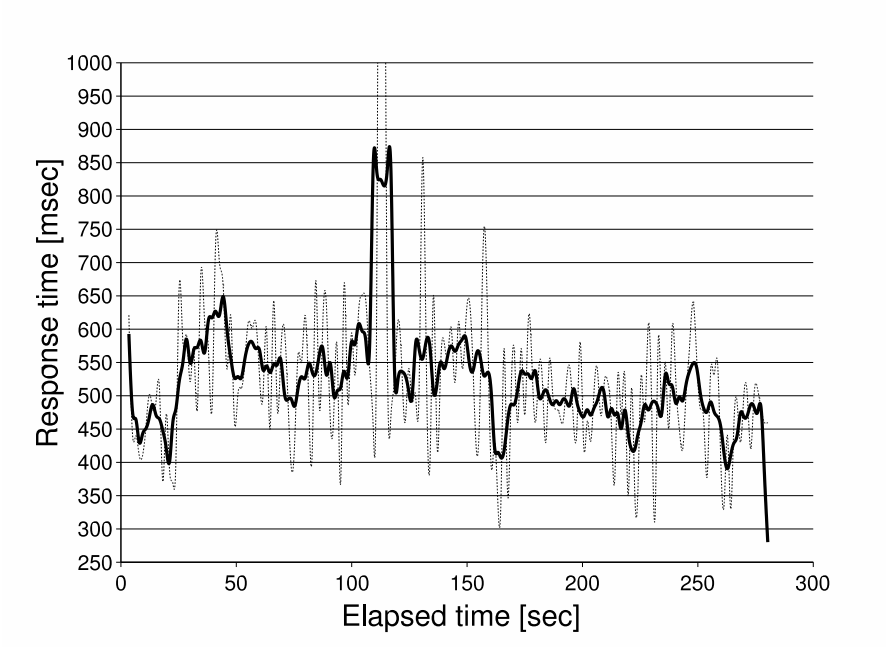

図 4.3　音声対話中の応答時間の例（細線は実測値 R，太線は移動平均値 R'）.

取る.

対話開始から終了までの R および R' の測定例を図 4.3 に示す．移動平均を取ることにより，特定の状態における負荷の大小が確認しやすくなっている.

対話状態ごとの応答時間の分布を全被験者について集計した結果を図 4.4 の箱ヒゲ図（中央値および 25%〜75% の領域を箱で示し，10% および 90% の位置に鬚を表示し，外れ値のみをプロットしたもの）に示す．特に外れ値に注目すると，S2 および S5 において応答時間が遅れる場合が目立つ.

しかし，被験者および対話状態 (S1〜S6) の 2 要因により R' の分散分析を行ったところ，交互作用が有意 (F=2.42) となった．これは，被験者の違いと対話状態ごとの応答時間が相互に影響を与えることを示している [46, 47]．各被験者・各対話状態ごとの応答時間の平均を図 4.5 に，対話状態の単純主効果における有意性を表 4.4 に示す（ただし F 比は偶然変動との比を表す）．これらの結果から，被験者 C1, C2 は状態 S2 の応答時間の平均が，被験者 C4, C5 は状態 S5 の応答時間の平均が，他の状態より有意に大きいことが確認できた.

4.5.5　考察

図 4.4 の結果より，対話状態 S2 および S5 において応答遅れが生じやすいことが確認された．これは，小島らの関連研究 [39, 40] と同じ応答遅れに注目した分析が，提案手法では対話状態ないしは対話パターンの比較においても可能であることを示している.

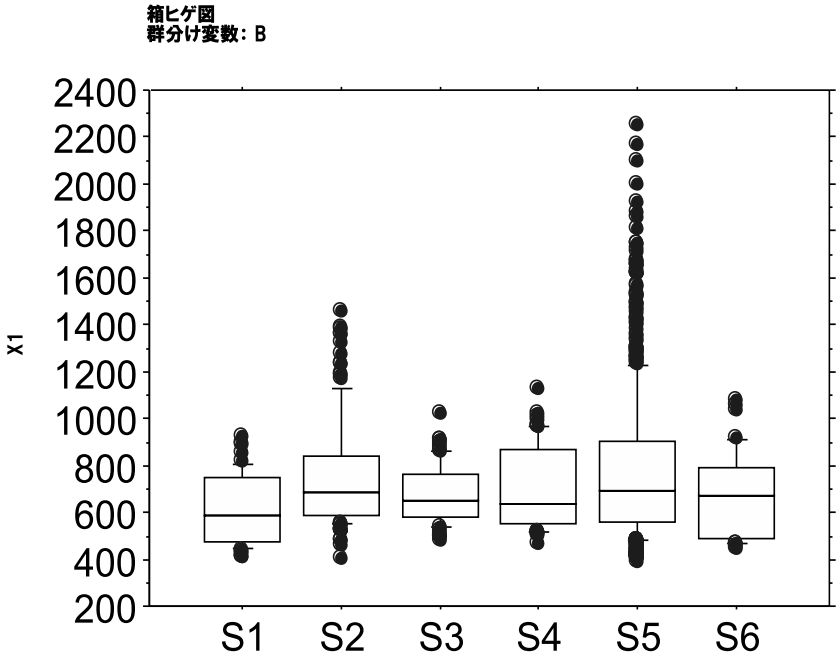

図 4.4　対話状態と応答時間 (msec) の関係（全被験者）.

表 4.4　対話状態の単純主効果における有意性.

被験者	F 比	p	有意差
C1	F(5,177)=3.380	p=.0061	**
C2	F(5,244)=7.222	p<.0001	**
C3	F(5,151)=1.591	p=.1658	
C4	F(5,157)=3.010	p=.0127	*
C5	F(5,222)=4.977	p=.0002	**

また，音声対話システムにおける実験結果には個人差があることが図 4.5 および表 4.4 から確認され，小島らの関連研究では有意ではなかった平均反応時間の差も有意となった．予備実験（4.4 節）から，負荷の大小が自明であれば実験結果の個人差は生じにくいことが確認されているため，実際の音声対話システムにおいては，個人の知識や経験に依存して対話負荷が異なると考えられる.

次に，有意に応答時間の長い被験者が 2 人ずつ存在した対話状態 S2 および S5 に関して検討し，実験結果の妥当性について考察する.

S2 は「ジャンル」および「所要時間」の 2 つのスロットに値を埋める対話である．プロンプトの詳細を図 4.6 に示す.「10 分」「レストラン」「レストランで 10 分」などの発話を受理し，受

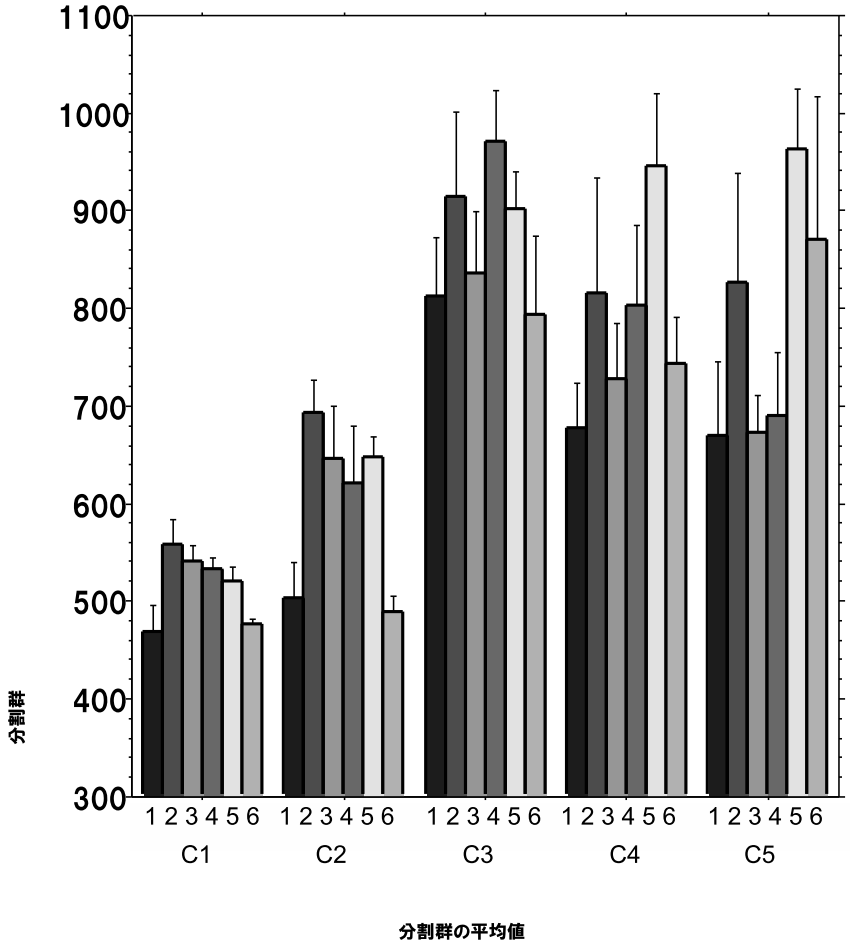

図 4.5　被験者 C1〜C5 と対話状態 S1〜S6 における R' (msec) の平均．エラーバーは 95% 信頼区間.

こちらは周辺情報サービスです．ジャンルと現在地からの所要時間で周辺情報を検索します．ジャンルはレストラン，アミューズメント，スポーツ，お得情報から選択できます．所要時間は 5 分，10 分，15 分から選択できます．ご希望のジャンルと所要時間を音声でお答えください．

図 4.6　対話状態 S2 におけるプロンプト.

理できなかった条件はシステム主導で再度質問される.

この対話は，読み上げられた候補の中から 1 つの単語を選んで復唱する他の典型的な対話と比較して，「何を言うべきか迷う」という状況が起こりやすく，対話負荷を大きくすることが予想される.

また S5 は，読み上げられたレストラン名を聴取して音声コマンドとして発話する対話状態で

現在地から 10 分以内でお一人様の予算が 1500 円以下の中華レストランは 4 件あります．レストラン名は「王将」「みよし」「クーロン」「天下一品」です．知りたいレストランを選んで音声でお答えください．

図 4.7　対話状態 S5 におけるプロンプト．

ある．プロンプトの詳細を図 4.7 に示す．この対話は，被験者がレストラン名をはじめて聞く場合や単語を聞き取りにくかった場合には，「何を言うべきかわかりにくい」という状況が起こりやすく，対話負荷を大きくすることが予想される．また，被験者に与えた「もっとも予算の安いレストランを選ぶ」という課題は，予算のもっとも安いレストラン候補の名前と予算を覚えておき，新たなレストランに関する情報を聞きながら，予算を比較する，という判断を必要とする．この作業も対話負荷を大きくする要因となり得る．

この実験の結果より，負荷が大きくなることが妥当と考えられる対話状態において，有意に値が大きくなるような定量可能な値 R' を得ることができた．したがって，対話負荷の大きい箇所を特定する手段として，提案手法の有効性は示されたと言える．

4.6　まとめ

本報告では，音声インタフェースを使用する課題の負荷を二重課題法によって測定するために，画面表示とキーボード入力による早押しゲームを用いた実験方法を提案し，単語復唱課題および音声対話システムを用いてその有効性を検証した．

本実験は自動車の運転を前提とした評価を行ってはいない点が多くの関連研究と異なっている．しかし，模擬運転課題を用いた Strayer らの関連研究 [42] と異なり，本実験では比較的負荷が低い単語復唱課題においても，負荷の大小を測定できた．さらに，関連研究 [39, 40, 41] では，対話パターンを比較したり対話中で負荷の大きい箇所を探すことは困難であったが，本実験では，特定の被験者が特定の対話状態において負荷が大きくなる，といった状況を見つけることができた．

これらの結果が得られた理由として，第二課題として使用した早押しゲームが，(1) 第一課題を優先するという教示を守りやすいタスクであること，(2) LED 刺激反応タスクよりも複雑であるため第二課題として適切な負荷の大きさであること，(3) 1 試行あたりの所要時間が 3 秒以下と短く，短時間に反復して実行できるため，比較的多くのサンプルを得やすいこと，(4) 反復しても慣れの効果が少ないこと，などが考えられる．これらの特長から本手法は，対話負荷の大きい箇所を高い時間分解能で探すことができると考えられ，所要時間の長い音声課題を評価したり，一人の被験者が繰り返し評価を行ったりする場合に特に有用であろう．

今後の課題としては，まず，被験者実験をより正確に効率よく行えるために，手法やツールの改良を行う必要がある．特に，本実験では問題にならなかったが，ゲーム課題と音声課題の類似性の影響（例えば，右左折などの道案内を音声で行なう対話においては，ゲーム課題と音声課題が左右方向の判断という点で類似してしまう）について検討したい．また，第 5 章での考察を発展させ，音声対話において負荷を大きくする要因を物理的負荷と認知的負荷のそれぞれについてさらに調査し，音声対話システムの改良に役立つガイドラインや対話パターンについて検討を行いたい．

さらに，関連研究で扱われていた，自動車の運転と音声対話を同時に行う場合の安全性などを論じるために，本手法で得られる対話負荷の尺度と，自動車の運転など実環境における負荷の関連性についても検討していく必要がある．

第 5 章

超早口音声の聴取に対する慣れの検討

5.1 はじめに

視覚障害者のコンピュータやインターネットの利用を目的として，画面に表示される情報やキーボード操作を音声で読み上げる技術が実用化されている．近年はパーソナルコンピュータ (PC) だけでなく携帯電話においても音声読み上げ機能を持つものが普及しつつある．こうした技術の利用においては，情報取得の効率性が重視されており，通常話速の 3 倍あるいは 4 倍といった，日常的な会話や放送の話速をはるかに超える早口の音声（超早口音声）が用いられる．

近年コーパスベース音声合成の技術が普及し，自然で明瞭な音声を任意のテキストから合成できるようになった．しかし超早口音声に関する視覚障害者のニーズに応えるために音声合成技術をどのように改良すべきなのか，明らかになっているとは言えない．

超早口音声は初めて聞く人にはほとんど聞き取れない．視覚障害者においてもこれは同じであると言われ，慣れによって聴取できるようになる．超早口音声への視覚障害者の慣れについては関連研究 [48, 49] がある．

聞き手の慣れの効果が大きいことは，超早口音声の品質評価を困難にしている．例えば先行研究 [50, 51] は HMM 方式で作られた 4 桁数字読み上げの超早口音声を聴取実験で評価しており，聞き手の慣れの効果が短時間で起こり定着しやすいことを示したが，慣れの効果がどのような場合に起こりやすいかは明らかにしていない．

一般に音声の聞き取りは，音という物理的な刺激の知覚（ボトムアップ情報）と，すでに聞き手が持っている語彙や聞き手が置かれた状況・文脈（トップダウン情報）の双方の影響を受ける．これらの要因を統制しつつ，日常生活の場面における音声聴取能力を正しく評価するために，親密度別単語了解度試験データベース FW03[52] を使う方法が提案されている．

本研究では，単語親密度（単語に対する主観的ななじみの程度）を統制した日本語の超早口音声の聴取を課題として用い，(1) 実験の途中での親密度条件の変化，(2) 親密度に関する教示の有無，の要因が聞き手の課題への慣れにどのような影響を与えるかを検証する．その際に，単語

了解度に加えて，被験者の主観的な評価としての心的負荷も検討対象とする．期待される成果は「超早口音声に慣れていくときに，実際に何が起きているのか」を示唆する知見であり，これは超早口音声の正しい評価に貢献することが期待される．

5.2 関連研究

5.2.1 視覚障害者の音声利用における要求

渡辺 [48] は日本国内で視覚障害を持つ PC 利用者がスクリーンリーダの音声をどのように設定しているか実状を調査した．その報告によると，多くの利用者はソフトウェアにおいて設定可能な最高の読み上げ速度を選択しており，これは一般的な読み上げ速度の約 2 倍であった．

視覚障害を持つ人々がどの程度の超早口音声を聞き取ることができるか，といった検討を浅川らが行っている [49]．浅川らは音声波形編集ソフトウェア CoolEdit の時間伸縮機能を用いて，日本語の文章を読み上げた音声を時間伸縮機能によって早口化し，視覚に障害を持ちスクリーンリーダの熟練者である複数の被験者がこれを聴取した．文章に含まれる単語の約 90% を聞き取ることができる話速を「最適速度」と定義し，約 50% の単語を聞き取ることができる話速を「最高速度」と定義し，合成音声の利用に熟練した被験者によって評価した結果，最適速度は約 18 モーラ／秒，最高速度は約 23 モーラ／秒であった．これに対して当時の一般的な音声合成エンジンの最高速度は 15 モーラ／秒以下であった．すなわち，多くの音声合成エンジンはユーザが望むような早い速度の音声を合成できておらず，視覚に頼らないユーザインタフェースの改善においては，より高速での読み上げに対応した音声合成エンジンが必要であるとされた．

その後，日本で広く使われているスクリーンリーダ（「IBM ホームページ・リーダー」）が音声合成の最高速度を引き上げるなど，超早口音声への要求に応える努力がなされている．しかし最高速度を引き上げたことの効果が客観的に検証されているとは言えない．

5.2.2 合成音声の聴取における慣れ

超早口音声に関する研究ではないが，自然性や明瞭性が十分ではないテキスト音声合成への学習効果については渡辺の報告 [53, 54] がある．合成音声の累積受聴単語数が 5000 単語くらいまでは聞き続けるほど単語了解度が上がり，1〜2 ヶ月の休止後の受聴テスト再開時も了解度が低下しない，とされている．

HMM 音声合成の手法で作られた超早口音声に関する検討 [50, 51] によると，晴眼の若年者に 4 桁の数字の超早口音声を聴取させると，超早口音声は最初は聞き取りにくいが，しばらく聞いていれば，提示した音声に対して正解を示さなくても，ある程度聞き取れるようになる．また，晴眼の若年者および 65 歳以上の高齢者における比較実験から，訓練前における若年者と高齢者は 4 桁の数字を同じ程度聞き取れていること，慣れによる了解度の向上は若年者でのみ顕

表 5.1　FW03 に含まれる単語の例.

親密度 7.0〜5.5	親密度 2.5〜1.0
アマグモ	アイキャク
イマフウ	イチハツ
ウチガワ	ウラジャク
オシダシ	エラブツ
オヤモト	オクデン

著であること，高齢者は若年者と比較して 4 つの数字の順序を誤って回答する割合が大きいこと，などの結果を得ている．

数字とは親密度の非常に高い，ごく小さな語彙だと考えることができる．したがって，これらの実験は，文章の超早口音声の聴取に対して単語の親密度が与える影響を調べるためには不十分である．

5.2.3　親密度別単語データベース FW03

従来の音声聴取能力の評価が，単音節や無意味三連音節などを対象としており，単語や文章の課題がこれらと比べて意味の統制が困難である，という問題を解決するために，親密度別単語了解度試験データベース FW03[52] が作られた．これは日常生活の場面における音声聴取能力を正しく評価することを目的としている．

FW03 の単語リストは，「日本語の語彙特性」[55] 第 1 巻「日本語親密度データベース」を用いて選定された．単語親密度とは単語に対する主観的ななじみの程度を表した数字であり，1（なじみがない）から 7（なじみがある）までの範囲を取る．この単語親密度は単語認知の正確さや速さと強い関係があり，単語親密度が高いほど単語の認知が正確であり，かつ素早く行われるとされる．単語親密度の評定においては 18 歳以上 30 歳未満の日本人 40 名の被験者が，音声提示，文字提示，文字音声同時提示の 3 種の提示モードで 7 段階尺度の一つの値を選択し，それらの平均値として単語親密度が算出された．

FW03 に含まれる単語の例を表 5.1 に示す．FW03 の単語リストは 4 モーラの単語群からアクセント型が 0 型および 4 型のものだけが選ばれ，音声提示モードにおける親密度に基づいて 4 段階 (7.0〜5.5, 5.5〜4.0, 4.0〜2.5, 2.5〜1.0) に分割され，さらに音韻バランスを考慮した各 50 単語の単語リスト 20 組として作られた．この単語リストを使った単語了解度試験では，親密度が高いほど正答率が高くなることが確認されている [56].

5.2.4 主観評価と心的負荷評価手法 NASA-TLX

一般に了解度は音声伝達性能の評価基準として重要ではあるが，主観評価も了解度を補完する手法となりうる．例えば「聞き取りやすさ」あるいは「聞き取りにくさ」といった主観評価値が，単語了解度を補完する尺度であることが指摘されている [57].

スクリーンリーダを日常的に使う視覚障害者からは「聞き疲れしない合成音声」を求める意見が出ている．そこで本研究では「疲れやすさ」という観点から超早口音声を主観的に評価する手法を検討する．例えば，単語了解度では差が生じにくいような条件間の比較が，内観としての疲れにくさの評定から可能になると期待される．

一般に心的負荷を評価する手法としての主観評価は，生理的尺度や客観評価（2 重課題法など）と比較して実験が簡便であり，客観評価と比較して感度の高い評価が行える可能性もある．一方で，被験者の内観によって信頼できる結果を得るために十分な配慮と結果の検証が必要となる．

ヒューマン・マシン・インタフェースにおける人間の負荷を総合的に評価できる手法として NASA-TLX (Task Load Index)[58, 59] がある．これは以下の 6 つの下位尺度によって心的負荷を評価する主観評価手法である．

- 知的・知覚的要求（小さい／大きい）
- 身体的要求（小さい／大きい）
- タイムプレッシャー（弱い／強い）
- 努力（少ない／多い）
- フラストレーション（低い／高い）
- 作業成績（良い／悪い）

被験者は負担度評価の対象となる作業を遂行する前に，下位尺度の重要度を評価する．この重要度評価のあとで，被験者は負担度評価の対象となる作業を遂行し，その後，6 つの尺度それぞれに対する評定を，目盛りのない直線の上に印を付けることによって行う．重要度の評定に基づいて重みをつけて，下位尺度の評定値の加重平均作業負荷 (mean weighted workload, WWL) 得点を求める．

メンタルワークロードの基盤となる注意資源は多次元的性質を持っており，単一の尺度（次元）だけで評価を行うことは適当ではないと考えられている [60]．NASA-TLX によって，ある被験者にとってのワークロードと関連性の強い尺度を大きく反映した得点を得ることができ，さらに被験者間の評価のばらつきを抑えることができる [61].

文献 [58] の手法では，尺度間の一対比較で重み付けを行い，各尺度の評定値に重みを与えてワークロード得点を算出する．その派生版として，一対比較を行わずに各尺度の平均値をワー

クロード得点とする方法や，各尺度の評定値の順位を重みとして用いる方法も提案されている [62, 63]．これらの方法によりメンタルワークロードを多次元的に評価できる．

NASA-TLX はさまざまな分野に汎用的に利用しやすい特長を持っているが，標準的な手法が必ずしも最良ではない場合もあり，例えば日本語版 NASA-TLX の下位尺度の説明文を参考にして新たにチェックリストを作り，メンタルワークロードを集団で測定する研究 [64] がなされている．

なお，得られるワークロード得点は，例えば「60 以上であれば作業継続が困難」のように絶対的な値として意味を持つと考える立場もあるが，被験者が感じる負荷の大小と評定値の対応には個人差があり，値の相対的な関係のみを信用すべき，という立場もある．

5.3 実験 1:NASA-TLX の有効性の確認

5.3.1 実験 1 の目的

予備的な検討として，単語親密度の異なる音声を話速を変えて聴取する実験を行い，了解度の検討と心的負荷の測定を行った．さらに主観評価手法としての NASA-TLX の有効性の検証を行った．この実験で期待される成果は，了解度の高低と心的負荷の大小の妥当性を確認すること，および，了解度の比較では得られない知見を心的負荷の比較から得ることである．

音声の聴取という比較的単純な課題に対しては，単に聞き取りにくさや疲労度を問うことでも負荷評価は可能と考えられる．本研究で NASA-TLX を用いることは必須ではないが，評価尺度を複数用いることで被験者が評定しやすくなる効果を期待した．また，将来スクリーンリーダを用いた課題遂行に関する負荷を総合的に比較する場合にも，同じ負荷評価手法が使えると期待される．

なお，メンタルワークロードの評価においては，適用すべき対象に合わせて予備実験を行うなどして，手法や結果の妥当性について事前に検証すべきである．本研究では，既存の日本語版 NASA-TLX の派生版である「各尺度の評定値の順位を重みとして用いる方法」を採用し，実験 1 で評価方法の有効性を確認し，実験 2 において慣れの効果を検証した．

5.3.2 NASA-TLX 用ソフトウェア

NASA-TLX 評価をコンピュータ上で行うための Windows 用ソフトウェアを作成した[*1]．当面は晴眼者を対象としたため，わかりやすさを重視してグラフィカルな操作を採用した．図 5.1 はメニューと各尺度の説明，図 5.2 は尺度の順序づけの画面である．下位尺度の説明は音声と文字で行われる．

[*1] NASA-TLX および後述する音声聴取実験用のソフトウェアは Delphi for Windows で作成した．

6 つの下位尺度の重要度の評価においては被験者に項目を並べ替えさせる方法を採用した．下位尺度の評定の操作では，目盛りのない直線の上に印を付ける代わりに 0〜100 のスクロールバー操作によって評定を行うことにした．また評定値の WWL 得点を計算する際には，重要度が最上位と評価された尺度から順に重みを 6〜1 とした．すなわち，各尺度の評定値を $S_1, S_2, ..., S_6$ ，各尺度の順位 (1〜6) を $R_1, R_2, ..., R_6$ とすると，各尺度の重み $W_1, W_2, ..., W_6$ は

$$W_n = 7 - R_n \tag{5.1}$$

であり，加重得点 WL は

$$WL = \frac{\sum_{n=1}^{6} S_n W_n}{\sum_{n=1}^{6} n} \tag{5.2}$$

である．

画面設計においては，課題間の値の大小関係を意識した評定を被験者に促すように考慮した．具体的にはリハーサルおよび複数の課題における各項目の評価の画面において図 5.3 に示すように過去の評定値を参照できるようにした．

尺度の説明においては本研究の課題に合わせて「作業成績＝すべて正しく聞き取って書き取る（入力する）ことが目標」「知的・知覚的要求＝聞くことを含む」「身体的要求＝聞くこと・書くこと・喋ることを含む」などを補足した．

5.3.3 実験 1 の手順

FW03 を用いて，単語親密度 (F) について H/L の 2 条件，話速 (S) について 1/2 の 2 条件を下記のように設定した．

- FH：親密度が 7.0-5.5
- FL：親密度が 2.5-1.0
- S1：FW03 の音声をそのまま使用
- S2：Adobe Audition 2.0 で 2 倍速に変換

FH-S1, FH-S2, FL-S1, FL-S2 の 4 群は，FW03 の異なるリストに含まれる各 50 個の音声の出現順序をランダムにしたものである．S1 で用いた音声ファイルは FW03 の男性話者 1 名分（48kHz サンプリング, 16bit, モノラル）をそのまま使用した．S2 を作成する際には Adobe Audition 2.0 でタイムストレッチ（比率 200%）を使用した．音声波形から実測したところ各発話の話速はほぼ一定で，S1 および S2 の平均話速はそれぞれ 5.3 モーラ／秒，10.8 モーラ／秒であった．また FH と FL の各群間の発話継続時間に有意差はなかった．

被験者は課題の音声を聴取したことがない大学生（3 年生および 4 年生，全員女性）で日本語を母国語とする晴眼者かつ健聴者であった．11 人の被験者が実験に参加した．被験者には決

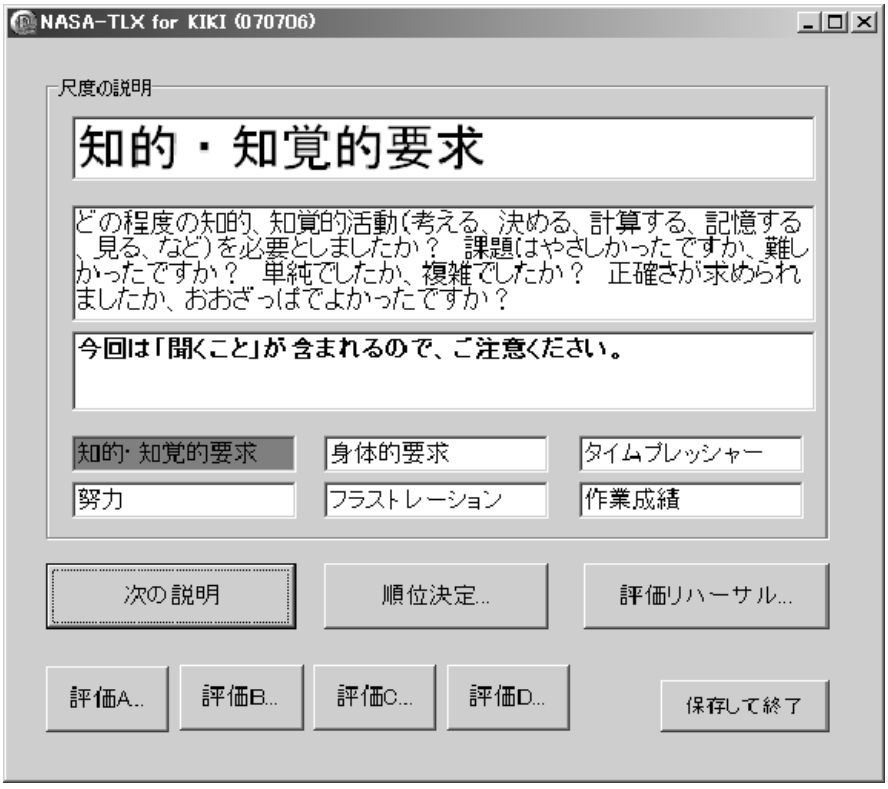

図 5.1　NASA-TLX ソフトウェアの画面（メインメニューと下位尺度の説明）.

まった金額の謝金が支払われた.

順序効果を避けるために被験者は 2 群に分け，片方の群では FH-S1, FH-S2, FL-S1, FL-S2 の順に，別の群は FL-S1, FL-S2, FH-S1, FH-S2 の順に実験を行った.

1 人 1 台のノート PC (Microsoft Windows XP) とヘッドフォン (Panasonic RP-H750-S) を使用した．4 モーラの音声を 10 秒間隔で 1 つずつ聴取させるために，音声提示用ソフトウェアを作成した．被験者自身にこの音声提示用ソフトウェアと NASA-TLX ソフトウェアを交互に操作させて，聞き取った単語を解答用紙にカタカナで記入させた.

被験者には，課題がすべて 4 モーラの単語であることを伝え，意味を成さないとおもっても聞こえてきた音を書き取るように教示した．単語親密度に関する教示は行わなかった．また，被験者の回答に対する判定や正解のフィードバックは行わなかった.

最初にリハーサル課題として，単語親密度が中程度の女性音声 15 問を聴取させ，各自が聞きやすいと感じるように音量を設定させ，可能な限りで聞き取りもさせた．次に，音声ガイドと文字表示で下位尺度の説明を行い，下位尺度の順位付けをさせ，ソフトウェアの操作に慣れるための練習の目的でリハーサル課題の負荷評定を行わせた.

続いて各群ごとに定めた順序で 4 条件について，(1) 50 問の聴取を行わせて（所要時間は 10 分弱），(2) 6 つの尺度の負荷評定を行わせ，(3) 疲労の影響を回避するために 5 分間の休憩を取

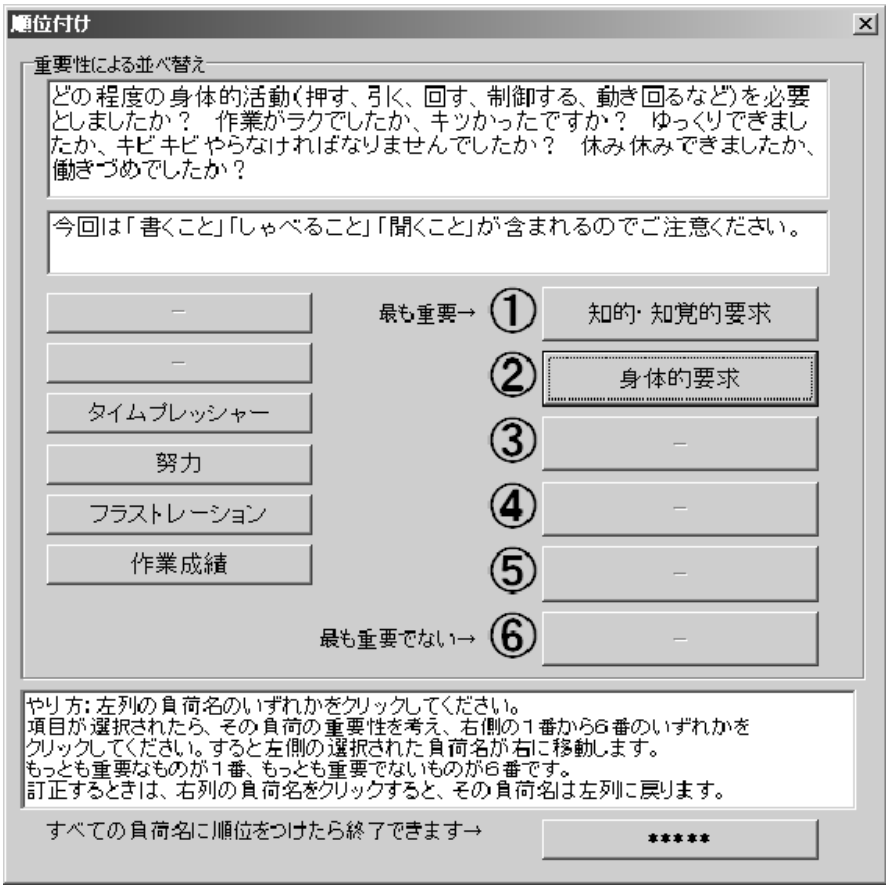

図 5.2　NASA-TLX ソフトウェアの画面（下位尺度の重要度評定）.

表 5.2　実験 1 の結果（了解度，WWL，正規化 WWL に関する被験者間の平均と標準偏差）.

Task	Intelligibility	WWL	N-WWL
FL-S1	97.4 (1.6)	53.1 (12.9)	47.1 (4.2)
FL-S2	65.8 (7.5)	63.5 (12.9)	64.9 (2.4)
FH-S1	99.8 (0.6)	47.9 (14.6)	39.8 (3.8)
FH-S2	96.8 (2.2)	53.4 (13.0)	48.3 (4.9)

る，という手順を繰り返した.

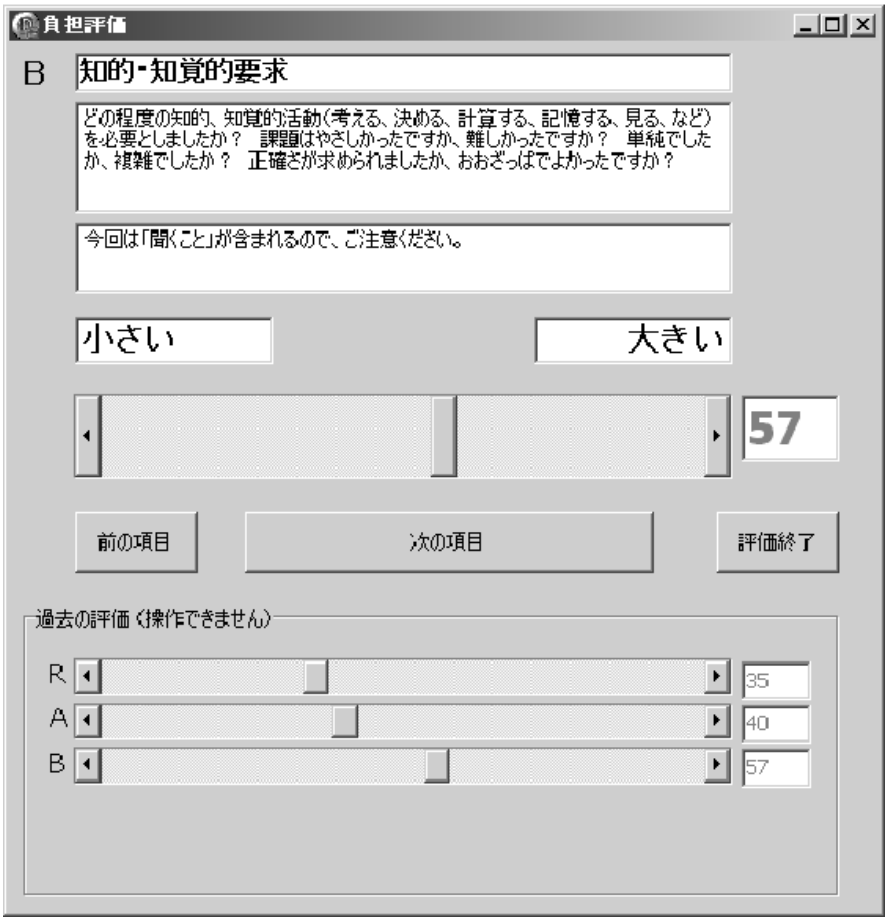

図 5.3　NASA-TLX ソフトウェアの画面（各下位尺度の負荷の値の入力）.

5.3.4　結果

10 名のデータを分析対象とした[*2]．図 5.4 に各被験者の単語了解度と WWL の分布を示す．また，被験者毎の WWL の 4 つの値の平均と標準偏差を 50 および 10 に正規化した WWL(以後 N-WWL と呼ぶ) を図 5.5 に示す．また，被験者 10 人の各課題群の平均値および標準偏差を表 5.2 に示す．JavaScript-STAR[47] による分散分析の結果，単語了解度については課題の効果が有意 ($P < 0.01, F = 140.24$) であり，多重比較で FL-S2 と他の 3 群それぞれの間の有意差が確認された.

WWL および N-WWL についても課題の効果は有意 ($P < 0.01$, $F = 24.13$ および $F = 49.09$) であった．さらに NASA-TLX の 6 個の下位尺度の評定値を重みづけせず平均した値 (Average Workload, AWL と呼ぶ) についても分析したところ，課題の効果はやはり有意

[*2] 「作業成績」の値の大小関係を誤解したと考えられる被験者 1 名のデータを除いた．後述する実験 2 では誤解されにくいようにソフトウェアの改良および教示の配慮を行った.

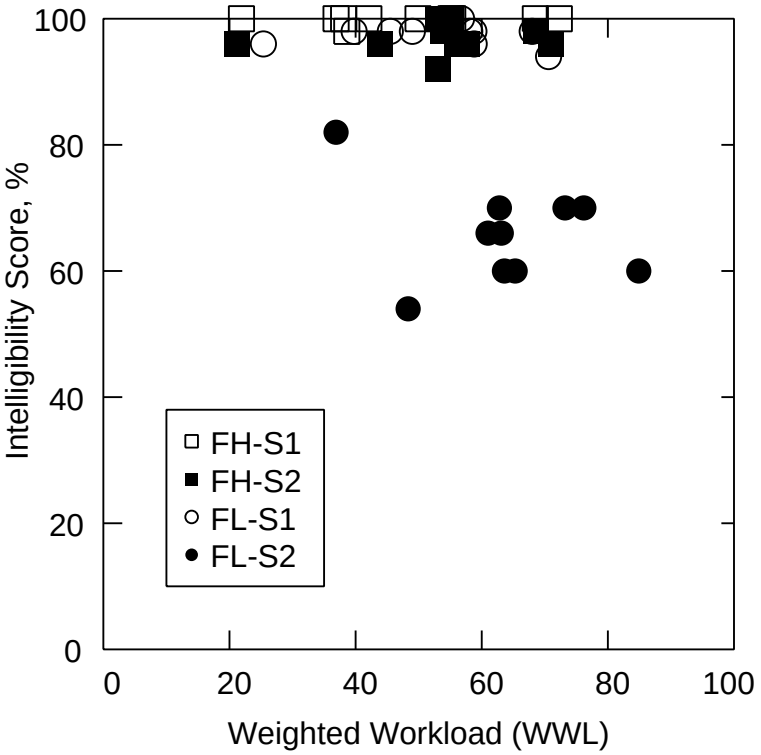

図 5.4　実験 1 の結果（被験者・課題ごとの WWL と了解度の分布）.

($P < 0.01, F = 32.77$) であった．多重比較で有意差が確認された条件の対を表 5.3 に示す (* は $P < 0.05$).

5.3.5　考察

実験 1 の結果について考察する.

まず，単語了解度が FL-S2 のみで有意に低かったことから，単語親密度が了解度に影響を与えるのは速度が早い場合に限られることが示唆された.

また，心的負荷の比較においては，単語親密度で有意差のなかった FL-S1–FH-S1 および FH-S1–FH-S2 の各群間で新たに有意差が確認できた．音声聴取において単語了解度では比較しにくい軽微な差の比較に NASA-TLX による主観評価が役立つことが示唆された.

図 5.4 の結果は群内のばらつきが大きく，WWL を課題の負荷の絶対値として用いることの難しさを示唆している．N-WWL を用いることでばらつきが減ることが確認できたものの，分散分析で被験者内比較を行う場合には大きな差はなかった．また重要度の重みを使わない AWL

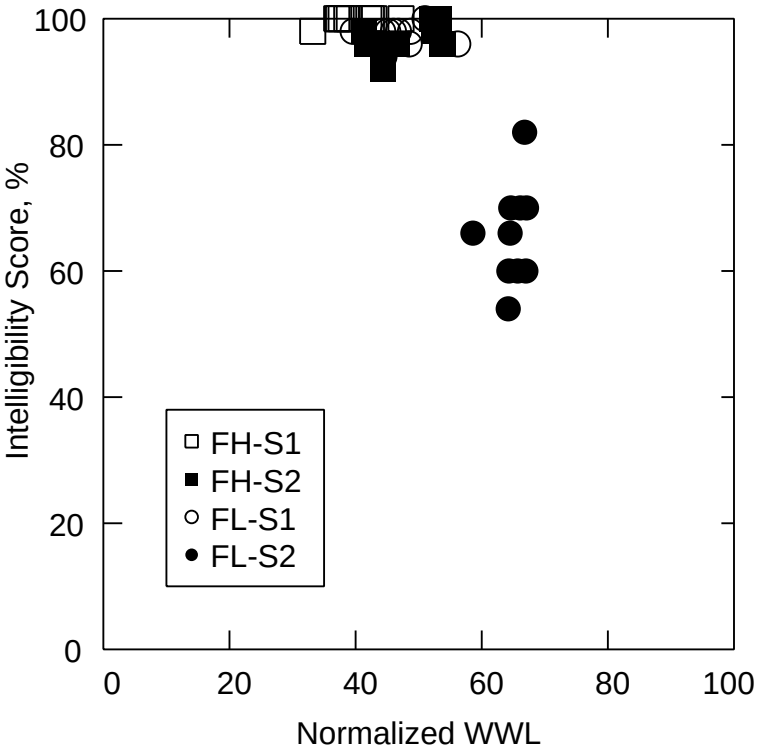

図 5.5　実験 1 の結果（被験者・課題ごとの正規化 WWL と了解度の分布）.

の場合でも，有意差の比較においては WWL と同等の結果を得られた．このことは「仮に被験者が重要度の重み付けの正しさに自信を持てない場合も，心的負荷の比較は信頼できる可能性が高い」と解釈でき，NASA-TLX による結果の信頼性が示された.

本実験の知見は，被験者の年齢および性別を限定したうえで得られたものである．今後は男女差の影響および加齢の影響を確認するための実験が必要である.

5.4　実験 2:慣れの効果の検討

5.4.1　実験 2 の目的

実験 1 の知見を踏まえて，超早口音声への慣れについて調べる実験を行った．50 個の音声を聴取する課題を 3 回繰り返し，被験者ごとの了解度および心的負荷の変化を見ることで慣れの検証を行った.

実験 2 の計画にあたり，超早口音声の聴取実験に関する以下の仮説を挙げる.

表 5.3　実験 1 の結果（WWL，N-WWL, AWL に関する群間の有意差）.

Conditions	WWL	N-WWL	AWL
FL-S1 < FL-S2	*	*	*
FL-S1 > FH-S1	*	*	*
FL-S1 = FH-S2			
FL-S2 > FH-S1	*	*	*
FL-S2 > FH-S2	*	*	*
FH-S1 < FH-S2	*	*	*

1. **了解度に関する仮説：**高親密度語彙が提示され，聞き手が「高親密度語彙である」と自覚できた場合に，心的辞書のアクセスに関するトップダウン情報が有効に活用され，これが了解度の上昇に貢献する．また，音の物理的な刺激の知覚（ボトムアップ情報）に関する慣れの効果は，了解度の上昇に大きくは貢献しない.
2. **心的負荷に関する仮説：**多くのトップダウン情報を活用できる条件では，心的負荷は少なくなる．トップダウン情報を活用できない状況と活用できる状況が切り替わることで，心的負荷の変化が生じる.
3. **親密度の知覚に関する仮説（高親密度語彙の場合）：**高親密度の超早口音声については，親密度の教示がない場合は，聞いているうちに聞き手が「高親密度語彙である」と気づくことができ，次第に「親密度の教示がある場合」と同じ聴き方に近づく.
4. **親密度の知覚に関する仮説（低親密度語彙の場合）：**低親密度の超早口音声については，親密度の教示があってもなくても「どのように聞けばよいか」という方策を獲得することが困難である.

実験 2 ではこれらの仮説を検討するために，親密度に関する 4 種類の条件の一部において，3 回の課題の 1〜2 回目と 3 回目の間で語彙の親密度の変更を行った.

また，これら 4 種類について，被験者が教示なしで実験条件の変化に気づく場合と，明確な教示によって被験者が実験条件を知っている場合の比較を行うこととした．つまり，教示の有無の 2 種類と親密度構成の 4 種類をすべて考慮した 8 群での実験を行った.

5.4.2　提示する音声

FW03 の録音音声（すべて同一の男性話者のもの）を独自の話速変換ソフトウェアで変換して 4.0 倍速音声（約 21 モーラ／秒）を作成した．サンプリング周波数は 48kHz のまま使用した．このプログラムは話速変換時のスペクトルのひずみを最小化するように繰り返し計算で振

表 5.4　実験 2 の構成.

群	教示	親密度	Trial 1	Trial 2	Trial 3	人数
G1	なし	L-L-L	FL-V1	FL-V2	FL-V3	16
G2	なし	H-H-L	FH-V4	FH-V5	FL-V3	15
G3	なし	L-L-H	FL-V1	FL-V2	FH-V6	14
G4	なし	H-H-H	FH-V4	FH-V5	FH-V6	14
G5	あり	L-L-L	FL-V1	FL-V2	FL-V3	8
G6	あり	H-H-L	FH-V4	FH-V5	FL-V3	8
G7	あり	L-L-H	FL-V1	FL-V2	FH-V6	7
G8	あり	H-H-H	FH-V4	FH-V5	FH-V6	7

幅と位相を最適化する手法を用いている.

浅川らの関連研究 [49] では,「最高速度」(約 50% の単語を聞き取ることができる話速) が約 23 モーラ／秒 (1400-1500 モーラ／分) であると報告されている. 本研究も話速についてはこれに近い条件を目標にした. 文献 [49] と本研究の音声刺激は, 録音された肉声の話速変換である点が共通している. しかし, 文献 [49] が文章の読み上げ音声 (日本音響学会研究用連続音声データベース内の ATR 音素バランス文) を用いたのに対して, 本研究では比較的丁寧に発声された 4 モーラの孤立発話音声 (FW03) を使用したため, 元の音声の話速が異なる. 目標に近い約 21 モーラ／秒に相当するのが FW03 の 4.0 倍速であったため, 今回の実験ではこの条件を用いた.

5.4.3　実験 2 の手順

被験者は課題の音声を聴取したことがない大学生 (1 年生から 4 年生, 全員女性, 全員が PC の操作に熟練している) で日本語を母国語とする晴眼者かつ健聴者である. 89 人の被験者が実験に参加した. 被験者には決まった金額の謝金が支払われた.

被験者を表 5.4 のとおり 8 群に分けて, 各被験者に 3 つのリストの聞き取りをさせた. FL および FH は親密度の高低を表す. 同じ親密度条件を繰り返す場合も異なる単語リスト (FL:V1〜V3, FH:V4〜V6, 各リストは 50 単語) の音声を用いた. 例えば G1 は FW03 における親密度 1.0〜2.5 の 3 種類の課題 FL-V1, FL-V2, FL-V3 を順番に聞き取らせ, その際に親密度に関する教示を行わなかったことを示す[*3].

1 人 1 台のノート PC (Microsoft Windows XP) とヘッドフォン (Panasonic RP-H750-S) を

[*3] 表 4 の実験構成において G1-G4 と G5-G8 の被験者数に隔たりがあるが, この理由と影響については 5.4.7 にて詳述する.

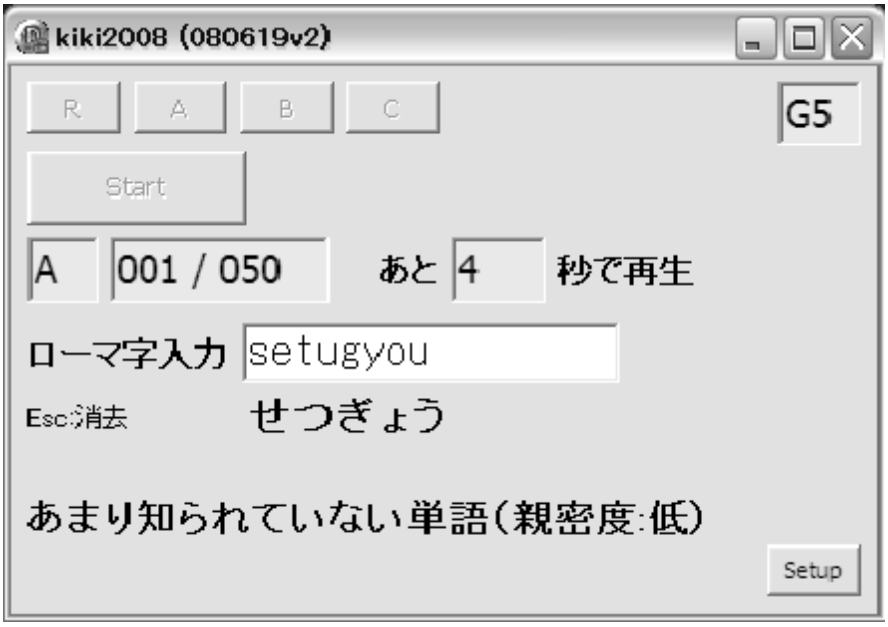

図 5.6　音声提示および回答入力のソフトウェアの画面.

使用した．NASA-TLX ソフトウェアは実験 1 とほぼ同じものを用いた．4 モーラの音声を 16 秒間隔で 1 つずつ聴取させて，親密度に関する提示を行い，解答を得るために，図 5.6 のような画面の音声提示用ソフトウェアを作成した．ローマ字で入力された文字列は逐次かな文字に変換して表示した．また親密度教示を行う場合は画面下に「あまり知られていない単語（親密度:低)」または「よく知られている単語（親密度:高)」と表示した.

被験者には，課題がすべて 4 モーラの単語であり，非常に早い音声であること，よく知られている単語も全く知られていない単語も含まれているが，聞こえてきたとおりに書き取るように，と教示した．また，被験者の回答に対する判定や正解のフィードバックは行わなかった.

最初にリハーサル課題として，単語親密度が中程度の女性音声 15 問を聴取させ，各自が聞きやすいと感じるように音量を設定させ，可能な限り聞き取りもさせた．次に，音声ガイドと文字表示で下位尺度の説明を行い，下位尺度の順位付けをさせ，ソフトウェアの操作に慣れるための練習の目的でリハーサル課題の負荷評定を行わせた.

その後 3 回の試行について，(1) 50 問の聴取を行わせて（所要時間は 15 分弱)，(2) 6 つの尺度の負荷評定を行わせ，(3) 疲労の影響を回避するために 5 分間の休憩を取る，という手順を繰り返した.

5.4.4　結果

了解度（モーラ単位での集計）についての被験者間の平均および標準偏差を表 5.5 に示す.

また NASA-TLX による WWL 得点を被験者ごとに平均 50，標準偏差 10 となるように正規

化した N-WWL 得点についての被験者間の平均および標準偏差を表 5.6 に示す．これらは被験者ごとに正規化されているため，課題群の間の比較はできない．

了解度の分散分析[*4]の結果を表 5.7 に示す．おおむね親密度の違いと慣れの効果を反映する結果となった．ただし G4 群と G8 群を比較すると，同じ課題 H-H-H にも関わらず，教示をしない場合のみ試行間の差が有意であった．

心的負荷を示す N-WWL 得点の分散分析の結果を表 5.8 に示す．課題 L-L-L である G1 群と G5 群は，教示の有無にかかわらず試行間の差は棄却された．課題 H-H-L である G2 群と G6 群は，教示をしない場合のみ試行間の差が有意であった．課題 L-L-H である G3 群と G7 群は，教示をした場合の方が有意差が強く支持される結果となった．課題 H-H-H である G4 群と G8 群に関しては了解度と同様，教示をしない場合のみ試行間の差が有意であった．

これらの値の変化を親密度教示の有無について比較するため，縦軸に了解度をとり横軸に N-WWL を取ったグラフを作成した．これを G1-G5, G2-G6 などの対にして図 5.7, 5.8, 5.9, 5.10 に示す．ただし四角は各試行の被験者間の平均を，上下のバーは標準偏差を表す．

5.4.5 仮説

実験 2 の結果を，5.4.1 で述べた仮説に基づいて解釈する前に，親密度の知覚に関して起こりうる状況をさらに具体的に述べる．

特に親密度の教示がない場合には，被験者は課題そのものによって語彙の親密度を知覚しなくてはならないため，以下の状況が予想される．

- 親密度を教示されず，親密度条件が途中で高 (FH) から低 (FL) に変化した場合には，それまで活用していたトップダウン情報が活用できなくなるため，被験者はすぐに変化を知覚する．
- 親密度を教示されず，親密度条件が途中で低 (FL) から高 (FH) に変化した場合には，被験者はしばらくトップダウン情報に頼らない聞き方を続けるかも知れないが，やがてトップダウン情報が活用可能であることに気づく可能性が高く，親密度条件の変化はゆっくり知覚される．

5.4.6 考察

前述の仮説に基づいて実験 2 の結果を考察する．

親密度を教示しない課題 H-H-H（G4 群）では，慣れの効果による了解度の上昇が特に顕著なように思えるが，他の条件の試行 3 と比較して了解度が高いとは言えない．しかし，この結果を

[*4] 実験 2 においても JavaScript-STAR を使用した．

「親密度が教示されないので，最初はトップダウン情報を活用しなかったが，次第に親密度の高い語彙であることに気づくことができ，トップダウン情報が活用されるようになり，試行全体にわたって了解度が上昇し心的負荷が減少した」と解釈すれば，前述の仮説は支持される．

逆に G8 群については「親密度が高いことを教示されたため，最初からトップダウン情報が活用され，了解度は最初から高く，心的負荷は最初から少なかった」と解釈できる．ただし G8 群の試行 3 についてはやや被験者間の分散が大きいことから，被験者群に何らかの偏りがあった可能性も否定できない．

課題 L-L-L である G1 群と G5 群について，心的負荷の試行間の差が棄却されたことは，「低親密度であることは聞き手にとって有効な情報ではなかった」と解釈できる．これらの群では了解度にも有意な変化はあったが，試行 2 から試行 3 にかけての変化は棄却されており，慣れの効果は限定的であると考えられる．

課題 H-H-L である G2 群の結果は「H-H と課題が続いて，聞き手がトップダウン情報を使う聞き方に慣れた後で，教示なしに L に切り替わると，それまでの方策が通用しなくなり，トップダウン情報に頼らない聞き方に切り替えるため，心的負荷が高くなる」と解釈できる．

同じく G6 群については前述の仮説からは「教示されて H から L に切り替われば，聞き手は即座にトップダウン情報に頼らない聞き方に切り替えるため，心的負荷が高くなる」可能性もあり，実験結果は仮説を支持していない．G6 群は試行 3 において，うまく方策を切り替えられなかった，あるいは，困難な課題と判断して無理な努力をしなかった，といった可能性がある．この点についてはさらなる検討を要する．

課題 L-L-H である G3 群については，「親密度が H になったことは教示がなくても気づきやすい」という仮説を支持している．

本実験の知見は，被験者の年齢および性別を限定したうえで得られたものである．今後は男女差の影響および加齢の影響を確認するための実験が必要である．

5.4.7 実験構成に関する課題

実験 2 の実験構成において G1-G4 と G5-G8 の被験者数に隔たりがあるが，ここではその理由について述べ，影響および課題について考察する．

実験 2 は当初，群 G1-G4 の 4 群の実験として計画され，被験者の募集が行われた．この時点では「被験者は教示がなくても親密度条件を容易に知覚できる」という前提であった．しかし実験を進めながら結果を分析し，また実験後に被験者に聞き取りをしたところ，この前提が自明ではない可能性が出てきた．そこで新たに 5.4.5 の仮説を検討し，G5-G8 の 4 群を追加した．前者 4 群と後者 4 群の被験者数が揃っていないのはこのような経緯による．一般に実験計画はあらかじめ設定した仮説に基づいて行われるべきであり，当初から 5.4.5 の仮説が検討されていれば，よりよい実験構成が可能であったと思われる．

本実験の結果から主張可能な点は限られている．しかし，表 7 および表 8 に示すように，G5-G8 についても，いくつかの条件において試行間の有意差が検出されたことから，個人差に左右されにくい数の被験者が確保されたと考えられる．またこれらの結果は 5.4.6 に述べたように，仮説を支持するひとつの結果として無視できない．

今後の課題として，親密度の教示の有無に関する被験者数を揃えて新たな実験を行い，了解度と心的負荷について，2 水準（教示の有無）× 3 水準（親密度の高低に関する順序）の分散分析を行うことによって，親密度教示の影響を直接比較できる可能性がある．

5.5 まとめ

単語親密度の異なる超早口音声の聴取実験を行い，単語了解度による評価と NASA-TLX の結果を比較した．実験 1 では被験者内比較によって，了解度と心的負荷が親密度の異なる超早口音声を評価する尺度として妥当であることを確認した．

実験 2 では被験者を複数の群に分けて，150 個の 4 モーラ単語を聴取させる実験を行い，被験者内比較によって慣れの効果を検討した．その結果，親密度が高いと教示された場合や，親密度が高いという条件を聞き手が自覚できた場合に，トップダウン情報としての心的辞書アクセスが促進され，了解度が高くなり心的負荷が少なくなる，という仮説を支持する結果が得られた．

本研究の実験結果は，「どういう内容なのか推測して聞くこと」「どういう内容なのかを早く適切に判断すること」が，一般的に「慣れ」といわれる現象と密接に関連しており，正しく楽に聞き取るための手がかりであることを示唆している．

本章で述べた知見は，超早口音声の評価手法の改良に役立つと期待できる．例えば，ひとりの被験者が超早口音声を大量に聞く場合の慣れの効果を抑制できれば，超早口音声を評価する実験は効率的になる．また，語彙を事前に教えて超早口音声で再認をさせるような課題であれば，聞き手の心的辞書が事前に活性化され，トップダウン情報に関する慣れが進んだ状態を作り出せる可能性がある．

今回の実験では，被験者はすべて女性で年齢の範囲も限られていた．しかし，今回得た知見に関して男女差の影響がないとは断定できない．聴覚については男性よりも女性のほうが可聴域が広いという知見 [65] もあるが，単語の聴取に与える影響は明らかではない．また，加齢変化による感音性難聴や認知機能などの低下も予想される．そこで今後，男女差および加齢の影響の検討が必要である．

さらに今後の課題として，時間をおいた場合の学習効果の持続性の検討，スクリーンリーダのための音声合成や録音図書プレイヤーの話速変換などへの応用などが挙げられる．また，本研究の知見は外国語学習などにも応用できる可能性がある．

表 5.5　実験 2 の結果（モーラ了解度 (%) の平均および標準偏差）.

Group	Trial 1	Trial 2	Trial 3
G1: L-L-L	54.7 (6.8)	62.1 (6.9)	62.2 (8.2)
G2: H-H-L	79.2 (5.6)	80.2 (3.9)	64.1 (6.0)
G3: L-L-H	55.1 (6.3)	62.4 (5.7)	83.8 (6.4)
G4: H-H-H	71.5 (8.1)	74.6 (10.3)	84.8 (4.6)
G5: L-L-L	53.6 (6.8)	61.4 (7.5)	62.9 (6.8)
G6: H-H-L	77.6 (5.4)	79.1 (7.7)	64.6 (5.8)
G7: L-L-H	51.6 (6.8)	58.5 (6.3)	83.3 (6.5)
G8: H-H-H	71.5 (6.6)	74.9 (6.9)	78.1 (12.4)

表 5.6　実験 2 の結果（N-WWL の平均および標準偏差）.

Group	Trial 1	Trial 2	Trial 3
G1: L-L-L	49.2 (10.8)	51.0 (9.6)	49.8 (9.4)
G2: H-H-L	48.5 (8.6)	43.0 (7.1)	58.6 (7.1)
G3: L-L-H	50.5 (6.7)	55.2 (9.2)	44.3 (10.6)
G4: H-H-H	58.3 (9.0)	49.4 (7.9)	42.4 (5.4)
G5: L-L-L	46.8 (10.4)	54.9 (4.1)	48.4 (11.8)
G6: H-H-L	46.8 (9.4)	50.8 (7.0)	52.5 (12.1)
G7: L-L-H	55.0 (8.2)	54.7 (5.9)	40.3 (7.4)
G8: H-H-H	49.8 (10.2)	54.0 (8.3)	46.2 (9.9)

表 5.7　実験 2 の結果（モーラ了解度の分散分析の結果）．**は $p < 0.01$, *は $p < 0.05$，ns は有意差なし.

Group	F 値	T1-T2	T1-T3	T2-T3
G1 (L-L-L)	$F = 30.45$ **	$<$ *	$<$ *	
G2 (H-H-L)	$F = 75.88$ **		$>$ *	$>$ *
G3 (L-L-H)	$F = 260.55$ **	$<$ *	$<$ *	$<$ *
G4 (H-H-H)	$F = 23.06$ **		$<$ *	$<$ *
G5 (L-L-L)	$F = 38.26$ **	$<$ *	$<$ *	
G6 (H-H-L)	$F = 25.05$ **		$>$ *	$>$ *
G7 (L-L-H)	$F = 150.81$ **	$<$ *	$<$ *	$<$ *
G8 (H-H-H)	$F = 2.11$ ns			

表 5.8　実験 2 の結果（心的負荷 N-WWL の分散分析の結果）．**は $p < 0.01$, *は $p < 0.05$, + は $p < 0.10$, ns は有意差なし．

Group	F 値	T1-T2	T1-T3	T2-T3
G1 (L-L-L)	$F = 0.08$ ns			
G2 (H-H-L)	$F = 9.96$ **		$<$ *	$<$ *
G3 (L-L-H)	$F = 3.28$ +			$>$ *
G4 (H-H-H)	$F = 9.53$ **	$>$ *	$>$ *	
G5 (L-L-L)	$F = 0.97$ ns			
G6 (H-H-L)	$F = 0.42$ ns			
G7 (L-L-H)	$F = 5.43$ *		$>$ *	$>$ *
G8 (H-H-H)	$F = 0.66$ ns			

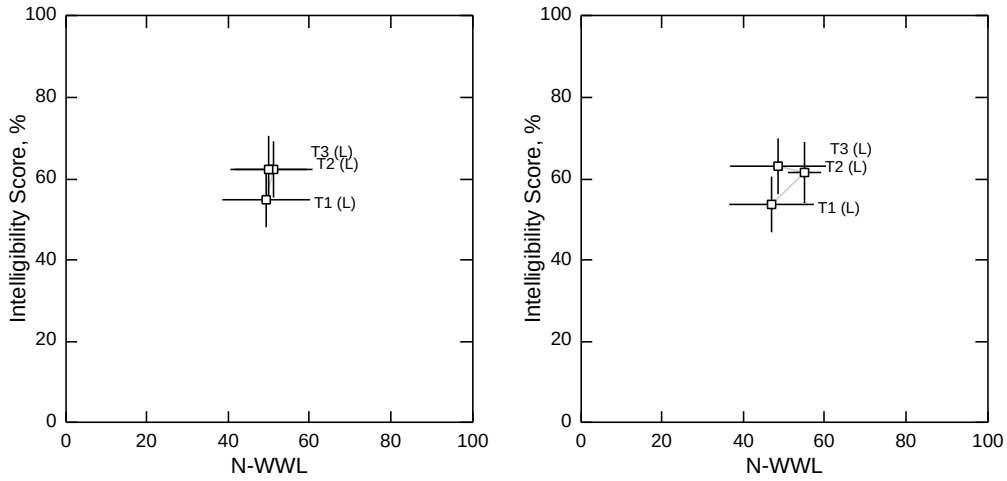

図 5.7　実験 2 の結果（親密度 L-L-L: G1（上），G5（下）における N-WWL と了解度の推移）．

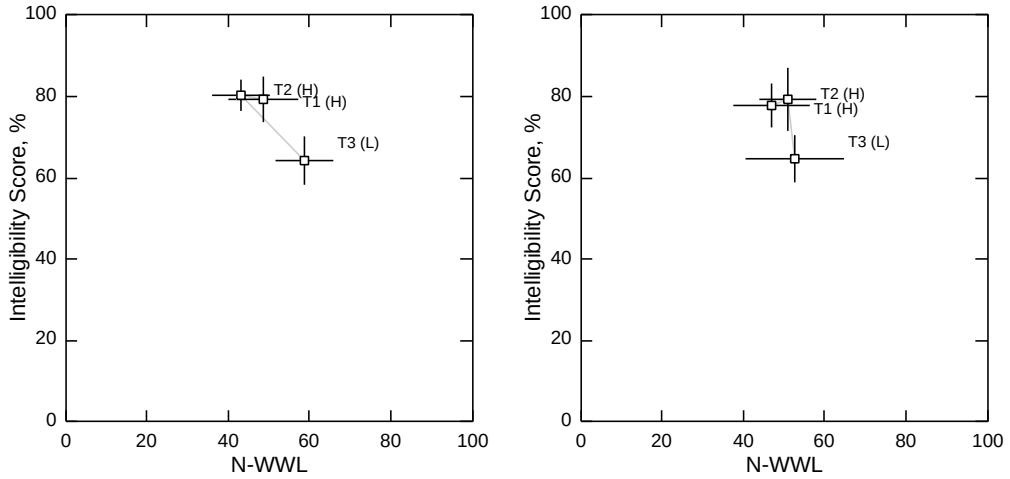

図 5.8　実験 2 の結果（親密度 H-H-L: G2（上），G6（下）における N-WWL と了解度の推移）.

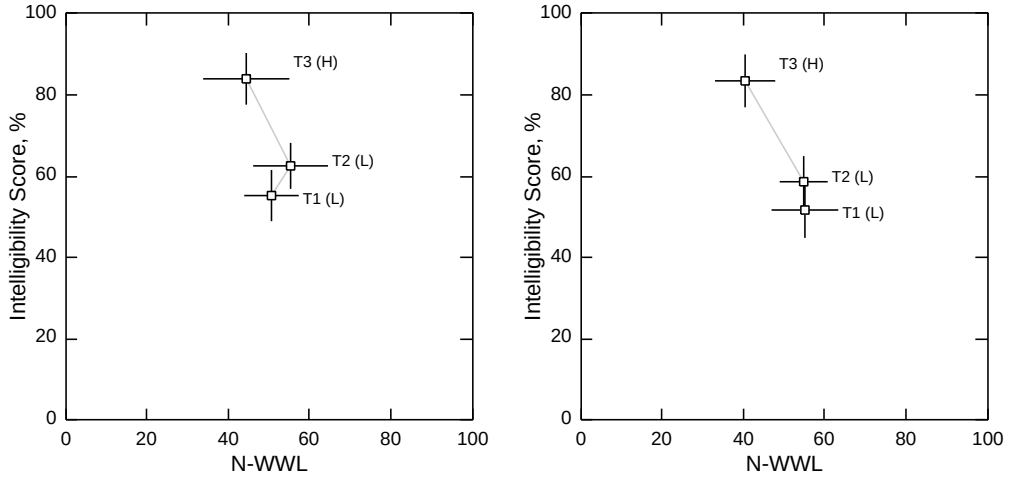

図 5.9　実験 2 の結果（親密度 L-L-H: G3（上），G7（下）における N-WWL と了解度の推移）.

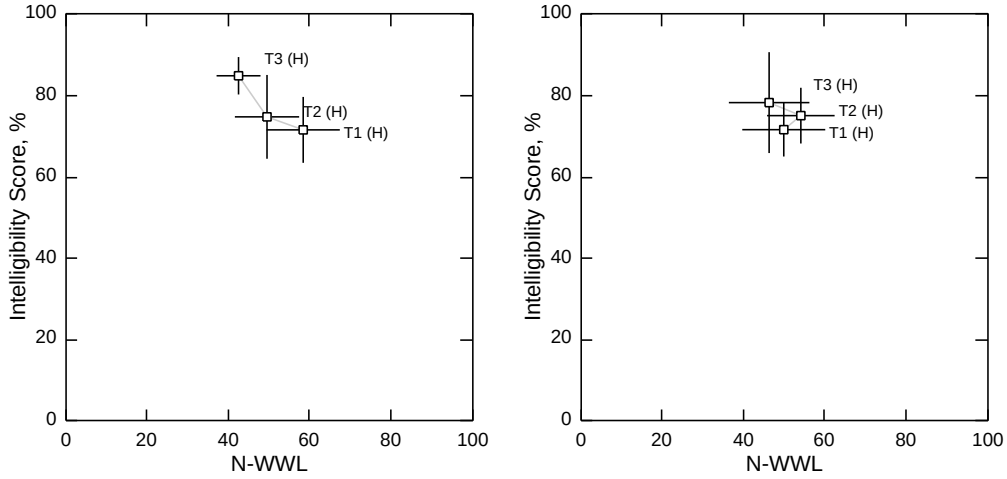

図 5.10　実験 2 の結果（親密度 H-H-H: G4（上），G8（下）における N-WWL と了解度の推移）.

第 6 章

さまざまな音声インタフェース研究

6.1 動機付けと楽しさの研究

インタフェース導入原則の「有用性の原則」は，ユーザに「使ってみたい」と感じさせるための「動機付け」の重要性を主張している．

例えばデザインの「美しさ」は近年インタフェース設計において重視されるようになった．Norman は「エモーショナル・デザイン」[72] の議論において，美しさ・楽しさなどを与えてくれるインタフェースシステムは，ユーザのタスク達成率が有意に高くなる，というエピソードを紹介している．この理由についてノーマンは，ユーザが楽しいと感じるほど，操作がわかりにくい場合や予期しない反応があったときに，積極的に試行錯誤を行って問題を解決するようになるため，という考えを示している．

視覚障害者向けタイピング練習ソフト「ウチコミくん」の開発 [73] においては，当初は視覚障害者のパソコン教室のニーズを踏まえた実用性が重視されたが，タイピング練習においては動機付けが重要という立場から，専門家が参加して「楽しさ」を盛り込んだコンテンツが制作された．

もっと直接的に「楽しさ」を扱った研究もある．

コンピュータとのインタラクションにおける楽しさを扱った先行研究として，山本ら [74] は，キーボード操作による「しりとり」ゲームを用いた実験を行い，特に「相手が人間であると思うこと」が被験者に楽しさを与えると指摘している．

チクセントミハイは「フロー体験」に関するモデルを提唱している [75, 76]．フロー体験とは内発的動機づけ（金銭的報酬など外部からの働きかけによらず，「対象自身が面白いから」など本人に内在する動機づけ）に基づく楽しさをもたらす体験であり，その成立条件は図 6.1 のようにモデル化されている．我々が行おうとしている課題に対して持っている技能レベルと，その課題を行うことの挑戦レベルがあるレベルで均衡するときに，快適に楽しく目標を達成できる状態が生じる．例えば，対戦ゲームを例に取れば，挑戦レベルは相手の技能レベルに依存するた

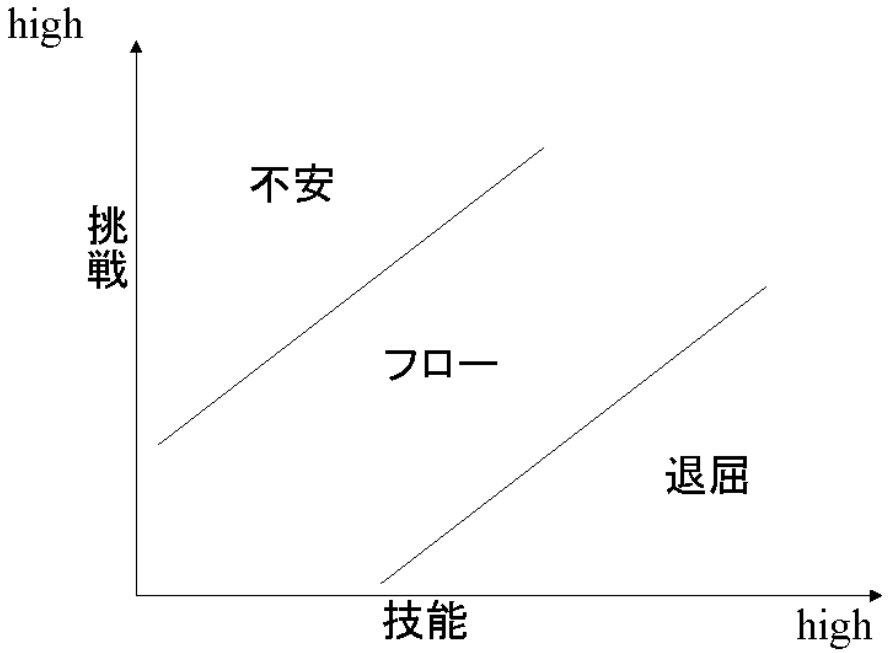

図 6.1　フロー体験モデル.

め，二人の能力が釣り合っていなければ楽しくない．能力の低い者は不安や緊張を感じるであろうし，高い者は退屈や不満を感じるであろう．二人のレベルが近ければ近いほど，お互いの技能レベルと挑戦レベルが釣り合うことになり，楽しいと感じることができる．

フロー体験には文献 [77] など各種の関連研究がある．フロー体験がもたらすインタラクションの快適性や効率性に注目した研究を以下に挙げる．

キーボードによるチャットを用いた研究 [78] においては，目的指向会話を挑戦の高いタスク，自由会話を挑戦の低いタスクとする．挑戦の高いタスクにおけるタイプスキルの高い被験者と挑戦の低いタスクにおけるタイプスキルの低い被験者がそれぞれ会話中にフロー体験をしていたこと，技能に比較して高い挑戦が与えられた被験者はスキルが向上したこと，などが検証されている．

Webster ら [79] は，コンピュータ操作における楽しさの構成要素を，自己の制御，注意の集中，好奇心，内発的動機づけ，の 4 項目に整理している．さらにフロー体験がソフトウェアの柔軟性に肯定的な印象を与える，ユーザの探険的操作を促す，将来の自発的なコンピュータ利用を促す，などの仮説を，表計算ソフトウェアおよび電子メールシステムを用いて検証している．

Wiedenbeck ら [80] は，コンピュータ操作の学習において，対話スタイルと学習経験の差異がフロー体験の有無に影響すること，フロー体験の有無がアプリケーションの使いやすさや有用性の知覚に影響すること，これらの知覚がタスク達成能力に影響すること，などのモデル化を行い，GUI (Graphical User Inteface) によるワードプロセッサとキーボード操作によるスク

リーンエディタを用いて検証している．

我々はフロー体験モデルを踏まえて，音声対話ゲームにおける音声認識の自己目的性の検証を試みた [81]．その結果によれば，音声認識は必ずしもシステムを親しみやすくする手段としては知覚されておらず，音声認識の使用は創造的で知的な問題解決，あるいは演技などの表現活動と類似した体験として知覚されていた．また，認識率を高めようとするユーザの問題解決的な努力が特に自己目的的な楽しみの対象になり得ること，音声認識に対する苦手感覚の克服が必要であること，などが確認された．

このような研究において楽しさを定量的に比較することは難しい．多くの先行研究では，被験者に対するアンケートに，楽しさや内発的動機づけや作業への没入度に対応する項目を設定し，アンケート回答に基づいて楽しさを評価している．しかし，楽しさには，内観により本人が自覚できるものだけではなく，明確には自覚されないが無意識に表情や行為に表れるものもあると考えられる．

関連研究 [84] ではディクテーションシステムのユーザに前向きな努力を促すために，ディクテーション作業を自己目的的に楽しむことが重要であると考え，ディクテーション作業の自己目的的な楽しさがどのような要因によってもたらされるか，また，ユーザの態度や認識性能等が楽しさにどのような影響を与えるのかを検討した．

また，Web ページに含まれるリンクやキーワードを音声コマンドで選択できる音声ウェブブラウザ VOXplorer [85, 86] の評価においても，未出版の実験ではあるが，音声コマンドの利用によって「挑戦と技能のバランス」を制御できる可能性が示唆されている．

このような研究は，エンタテインメント性を持つ音声応用システムの実現において特に有用であり，今後の発展が期待される．

6.2 音声インタフェースにおける 7 段階モデル

Norman の 7 段階モデル [6] は本研究においてはインタフェース原則の基礎となる理論の一つである．7 段階モデルにおける「実行の淵」「評価の淵」はインタフェースシステムにおける現象を述べたものであるが，これは具体的なインタフェース設計に当てはめると，インタフェース基本原則の「手順連想容易性」「理解容易性」に対応している．

本節では，音声インタフェースのように，時系列上のイベントとして情報を受け取って操作を行うシステムを，この 7 段階モデルに基づいていかに解釈すべきかという問題を論じる．

電話音声応答システムの試作と評価 [87] からは，Balentine らの主張 [34] とも重なる知見が得られた．この研究においては，図 6.2 のような 7 段階モデルの拡張が検討された．

音声認識と音声合成を用いた電話音声応答システムにおいて，利用可能な音声認識や自然言語理解の性能は十分ではない場合には，ユーザの自由な発話を許すことは困難であり，音声入力をどのように使うことが可能であるかを，システムに不慣れなユーザにもわかりやすい方法で

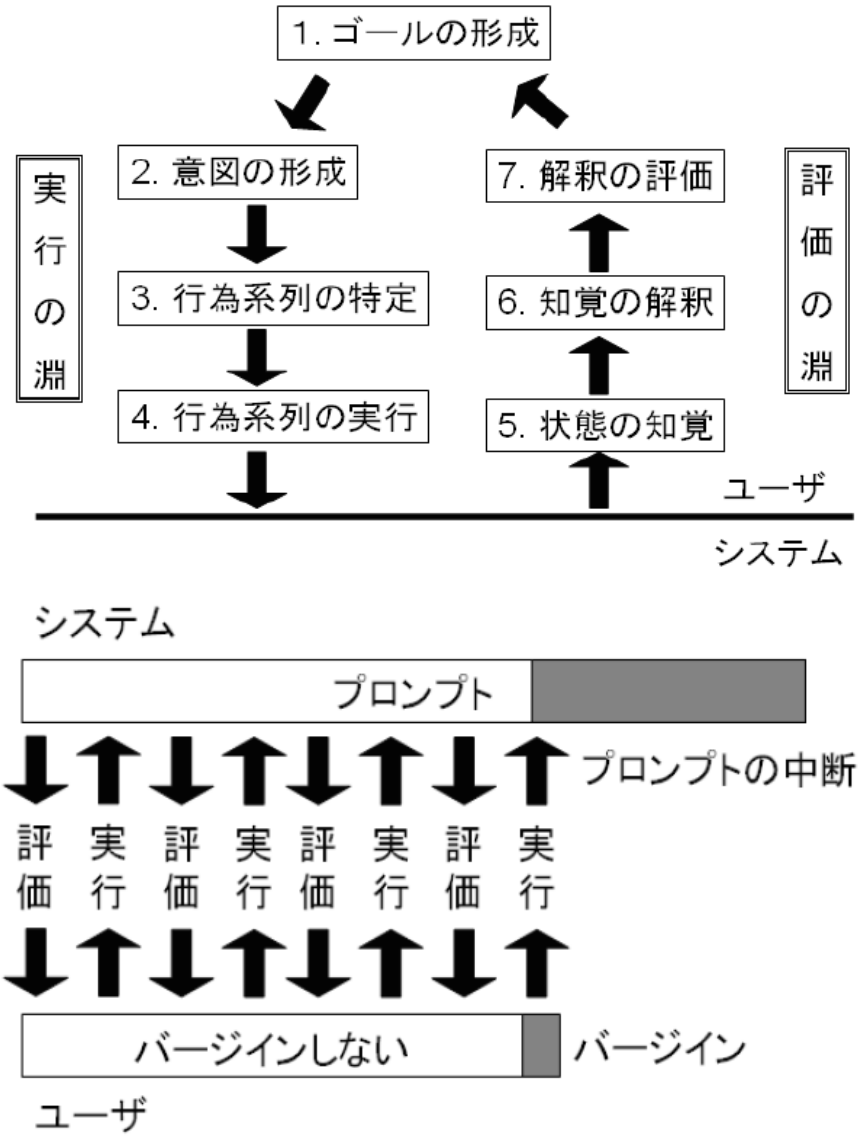

図 6.2　7 段階モデル（上）と音声インタフェース（下）.

システムが提示することが重要である.

このような場面で起こりうるユーザの困難を，Norman のインタフェース行為モデルに基づいて整理すると,「実行の淵」とは音声入力を行うにあたっての問題であり，評価の淵とは音声出力（プロンプト）を聞くにあたっての問題である．ユーザがプロンプトを聞く過程を詳細化すると次のようになる.

- システムの状態を知覚する： 出力された音声を（例えばテキストとして）知覚する.
- 状態を解釈する： 知覚したプロンプトを意味解釈する（例えばどんな意味的内容の選択肢を提示しているか）.

- システムの状態を評価する： 解釈した意味を，入力実行時に意図した目標と比較評価する（例えばタレントの名前を探していたのであれば，探していたタレントの名前か，など）．提示された情報に期待した項目があったか，いままで行った自分の入力操作は正しかったか，といった判断を行う．

ユーザの 1 発話とシステムのプロンプト 1 つを単位として実行と評価が繰り返すモデルを，本研究では「マクロな実行・評価モデル」と呼ぶ．一方で，バージインを前提とする場合は，ユーザは少しずつプロンプトを聞き，任意のタイミングで発話することができる．従ってユーザは「バージインするかしないか」という判断も含めて，1 つのプロンプトを聞きながら，何度も評価と実行を繰り返している．これを本研究では「ミクロな実行・評価モデル」と呼ぶ．

6.3 プロンプトにおける効果音と言語情報の役割

6.3.1 概要

音声応答システムにおいて，プロンプトがもたらす情報の役割を整理し，特に非言語情報としての効果音と言語情報の役割分担に関する考察を行う [87].

ただし本研究では扱う問題を単純化しており，例えば音楽を音声の BGM として用いることや，複数の効果音に異なった意味を持たせる，といった音楽や効果音の利用は扱わない．

また，本研究では実現の容易さという観点から，非言語情報としての効果音の利用を提案する．しかし，音声を部分的に強調する，男性音声と女性音声を併用する，無音を挿入する，といった非言語情報の利用も考えられる．本研究は，効果音でなければ実現できない用法を提案するものではない．

6.3.2 実行の淵を埋める役割

まず，実行の淵を埋める役割について検討する．プロンプトがコマンド語彙を含むことによって，ユーザは有効なグラマーを知ることができる．

> 例： スタートメニューです．交通アクセス，イベント情報，チケット情報のうち，知りたい情報を音声でお答え下さい．

しかし，例えば以下のプロンプトはコマンド語彙を何も含んでいないが，そのことを情報として伝えることはできない．

> 京阪線から，三条駅で地下鉄の東西線二条行きに乗り換えてください．それから，烏丸御池駅で地下鉄烏丸線国際会館行きに乗り換えてください．松ヶ崎駅で下車してください．１番出口より東へ徒歩５分です．

プロンプト中のある語彙が有効なコマンドであるか否かを，より明示的に伝えることで，ユーザの無効な発話を防ぐことができる．

我々は，コマンド語彙の直前に，効果音 (SE1) を挿入することを提案する（以下の例で (SE1)(SE2) などは効果音を示す）．

> 例： スタートメニューです．(SE1) 交通アクセス，(SE1) イベント情報，(SE1) チケット情報，のうち知りたい情報を音声でお答えください．

これにより，単にコマンド語彙を強調するだけでなく，効果音が付与されていない語彙がコマンドでないことも容易に知覚できる．

6.3.3 評価の淵を埋める役割

次に，操作に対するフィードバックを充実させ，評価の淵を埋めるようなプロンプトの用法について検討する．

ある操作を行う前に，操作がもたらす結果を予想させることができれば，早い段階で自らの操作を評価し，誤った操作を防ぐことができる．次の例では，「イベント情報」を選択したユーザに対して，さらに深い階層に移動した後にどんな項目が選択可能であるかを，前もって提示することを試みている．

> 例： イベント情報です．**各イベントの時間・内容と，整理券についてお知らせします．**連日のイベント，23 日のイベント，24 日のイベント，25 日のイベントから音声でお答えください．

操作に伴ってどの状態に遷移したのかを知らせたり，入力が可能であるかを示すプロンプトも，ミクロな実行・評価におけるフィードバックへの配慮となる．

> 例： スタートメニューです．
> 例： お待ち下さい．

ところで，プロンプトの途中に命令文があると，そこで入力可能であるとユーザが判断してしまい，開発者の意図に反して発話が促されることがある．次の例ではユーザは，強調部分を聞かずにバージインしてしまう可能性が高い．

> 例： イベント情報です．連日のイベント，23 日のイベント，24 日のイベント，25 日のイベント，から音声でお答え下さい．**各イベントの時間・内容と整理券についてお知らせします．**

初心者の立場からは，未知のプロンプトにはバージインせず，なるべく最後まで聞くことが望

ましい．一方で，熟練者の立場からは，プロンプトに対して，バージイン操作をしてもよいかどうか，なるべく早く判断できることが望ましい．また，プロンプトの内容は，例えば，すでに聞いたことがあるかどうか，といった判断により，なるべく早い段階で予想できることが望ましい．

我々は，プロンプトが終了したことを効果音 (SE2) で示すことを提案する．

> 例： (SE1) 連日のイベント，(SE1)23 日のイベント，(SE1)24 日のイベント，(SE1)25 日のイベントから音声でお答えください．(SE2)

この効果音 (SE2) は，次の音声入力が可能であることを示すと同時に，効果音が鳴っていない場合はプロンプトがまだ継続している，ということも示している．

次に，プロンプトを冗長化することで，フィードバックを提供する例を挙げる（S はシステム，U はユーザを表す）．

> S: イベント情報です．各イベントの時間内容と整理券についてお知らせします．(SE1) 連日のイベント，(SE1)23 日のイベント，(SE1)24 日のイベント，(SE1)25 日のイベント，から音声でお答えください（SE2)
> U: 連日のイベント
> S: **23 日から 25 日までの連日**，(SE1) フリーマーケット，(SE1) はしっこ企画，(SE1) 模擬店・教室展示，が行われます．知りたい情報を音声でお答えください．

強調部分は，ユーザが選択した項目を復唱したり，言い換えたりすることで，ユーザの意図とシステムの状態を比較しやすくしている．ミクロな実行・評価を助けるためには，バージイン発話によってプロンプトが瞬時に停止すること，認識が成功したことを効果音で示すこと，なども重要である．

我々は，プロンプトの先頭に効果音 (SE3) を挿入することを提案する．これによって直前の音声認識が成功したことを明確に示せる．

> U: イベント情報.
> S: (SE3) イベント情報です．各イベントの時間・内容と，整理券についてお知らせします．(SE1) 連日のイベント，(SE1)23 日のイベント，(SE1)24 日のイベント，(SE1)25 日のイベントから音声でお答えください．(SE2)

これらの提案をシステムに適用した実験の結果，適切に効果音を用いることで，効果音を用いないシステムと比較して，語彙外の発話を削減し，より適切なバージインの使い方を促すことが可能となった．

6.4 視覚障害者のための音声インタフェースの検討

視覚障害者がスクリーンリーダを介して Web や電子メールを使用する状況は，音声インタフェースの観点から検討できる [88].

例えば一般的なスクリーンリーダは，晴眼者向けのアプリケーションに音声読み上げ対応の機能を付加するものである．このようなアプリケーション環境の操作においては，Norman の 7 段階モデルにおける「実行」と「評価」の不自然な対応が，例えば以下のような形で発生し，「使いにくさ」につながっている．

- キー操作に対する音声フィードバックの遅れ・「音切れ」の悪さ
- 超早口化されて聞き取り困難な合成音声
- キー操作と読み上げ音声の不自然な因果関係からシステムの状況を知覚する必要性
- 本文にジャンプするためにコンテンツを後ろからさかのぼってたどる裏技

音声インタフェースに固有の快適性や性能を十分に追求することは，このような問題の解決に繋がる．第 5 章で述べた超早口音声の研究もその一つである．

さらに，GUI の背後に存在するタスクの本質的な構造を抽出し，音声インタフェースとして実装することも有効である．対面朗読者と視覚障害者の対話を分析し，その主たる要素を「ショッピングカートモデル」として Web アプリケーションに実装し，音声ブラウザを用いて評価した研究 [89] はその一例である．こうした観点から Web アクセシビリティにおける「インタフェースの原則」の適用可能性を検討することは有効である．

アクセシビリティとセキュリティの両立は，支援技術において難しい課題である．我々は第 5 章で用いた NASA-TLX 法を応用して，人間と機械を認証する CAPTCHA 技術の音声版について，安全性と心的負荷の両面から検討を進めている [90].

6.5 頭部モーションセンサと音声を用いたインタフェース

6.5.1 研究の目的

本節では，音声対話に伴って人間が頭部を動かす動作に着目し，3 次元モーションセンサを用いて頭部動作を測定し，これを音声入力と組み合わせて利用する，という入力インタフェースについて検討する [91]．必ずしも自然な人間の動作の認識を対象とするのではなく，習熟が容易で，頑健で確実な入力手段を検討する．具体的には，ユーザの音声入力に対するシステムからの応答を聞きながら頭を縦や横に振るなどして，肯定や否定などの意志表示を行なう入力インタフェースの提案を行う．また，効果音によって動作認識システムの内部状態をユーザに提示す

ることの有効性についても合わせて検討する．

音声入力によってシステムに指示を行ない，システムからの音声出力のみによってフィードバックや情報を得る，といったインタラクションは，特に車載音声インタフェースや視覚障害者の支援技術など，ユーザが視覚的なフィードバックを得にくい場合に有効である．

入力手段としての音声認識には，手や視覚を拘束されないという利点がある．また，大語彙かつ自然な発話を正しく認識できれば，効率的で自然な入力インタフェースとなることが期待される．しかし多くの場合には，音声による入力に加えて，発話の開始をシステムに知らせたり，認識結果として得られる複数の候補から適切な項目を選択したり，入力された内容をキャンセルしたり，誤認識の訂正などを行なう，といった操作が必要となる．音声認識の結果がたとえ完璧であっても，入力途中でユーザの気が変わった，といった場合もありうる．したがって，このような補助的な入力手段は，将来，音声認識性能が十分に向上した場合であっても重要であろう．

6.5.2 関連研究

音声認識と他の入力手段を併用するシステムとしては，音声認識とセンサを併用する MIT の Put-that-there システム [92] をはじめ，タッチパネルの併用 [93, 94]，マウスやキーボードの併用 [118] など，さまざまな提案がなされている．これらは，音声認識という入力手段の利点を生かしつつ，音声入力では困難な空間的な位置や座標の入力，誤認識時の訂正など，音声入力の弱点を補うものである．また，モーションセンサをヒューマン・マシン・インタフェースに利用するシステムとして，Toss-it[95] がある．これは，相手の PDA にボールをトスするかのように自分の PDA を振ることで，直感的な情報の移動を行うものである．

6.5.3 頭部運動を用いた対話の提案

人間同士の対話において，肯定・否定の頭部動作は，「はい」「いいえ」などの発話と同時に生じることが多い．また，相手が自分の声を聞き取れない場合などには頭部動作のみでも情報が伝わる．対話における確認のやりとりをマルチモーダル化することにより，ユーザにとって信頼できる入力システム，つまり，効率性や確実性の高いシステムを実現できる可能性がある．

インタフェースシステムに動作認識を用いる場合は，

1. 人間の自然な動作をできるだけ高精度に認識する．
2. システムが確実に認識できるような動作を対象とし，ユーザがそのような動作を確実に行えるように練習をする．

という 2 つのアプローチが考えられる．ユーザに強いる負担をできるだけ減らす，という立場からは，前者が理想的である．しかし，入力そのものが自然な行為であっても，誤認識によっ

て使い勝手が損なわれる場合には，ユーザの負担は大きくなる．これはすでに音声認識の応用において生じている問題である．

- 人間同士の対話における自然な頭部動作は，同じ意図の動作においても個人差やバリエーションが大きく，頑健なパターン認識は容易ではない．

本研究は音声認識を補完する入力手段を目指すために，後者のアプローチを選ぶ．その際，ユーザの負担を軽減するために，以下の配慮が重要ではないかという仮説を立てた．

1. 人間の自然な動作にできるだけ近い動作を用いることで，ユーザの学習が容易になる．
2. 用いる動作を，ユーザが意識しやすい複数の状態に分割することで，ユーザの学習が容易になる．（例：「頭を下げる」「しばらく待つ」「頭を上げる」）
3. 特に重要な状態遷移が起きた場合には効果音を提示することで，ユーザの学習が容易になり，確実な入力が可能になる．

6.5.4 頭部運動データの予備的検討

本研究で使用する NEC Tokin 製 3D モーションセンサ MDP-A3U9S の仕様を以下に示す．

- ロール角 (X 軸) 検出範囲：± 180 deg
- ピッチ角 (Y 軸) 検出範囲：± 90 deg
- ヨー角 (Z 軸) 検出範囲：± 180 deg
- インタフェース：USB 1.1
- 外形寸法：20 mm × 20 mm × 15 mm
- 重量：6 g
- 対応 OS：Windows 98/Me/2000/XP

本研究では，モーションセンサを図 6.3 のように帽子の頭頂部に付け（以下，「センサーハット」と呼ぶ），29.97 サンプル／秒でデータを取得した．このセンサは，ロール角（X 軸），ピッチ角（Y 軸），ヨー角（Z 軸）を取得することができる．

否定および肯定の頭部動作を順番に行った場合の角度とその差分・二次差分の値の例を図 6.4〜6.6 に示す．図 6.4 （角度値）においては，否定の動作（首を横に振る）でヨー角が正および負に 1 回ずつ変化する．また，肯定の動作（首を縦に振る）でピッチ角が負に 1 回変化する．得られる値は連続的であり，センサの出力値にはノイズがほとんどない．図 6.5 （角度の差分）においては，否定の動作でヨー角が「正→負→正」のパターンで変化する．また，肯定の動作でピッチ角が「負→正」のパターンで変化する．

図 6.3　帽子に取り付けた 3D モーションセンサ.

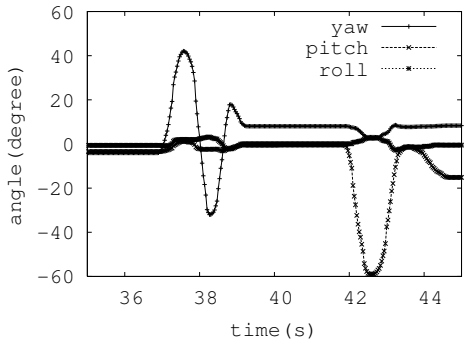

図 6.4　3D モーションセンサの出力例（角度）.

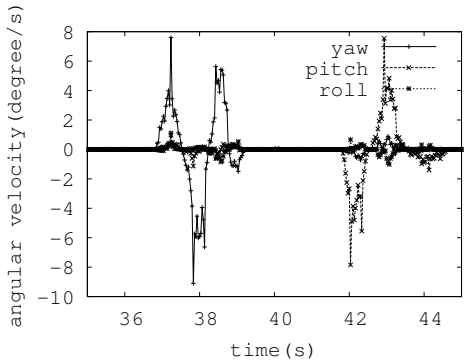

図 6.5　3D モーションセンサの出力例（角速度）.

6.5.5　頭部運動データの認識

我々は頭を縦または横に傾けて戻すという一連の動作をいくつかの状態に分割し，それぞれの状態遷移に時間や角度の閾値を設定する，という状態遷移モデルを用いて，Toss-it での速度

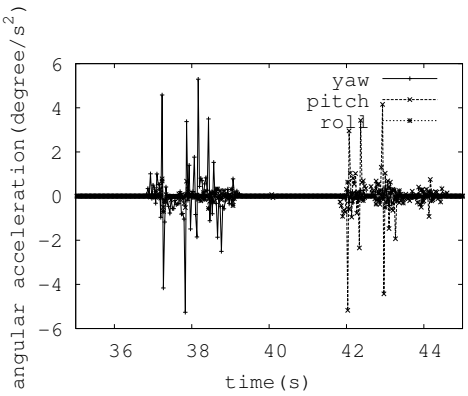

図 6.6　3D モーションセンサの出力例（角加速度）.

の積分値との比較と同等な頑健性を目指すことにした．その際,「意図した入力を確実に行う」「意図しなかった入力を確実に回避する」の両者についてユーザがコツをつかみやすくすることを重視する.

図 6.7 にこのアルゴリズムで用いた状態遷移モデルの概要を示す．このモデルでは，頭部の角度および角速度によって状態が遷移する．図中の角度値，角速度値における矢印は閾値を表しており，縦の矢印は角度値，横の矢印は時間を意味している．なお，簡略化のため角度値，角速度値は等速度運動に単純化してある.

遷移の条件を表 6.1 に表す．現在の角度値を「D」と表し，時間値を「T」と表す．なお，状態が遷移する際の最終的な角度値を「D_状態番号」として表し，角速度値を「V_状態番号」として表す．例えば，状態が 1 から 2 に遷移する際の最終的な角度値は「D_1」とする．また，遷移時の条件として設定した角度値における閾値は「d_閾値番号」として表し，時間値における閾値は「t_閾値番号」として表す．初期化時の状態番号は 0 とする.

図 6.7 の状態遷移モデルについて，状態 5 は受理を意味している（以下,「Accept」と呼ぶ）. また，状態が 2 から 3 に移る際に（表 6.1 におけるイベント 5），状態 4 に移ることができる合図である効果音（以下,「効果音 A」と呼ぶ）が鳴る．全ての状態は初期化を経て状態 1 に遷移することができる（以下,「Reject」と呼ぶ）.

このアルゴリズムを用いて, 頭の頷き（以下,「Positive」と呼ぶ）とかしげ（以下,「Doubtful」と呼ぶ）のジェスチャーの認識システムを実装した．開発にはモーションセンサの SDK が提供する API を使用し，Microsoft Visual C++ 6.0 を用いてコマンドラインのプログラムを作成した．動作環境は Microsoft Windows XP である.

Positive の認識には Pitch の角度値，角速度値を用い，Doubtful には Roll の角度値，角速度値を用いた．利用者は頷きをし，効果音 A が聞こえたときに頭を上げることで入力することができる．もし，途中で入力の意思を変えたときや誤入力が起きたときは，少しの間頭を動かさないことで入力をキャンセルすることができる.

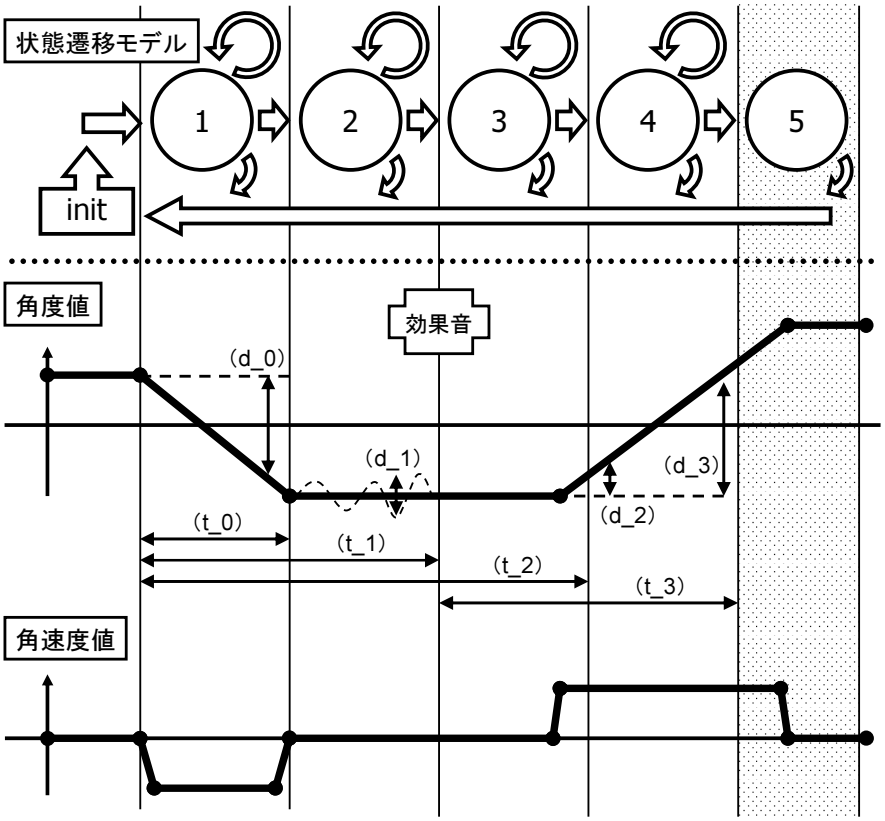

図 6.7　頭部の角度および角速度の状態遷移モデル.

なお，状態 3 および 4 から動作をキャンセルした場合（表 6.1 におけるイベント 10,11,12）には Reject を示す効果音（効果音 B）が鳴る．また，状態 5 に到達した場合（表 6.1 におけるイベント 13）には Accept を示す効果音（効果音 C1/C2，それぞれ Positive と Doubtful に対応）が鳴る．これらにより，Accept や Reject が確実に行われたことを，ユーザは効果音によって知ることができる．

6.5.6　状態遷移モデルを用いた実験

肯定や否定の意思をコンピュータに入力するタスクを設定した．入力手法は Positive と Doubtful の二種類とする．POI を照合するタスクとして，以下のような課題を行わせる．

1. 被験者はセンサーハットをかぶり，PC の前に座る.
2. 被験者は 1 分間程で自分にあった入力のタイミング（閾値）を選ぶ.
3. モニターに地名が表示され，スピーカーから地名を表す合成音声が流れる.
4. 被験者は表示された地名と聞えた音が同じかどうかを，頭の動きだけで PC に入力する．同じであった場合は Positive を，異なっていた場合は Doubtful を入力する.

表 6.1　頭部モーションセンサシステムにおける状態遷移の条件.

イベント番号	状態の変化	条件
1	1 → 初期化 → 1	T - T_0 が t_0 より大きいとき
2	1 → 2	イベント 1 でなく，D - D_0 の絶対値が d_0 より大きく，T - T_0 が t_0 より小さいとき
3	1 → 1	イベント 1,2 でなく，T - T_0 が t_0 より小さいとき
4	1 → 初期化 → 1	イベント 1,2,3 でないとき
5	2 → 3	T - T_1 が t_1 より大きいとき
6	2 → 初期化 → 1	イベント 5 でなく，D - D_1 の絶対値が d_1 より大きいとき
7	2 → 2	イベント 5,6 でないとき
8	3 → 4	D - D_2 の絶対値が d_2 より大きいとき
9	3 → 3	イベント 8 でなく，T - T_2 が t_2 より小さいとき
10	3 → 初期化 → 1	イベント 8,9 でないとき
11	4 → 初期化 → 1	T - T_0 が t_2 より大きいとき
12	4 → 初期化 → 1	イベント 11 でなく，T - T_2 が t_3 より小さいとき
13	4 → 5	イベント 11,12 でなく，D - D_3 が d_3 より大きいとき
14	4 → 4	イベント 11,12,13 でない場合
15	5 → 初期化 → 1	無条件

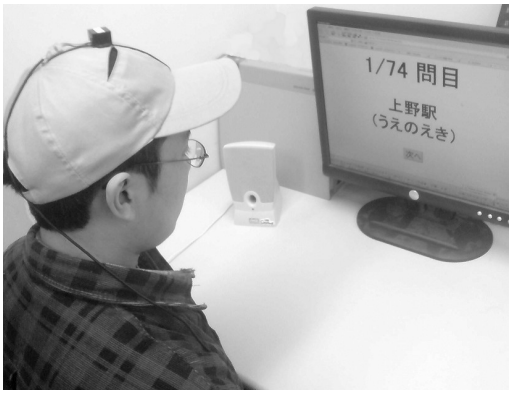

図 6.8　頭部モーションセンサを用いた実験の様子.

5. 3, 4 を全ての問題が終わるまで繰り返す.

実験の様子を図 6.8 に示す．実験に用いた地名は，事前に合成音声として作成したものを用いる．また，被験者にはウェブブラウザ上で入力を行ってもらう．被験者の入力を JavaScript を用いて検知することで，被験者は自分自身で問題を進めることができる.

問題は 74 問あり，答えの半分が〇で半分が×である．被験者は 21 歳から 25 歳までの今までに実験に参加したことのない理工系学生 8 人（男性 6 人，女性 2 人）である.

表 6.2　状態遷移モデルにおける結果数.

	結果 Positive	結果 Doubtful
正解 Positive	296	0
正解 Doubtful	0	296

表 6.3　状態遷移モデルにおける Reject の発生回数.

	Positive	Doubtful
効果音 A 前 Reject	49	160
効果音 A 後 Reject	48	190

6.5.7　実験結果

8 人の被験者から取得したデータは合計約 35 分，各被験者のセッションは平均約 4 分である．各被験者は 1 回の実験で 74 回の入力を行った．得られた動作サンプルがどのように認識されたかを表 6.2 に示す．また，Reject がどのような状態で発生したかを表 6.3 に示す.

表 6.2 の結果より Positive，Doubtful 共に意図しない誤入力は発生することはなく，頑健に動作した.

表 6.3 の結果より Doubtful は Positive と比べ Reject の回数が多くあった．被験者の中には，入力の Reject を自分自身の意思で行えない被験者がいたが，意図しない入力の際に，頭の動きが無意識に止まっていたために，正しく Reject することができていた.

被験者ごとの Reject の回数を図 6.9 に示す．Reject が発生する回数は被験者によって大幅な差があり，少ない人で 1 回，多い人で 205 回であった．全ての被験者は初心者であるため，Reject が多い人はセンサーハットのコツと最初に触ったときのイメージが離れていると考えられる.

練習によるなれの効果を知るため，Reject の回数が多い被験者 2 人に対して約 5 分の練習を行ってもらい再実験を行った．その結果を図 6.10 に示す．練習では我々が被験者に口頭でコツの教示のみを行った.

本結果では，Reject の回数が 8%, 20% と非常に減った．このことより，僅かな練習で頑健に入力できるようになることがわかる.

6.5.8　フィードバックの必要性

次に，効果音 A の有無によって，認識にどのような影響を与えるかを実験した．被験者は Reject の少ない 2 名を選び，状態遷移モデルでの状態が 2 から 3 に遷移する際の効果音 A を

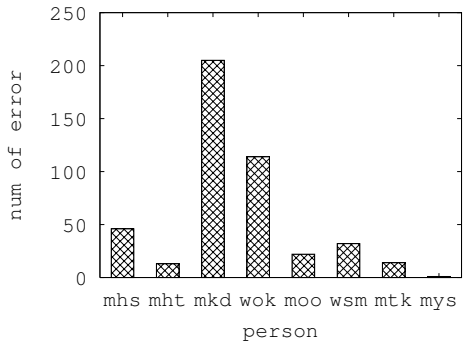

図 6.9　74 回の入力に対して発生した Reject の回数.

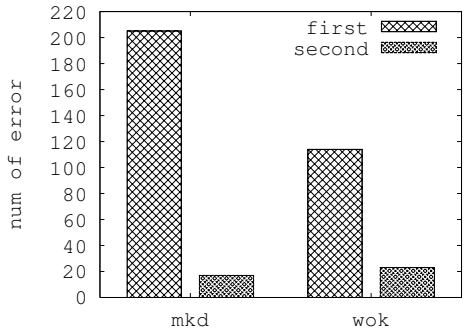

図 6.10　練習を行った際の Reject の回数の推移.

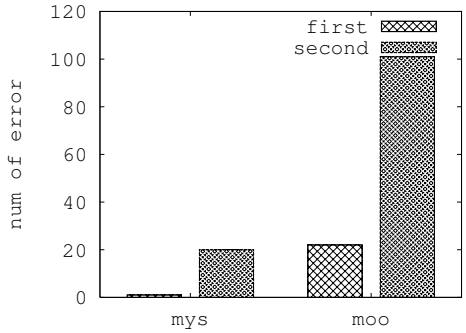

図 6.11　効果音 A の有無による Reject 数の推移.

鳴らさずに再実験を行った．再実験を行うにあたって，約 5 分間被験者に自由に練習を行わせた．その実験結果を図 6.11 に示す．また，被験者 mys の Reject の回数の変化を図 6.12 に示す．図 6.12 の縦軸は一回の入力あたりの Reject の回数を表している．

効果音 A を鳴らさないことによって Reject の回数が 20.0 倍, 4.6 倍と増加した．また，図 6.12 より Reject が発生する場所に偏りがあり，この傾向は被験者 moo にも見られた.

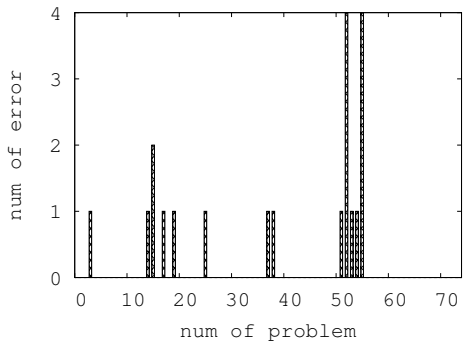

図 6.12　被験者 mys の Reject の回数の変化.

Reject が偏ってしまった原因は,

1. 被験者は効果音 A を聞くことが出来ないため，図 6.7 における状態 2 から 状態 3 へ遷移したタイミングを知ることが出来ない.
2. 頭を戻すタイミングを被験者自らの勘で判断し入力を行わなければならず，問題を繰り返すうちにタイミングがずれてしまい Reject が発生してしまう.
3. 効果音 A が無いために被験者は Reject の発生した原因を知ることが出来ず，再び同じ原因で Reject を発生させてしまう.

ということだと考えられる．被験者の勘によりタイミングを判断することは不可能ではなかったため，慣れの効果により効果音 A を除去しても思い通りに入力できるかもしれないが，現段階では内部の動作を知るための効果音 A は，確実性を保障するために必要である.

現在の状態をユーザが知ること，つまりユーザがフィードバックを得ることは学習を容易にさせ，認識率を向上させるために重要である.

6.6　インクリメンタル音声検索の研究

6.6.1　研究の目的

7 段階モデルおよび「フィードバックの原則」が示唆するのは，たとえ音声入力であっても「リアルタイムに反応すること」は重要だということである.

このような観点から，我々は，オンラインショッピングなどにおいて候補の検索や比較検討などの要素を含む探索的検索を，音声インクリメンタル検索とドラッグ&ドロップ操作によって直感的に行うインタフェースを提案・試作した [98].

我々がコンピュータや情報機器によって行う作業の多くは探索的検索 (exploratory search) [99] の要素を含んでいる．つまり選択肢の中から既知の項目を効率よく指定するのではなく，複

数の候補を比較検討しながら自分が望む項目を選んでいく場合が多くなっている．例えばウェブでショッピングを行う場合には，条件を指定していくつかの候補を絞り込み，それぞれの候補の価格や仕様や評判情報などを閲覧し，場合によっては条件を修正しながら最終的な候補を選びとっていく作業となる．また，情報機器や携帯電話の操作においても，何かを行いたい場合に階層メニューをたどるだけでは所望の操作を見つけ出すことができなかったり，所望の機能がどのような名前で登録されているのかが分からなかったりするため，探索的検索を必要とする場合が多い．

キーボード入力によるキーワード検索は探索的検索の有用な手段と考えられる．例えば PC の操作においてもデスクトップ検索は必須の機能となりつつある．しかし，情報家電やモバイル機器などではキーボードを用いることが困難な場合もあり，このような場面では音声入力が有力な手段になりうる．

6.6.2 インクリメンタル検索の有効性

ヒューマンインタフェースの原則 [118] では，インタフェースの基本原則として「操作労力」「システムの透過性」「頑健性」の 3 つを挙げている．マウスとキーボードを用いたインタフェースにおいては近年「操作労力」「システムの透過性」に関する配慮が多くなされるようになった．いわゆるインクリメンタル検索は，GNU Emacs などのテキストエディタだけでなく Apple iTunes や Google Suggest などのシステムで用いられている．Raskin[100] はインクリメンタル検索の有効性を特に強く主張しており，音声入力によるインクリメンタル検索の実現可能性も示唆している．

インクリメンタル検索は

- 操作労力の原則：ユーザの操作が妥当であることを，絞り込まれた結果や候補数の提示などによってフィードバックする
- システムの透過性の原則：所望する候補が選択可能になれば入力を打ち切ることができるため，ユーザに操作労力削減の機会を与える

などによりインタフェースの基本原則に適合している．音声によるインクリメンタル検索もインタフェースの基本原則に適合すると考えられるが，そのためには

- 発話中に候補やその個数を逐次表示する
- 発話の途中で内容が確定したら，発話を中断しても許容される

という要件を満たす必要がある．

6.6.3 音声入力のリアルタイム性

従来の音声認識システムは性能を重視する一方で認識結果を出力する際の遅延はやむを得ないものと見なす場合が多い．多くの大語彙連続音声認識システムが，trigram 言語モデルを適用したり N-best 候補を得るために 2 パスの探索を行っており，入力同期の第 1 パス処理だけでは最終結果を得られない．しかし，音声認識システムに対するしゃべり方や発話内容が妥当であるか否かを発話終了まで知ることができないのは，音声入力システムの「透過性」においては不十分である．これに対して，人間同士の対話において，人間は相手の表情から反応を読むことができる．人間は一方が話している間も頷いたり首をかしげたりするだろうし，相手の発話が聞き取りにくければ直ちに「え？」などと聞き返すことができる．自分の発話を理解しているのか理解できないのかを示してくれる相手とは，会話がしやすいのではなかろうか [101].

音声認識技術におけるリアルタイム性の検討は過去にも行われてきた．例えば，放送字幕作成のための音声認識システムにおいて，ニュース音声の認識結果を 2 秒以内に得ることを目標とし，2 パスデコーダを改修して一定フレームごとに結果を確定する，といった手法 [102] が提案されている．また音声認識エンジンからの情報に基づいてユーザの発話中に頷いたり相槌を打つ音声対話システムが NTT や早稲田大学などで試作されており [103, 104]，従来の HMM による音声認識に限らず WFST などの利用も提案されている．逐次的な音声認識に韻律解析を組み合わせる手法 [105]，発話の終了判定において無音継続時間の閾値を短くし，発話中のポーズで音声認識結果を出力する手法 [106] も提案されている．これらはリアルタイム音声インタフェースの可能性を示すデモであり，自然な音声対話を実現したとされているが，音声インタフェースの汎用的な枠組みの提案には至っておらず，操作の効率性に貢献する提案とは言えない．

6.6.4 音声入力と効率性

音声認識は効率的な入力手段として期待されるが，スイッチやボタンなど何らかの代替手段が利用できる場面では，入力モダリティの適切な組み合わせを考慮する必要がある．音声入力においても「長い文章やコマンドは発話することに多くの労力を要する」と考えなくてはならない．ユーザに無駄な入力や発話をさせないために，音声インクリメンタル検索をマルチモーダルインタフェースと効果的に組み合わせることが望まれる．

6.6.5 直接操作型インタフェース

音声入力を操作手段とするシステムは一般的に，「対話」あるいは「対話エージェント」をモデル化する「秘書型」のインタフェースである．しかしユーザの興味が操作対象そのものである場合には間接的操作を強いられることは望ましくない [107]．これに対して，操作対象を直接の

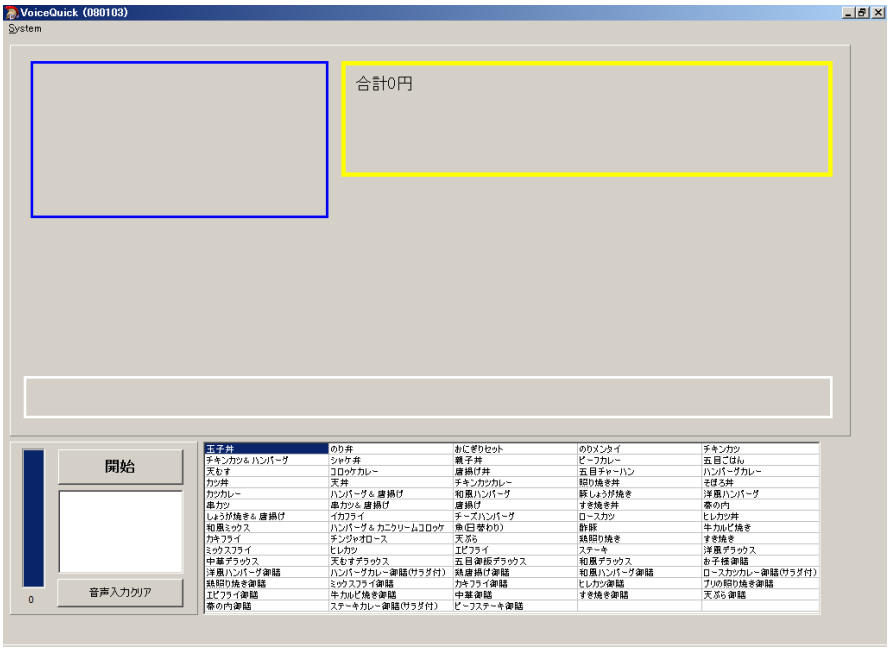

図 6.13　インクリメンタル音声検索システムの全体画面（初期状態）.

操作対象にする「道具型」のインタフェースは，アプリケーション全体としてシステム透過性の向上に貢献している.

6.6.6　プロトタイプシステムの設計

我々はお弁当購入タスク [89] を対象として，音声インクリメンタル検索と直接操作インタフェース [117] の要素を盛り込んだプロトタイプシステムの設計を行った.

画面構成を図 6.13 に示す．画面右下には商品名（73 個）の候補を示すリストボックスがある．画面左下にはタスクを初期化する「開始」ボタン，音声で入力された文字列を示すリストボックス，その内容を初期化する「音声入力クリア」のボタンがある．その左のプログレスバーは音声発話が終了してからの経過時間を示しており，10 秒以内に次の操作を行わないと音声入力クリアが実行される.

画面上部には「インスペクタ」「カート」「候補パレット」が並んでいる．音声発話によって候補が一定数（5 個）以下に絞り込まれると，商品が候補パレットに並ぶ．候補パレットの商品はドラッグ＆ドロップ操作でパレットの外に取り出すことができる．取り出し操作はパレット外部へのオブジェクト複製の操作となっており，同じ商品を複数回取り出すことができる．前述した 10 秒が過ぎると候補パレットは初期化される.

取り出された商品はインスペクタ内部に配置することで，詳細情報を知ることができる（現在は価格のみ表示）．最終的に商品をカート内部に移動して決定を行うことでタスクは完了する.

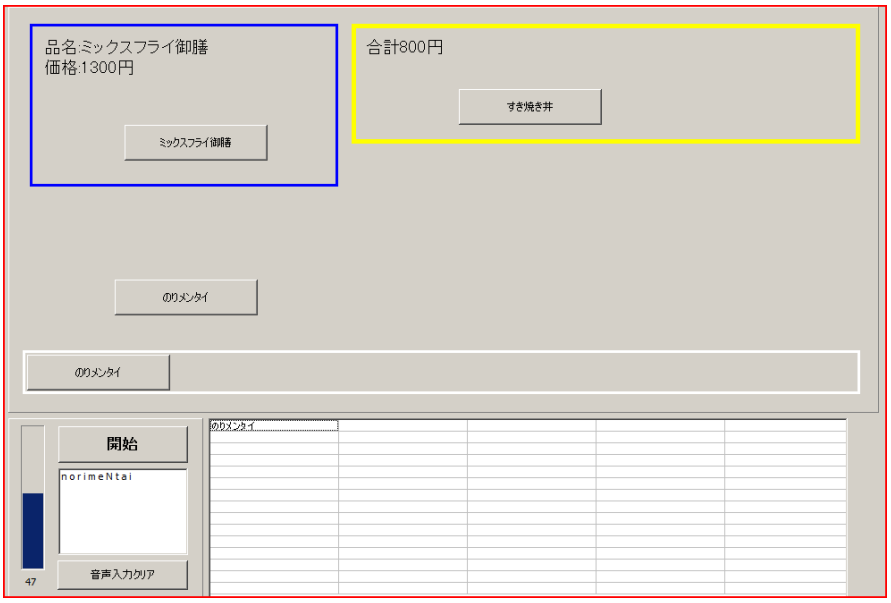

図 6.14　システムの全体画面（インスペクタおよびカートに商品を乗せた状態）.

図 6.15　システムの画面（「ハンバーグ」と発話した直後）.

現在の実装では最終決定操作は実装していない．カート内部の商品の合計価格を動的に表示している．これらの作業の様子を図 6.14 に示す．

音声インクリメンタル検索の様子を図 6.15, 図 6.16 に示す．例えばユーザが「ハンバーグ」と発話しはじめると，発話終了を待つことなく候補リストの更新が始まる．発話が終了した時点では候補リストは「ハンバーグ」を含む商品 10 個に絞り込みが行われている．さらにユーザが（10 秒以内に）「チキン」と発話することで，「チキンカツ＆ハンバーグ」のみが候補として残り，候補パレットにこの商品が表示される．

このシステムの構成要素はインタフェース基本原則への適合性において，以下の配慮がなされている．

操作労力最小化の原則　部分発話を許容することで，音声入力における労力を最小化する．

システム透過性の原則　従来システムと比較して低遅延で多くの情報を提示することで，より適

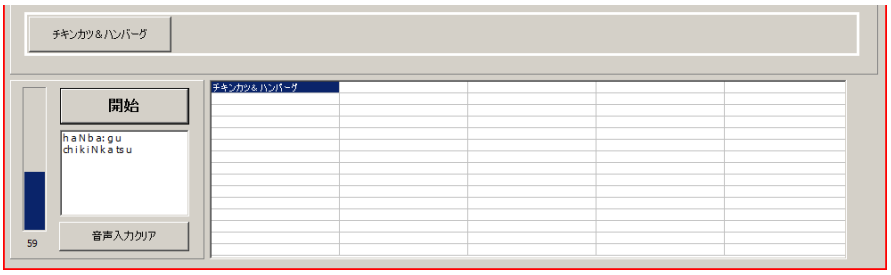

図 6.16　システムの画面（「ハンバーグ」「チキン」と発話した状態）.

切なフィードバックを提供する.

頑健性の原則　頑健性の不十分な音声入力を候補選択のみに使用し，他のモダリティで決定操作を行う．これによりシステム全体として頑健性を確保している.

またインタフェース構成原則への適合性においては，以下の配慮がなされている.

初心者保護の原則　直感的な操作体系と画面構成により，マニュアル不要の使いやすさを提供する.

熟練者優遇の原則　発話において省略を許す，というインクリメンタル検索の利点を生かしている.

上級利用移行支援の原則　発話中に逐次画面更新を行うことで，語尾などの発話省略が可能である，ということを直感的に示している.

本節の研究の一つの発展形は，機械と人間の音声対話におけるフィードバックの遅延を小さくすることである [108, 109]．例えば，音声認識エンジンの第一パス認識結果からより早く情報を取り出すために，モーラ単位の認識辞書を利用し，強化学習 [110] によってタスク依存の学習を効率的に行うことで，ユーザの発話終了タイミングを予測する応答生成が実現できる [111].

6.7　音声インタフェースとしてのラジオ

第 3 章で述べた非同期音声会議システムの試作からの知見として，情報メディアとしての「ラジオ放送」からは，音声技術が学ぶべき点がたくさんある．本節ではこのような着想から行われたいくつかの検討について述べる.

例えば AVM (VoiceCafe) クライアントの操作は，ラジオ番組を聞き流すような感覚で音声メッセージを連続再生でき，興味をひいた話題が聞こえてきたときには能動的な操作ができる，といった使い方が想定された.

RadioDoc[112] は,「聞き流し」を想定した音声コンテンツ記述言語の提案である．VoiceXML など既存の記述言語が音声コマンド入力によって状態遷移を記述しているのに対して，Ra-

dioDoc は「章・節」など書籍の論理構造が用いられ，どの状態からでも音声コマンドによって任意の章や節の先頭にジャンプすることを許容している．このシステムは VoiceXML へのコンバータとしてひとたび試作され，その後十分な検討がなされていない．しかしこの提案は，読書に障害を持つ人を対象としたマルチメディア技術 DAISY (Digital Accessible Information System) [*1] と親和性が高いと考えられ，DAISY プレイヤー実装の一形態として受け入れられる可能性がある．

ラジオにおけるアナウンサーの「伝える技術」に着目した検討 [114] も行われた．ラジオ放送におけるスポーツ実況中継では，音声のみを用いて聞き手に視覚的なイメージを与えるための配慮がなされている．このような配慮は視覚障害者支援技術にも役立つことが期待される．そこで，競馬，野球，サッカーなどの実況中継におけるテレビおよびラジオのアナウンサーの発話内容を比較・分析した．例えば，競馬においては，レースの序盤，中盤，終盤，といった状況ごとに注目される対象の遷移が見られた．野球においては，テレビでは常時画面に表示されている得点やボールカウントなどの試合状況は，ラジオ中継では頻繁に音声で伝えられており，重要な情報ほど高頻度で発話されていた．またサッカーにおいては，連続的なゲームの展開を，間投詞を用いてボールを取ってからシュートを試みるまでをシーンに分割して伝えていた．

6.8 ラジオ放送のための音声投稿システムの開発

6.8.1 研究の概要

音声コンテンツ，あるいはインターネットラジオ番組を個人が手軽に制作できる環境を整備することは，音声インタフェースの応用においても新たな挑戦の場であると考えられた．このような観点から一連の検討と試作 [115, 116] がなされた．これらを踏まえつつ，IPA 未踏ソフトウェア創造事業の支援を得て，ラジオ放送局での利用を目的とした音声投稿システム「オラビー」の開発が行われた [117].

本システムの全体構成を図 6.17 に示す.

ラジオ放送支援システム「オラビー」の開発は，当初「ソーシャル・ネットワーキング型ラジオ番組のシステム開発」として，個人によるポッドキャスト番組の製作支援ツールとして構想された．しかし IPA プロジェクトとしての採択が決定した 2005 年 6 月以降，ラジオ放送局の現場を支援するツールに開発目標を変更した．これに伴い，関係者からの聞き取り調査に基づくニーズ把握を行った.

ラジオ放送局で実際に番組制作の業務を行っているディレクターから，ラジオ放送ビジネスの現状，ラジオ放送における音声投稿の重要性，ラジオ番組制作現場における技術的な課題などを聞き取り調査した．また，埼玉県入間市のコミュニティ FM 放送局「エフエム茶笛（チャッ

[*1] http://www.daisy.org/

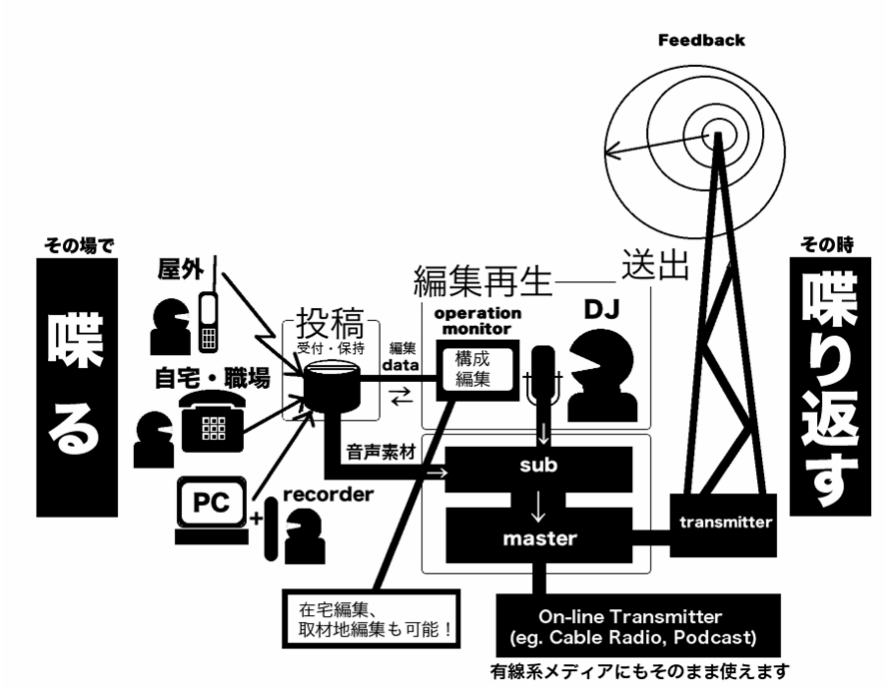

図 6.17　オラビーの全体構成.

ピー）を」を訪問し，コミュニティ FM 放送局の抱える人材不足や番組のマンネリ化，といった現状について理解を深めた．これらの過程を通じて，本プロジェクトが開発すべき主要な機能を「音声投稿番組の制作支援」に絞り込み，直感的なユーザインタフェースでこれを実現する，という目標の明確化が行われた.

2005 年 11 月に，オラビーの基本機能（電話による音声投稿とその再生）を実装し，エフエム茶笛において電話投稿受付機能の試験運用を行った．さまざまな環境からの音声投稿を行い，誤操作の起きやすい部分など仕様上の問題点を抽出し，逐次改良を行った．それ以降，本システムを放送局に常時設置し，さまざまな番組で担当者に運用してもらい，システムの不都合や仕様の不備などの報告を受けて，改良を行った.

さらに，2006 年 2 月に，エフエム茶笛にて 4 週間にわたって音声投稿を中心とした実験番組「なべやかんの居留守放送局」を制作した．1 時間番組の生放送を行うと同時に，インターネットでも再配信を行い，番組のリスナーから幅広く音声投稿を受け付けた．また，出演者にボイスレコーダで取材させた音声を自宅からサーバにアップロードさせる，といった実験を行った．放送前の準備および生放送中にディレクターに本システムを使用させ，不都合の報告や機能の要望などの意見を聞いた．これらの実験を通じて，ラジオ放送の現場での要求に合わせたシステムの改良が行われた.

6.8.2 オラビーの構成

電話によって音声投稿を受け付けるために，音声対話記述言語の標準規格である VoiceXML を使用した．VoiceXML 処理系 Plum Voice Portal を使用し，PHP 言語で実装されたサーバサイドスクリプトが動的に VoiceXML ドキュメントを生成する．今回用いた VoiceXML の機能は，録音済み音声ファイルの再生 (<audio>)，数字キー (DTMF) による値の入力 (<field>)，ファイルへの音声録音 (<record>) などである．

処理は以下の流れで構成されている：

1. **グリーティング：**「オラビーへようこそ」という音声メッセージを再生する．
2. **番組アクセス番号入力：**「5 桁の番組アクセスナンバーを入力してください」という音声ガイドに続いて，電話機による数字入力を行わせる．
3. **マイクテスト開始：**「これからマイクテストを行います」という音声ガイドに続いて，電話音声を録音させる．
4. **マイクテスト終了と音量チェック：**録音された音声ファイルを分析し，録音時間を「約 30 秒でした」のように音声で提示する．レベルオーバーが生じた場合には，「声が大きすぎて音が割れています」という音声ガイドを行う．続いて，録音された音声を再生し，確認させる．レベルオーバーが起こらなくなるまで繰り返しマイクテストを行わせる．
5. **本番録音開始：**「次は本番です」「あなたのお名前，あなたの今いる場所，メッセージの順番でお話ください」などの音声ガイドに続いて，本番の録音を行わせる．
6. **本番録音終了：**「あなたのメッセージが登録されました」という音声ガイドを行い，録音時間を報告し，録音された音声を再生して検聴させる．作業が終了したら電話回線を切断する．

録音時間の報告とレベルオーバーの検出は，試験運用を通じて必要となった機能である．本システムでは録音中に話者に画面を見せることができないために，録音時間を音声でフィードバックすることが重要になる．また，レベルオーバーによる音のひずみは放送時には非常に耳障りであるが，電話の音質では本人が聞き返しても判断することができない．従って，レベルオーバーをシステムが判定して再録音を促すことは非常に有効である．

ウェブブラウザで投稿された音声素材の確認，メモ付与，分類，編集などができる環境を実現した．また，ボイスレコーダ等で録音された音声ファイルの投稿も可能である．PHP 言語で実装されたサーバサイドスクリプトが動的に HTML ドキュメントを生成する．

図 6.18 は HoldStation の画面である．上部には素材送出機能 (CastStudio) の起動ボタンがある．また，右側の番組アクセス番号入力ボックスを使用し，他の素材リストに移動することができる．素材リストには左から順に，投稿された日時と投稿者の電話番号，音声再生機能，サ

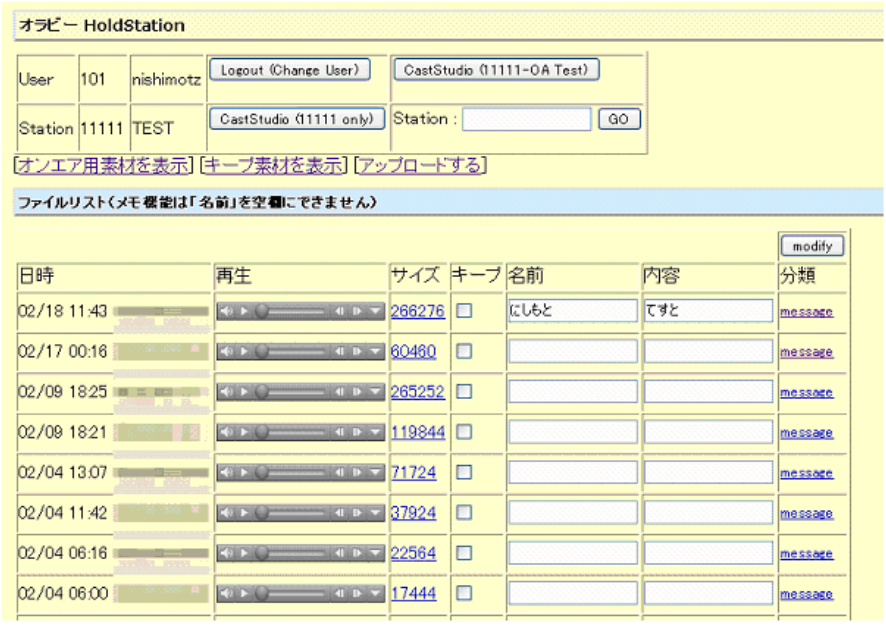

図 6.18　HoldStation の画面.

イズ，キープ機能，名前入力ボックス，内容入力ボックス，分類が表示されている．サイズはサーバに蓄積されている音声ファイルのバイト数を示すと同時に，音声ファイルへのリンクになっており，ブラウザのファイル保存操作によって音声ファイルをダウンロードすることができる．キープ機能はチェックボックスとなっており，ひとつまたは複数の素材にチェックをして「modify」ボタンを押すことにより，指定された素材を「キープ状態」にすることができる．キープ状態の素材は CastStudio では表示されない．名前入力ボックスおよび内容入力ボックスは，素材に投稿者の名前と内容のコメントを付与するための機能である．ひとつまたは複数の素材について入力を行った後「Modify」ボタンを押すことによって更新が行われる．入力された内容は全ユーザに共有される．分類は”message”または”sticker”の 2 種類が存在する．sticker に分類された素材は，CastStudio において通常の素材 (message) と比べて小さな箱として表示され，起動時に自動的に音声データの読み込みが行われる．

この画面の他に，個々の音声素材に関する詳細画面があり，マイクテストの音声を聴取することができる．また，素材を他の番組アクセス番号のフォルダに移動することができる．また，アップロード画面ではボイスレコーダで録音した mp3 形式のファイルなどをアップロードすることができる．

HoldStation の画面から素材リストに対応した素材送出機能（CastStudio）を呼び出せる．CastStudio は Java 言語で開発されている．素材ボックスをマウスでドラッグして配置し，キューシートなどの画面構成要素をクリックする，といった単純な操作のみで実行できる．

図 6.19 は，CastStudio においてインスペクタ上のアイテムの再生（検聴）を行っている状態

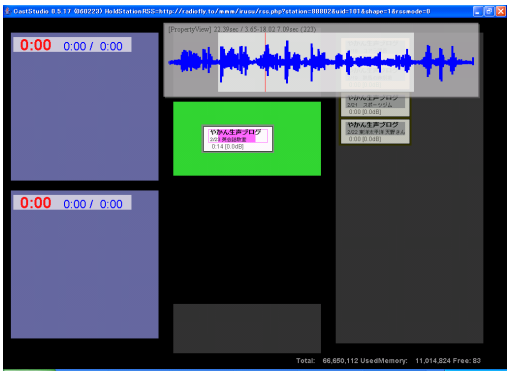

図 6.19　CastStudio でアイテムの検聴を行っている状態.

図 6.20　CastStudio でキューシートを再生した状態.

である．左側の 2 つの矩形は「キューシート」，中央の矩形は「インスペクタ」，右側の矩形は「ポーター」と呼ばれる．ポーターはサーバとの通信の役割を担っており，起動時にサーバに蓄積された素材を「アイテム」と呼ばれる矩形として表示する．アイテムには HoldStation で入力された「投稿者名」および「内容」が表示される．起動直後には各アイテムの音声データは読み込みが行われておらず，再生することはできない．そのことを示すために，アイテムは灰色がかった色で表示される．

アイテムをインスペクタに乗せると，アイテムの音声再生に必要な情報がサーバから取得される．それと同時に，サーバから音声波形の表示に必要な情報をダウンロードする．インスペクタの色は，アイテムの読み込み中は赤に，読み込みが完了すると明るい灰色に変化し，インスペクタの上部に音声波形シートが描画される．

音声の再生は，インスペクタの余白（アイテム以外の部分）をクリックすることで開始し，再度クリックすることで停止する．音声波形シート上で左クリックを行うと再生開始ポイントを，右クリックを行うと再生終了ポイントを，それぞれ指定することができる．範囲指定されたアイテムの再生時間はアイテムの下部に表示される．

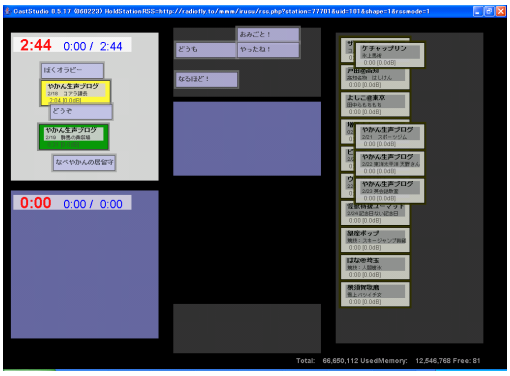

図 6.21　CastStudio をミックスモードで実行している状態.

インスペクタに同時に置くことができるアイテムは 1 つだけであり，2 個目のアイテムをインスペクタに乗せようとした場合，その移動操作はキャンセルされ，アイテムは自動的に元の位置に戻される．

アイテムを右クリックすることで，アイテムの色を 5 種類に変更することができる．アイテムの色は再生機能とは関係がなく，色分けはユーザがアイテムの識別を容易にするための機能である．インスペクタおよびキューシートと一部分でも重なりを持たないような箇所であれば，作業途中のアイテムを任意の場所に自由に置いておくことができる．

画面右側の矩形（ポーター）の余白（アイテムがない部分）を左クリックすると，マウスカーソルが一瞬砂時計になり，サーバとの通信が行われ，マウスカーソルが元の矢印の形状に戻る．サーバとの通信の結果，新たなアイテムの存在が確認された場合は，そのアイテムをオレンジ色でポーターに表示する．

図 6.20 は，CastStudio においてキューシートを再生した状態である．

アイテムをキューシートに完全に包含されるような位置に移動すると，そのアイテムはキューシートの構成要素として判定される．キューシート上部には「残り時間」と「経過時間／合計時間」の表示がある．キューシートにアイテムを配置すると自動的に残り時間および合計時間が再計算されて表示される．

上下 2 つのキューシートの機能はまったく同じである．キューシート内の素材は，アイテムの矩形の上辺の上下の位置関係によって再生順序が決定される．キューシートに 1 つでも構成要素が存在する場合には，キューシートの色は青から灰色に変化し，キューシートが再生可能であることをユーザに示す．インスペクタにて読み込みを行った素材はキューシートに配置可能であるが，読み込みを行っていない状態の素材はキューシートに置くことができない．そのような移動操作はキャンセルされ，アイテムは元の位置に戻される．

キューシートの余白（アイテムのない部分）を左クリックすることで，キューシートの再生が開始され，アイテムが上から下に向かって順番に再生される．再生中のキューシートの色は緑

色に変化する．再生中のキューシートをさらに左クリックすることで，再生を中止することができる．なお，上下のキューシートは同時に再生することも可能である．キューシート内のどの素材が現在再生されているかは，アイテムの範囲表示ボックスの色が灰色からピンク色に変化することで示される．

すべてのアイテムの再生が終了したことによってキューシート全体の再生が完了した場合は，キューシート内のアイテムは画面中央下部の「リサイクラー」に自動的に移動する．リサイクラーに移動したアイテムは再度利用することが可能である．

再生中のキューシートの未再生アイテムはキューシートの外に移動することができ，当該アイテムは再生をキャンセルできる．再生中のキューシートにおける再生済みのアイテムは自由にキューシートの外に移動することができる．

再生中のキューシートの再生中アイテムより上の位置に，当該キューシートの外からアイテムを移動することはできない．再生中のキューシートの再生中アイテムよりも下の位置に，当該キューシートの外からアイテムを移動した場合は，その移動は有効となり，残り時間および合計時間の再計算が行われる．また，追加されたアイテムを含めて，未再生のアイテムは配置された位置に応じた順序で再生される．

再生中のキューシートの中で素材の上下関係を並べ替えることはできない．いったんキューシートの外に出したアイテムを，元のキューシートの再生中のアイテムよりも下に配置することは可能である．

図 6.21 は，「ミックスモード」で CastStudio を起動した状態の画面である．

ミックスモードとは，複数の素材ボックスのアイテムを同時に CastStudio で使用するためのモードである．その際，sticker に分類されるアイテムが含まれている場合は，CastStudio は起動時に sticker の読み込みを行い，画面中央上部の「ステッカーホルダーシート」に「ステッカーアイテム」として表示する．ステッカーアイテムは通常のアイテム（メッセージアイテム）と同様にキューシートに配置し再生することができる．キューシートの再生が完了した場合に，メッセージアイテムはリサイクラーに自動的に移動するが，ステッカーアイテムはステッカーホルダーシートに自動的に移動する．

ステッカーは，事前に作成された音楽や効果音であり，投稿されたアイテムの前後や間に挟むことによって，投稿素材であることをわかりやすくしたり，次の素材に切り替わったことを直観的に示す役割を持つ．

6.8.3 試験運用

オラビーを使用して，2006 年 2 月に音声投稿を中心とした 1 時間の実験番組を 4 回制作した．この番組制作を通じて電話投稿受付機能，音声素材管理機能，素材送出機能のテストを行った．無線 LAN 環境が使用できる放送局スタジオにノートパソコンを設置し，ノートパソコンの

音声出力を，音源のひとつとしてミキサーにつないだ．

システムの改良を重ねた結果，本システムが，ボイスレコーダで取材された４～５本の高音質音声素材，電話によって投稿された 10 本前後の音声素材，あらかじめ用意されたステッカーや効果音などを組み合わせて，1 時間の生放送番組での使用に耐える機能と性能を有することが確認された．特に，本システムの電話投稿受付機能が，不特定多数のユーザが失敗せずに音声投稿を行えるだけの操作性を有すること，また，本システムの素材送出機能が，一人のディレクターの手によって番組進行管理，音楽の送出，ミキサー制御などと並行して操作できるだけの信頼性と操作性を有することなどが確認された．

実験番組を担当したラジオディレクターからは，特に以下の点が好評であった．

- 番組オンエアの直前に，HoldStation のウェブ画面を用いて，投稿済み素材の検聴を行い，不要な素材の消去などの作業ができる．特に事前に自宅で番組の準備ができるのは便利である．
- 葉書やファクスなどの紙を並べ替えるような感覚で，音声素材を CastStudio の画面内に自由に並べて整理することができる．アイテムの色分けは，どのコーナーでどの素材を使うか，といった判断を直観的に行うために役立つ．
- キューシートによる音声素材の連続再生機能は，再生中に次のコーナーの打ち合わせや楽曲の頭出しなど，別の作業を行う時間をディレクターが確保できるため，有効である．
- CastStudio はドラッグ＆ドロップとクリックだけですべての操作が行えるように設計されており，メニューからのコマンド選択操作を排除している．また，アイテムを画面内で置く位置にも自由度がある．これらの結果，万一送出すべき素材を間違えたり，急遽内容を変更したい，という場合でも，混乱せず，効率的にキューシートの組み替えが行える，という利点がある．
- ラジオ放送の魅力は生放送であり，ラジオ番組の制作は作品制作ではなく，リスナーと共有できる「時間と空間」を作ることである，と考える．オラビーは生放送に特に適した音声投稿システムである．

従来，留守番電話でリスナーからの投稿を受け，これを放送するためには「電話機の記録装置から放送用の一時固定媒体（テープ，MD など）に吐き出させ，これをダビングによって並べ替え，トリミングし，送出の構成にあわせて再度ダビングし，本番に備える」という手間を要した．本技術は，この工程を電子化し，ダビングの待ち時間や機器接続の時間を節約するとともに，音声の移し替えによる音質の劣化やトラブルをなくすことを実現した．さらに，番組の展開や周辺状況の変化などに応じて，急遽，素材の入れ替えをすることを可能にした．

本開発により，現在，日本のラジオ放送が最も苦手とする「肉声による投稿参加型番組」が容易になり，ラジオを中心としたコミュニティの知財共有がすすむ．例えば，現在ラジオ放送の半分以上の時間を占める生ワイド番組に，機動力と参加性を飛躍的に向上させ，市井の声をより積

極的に反映させることが可能になる．オラビーは 2009 年より Ruby on Rails でサーバ処理系の再実装を進めており，Web サイト[*2]を通じてソースコードを公開しながら開発を続けている．

2010 年に入ってラジオ番組がマイクロブログサービス Twitter や動画配信サービス Ustream と連動する事例が増えてきた．ラジオ番組にリスナーがリアルタイムに参加したいという要求は高まっており，放送の送り手も音声投稿サービスの導入に向けて体制を整えつつあると言える．

一方で，ラジオ放送そのものをインターネット配信で聴取することが一般化し，個人が Ustream で配信するコンテンツもラジオ番組のフォーマットに非常に近いものになりつつある．さらに考察すると Twitter というメディアそのものが「放送番組のようなソーシャルメディア」といえる．Twitter は流し読みができるコンテンツでありつつ，気になった情報については積極的な検索ができたり，双方向の交流がシームレスに行えるメディアなのである．

このことから，オラビーのようなシステムは，当初の我々の主張 [115] どおり，音声メディアによる放送という情報発信手段を使いこなしたい個人を支援するものとして，ますます着目されるのではなかろうか．

6.9 オープンソースプロジェクトと音声技術

インタフェース導入原則のうち「有用性の原則」から導かれるのは，「有用性の発見を促すために，音声技術に関するオープンソースプロジェクトを推進すべきである」という主張である．

本節では 2 つのプロジェクトについて簡単に紹介する．

6.9.1 Galatea プロジェクト

Galatea Project は，擬人化音声対話エージェントのツールキット Galatea Toolkit を開発し， オープンソース，ライセンスフリーで公開提供するプロジェクトで，国内の十数大学などの音声・言語・画像研究者が参加して進められた．顔，声，音声合成テキスト，認識文法，対話の流れなどはカスタマイズ可能で，これを用いて容易に人間の顔と表情を持ち，音声で対話する自分独自のエージェントを作成することができる．また，構成要素 (音声認識，音声合成，顔画像合成など) を別々に無償で利用することもできる．商用利用も可能である．このツールキットは 2003 年 8 月から公開され，現在もオープンソースコミュニティ[*3] として活動をしており，最新の Linux 環境への対応などが行われている．

[*2] http://ora-be.nishimotz.com/

[*3] http://sourceforge.jp/projects/galatea/

6.9.2 NVDA 日本語化プロジェクト

視覚に障害がありパソコンの画面を見ることができない人は，マウスを使わずキーボードだけでパソコンを操作し，音声出力や点字ディスプレイで文字情報を読み取っている．OS の機能を拡張してこのような使い方を可能にするのがスクリーンリーダと呼ばれるソフトウェアである．日本語の Microsoft Windows 環境に対応した製品としては 95Reader, JAWS, PC-Talker, FocusTalk, xpNavo などが知られており，複数のソフトウェアを目的に応じて使い分ける人もいる．さらに Web サイトを読み上げるための専用ソフトウェア (ホームページ・リーダーなど) も併用される．

この分野で注目されているのは NVDA (NonVisual Desktop Access) というオープンソースソフトウェアである．オーストラリアの Michael Curran 氏をはじめとした多くの人たちによって，2006 年より開発され，20 ヶ国語以上の言語に翻訳されている．

NVDA は無償で配付されているので，障害がある人に加えて，例えば Web サイトの制作者に「アクセシビリティに配慮するための確認用」として利用されつつある．

もう一つ NVDA が注目されている理由は，高機能であることだ．例えば最近の Web サイトでは，ページを遷移しないで内容をどんどん書き換える AJAX(Asynchronous JavaScript + XML) という技術が多用されている．これまでは「動的に書き換わるサイトはスクリーンリーダで操作できない」のが常識であった．しかし WAI-ARIA という規格が提案され，この問題が解決されつつある．このような最新技術に NVDA は積極的に対応している．日本の Web アクセシビリティ規格の最新版 JIS X8341-3:2010 への対応のためにも NVDA の活用が期待されている．

NVDA 日本語化プロジェクト (NVDAjp) [*4] は，ユーザインタフェースやヘルプの日本語化，日本語キーボードへの対応などを行ったソフトウェアを公開している．

主要部分が Python というスクリプト言語で実装されている NVDA は，この分野のソフトとしては開発者にとっての敷居が低くなっている．NVDAjp では「かな漢字変換操作の読み上げ」について急ピッチで作業を進めている．また無償配付できる音声合成エンジンの組み込みにも取り組んでおり，2010 年 9 月にオープンソース日本語音声合成エンジン Open JTalk の技術を利用した開発版が初めて公開された．

6.10 マルチモーダル対話システムのアーキテクチャ

マルチモーダル対話システムをさまざまなタスクに効率よく導入するためには，マルチモーダル対話を実現するための汎用ツールキットと，タスクやインタフェースの詳細に関する汎用

[*4] http://sourceforge.jp/projects/nvdajp/

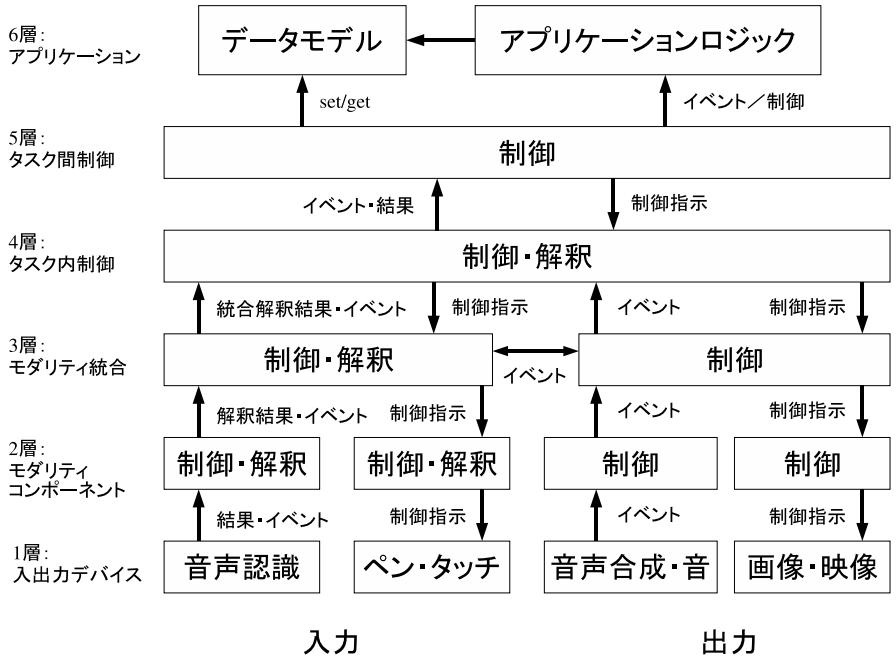

図 6.22　MMI システムのアーキテクチャ階層化

の記述言語が必要となる．特に，モダリティに依存する処理をモダリティごとに集約すること，使用するモダリティの追加や変更を柔軟にすること，きめ細かいタイミングでの入出力の制御を行うことなどがアーキテクチャと記述言語への要求として挙げられる．

音声対話技術コンソーシアム (ISTC) [*5] では，音声対話システムおよびマルチモーダルインタフェース (MMI) のツールキット (Galatea Toolkit) の開発と並行して，マルチモーダル対話記述言語の仕様を作成するためのワーキンググループ（SIG-MMI-WG）を 2004 年に設置し，活動を行った[*6] ．この活動は 2006 年より，情報処理学会・学会試行標準専門委員会・WG4 小委員会（主査・新田恒雄 豊橋技術科学大学 教授）における音声言語インタフェースの標準化活動とも連携しており，World Wide Web Consortium (W3C) の標準に準拠した MMI 記述言語の策定を目指している．

ソフトウェア開発においては Model, View, Controller (MVC) を分離することが推奨される．Model とはアプリケーションが内部的に保持するデータモデルを，View とはユーザに提示する画面表示や入力イベントなど UI (User Interface) を，Controller はこれらを対応づけてユーザインタフェースを実現するロジックを意味している．UI とロジックを分離することで柔軟なシステム開発が可能になる．このような検討を踏まえて，MMI システムのアーキテクチャ階層化（図 6.22）が提案された [125].

[*5] http://www.astem.or.jp/istc/
[*6] http://www.astem.or.jp/istc/ISTC-SIG-MMI/

表 6.4　MMI 階層構造

第 1 層	入出力デバイス層
第 2 層	モダリティコンポーネント層
第 3 層	モダリティ統合層
第 4 層	タスク内制御層
第 5 層	タスク間制御層
第 6 層	アプリケーション層

MMI 階層構造は表 6.4 に示す 6 階層によって構成される．

6 層はアプリケーション層である．ここではデータモデルとアプリケーションロジックを実装する．

5 層はタスク間制御層である．ここでは対話タスクの全般的な制御を行う．また，アプリケーション層との入出力通信を行う．

4 層はタスク内制御層である．ここでは各タスク内の対話の制御と応答内容の決定を行う．フォーム処理における充足性判定や状況判断，フォーム充足のための次応答処理のモダリティ制御，バージインやシステム割り込みなどタスク内の対話遷移処理などが含まれる．

3 層はモダリティ統合層である．ここでは入力の統合，出力の分解，入出力同期制御などを行う．逐次入力や同時入力の解釈，逐次出力や同時出力の同期，モダリティ拡張などが含まれる．

2 層はモダリティコンポーネント層である．ここでは画面表示，音声入出力，擬人化エージェント制御など，個々のモダリティの制御を行う．入出力の取扱いは分離し，レベルメータや音声認識状態表示など，モダリティ間の連携は 3 層を介して実現される．

1 層は入出力デバイス層である．それぞれのデバイスに依存する API やイベントなどが用いられる．

各層の役割は記述言語の仕様と以下のように対応できる．

5 **層**:**タスク間制御層**　SCXML, VoiceXML
4 **層**:**タスク内制御層**　VoiceXML の FIA アルゴリズムや HTML の form 要素
3 **層**:**モダリティ統合層**　SMIL 2.0 （出力）, XISL （出力，統合解釈）
2 **層**:**モダリティコンポーネント層**　SVG の rect 要素，VoiceXML の prompt 要素と SSML, SALT の listen 要素など

また，各層におけるデータ交換形式は EMMA, XForms などと対応する．

第 2 層から第 5 層までのすべてで記述言語を用いて制御を行うことは繁雑であり，特に下位レイヤではデフォルト動作を設定することが望ましい．例えば，フォーム充足のための応答手順決定や，照応解決などの知的な処理，エージェントの頷きなどの自律的動作などが挙げられる．

音声対話および MMI アプリケーションの開発においては，モダリティ独立な第 5 層を記述し，第 2 層（音声認識・音声合成など）の動作に必要なデータの準備を行えばよいような記述言語とし，必要に応じてデフォルト動作の変更が行えるような仕様が望ましい.

第 7 章

インタフェース原則の検証

7.1 音声研究におけるインタフェース原則の役割

本章では「幅広い視点に基づく複数の問題を扱っている音声応用システムの開発者や評価者にとって，本論文のインタフェース原則は有用である」という仮説を検証するために，前述した 4 つの研究においてインタフェース原則が果たした役割を論じる．

7.1.1 音声作図システム研究におけるインタフェース原則

音声作図システム研究（第 2 章）においては，音声インタフェースの特徴を明らかにして，音声コマンド入力と他の入力手段を効果的に協調させることが重要であった．

結果的に「作図システム」というアプリケーションは，音声コマンド入力という要素技術に着目した場合の，「インタフェース構成原則」における「有用性」「適合性」の視点から導きうるものであった．

作図システムとは「ユーザがやりたいこと」と「やりたいことを実現する手段」の間にギャップができやすいアプリケーションである．例えば Microsoft PowerPoint や Adobe Illustrator など，作図機能を備える多くのアプリケーションは，画面を構成する要素が多数になってしまい，使いにくさを感じさせる原因となっている．このギャップを埋めるために，このようなソフトウェアでは，メニューを階層化したり，内部状態に応じて表示を切り替えたりするなど，さまざまな配慮がなされている．このような状況を改善しうる一つの有力なモダリティとして，音声の有用性は発見可能である．

さらに，作図システムにおける作業のほとんどは音声と無関係であるため，適合性においても音声は問題にならない．

システムの具体的な設計において，「インタフェースの基本原則」の「操作労力に関する原則」は，

- 音声入力を用いることで距離や操作回数を少なくする可能性があること
- 音声コマンド入力においては語彙サイズを増やすことが容易であり，キーボードショートカットに比べて多くの選択肢を提供できる可能性があること

などの発見を支援した．また「システムの透過性に関する原則」「頑健性に関する原則」は，

- 図形の色などの属性の変更や，描画モードの変更など，コマンド名を連想しやすい場面で，音声入力はシステムの透過性を高める可能性があること
- フィードバックや誤入力防止などの点で音声入力は不利になりやすいため，これらを補う配慮が必要であること

などの発見を支援した．

これらの発見に基づいて設計されたマルチモーダル作図システム S-tgif では，作図システムにおいてマウスを描画エリアのポインティングに専念させるために，その他の操作を音声入力で行う，という提案が行われた．「インタフェースの構成原則」における「初心者保護」「熟練者優遇」の観点から，どのようなユーザがどのような手段を好むかを予測することができ，これに基づいたシステム評価が可能となった．「慣れた被験者は左手でキーボードショートカットを操作し，右手でマウス操作を行うので，熟練すると音声の優位性は小さくなる」という見方もあったが，「しばらく利用をやめるとショートカットの利用は思い出すことが難しく，そういった場合に音声は再度有効性を発揮する」ということもあわせて示された．

さらに「上級利用移行支援」の観点から，音声コマンド一覧を画面に設ける必然性も明らかになった．

7.1.2 非同期型音声会議システム研究におけるインタフェース原則

非同期型音声会議システム研究においてインタフェース原則が果たした役割を考察する．この研究における音声の役割はコマンドではなく，音声は人間と人間のコミュニケーションに用いられるメディアであった．

非同期型音声会議とインタフェース基本原則

まず「インタフェースの基本原則」は人間と人間のコミュニケーションに対しても当てはめることができる．

例えば操作労力については，「音声の録音は，キーボードによる文字入力と比較して，少ない労力で効率的に情報を伝えられる」という仮説に帰着する．この仮説は文字による掲示板との比較実験によって支持された．さらに「操作労力」を増やさないようにしつつ，メッセージの相互参照というメタ情報を生成するために，「再生と録音のオーバラップ」「相槌や割り込み」と

いった本システム特有の機能が導かれた．

また「透過性」の原則は，ディクテーション結果としての文字情報の活用方法を導いた．すなわち，文字情報は，情報の透過性のうえで「音声」という見えないものを可視化する手段として有用であった．そして「文字の掲示板」のためのスレッド表示を利用しつつ，「録音ボタン」「再生ボタン」といった音声レコーダーの操作系も取り入れたことが，本システムを成功させた一因である．

「頑健性」については，音声メッセージそのものを主たるコンテンツとすることで，副次的な情報としてのディクテーション結果を有用なものにしている．このような補完関係は必ずしも一般的なアプリケーションで成り立つとは言えないが，音声技術を生かすために不可欠な視点である．

非同期型音声会議とインタフェース構成原則

「インタフェースの構成原則」は本研究においては，音声コミュニケーションの「上級利用移行」を促すという発想をもたらした．すなわち「音声会話の漸次性と相槌」は「人間が経験する効率的なコミュニケーション様式」である．この経験がそのまま非同期型音声会議で活用できるように，本システムは設計された．BISP 機能の実装方法の課題は残されたものの，この試みによって例えば被験者の「反論・同意といった意見を出しやすい」という評価が得られたと考えられる．

非同期型音声会議とインタフェース導入原則

最後に「インタフェースの導入原則」について考察する．「音声メッセージシステム」そのものがすでに様々な形で受け入れられている以上，このようなシステムを使いやすくする提案は「有用性の原則」にかなっている．

特に「動機付け」について言及しておきたい．このシステムは「時間と空間を越えて擬似的にリアルタイム音声会話の感覚が得られる」というユニークな特性によって，音声技術を積極的に使ってみたいという「動機付け」をユーザに与えた．この研究ののち，2000 年代中頃から Web 2.0 というキーワードとともに Web の新しい利用法が広まった．受け手と送り手が流動化し，誰もが Web を通じて情報を発信できる時代が到来し，例えば本研究が着目した「相互参照」の概念は，ブログにおけるトラックバック機能として一般化した．

さらに 2006 年に日本でサービスが開始された「ニコニコ動画」[*1]は，動画の特定の再生時間上にユーザがコメントを投稿できる機能を持ち，まさに「時間と空間を越えて擬似的にリアルタイム通信の感覚が得られる」ことが人気を呼んでいる．

「ニコニコ動画」で非同期リアルタイムのコミュニケーションの有用性が知られるようになっ

[*1] http://www.nicovideo.jp/

た現在こそ，本研究が提案した非同期音声会議のコンセプトは，広く受け入れられる可能性があると言える．

「適合性の原則」については第 3 章では言及できなかったが，例えば「ユニバーサルデザイン」については，このシステムは聴覚に障害をもつ人や高齢者と，健聴者の間のコミュニケーションに役立ちうるシステムである．

また「状況と環境の考慮」については，PC ではなく電話応答システムや PDA に対応したインタフェースを有する非同期音声会議システム VoiceCafe を (株) ネットイン京都などと共同で開発し [66]，さまざまな状況やデバイスに対応した拡張が可能であることを示した．

「妥当性の原則」から要求されて行われた評価実験では，「目的指向の話し合いを行わせて競わせる」という手法が導入され，評価方法を目的に合わせる配慮がなされた．

7.1.3 対話負荷の評価におけるインタフェース原則

対話負荷の研究（第 4 章）の必要性を示唆した原則は「インタフェース基本原則」の「システムの透過性」，特に「理解容易性」の項である．過去に「インタフェース基本原則」における客観的な評価は，音声作図研究（第 2 章）では「操作労力」についてのみ行われていた．しかし「透過性」が低いインタフェースにおいて情報を理解したり手順を連想することは，「操作」とは異なるものの何らかの負荷を伴っている．「システムの透過性」に関する客観評価を数量化する手法として，従来用いられていた二重課題法を，音声インタフェースに適した方法に改良したのが第 4 章の研究である．

さらにこの研究の必要性は，「インタフェース導入原則」の「適合性」および「妥当性」の項からも導かれたものである．すなわち「ユーザが行っている他のタスクに悪影響を与えない」という「適合性」の主張は，車載機器との音声インタフェースに特有な要求である．第 4 章の研究は，同時期に (財) 日本自動車研究所などで実施された研究および標準化活動の成果として「テレマティックス音声対話ガイドライン」[67, 68] の作成につながった．

7.1.4 超早口音声の評価におけるインタフェース原則

超早口音声の研究（第 5 章 [123, 124]）を示唆した原則は「インタフェース基本原則」の「システムの透過性」，特に「理解容易性」の項である．すなわち「超早口音声」に慣れることで短時間で情報を得られれば，それは理解が容易になったと考えられる．

超早口音声という手段は「インタフェース構成原則」からみれば「熟練者優遇」の手段である．さらにユーザが慣れの効果をいかに得ているのか，ユーザの慣れをいかに促すか，といった観点は「上級利用移行支援」のために重要である．

そして「インタフェース導入原則」の観点から，スクリーンリーダなどの音声読み上げ技術

は「有用性」（視覚に頼らないという必然性）においても「適合性」（ユニバーサルデザインの観点からの必要性）においてもインタフェース原則の有効性を考える上で避けて通れない研究であった．

インタフェースの心的負荷を扱うという観点では，本研究の立場は第 4 章の研究と重複している．しかし二重課題法の研究がタスク成績という客観的な尺度を利用していたのに対して，第 5 章の研究では主観評価が用いられた．これは「聞こえた音声を書き起こす」という課題が二重課題法と両立しにくいためであったが，結果的に心的負荷評価の方法を幅広く扱うことができた．

この分野の先行研究が了解度やアンケート結果に頼っていたのに対して，本研究では「音声を聞くことが楽であるか」という観点から，インタフェースの主観評価法 NASA-TLX の利用を着想し，有効性を確認した．さらに「楽に聞けるかどうかは聞き手の意識に依存する」という知見を定量的に検証することに成功した．

7.2 既存ガイドラインとインタフェース原則の比較

本節では，一般的なインタフェースシステムのためのガイドライン [14] および音声認識インタフェースのためのガイドライン [34] などを概観し，これらに対して，本論文の提案するインタフェース原則に優位性があることを述べる．

7.2.1 インタフェース設計の 8 つの黄金律

Shneiderman [14] の “Eight Golden Rules of Interface Design” は優れたインタラクションシステム実現の手引きとして広く用いられており，以下の項目から構成される．

1. 一貫性を保つように努めよ．Strive for consistency.
2. 頻繁に使うユーザにショートカットを提供せよ．Enable frequent users to use shortcuts.
3. 有益なフィードバックを提供せよ．Offer informative feedback.
4. 完了感を与えるための対話を設計せよ．Design dialog to yield closure.
5. エラーを単純に処理できるようにせよ．Offer simple error handling.
6. 簡単にやり直しができるようにせよ．Permit easy reversal of actions.
7. 内部の動きが簡単に把握できるようにせよ．Support internal locus of control.
8. 人間の短期記憶の負担を軽減せよ．Reduce short-term memory load.

本論文の提案する原則の多くは，以下に示すように，Shneiderman の提案と重複する主張である．

- 第 1 項（一貫性）は基本原則「システムの透過性」（手順連想容易性）の一部である．
- 第 2 項（ショートカットの提供）は基本原則「操作労力」および構成原則「熟練者優遇」

に対応する．

- 第 3 項（フィードバック）は基本原則「システムの透過性」（フィードバックの原則）に対応する．
- 第 4 項（適切なインタラクションの構成）は基本原則「システムの透過性」（理解容易性，手順連想容易性，フィードバック）に対応する．
- 第 5 項（エラー処理）は基本原則「頑健性」（誤入力防止，修復容易性）に対応する．
- 第 6 項（操作の可逆性）は基本原則「頑健性」（修復容易性）に対応する．
- 第 7 項（主体的な操作感の提供）は基本原則「システムの透過性」（手順連想容易性，フィードバック）に対応する．
- 第 8 項（短期記憶の軽減）は基本原則「システムの透過性」（理解容易性）に対応する．

Shneiderman の提案は，本研究が提案する原則と対応させると，（第 5 項と第 6 項のように）互いに内容が重複しているものがある．本研究の原則は，Shneiderman の提案と比べるとより体系化され，重複が少なく，網羅的である．

一方で，本研究が提案する「構成原則」には Shneiderman の提案に含まれていない「上級利用移行支援」の観点が含まれている．これは本研究で扱った音声応用システムにおいて非常に重要な概念であった．

さらに「導入原則」は Shneiderman の原則が論じていない応用分野の発見や提案，不適切な応用の回避，望ましい開発プロセスなどを論じるものである．

これらの考察によって，本研究でなされた音声インタフェース技術の検討は，Shneiderman の提案に基づく検討だけでは不十分であったことが示された．

7.2.2 音声インタフェースのガイドライン

近年，音声技術に特化したインタフェース設計ガイドラインが発表されている．

Weinschenk らの著書 [69] には以下のような章があり，音声インタフェースのための法則とガイドラインについて述べられている．

- Chapter 9 - Laws of Interface Design
- Chapter 10 - Speech Guidelines
- Chapter 11 - Userbility Processes and Techniques
- Chapter 12 - Universal Design

Chapter 10 の Errors という項目について具体的には以下のような提案がなされている．

- 具体的なエラーメッセージを提供せよ．use specific error messages
- 背景ノイズを除去せよ．limit background noise

- 入力デバイスを無効化できるようにせよ．allow the user to turn off the input device
- 操作を取り消す機能を提供せよ．provide an undo capability
- 聴覚アイコンを使用せよ．use an auditory icon
- 可能であればエラーにマルチモーダルのキューを使用せよ．use multi-modal cues for errors if appropriate
- 聞き返し操作を提供せよ．consider offering replay
- 人はすべてを聞き取れるという前提を捨てよ．don't assume people hear everything

これらは非常に具体的な助言ではあるが，この章にはこのような記述が Errors, Feedback, Confirmations, ... と合計 11 項目続いている．内容の重複と思われる箇所もあり，必要な要素が網羅されているかどうかも判断しにくい．

Balentine らは主に電話音声応答システムの設計者に向けて，音声認識技術を有効に活用するための具体的な手引きを書いている [34]．音声認識に限らず数字ボタン (DTMF) 入力を前提とした音声応答システムの実装についても有用な知見が多数書かれている．特に提示する音声メッセージと音声認識の語彙選択が非常に密接な関係にある，という立場は興味深い．しかし残念ながらいくつかの主張は英語という言語の語順に依存しており，音声以外のモダリティを組み合わせた場合についての議論もなされていない．全体的に，音声の聞き取りにくさ，音声認識の失敗しやすさ，という音声技術の否定的な部分が強調され，それをいかに補うべきか，といった議論が多い．

Balentine らは新たな著書 [34] において，その立場を推し進めて「機械を擬人化する音声対話技術は人間を裏切ってしまう」「機械が機械的に振る舞えるような音声対話を実現すべき」といった趣旨の主張をしている．このような Balentine らの主張には傾聴すべき点もある．

しかし本研究が目指したのは「音声に依存しないインタフェースのガイドライン」から出発し，「新しい音声応用システムの提案」「音声技術の評価手法の確立」を行うことであった．音声インタフェースに特化したガイドラインは，この研究目標には不十分であった．

また，既存のガイドラインには，音声技術に依存する主張とインタフェース設計に関する一般的な主張が混在している．達成すべき状態と具体的な実現方法が区別されていないため，新しいアプリケーションや新しい技術要素の組み合わせについて検討をすることが困難である．

7.2.3 ユニバーサルデザインの 7 原則

NC State University, The Center for Universal Design のユニバーサルデザイン提唱者[*2]による「ユニバーサルデザイン 7 原則」[*3]は，情報技術に限らず幅広い分野への応用を想定した原

[*2] Bettye Rose Connell, Mike Jones, Ron Mace, Jim Mueller, Abir Mullick, Elaine Ostroff, Jon Sanford, Ed Steinfeld, Molly Story, GreggVanderheiden

[*3] http://www.design.ncsu.edu/cud/about_ud/udprinciples.htm

則である．

この原則には，本論文で考察したインタフェースの基本原則と深く関連する主張が含まれている．また，ユニバーサルデザインの概念は，インタフェースの導入に関する原則における重要な視点でもある．

ユニバーサルデザインとは，様々な人にとって，できる限り利用可能であるように，製品，建物，環境をデザインすることである[*4]．

この原則は以下から構成される．

原則　簡潔でかつ覚えやすく表現された基本的な考え方
定義　原則に沿ったデザインをするための簡潔な方向付け
ガイドライン　原則に忠実であるために必要とされる基本要件

原則として下記の 7 項目が挙げられている．

- 原則 1: 誰にでも公平に利用できること
- 原則 2: 使う上で自由度が高いこと
- 原則 3: 使い方が簡単ですぐわかること
- 原則 4: 必要な情報がすぐに理解できること
- 原則 5: うっかりミスや危険につながらないデザインであること
- 原則 6: 無理な姿勢をとることなく，少ない力でも楽に使用できること
- 原則 7: アクセスしやすいスペースと大きさを確保すること

例えば原則 1 に対応する定義とガイドラインは以下のとおりである．

定義　誰にでも利用できるように作られており，かつ，容易に入手できること．
ガイドライン　1a. 誰もが同じ方法で使えるようにする．それが無理なら別の方法でも仕方ないが，公平なものでなくてはならない．
1b. 差別感や屈辱感が生じないようにする．
1c. 誰もがプライバシーや安心感，安全性を得られるようにする．
1d. 使い手にとって魅力あるデザインにする．

7.2.4　Web ユーザビリティとアクセシビリティ

21 世紀にはいって，ヒューマンインタフェース設計は特に Web ユーザビリティの基礎理論として注目を浴びるようになった．例えばニールセン [71] の主張は，過剰なデザインではなく，

[*4] 本項の記述は株式会社ユーディットのサイトに基づく．http://www.udit.jp/ud/ud/ud_7rules.html

定量的に評価できる使いやすさを重視しており，本研究の「基本原則」と重なる部分も多い．

しかし，本来の Web は特定のモダリティやデバイスに依存しないユニバーサルなものである．視覚に障害がある人，身体に障害がある人などを考慮した Web アクセシビリティや，マルチタッチ型モバイルデバイス (Apple iPhone, Google Android など) のための Web デザイン，VoiceXML 技術などを用いたボイスウェブのシステムは，より普遍的な原則に基づいて考察しないと最適化できない．

本研究の提案する原則は，インタフェース設計に関する一般的な主張に特化しつつ，音声やタッチパネルなど特定のモダリティ技術を当てはめて柔軟に適用することができる．すなわち，特定のモダリティやデバイスに依存しないユニバーサルな Web 技術の推進において有用である．

7.3 まとめ

4 つの研究においてインタフェース原則が果たした役割を概観し，関連する既存のガイドラインについて考察を行った．その結果として，幅広い視点に基づく複数の問題を扱っている音声応用システムの開発者や評価者にとって，本論文のインタフェース原則は必要不可欠であることを示した．

第 8 章

結論

本論文は，音声合成および音声認識を用いて構成される情報通信システムを，誰にでも使いやすいものにするための体系的な方法論を提案した．また，提案する方法論に基づいて行われた応用システムの設計，実装，評価について述べた．

第 1 章では，ヒューマンインタフェースに関する基礎理論を踏まえて，独自の視点を加えた「基本原則」「構成原則」「導入原則」の 3 つのインタフェース原則を提示した．これらの提案は音声技術に依存していない一般的な原則でありつつ，適切に構成され，特定のモダリティやデバイス，アプリケーションに当てはめて考察することが容易になっている．

第 2 章以降の 4 つの章では，音声応用システムの発見，設計，評価において，本研究で提案した原則論が有効であることを示した具体的な成果について述べた．その主要な部分は，音声技術の応用分野を発見し，その有効性を検証する 2 つの研究と，特に心的負荷（メンタルワークロード）に着目した音声インタフェース評価に関する 2 つの研究であった．

第 2 章の具体的なシステムは，音声認識・マウス・キーボードを併用した作図アプリケーションであった．第 3 章で論じたのはインターネットを介した非同期の音声コミュニケーションを実現した音声会議システムであった．

第 4 章では，車載情報システムを想定した音声対話において二重課題法を適用し，比較的高い時間分解能でワークロードの増加する箇所を特定できることを示した研究を論じた．第 5 章では視覚障害者のためのスクリーンリーダを想定した超早口音声の聞き取りにおいて NASA-TLX 法を適用し，被験者への教示によって被験者の知識や慣れの効果が制御されることを示した．

第 6 章ではその他のさまざまな音声インタフェース研究について概説した．フロー体験理論などの観点からインタフェースシステムにおける動機付けと楽しさを論じた考察，音声インタフェースにおける 7 段階モデルの詳細な検討と電話音声応答システムおよびスクリーンリーダに当てはめた考察，頭部モーションセンサと音声の有効な組み合わせ方の考察，インクリメンタル音声検索の試作，そしてラジオを音声インタフェースという観点からとらえた考察とラジオ放送そのものの支援を行うインタフェースシステムの提案，さらにオープンソースプロジェク

トの紹介をした．また，モダリティの多様性を支える知見として，マルチモーダルシステムのアーキテクチャについて論じた．

第 7 章では主に第 2 章から第 5 章までの研究を振り返って，これら 4 つの研究においてインタフェース原則が果たした役割を概観した．そして，本論文のインタフェース原則は幅広い視点に基づく多数の問題を扱うべき音声応用システムの開発者や評価者にとって必要不可欠であることを示した．

本研究の主たる貢献は，ヒューマンインタフェース技術の幅広い要素を網羅した原則の考察と，音声認識および音声合成の利用に特化した具体的な検討が，明確に分離されて議論されている点である．既存のガイドラインは従来のインタフェースデバイスに特化していたり音声技術に特化していた．これに対して本研究は，インタフェース技術の普遍的なガイドラインから自然に導かれる音声応用システムの設計や評価を論じた．

特に現在の情報通信技術に不可欠な要素である World Wide Web は，テキスト情報を中心とするユニバーサルな技術として設計されている．テキストはディスプレイに画像として表示することも，合成音声として提示することも，点字ディスプレイに出力して触覚情報として提示することもできる．音声入力，音声出力，および音声技術と共通な特質を持つ実時間メディアは，モダリティの多様化を支える要素技術である．例えば情報アクセシビリティのための音声や触覚の応用システムは，特別な分野の特別な技術として扱われることが多かった．しかし，本研究の提案する枠組みは，ソフトウェア工学，ヒューマンインタフェース，心理学，人間の視聴覚認知など，普遍的な知見や方法論の応用である．たとえ障害を持つ人のための支援技術の設計であっても，モダリティを多様化する観点から，普遍的な知見を適用し，的確に議論できる．

このように，本研究は新たなインタフェース技術や応用分野に対して柔軟に適用できる枠組みを提供したと言えよう．

謝辞

本論文は，筆者が 1993 年から 1996 年まで早稲田大学 理工学部および大学院 理工学研究科に在籍している間に行った研究，1996 年から 2002 年まで京都工芸繊維大学 工芸学部 電子情報工学科に勤務しながら行った研究，および 2002 年から 2011 年まで東京大学 大学院 情報理工学系研究科に勤務しながら行った音声インタフェースに関する研究を，早稲田大学 理工学術院における適切な指導のもとに取りまとめたものです．

インタフェースの原則およびマルチモーダル作図システムの最初の検討は，1993 年 3 月「音声・マウス・キーボードによる統合的入力環境」(信学技報 HC92-68) として小林 哲則 氏, 竹内 陽児 氏, 白井 克彦 氏によって発表されました．また音声作図システム S-tgif の研究は志田 修利 氏, 小林 哲則 氏, 白井 克彦 氏と共同で行われました．

非同期音声会議システム AVM の研究は京都工芸繊維大学にて，幸 英浩 氏, 川原 毅彦 氏, 荒木 雅弘 氏, 新美 康永 氏と共同で行われました．また AVM に関する研究の一部は科学技術振興事業団の委託事業「コミュニティ形成支援機能を持つ音声会議システム」として株式会社ネットイン京都と共同で行われました．

二重課題法の研究は京都工芸繊維大学にて，高山 元希 氏, 櫻井 晴章 氏, 荒木 雅弘 氏と共同で行われました．また，この研究においては（財）自動車走行電子技術協会 ネットワーク型音声利用システム研究委員会における活動から多くの示唆を得ました．委員長の 小林 哲則 早稲田大学教授，委員会メンバーの皆様，およびワーキンググループのメンバーの皆様に感謝します．また，助言および協力を頂いた滋賀県立大学 細馬 宏通 先生，京都工芸繊維大学 新美 康永 名誉教授およびパターン情報処理研究室の皆様に感謝します．

超早口音声の研究は 渡辺 隆行 先生（東京女子大学）と共同で行われました．渡辺 哲也 氏（現在・新潟大学）からは超早口音声の研究を始めるきっかけを与えていただきました．心的負荷評価ソフトウェアの作成においては大阪大学 篠原 一光 先生の「日本語版 NASA-TLX」を参考にしました．また本研究の一部で東京大学 小野 順貴 講師の作成した音声速度変換ソフトウェアを使用しました．東京女子大学 小田 浩一 教授，東京大学 嵯峨山 茂樹 教授，酒向 慎司 氏（現在・名古屋工業大学），早稲田大学 小林 哲則 教授，音声・音楽研究会 田中 章浩 氏（現在・早稲田大学）ほかの皆様から多くの助言や示唆を得ました．

また，本研究の一部は東京女子大学の小川 佳奈子 さん，大島 一恵 さん，小野 友理子 さん，光部 杏里 さん，藤原 扶美 さん，下永 知子 さん，福岡 千尋 さん, 狩谷 幸香 さん，瀬川 亜希 さん，林 美希 さん，望月 麻衣 さんの卒業研究として実施されました．本研究の一部は，千葉工業大学 會田 卓也 氏の修士研究として実施されました．

また，本論文で述べた研究の一部は以下の科学研究費補助金を受けました：研究課題番号 05241107 「音声言語による対話過程のモデル化に関する研究」，研究課題番号 16091210 「視覚障害者の聴覚認知の解明と音声対話への利用」，研究課題番号 17700173 「テレマティックス音声対話における安全性と快適性の評価に関する研究」，研究課題番号 17650045 「複素スペクトル円心法 (CSCC 法) によるマイクロホンアレー信号処理に関する研究」，研究課題番号 17300054 「音楽解析・認識・生成のための多重音の信号処理と情報処理の研究」，研究課題番号 20240017 「数理モデルに基づく音楽信号と音楽情報の解析・認識・加工・生成の研究」，研究課題番号 14380156 「連続音声認識手法を用いた音楽情報処理の研究」，研究課題番号 12780279 「感性自律ロボットとの自己目的的な音声対話に関する研究」，研究課題番号 11480086 「コミュニケーションの場における雰囲気情報の分析と合成に関する研究」，研究課題番号 08837011「音声対話における情報伝達能率と対話制御に関する研究」．

本研究の一部は，情報処理振興事業協会 (IPA) による独創的情報技術育成事業「擬人化音声対話エージェント基本ソフトウェアの開発」および，情報処理推進機構 (IPA) による未踏ソフトウェア創造事業 (2005 年度上期，竹林 洋一 プロジェクトマネージャ) 「ソーシャル・ネットワーキング型ラジオ番組のシステム開発」の支援を受けました．

本研究の一部は，東京大学 21 世紀 COE「情報科学技術戦略コア」の支援を受けました．

本論文の LaTeX による製版作業においては，中野 拓帆 氏にご協力いただきました．

上記に加えて，研究を支えてくださった早稲田大学，京都工芸繊維大学，東京大学，東京女子大学，千葉工業大学の皆様，共同研究を通じてご支援いただいた企業の皆様，電子情報通信学会福祉情報工学研究会をはじめとする学会・研究会・国際会議，あるいはソーシャルネットワークを通じて，貴重な示唆を与えてくださった皆様に感謝します．

また，これらの研究においては，Linux, Galatea Toolkit, 音声認識エンジン Julius, 音声合成エンジン Open JTalk, HTK をはじめとする多くのオープンソースプロジェクトの成果を利用させていただきました．これらのソフトウェアの開発者の方々にも感謝の意を表します．

最後に，筆者の研究を陰ながら支えていただいた友人・知人の皆様，そして家族に感謝の気持ちを伝えたいと思います．

付録 A

研究実績

A.1 学術誌原著論文 (第一著者)

1. 西本 卓也, 渡辺 隆行: “単語親密度を統制した超早口音声の聴取に対する慣れの検討,” 電子情報通信学会論文誌 D, Vol.J94-D, No.1, pp.209-220, Jan. 2011.
2. 西本 卓也, 高山 元希, 櫻井 晴章, 荒木 雅弘: “音声インタフェースのための対話負荷測定法,” 電子情報通信学会論文誌 D-II, Vol.J87-D-II, No.2, pp.513-520, Feb. 2004.
3. 西本 卓也, 幸 英浩, 川原 毅彦, 荒木 雅弘, 新美 康永: “非同期型音声会議システム AVM の設計と評価,” 電子情報通信学会論文誌 D-II, Vol.J83-D-II, No.11 pp.2490-2497, Nov. 2000.
4. 西本 卓也, 志田 修利, 小林 哲則, 白井 克彦: “マルチモーダル入力環境下における音声の協調的利用—音声作図システム S-tgif の設計と評価—,” 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.12, pp.2176-2183, Dec. 1996.

A.2 学術誌原著論文 (第一著者でないもの)

1. 西亀 健太, 和泉 洋介, 渡部 晋治, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “スパース性に基づくブラインド音源分離を用いたステレオ入力音声認識,” 電子情報通信学会論文誌 D, Vol.J93-D, No.3, pp.303-311, Mar. 2010.
2. Shoichiro Saito, Hirokazu Kameoka, Keigo Takahashi, Takuya Nishimoto, Shigeki Sagayama: “Specmurt Analysis of Polyphonic Music Signals,” IEEE Trans. on Audio, Speech, and Language Processing, Vol.16, No.3, pp.639-650, 2008.
3. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: “A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering,” IEEE Trans. on Audio, Speech and Language Processing, Vol. 15, No. 3, pp.982-994, Mar., 2007.

4. 武田 晴登, 西本 卓也, 嵯峨山 茂樹: “テンポ曲線と隠れマルコフモデルを用いた 多声音楽 MIDI 演奏のリズムとテンポの同時推定,” 情報処理学会論文誌, Vol.48, No.1, pp.237-247, Jan. 2007.
5. 鎌本 優, 守谷 健弘, 原田 登, 西本 卓也, 嵯峨山 茂樹: “ISO/IEC MPEG-4 Audio Lossless Coding（ALS）におけるチャネル内とチャネル間の長期予測,” 電子情報通信学会論文誌 D, Vol.J89-B, No.2, pp.214-222, Feb. 2006.
6. 鎌本 優, 守谷健弘, 西本卓也, 嵯峨山茂樹: “チャネル間相関を用いた多チャネル信号の可逆圧縮符号化,” 情報処理学会論文誌, Vol.46, No.5, pp.1118-1128, May 2005.
7. 武田 晴登, 西本 卓也, 嵯峨山 茂樹: “確率モデルによる多声音楽演奏の MIDI 信号のリズム認識,” 情報処理学会論文誌, Vol.45, No.3, pp.670-679, Mar. 2004.
8. 奥 智岐, 西本 卓也, 荒木 雅弘，新美 康永: “タスクに依存しない音声対話の制御方式,” 電子情報通信学会論文誌 D-II, Vol. J86-D-II, No.5, pp.608-615, May 2003.
9. 川本 真一, 下平 博, 新田 恒雄, 西本 卓也, 中村 哲, 伊藤 克亘, 森島 繁生, 四倉 達夫, 甲斐 充彦, 李 晃伸, 山下洋一, 小林 隆夫, 徳田 恵一, 広瀬 啓吉, 峯松 信明, 山田 篤, 伝 康晴, 宇津呂 武仁, 嵯峨山 茂樹: “カスタマイズ性を考慮した擬人化音声対話ソフトウェアツールキットの設計,” 情報処理学会論文誌, vol.43, no.7, pp.2249-2263, Jul. 2002.

A.3 学術誌論文 (翻訳)

1. Takuya Nishimoto, Motoki Takayama, Haruaki Sakurai, Masahiro Araki: “Measurement of Workload for Voice User Interface Systems,” Systems and Computers in Japan, Volume 36, Issue 8, pp.81-89, May 2005.
2. Tomoki Oku, Takuya Nishimoto, Masahiro Araki, Yasuhisa Niimi: “A task-independent control method for spoken dialogs,” Systems and Computers in Japan, Volume 35, Issue 14, pp.87-95, 2004.
3. Takuya Nishimoto, Hidehiro Yuki, Takehiko Kawahara, Masahiro Araki, Yasuhisa Niimi: “Design and evaluation of the asynchronous voice meeting system AVM,” Systems and Computers in Japan, Volume 33, Issue 11, pp.61-69, 2002.

A.4 総説 (学術誌の解説，講座等)

1. 荒木 雅弘, 西本 卓也: “基礎講座 音響・音声インタフェース 第 2 回 音声対話システムの開発方法論とプラットフォーム,” ヒューマンインタフェース学会誌 Vol.12 No.2, pp.51-54, May 2010.

2. 西本 卓也, 西田 昌史: “インターネットと音声合成,” 電子情報通信学会誌, Vol.91, No.12, pp.1030-1035, Dec 2008.
3. 新田 恒雄, 松浦 博, 西本 卓也, 西村 雅史: “音声言語インタフェースのための情報処理学会試行標準,” 情報処理学会誌 Vol.47 No.7, pp.762-767, Jul 2006.
4. 嵯峨山 茂樹, 武田 晴登, 亀岡 弘和, 西本 卓也: “音声認識技術を用いた音楽情報処理,” 日本音響学会誌, Vol. 61, No. 8, pp. 454-460, Aug. 2005.
5. 嵯峨山 茂樹, 西本 卓也, 中沢 正幸: “擬人化音声対話エージェント,” 情報処理学会誌, Vol.45, No.10, pp.1044-1049, Oct. 2004.

A.5 講演 (査読つき国際会議予稿)

1. Miquel Espi, Shigeki Miyabe, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: “Analysis on Speech Characterization for Robust Voice Activity Detection,” Proc. of IEEE SLT Workshop, Dec., 2010.
2. Siwei Qin, Satoru Fukayama, Takuya Nishimoto, Shigeki Sagayama: “Lexical Tones Learning with Automatic Music Composition System Considering Prosody of Mandarin Chinese,” INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology, p.4, Sep., 2010.
3. Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: “Automatic Song Composition form Lyrics with Singing Voice Synthesizer,” Proc. of Intersinging, 2, pp.1-4, Oct., 2010.
4. Jun Wu, Yu Kitano, Stanislaw Andrzej Raczynski, Shigeki Miyabe, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: “Musical Instrumental Identification Based on Harmonic Temporal Timbre Features,” Proc. Workshop on Statistical and Perceptual Audition (SAPA), pp.7-12, Sep., 2010.
5. Tsubasa Tanaka, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: “Automatic Music Composition based on Counterpoint and Imitation using Stochastic Models,” Proceedings of SMC 2010, in CD-ROM, PS2-3, pp.1-8, Jul., 2010.
6. Tae Hun Kim, Satoru Fukayama, Takuya Nishimoto, Shigeki Sagayama: “Performance rendering for polyphonic piano music with a combination of probabilistic models for melody and harmony,” Proceedings of SMC, pp.23-30, Jul., 2010.
7. Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Takuya Nishimoto, Shigeki Sagayama: “Automatic Song Composition from the Lyrics exploiting Prosody of the Japanese Language,” Proceedings of SMC, pp.299-302, Jul., 2010.

8. Takuya Nishimoto, Takayuki Watanabe: "The Comparison Between the Deletion-Based Methods and the Mixing-Based Methods for Audio CAPTCHA Systems," Proceedings of Interspeech 2010, pp.266-269, 2010-09-27, Makuhari, Chiba, Japan.
9. Takuya Nishimoto, Takayuki Watanabe: "The evaluation of deletion-based method and mixing-based method for audio CAPTCHAs," K. Miesenberger et al. (Eds.) ICCHP 2010, Part I, LNCS 6179, pp.368–375, 2010.
10. Yushi Ueda, Yuuki Uchiyama, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: "HMM-Based Approach for Automatic Chord Detection Using Refined Acoustic Features," Proc. of ICASSP, SS-L5.4, pp.5518-5521, Mar. 2010.
11. Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Yu-ichiro Yonebayashi, Tae Hun Kim, Qin Si Wei, Takuho Nakano, Takuya Nishimoto, Shigeki Sagayama: "Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics," Entertainment Computing - ICEC 2009, pp.309-310, Sep. 2009.
12. Kouichi Katsurada, Akinobu Lee, Tetsuya Kawahara, Tatsuo Yotsukura, Shigeo Morishima, Takuya Nishimoto, Yoichi Yamashita, Tsuneo Nitta: "Development of a toolkit for spoken dialog systems with an anthropomorphic agent: Galatea," Proceedings of APSIPA ASC 2009, MP-SS1-5 0226, Sapporo, Oct 2009.
13. Yosuke Izumi, Kenta Nishiki, Shinji Watanabe, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: "Stereo-input Speech Recognition using Sparseness-based Time-frequency Masking in a Reverberant Environment," Proceedings of Interspeech 2009, Wed-Ses2-O4, Brighon, Sep. 2009.
14. Yutaka Kamamoto, Noboru Harada, Takehiro Moriya, Nobutaka Itou, Nobutaka Ono, Takuya Nishimoto, Shigeki Sagayama: "An efficient lossless compression of multichannel time-series signals by MPEG-4 ALS, " Proc. ISCE, pp.159-164, May. 2009.
15. Ikumi Ota, Ryo Yamamoto, Takuya Nishimoto, Shigeki Sagayama: "On-Line Handwritten Kanji String Recognition Based on Grammar Description of Character Structures," Proc. of ICPR 2008, Dec. 2008.
16. Takuya Nishimoto, Takayuki Watanabe: "An analysis of human-to-human dialogs and its application to assist visually-impaired people," Computers Helping People with Special Needs, LNCS 5105, Springer, (Proceedings of 11th International Conference, ICCHP 2008, Linz, Austria), pp.809-812, 10th Jul 2008.
17. Takuya Nishimoto, Yukika Kariya, Takayuki Watanabe: "The effect of learning on listening to ultra-fast speech," Proceedings of Acoustics '08 Paris, pp.6119-6124, Jul. 2008.

18. Ken-ichi Miyamoto, Hirokazu Kameoka, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: "Harmonic-Temporal-Timbral Clustering (HTTC) For the Analysis of Multi-instrument Polyphonic Music Signals," Proc. of ICASSP, pp.113-116, Apr. 2008.
19. Takuya Nishimoto, Shinji Sako, Shigeki Sagayama, Kazue Ohshima, Koichi Oda, Takayuki Watanabe: "Effect of Learning on Listening to Ultra-Fast Synthesized Speech," Proceedings of the 28th IEEE Engineering in Medicine and Biology Society Annual International Conference (EMBC2006), pp.5691-5694, New York, Sep 2006.
20. Juhei Takahashi, Makoto Shioya, Takuya Nishimoto, Hideharu Daigo: "A Study on Dialog Management Corresponding to the Driver's Workload and Other Factors," 12th ITS World Congress (San Francisco), Nov. 2005.
21. Takuya Nishimoto, Makoto Shioya, Juhei Takahashi, Hideharu Daigo: "A study of dialogue management principles corresponding to the driver's workload," Biennial Workshop on Digital Signal Processing for In-Vehicle and mobile systems, Sesimbra, Portugal, Sep. 2005.
22. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: "Maximum Likelihood Based HMM State Filtering Approach to Model Adaptation for Long Reverberation," IEEE ASRU'05, Mexico, Nov 2005.
23. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: "Harmonic-temporal structured clustering via deterministic annealing EM algorithm for audio feature extraction," in Proc. International Conference on Music Information Retrieval (ISMIR2005), pp. 115-122, 2005.
24. Shoichiro Saito, Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: "Specmurt Analysis of Multi-Pitch Music Signals with Adaptive Estimation of Common Harmonic Structure," in Proc. International Conference on Music Information Retrieval (ISMIR2005), pp.84-91, Sep. 2005.
25. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: "Model Adaptation by State Splitting of HMM for Long Reverberation," Proc. Interspeech 2005 (Lisbon, Portugal), pp. 277–280, Sep. 2005.
26. Shigeki Sagayama, Hirokazu Kameoka, Shoichiro Saito, Takuya Nishimoto: "Specmurt Anasylis' of Multi-Pitch Signals," Proc. IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing, (May 18-20, 2005, Sapporo Convention Center, Sapporo, Japan), 2005.
27. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: "Audio Stream Segrega-

tion Based on Time-Space Clustering Using Gaussian Kernel 2-Dimensional Model," Proc. IEEE, International Conference on Acoustics, Speech and Signal Processing (ICASSP 2005) (Philadelphia, PA, USA), AE-L1.2, Vol.3, pp. 5-8, Mar 2005.

28. Kazuhito Inoue, Takashi Okajima, Yutaka Kamamoto, Takuya Nishimoto, Shigeki Sagayama: "Complex Spectrum Circle Centroid Method for Noise Reduction in Array Signal Processing," Workshop on Hand-free Speech Communication and Microphone Arrays, New Jersey, USA, Mar 2005.
29. Makoto Shioya, Nobuo Hataoka, Takuya Nishimoto, Juhei Takahashi, Hideharu Daigo: "Study on Reference Models for HMI in Voice Telematics to Meet Driver's Mind Distraction," 11th ITS World Congress (Nagoya), Nov. 2004.
30. Haruto Takeda, Takuya Nishimoto, Shigeki Sagayama: "Rhythm and Tempo Recognition of Music Performance from a Probabilistic Approach," Proc. 5th International Conference on Music Information Retrieval (ISMIR) (Barcelona, Spain), pp.357-364, Oct. 2004.
31. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: "Multi-Pitch Trajectory Estimation of Concurrent Speech Based on Harmonic GMM and Nonlinear Kalman Filtering," Proc. International Conference on Spoken Language Processing (ICSLP2004) (Jeju, Korea), Oct. 2004.
32. Shigeki Sagayama, Takashi Okajima, Yutaka Kamamoto, Takuya Nishimoto: "Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition," Proc. International Conference on Spoken Language Processing (ICSLP2004) (Jeju, Korea), Oct. 2004.
33. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: "Model Composition by Lagrange Polynomial Approximation for Robust Speech Recognition in Noisy Environment," Proc. International Conference on Spoken Language Processing (ICSLP2004) (Jeju, Korea), Oct. 2004.
34. Shigeki Sagayama, Hirokazu Kameoka, Takuya Nishimoto: "Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signal by Deconvolution of Log-Frequency Spectrum," Proc. 2004 ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA2004), (2 October 2004 - 3 October 2004, Jeju, Korea), Oct. 2004.
35. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: "Separation of Harmonic Structures Based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds," Proc. IEEE, International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004) (Montreal, Canada), May 2004.

36. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: “Extraction of Multiple Fundamental Frequencies from Polyphonic Music Using Harmonic Clustering,” Proc. International Congress on Acoustics (ICA) (Kyoto, Japan), Apr. 2004.
37. Haruto Takeda, Takuya Nishimoto, Shigeki Sagayama: “Maximum Likelihood Method for Estimating Rhythm and Tempo,” Proc. International Symposium on Music Acoustics (ISMA) (Nara, Japan), in CD-ROM, Mar. 2004.
38. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: “Multi-pitch Detection Algorithm Using Constrained Gaussian Mixture Model and Information Criterion,” Proc. Speech Prosody 2004 (Nara, Japan), pp.533-536, Mar. 2004.
39. Hirokazu Kameoka, Takuya Nishimoto, Shigeki Sagayama: “Accurate F0 Detection Algorithm for Concurrent Sounds Based on EM Algorithm and Information Criterion,” Proc. Special Workshop in Maui (SWIM) (Maui, USA) in CD-ROM, Jan. 2004.
40. Hitoshi Yamamoto, Takuya Nishimoto and Shigeki Sagayama: “Frame-by-frame HMM Adaptation for Reverberant Speech Recognition,” Proc. Special Workshop in Maui (SWIM) (Maui, USA) in CD-ROM, Jan. 2004.
41. Makoto Shioya, Nobuo Hataoka, Takuya Nishimoto, Juhei Takahashi, Takashi Odajima: “Research on Network-Distributed Voice-Activated System Architectures for Telematics,” 10th ITS World Congress (Madrid), PS156, Nov. 2003.
42. Shin-ichi Kawamoto, Hiroshi Shimodaira, Tsuneo Nitta, Takuya Nishimoto, Satoshi Nakamura, Katsunobu Itou, Shigeo Morishima, Tatsuo Yotsukura, Atsuhiko Kai, Akinobu Lee, Yoichi Yamashita, Takao Kobayashi, Keiichi Tokuda, Keikichi Hirose, Nobuaki Minematsu, Atsushi Yamada, Yasuharu Den, Takehito Utsuro, Shigeki Sagayama: “Open-source Software for Developing Anthropomorphic Spoken Dialog Agent,” Proc. of PRICAI-02, International Workshop on Lifelike Animated Agents, pp.64-69, 2002.
43. Takuya Nishimoto, Masahiro Araki, Yasuhisa Niimi: “RadioDoc : A Voice-Accessible Document System,” ICSLP2002, pp.1485-1488, Denver, 2002-09.
44. Takuya Nishimoto, Takayuki Kawasaki: “Support tools for broadcasting self-created radio programs for the visually impaired,” CSUN 17th Annual International Conference, Technology and Persons with Disabilities, March 2002.
45. Takuya Nishimoto, Masahiro Araki, Yasuhisa Niimi: “The Practical Side of Teaching the Elderly Visually Impaired User to Use the E-Mail,” Proceedings of the UAHCI 2001 Conference, (Universal Access in Human-Computer Interaction) New Orleans, Louisiana USA, Vol.3, pp.963-967, 2001-08.

46. Takuya Nishimoto, Hidehiro Yuki, Takehiko Kawahara, Yasuhisa Niimi, “An Asynchronous Virtual Meeting System for Bi-Directional Speech Dialog,” Eurospeech’99, Vol.6, pp.2471-2474, 1999.
47. Takuya Nishimoto, Yutaka Kobayashi, Yasuhisa Niimi: “Spoken Dialog System for Database Access on Internet,” AAAI Spring Symposium SS-97-05, pp.146 – 153, 1997.
48. Takuya Nishimoto, Nobutoshi Shida, Tetsunori Kobayashi, Katsuhiko Shirai: “Improving Human Interface in Drawing Tool Using Speech, Mouse and Key-Board,” Proceedings of 4th IEEE International Workshop on Robot and Human Communication, ROMAN95, pp.107-112, 1995.
49. Takuya Nishimoto, Nobutoshi Shida, Tetsunori Kobayashi, Katsuhiko Shirai: “Multimodal Drawing Tool Using Speech, Mouse and Key-Board,” Proceedings of ICSLP94, S22-12, pp.1287-1290, 1994.

A.6 講演 (研究会)

1. 盧 迪, 深山 覚, 西本 卓也, 嵯峨山 茂樹: “音声入力への応答タイミング決定のための強化学習の検討,” 電子情報通信学会技術報告 (音声研究会)(共催 日本音響学会 聴覚研究会), Mar. 2011.
2. 深山 覚, 西本 卓也, 嵯峨山 茂樹, “歌唱曲自動作曲の需要と今後 - 2 年間の Orpheus 運用を通じて,” 情報処理学会研究報告, 86, 2, pp.1-6, Jul. 2010.
3. 金 泰憲, 深山 覚, 西本 卓也, 嵯峨山 茂樹, “単旋律と和音の確率モデルの組み合わせによるピアノ曲演奏の自動表情付け,” 情報処理学会研究報告, 85, 2, pp.1-6, May. 2010.
4. 山口 俊光, 西本 卓也, 四方田 正夫, 渡辺 哲也: “第 53 回 WIT 研究会におけるリアルタイム映像配信の報告,” 電子情報通信学会技術報告 (共催：ヒューマンインタフェース学会研究会), Vol.110, No.164, 福祉情報工学研究会 (SIG-WIT), pp.73-78, Aug. 2010.
5. 田中 翼, 西本 卓也, 小野 順貴, 嵯峨山 茂樹, “確率モデルを用いた対位法および模倣に基づく自動作曲,” 日本音響学会音楽音響研究会, 28, 8 (MA2009-84), pp.13-18, Mar., 2010.
6. 西本 卓也, 松村 瞳, 渡辺 隆行: “音声 CAPTCHA システムにおける削除法と混合法の比較,” 電子情報通信学会技術報告 (福祉情報工学研究会・音声研究会), WIT2009-64/SP2009-58, pp.55-60, Aomori, Oct 2009.
7. 荒木 雅弘, 西本 卓也, 桂田 浩一, 新田 恒雄: “階層的 MMI アーキテクチャに基づくプラットフォーム実装方法の検討,” 情報処理学会 研究報告 音声言語情報処理（SLP）,

Vol.2009-SLP-078 No.5, Tokyo, Oct 2009.

8. 水野 優, 小野 順貴, 西本 卓也, 嵯峨山 茂樹: “パワースペクトログラムの伸縮に基づく多重音信号の再生速度と音高の実時間制御,” 日本音響学会聴覚研究会資料, 39, 6, pp.447-452, Oct., 2009.
9. 深山 覚, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “非和声音規則に基づく経路制約を用いた自動旋律生成,” 情報処理学会研究報告, 2009-MUS-81(15), pp.1-6, Jul., 2009.
10. 長谷川 隆, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “音楽知識に基づく音高・音長の組合せ特徴量を用いた MIDI データからの作曲家判別,” 情報処理学会研究報告, 2009-MUS-79(10), pp.47-52, Feb., 2009.
11. 西亀 健太, 和泉 洋介, 渡部 晋治, 小野 順貴, 西本 卓也, 嵯峨山 茂樹: “スパース性に基づくブラインド音源分離を用いた 2 チャンネル入力音声認識,” 電子情報通信学会技術研究報告 (SP), Vol.108, No.338, pp.1-6, Dec., 2008.
12. 瀬川 亜希, 西本 卓也, 渡辺 隆行: “超早口音声の聴取における単語親密度の教示効果,” 電子情報通信学会 技術報告 (福祉情報工学研究会), WIT2008-55, pp.89-94, Dec 2008.
13. 福岡 千尋, 西本 卓也, 渡辺 隆行: “音韻修復効果を用いた音声 CAPTCHA の検討,” 電子情報通信学会 技術報告 (福祉情報工学研究会), WIT2008-54, pp.83-88, Dec 2008.
14. 山下 洋一, 李 晃伸, 河原 達也, 四倉 達夫, 西本 卓也, 桂田 浩一, 新田 恒雄: “音声対話技術コンソーシアム (ISTC) の活動成果報告,” 情報処理学会研究報告 2008-SLP-73(9), pp.47-52, Oct 2008.
15. 西本 卓也, 西亀 健太, 嵯峨山 茂樹, 福岡 千尋, 渡辺 隆行: “音声 CAPTCHA のための音韻修復効果の検討,” 日本音響学会聴覚研究会資料,Vol.38, No.6, H-2008-111, pp.639-644, Oct 2008.
16. 内山 裕貴, 宮本 賢一, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “調波音・打楽器音分離手法を用いた音楽音響信号からの自動和音認識,” 情報処理学会研究報告, 2008-MUS-76(23), pp.137-142, Aug., 2008.
17. 深山 覚, 中妻 啓, 米林 裕一郎, 酒向 慎司, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “Orpheus 歌詞の韻律に基づいた自動作曲システム,” 情報処理学会研究報告, 2008-MUS-76(30), pp.179-184, Aug., 2008.
18. 西亀 健太, 渡部 晋治, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “複数残響特性下の音声を単一モデル学習に用いた未知残響環境に頑健な音声認識の検討,” 電子情報通信学会技術研究報告 (SP), 108, 66, pp.43-48, May., 2008.
19. 大田 郁実, 山本 遼, 西本 卓也, 嵯峨山 茂樹: “文字構造の文法記述に基づくオンライン手書き漢字列認識,” 電子情報通信学会技術研究報告 (PRMU), 107, 491, pp.75-80, Feb., 2008.
20. 諸岡 孟, 西本 卓也, 嵯峨山 茂樹: “確率文脈自由文法による和声学規則の表現と楽曲の自

動和声解析," 情報処理学会研究報告, 2008, 12, pp.77-82, Feb., 2008.

21. 西本 卓也, 岩田 英三郎, 櫻井 実, 廣瀬 治人: "探索的検索のための音声入力インタフェースの検討," 情報処理学会研究報告 2008-HCI-127(2), pp.9-14, Jan 2008.
22. 西本 卓也, 狩谷 幸香, 渡辺 隆行: "早口音声の聴取訓練における単語親密度の影響," 電子情報通信学会技術報告 Vol.107 No.406 (NLC2007-53, SP2007-116, SLP 共催), pp.119-124, Dec 2007.
23. 狩谷 幸香, 西本 卓也, 渡辺 隆行: "早口音声聴取における単語親密度と学習効果の検討," 電子情報通信学会技術報告, WIT2007-44, pp.67-72, Dec 2007.
24. 新田 恒雄, 桂田 浩一, 荒木 雅弘, 西本 卓也, 甘粕 哲郎, 川本 真一: "マルチモーダル対話システムのための階層的アーキテクチャの提案," 情報処理学会研究報告, 2007-SLP-68(2), pp.7-12, Oct 2007.
25. 西本 卓也, 嵯峨山 茂樹, 藤原 扶美, 下永 知子, 渡辺 隆行: "対面朗読者と視覚障害者の対話の分析とその応用," 情報処理学会研究報告, 2007-SLP-11, pp.55-60, Feb 2007.
26. 會田 卓也, 西本 卓也, 大川 茂樹, 嵯峨山 茂樹: "頭部モーションセンサと音声を用いた対話インタフェースの検討," 電子情報通信学会技術報告, WIT2006-87, pp.85-90, Jan 2007.
27. 藤原 扶美, 下永 知子, 渡辺 隆行, 西本 卓也: "対話分析に基づく視覚障害者用音声対話システム," 電子情報通信学会技術報告, WIT2006-67, pp.93-102, Dec 2006. (第 41 回ヒューマンインタフェース学会研究会 共催)
28. 小野 友理子, 渡辺 隆行, 西本 卓也: "早口合成音声に対する高齢者の慣れ," 電子情報通信学会技術報告, WIT2006-70, pp.115-120, Dec 2006. (第 41 回ヒューマンインタフェース学会研究会 共催)
29. 西本 卓也, 川崎 隆章: "ラジオ放送支援システム「オラビー」の開発," 電子情報通信学会技術報告, WIT2006-25, pp.49-54, Jul 2006.
30. 西本 卓也, 光部 杏里, 渡辺 隆行: "ラジオ放送番組におけるスポーツ実況中継の分析," 電子情報通信学会技術報告, WIT2006-6, pp.27-32, May 2006.
31. 西本 卓也, 小川 佳奈子, 渡辺 隆行: "対面朗読者と視覚障害者の対話の分析," 電子情報通信学会技術報告, WIT2005-75, pp.7-12, Mar 2006.
32. 山本 遼, 酒向 慎司, 西本 卓也, 嵯峨山 茂樹: "ストローク単位の確率文脈自由文法を用いたオンライン手書き数式認識," 電子情報通信学会技術研究報告, PRMU2005-221, pp.111-116, Feb. 2006.
33. 大島 一恵, 西本 卓也, 渡辺 隆行: "視覚障害者用早口合成音声による慣れの効果," 電子情報通信学会技術報告, WIT2005-43 / SP2005-81, pp.19-24, Oct 2005.
34. 西本 卓也, 酒向 慎司, 嵯峨山 茂樹, 小田 浩一, 渡辺 隆行: "早口合成音声の聴取実験によるテキスト音声合成の評価," 電子情報通信学会技術報告, WIT2005-5, pp.23-28, May

2005.

35. 渡辺 隆行, 安村 通晃, 小田 浩一, 西本 卓也: “視覚障害者の聴覚認知の解明と音声対話への利用に向けて,” 電子情報通信学会技術報告, WIT2004-74, pp.7-12, Mar 2005.
36. 西本 卓也, 西村 雅史, 赤堀 一郎, 石川 泰, 磯谷 亮輔, 伊藤 克亘, 大淵 康成, 金澤 博史, 國枝 伸行, 外山 聡一, 新田 恒雄: “音声認識応用に関する学会試行標準,” 情報処理学会研究報告, 2005-SLP-55, pp.47-52, Feb. 2005.
37. 酒向 慎司, 西本 卓也, 嵯峨山 茂樹: “実世界環境における視聴覚情報を統合した擬人化対話エージェントシステムの検討,” 人工知能学会研究報告 (SIG-SLUD), 2005-SLUD-A502, pp. 35-40, Nov. 2005.
38. 米田 隆一, 西本 卓也, 嵯峨山 茂樹: “マルコフ確率場を用いた調認識，自動和声付け，および自動対位法,” 情報処理学会研究報告 (MUS), 2005-MUS-63, pp. 31-36, Dec. 2005.
39. 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “調波時間構造化クラスタリング (HTC) による音楽の音響特徴量同時推定,” 情報処理学会研究報告, 2005-MUS-61-12, pp. 71-78, Aug. 2005.
40. 齊藤 翔一郎, 武田 晴登, 西本 卓也, 嵯峨山 茂樹: “specmurt 分析と chroma vector を用いた HMM による音楽音響信号の調認識,” 情報処理学会研究報告 (MUS), 2005-MUS-61, pp. 85-90, Aug. 2005.
41. 井上 和士, 鎌本 優, 岡嶋 崇, 西本 卓也, 嵯峨山 茂樹: “複素スペクトル円心 (CSCC) の推定に基づくマイクロホンアレーによる雑音抑圧,” 電子情報通信学会技術研究報告, SP2004-145, pp. 1-6, Jan. 2005.
42. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: “Model Adaptation for Reverberant Speech by HMM State Splitting and Convolution of Distributions,” 電子情報通信学会技術研究報告, SP2004-151, pp. 37-42, Jan. 2005.
43. 槻 武也, 西本 卓也, 嵯峨山 茂樹: “音響モデル変換による残響環境中の音声認識,” 電子情報通信学会技術研究報告, SP2004-150, pp. 31-36, Jan. 2005.
44. 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “擬人化音声対話エージェントにおける視線制御モデルの提案,” 人工知能学会 SIG-SLUD-A303, pp.21-26, Mar 2004.
45. 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “擬人化音声対話エージェントにおける視線制御方法の検討,” 情報処理学会研究報告, 2003-SLP-50, pp.63-68, Feb 2004.
46. 鎌本 優, 守谷 健弘, 西本 卓也, 嵯峨山 茂樹: “多チャンネル時系列信号のロスレス符号化,” Proc. 27th Symposium on Information Theory and Its Applications (SITA2004), (Gero, Gifu, Japan), Vol. 2, pp. 819-822, Dec. 2004.
47. 亀岡 弘和, 齊藤 翔一郎, 西本 卓也, 嵯峨山 茂樹: “Specmurt における隼最適共通調波構造パターンの反復推定による多声音楽信号の可視化と MIDI 変換,” 情報処理学会研究報告 (MUS), 2004-MUS-56, pp. 41-48, Aug. 2004.

48. 中潟 昌平, 西本 卓也, 嵯峨山 茂樹: “動的計画法と音列出現確率を用いた対位法の対旋律の自動生成,” 情報処理学会研究報告 (MUS), 2004-MUS-56, pp. 65-70, Aug. 2004.
49. 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “調波スペクトル分離の原理 Harmonic Clustering と赤池情報量規準による多声部楽曲音響信号の同時発音数および多重ピッチの推定,” 日本音響学会 7 月音楽音響研究会資料, Jul, 2004.
50. 武田 晴登, 西本 卓也, 嵯峨山 茂樹: “リズム語彙を用いた HMM による MIDI 演奏のリズムとテンポ推定,” 情報処理学会研究報告 (MUS), 2004-MUS-54, pp.51-56, Mar. 2004.
51. 嵯峨山 茂樹, 伊藤 克亘, 宇津呂 武仁, 甲斐 充彦, 小林 隆夫, 下平 博, 伝 康晴, 徳田 恵一, 中村 哲, 西本 卓也, 新田 恒雄, 広瀬 啓吉, 峯松 信明, 森島 繁生, 山下 洋一, 山田 篤, 李 晃伸: “擬人化音声対話エージェント基本ソフトウェアの開発プロジェクト報告, ” 信学技報, Vol.103, No.520, SP2003-168, pp.73-78, (情報処理学会研究報告, 2003-SLP-49), Dec. 2003.
52. 西本 卓也, 中沢 正幸, 嵯峨山 茂樹: “音声対話における擬人化エージェントの利用効果の検討,” 情報処理学会研究報告, 2003-SLP-47, pp.25-30, Jul. 2003.
53. 西本 卓也, 嵯峨山 茂樹: “擬人化エージェント Galatea のための VoiceXML 処理系,” 第 17 回人工知能学会全国大会, 2C2-04, Jun. 2003.
54. 住吉 悠希, 荒木 雅弘, 西本 卓也: “ラジオ番組収録のための音声インタフェースの設計と評価,” 人工知能学会 SIG-SLUD-A203, pp.151-156, Mar. 2002.
55. 西本 卓也, 高山 元希, 荒木 雅弘: “音声インタフェースにおける認知的負荷測定法とその評価,” 情報処理学会研究報告, 2002-SLP-45-5, pp.29-34, Feb. 2003.
56. 嵯峨山 茂樹, 川本 真一, 下平 博, 新田 恒雄, 西本 卓也, 中村 哲, 伊藤 克亘, 森島 繁生, 四倉 達夫, 甲斐 充彦, 李 晃伸, 山下 洋一, 小林 隆夫, 徳田 恵一, 広瀬 啓吉, 峯松 信明, 山田 篤, 伝 康晴, 宇津呂 武仁: “擬人化音声対話エージェントツールキット Galatea,” 情報処理学会研究報告, 2002-SLP-45-10, pp.57-64, Feb. 2003.
57. 松浦 博, 西本 卓也, 金子 宏, 磯谷 亮輔, 石川 泰, 西村 雅史, 伊藤 克亘, 新田 恒雄: “音声認識読み記号および音声関連ソフトウエアに係わる用語の試行標準案,” 情報処理学会研究報告, 2002-SLP-45-11, pp.65-70, Feb. 2003.
58. 西本 卓也: “音声インタフェースは本当に人に優しいか？,” 人工知能学会研究会資料, SIG-SLUD-A202-05(11/7), pp.27-32, Nov. 2002.
59. 西本 卓也, 伊勢 史郎, 大村 皓一, 高木 治夫: “3 次元キーエコーを用いたタイピング練習,” 電子情報通信学会技術研究報告, WIT2002-4, pp.19-24, Jun. 2002.
60. 岐津 三泰, 西本 卓也, 荒木 雅弘: “擬人化エージェントのための VoiceXML 処理系の開発,” 人工知能学会 SIG-SLUD-A201-01, pp.1-6, Jun. 2002.
61. 西本 卓也, 櫻井 晴章, 荒木 雅弘, 新美 康永: “ディクテーション作業における楽しさの

分析," 人工知能学会 SIG-SLUD-A103-01(3/8), pp.01-06, Mar. 2002.
62. 西本 卓也, 新田 恒雄, 足立 裕秋, 桂田 浩一: "対話システムにおけるタスク記述とプロトタイプ作成支援," 情報処理学会 SLP-40-13, HI-97-13, pp.73-78, Feb. 2002.
63. 高山 元希, 西本 卓也, 荒木 雅弘, 新美 康永: "電話音声応答システムにおける効果音の役割," 電子情報通信学会技術研究報告, SP 2001-132, pp.55-62, Jan. 2002.
64. 植田 喜代志, 秋田 祥史, 荒木 雅弘, 西本 卓也, 新美 康永: "VoiceXML のマルチモーダル化の検討," 情報処理学会研究報告，2001-SLP-38-7, pp.43-48, Oct. 2001.
65. 西本 卓也, 園 順一, 浅野 令子, 高木 治夫: "視覚障害者のためのタイピング練習ソフトの設計と評価," 電子情報通信学会技術研究報告, WIT 2001-11, pp.7-12, Aug. 2001.
66. 西本 卓也, 宮川 祥子, 川崎 隆章: "インターネットラジオによる情報発信支援ツールの設計," 電子情報通信学会技術研究報告, WIT 2001-7, pp.35-40, May 2001.
67. 西本 卓也, 住吉 悠希, 荒木 雅弘, 新美 康永: "視覚障害者のための電子メール環境における操作性の検討," ヒューマンインタフェース学会研究報告集, Vol.2 No.5, pp.33-38, Dec. 2000.
68. 板口 晋也，西本 卓也: "視覚障害者のためのタイピング練習環境," 日本音響学会関西支部若手研究者交流研究発表会, Dec. 2000.
69. 西本 卓也: "視覚障害者のための電子メール環境の操作性について," 第 8 回 ITRC 総会・研究会, Nov. 2000.
70. 西本 卓也, 荒木 雅弘, 新美 康永: "視覚障害者のためのタイピング練習ソフト「打ち込み君」の改良," 電子情報通信学会技術研究報告（福祉情報工学研究会）, WIT00-21, pp.55-59, Aug. 2000.
71. 西本 卓也, 園 順一, 浅野 令子, 高木 治夫: "視覚障害者のためのタイピング練習機「打ち込み君」の開発," 第 7 回 ITRC 総会・研究会, May 2000.
72. 高城 敏弘, 西本 卓也, 園 順一, 浅野 令子, 高木 治夫: "視覚障害者のためのモニターレス・キーボード練習環境," 電子情報通信学会技術研究報告（福祉情報工学研究会）, WIT99-30, pp.41-46, Mar. 2000.
73. 西本 卓也, 新美 康永: "音声認識の自己目的的な楽しさ," 人工知能学会研究会資料, SIG-SLUD-9804, pp.13-18, Feb. 1999. (1998 年度人工知能学会研究会奨励賞)
74. 幸 英浩, 西本 卓也, 新美 康永: "非同期型音声メッセージシステムの提案," 情報処理学会 (音声言語情報処理), SLP-22-13, Jul. 1998.
75. 西本 卓也, 新美 康永: "非同期型音声メッセージシステムの設計," 電子情報通信学会マルチメディア・仮想環境基礎研究会（映像情報メディア学会／計測自動制御学会と共催), MVE-97-98, pp.39-46, Feb. 1998.
76. 西本 卓也, 新美 康永: "ネットサーフィンにおける音声入力語彙とその役割," 情報処理学会 (音声言語情報処理), SIG-SLP 20-13, pp.75-80, Feb. 1998.

77. 西本 卓也, 小林 豊, 新美 康永: “ネットサーフィンにおける音声コマンド候補の生成について,” 信学技報, SP97-59, pp.13-18, Nov. 1997.
78. 西本 卓也, 志田 修利, 小林 隆, 春山 智, 小林 哲則: “音声利用効果の経時変化と顔向認識による不要発話の棄却 –マルチモーダル作図システム S-tgif における評価–,” 信学技報, SP96-32, pp.89-96, Jun. 1996.
79. 西本 卓也, 志田 修利, 山岡 紳介, 小林 哲則, 白井 克彦: “音声・マウス・キーボードを用いたマルチモーダル作図システム,” 電子情報通信学会技術研究報告, HC93-83, pp.25-30, Mar. 1994.
80. 志田 修利, 西本 卓也, 小林 哲則, 白井 克彦: “音声・マウス・キーボードを併用した作図システム S-tgif とその評価,” 電子情報通信学会技術研究報告, SP94-29, pp.49-56, Jun. 1994.

A.7 講演 (全国大会・シンポジウム)

1. Jun Wu, Yu Kitano, Stanislaw Andrzej Raczynski, Shigeki Miyabe, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama: “Statistical Harmonic Model with Relaxed Partial Envelope Constraint for Multiple Pitch Estimation,” 日本音響学会秋季研究発表会講演集, pp.909-910, Sep. 2010.
2. 秦 思為, 深山 覚, 西本 卓也, 嵯峨山 茂樹: “歌詞の韻律を考慮した中国語声調学習支援のための自動作曲システムの試作,” 日本音響学会秋季研究発表会講演集, p.2, Sep. 2010.
3. 盧 迪, 久保 伸太郎, 深山 覚, 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “マルチモーダル入力と強化学習による擬人化エージェントの対話制御の検討,” 2010 年度人工知能学会全国大会（第 24 回）論文集, 1J1-OS13-5, pp.1-4, Jun. 2010.
4. 西本 卓也: “学会・研究会の情報保障におけるソーシャルネットワークの役割,” 人工知能学会全国大会（第 24 回）, 1D3-1, Nagasaki, Jun. 2010.
5. 西本 卓也, 松村 瞳, 渡辺 隆行: “音声 CAPTCHA における了解度と心的負荷の検討,” 日本音響学会 2010 年春季研究発表会 講演論文集, 3-4-3, pp.1487-1490, Tokyo, Mar. 2010.
6. 深山 覚, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “非和声音規則を語彙とする確率的旋律モデル,” 日本音響学会春季研究発表会講演集, 2-8-15, pp.981-982, Mar. 2010.
7. 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “力学モデル駆動による音声対話エージェントの動作生成,” HAI シンポジウム 2009, 2C-1, Tokyo, Dec. 2009.
8. 盧 迪, 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “擬人化エージェントとの円滑なマルチモーダル対話のための強化学習を用いた割り込み制御の検討,” HAI シンポジウム 2009, 2D-1,

Tokyo, Dec. 2009.

9. 西本 卓也, 瀬川 亜希, 渡辺 隆行: “超早口音声の聴取訓練における単語親密度とメンタルワークロードの検討,” 日本音響学会 2008 年秋季研究発表会講演論文集, 2-7-6, pp.1583-1586, Sep. 2008.
10. 宮本 賢一, 亀岡 弘和, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “Source-Filter モデルを含めた調波構造・時間包絡・音色の統合的クラスタリング（HTTC）による楽音分析,” 日本音響学会春季研究発表会講演集, pp.907-908, Mar., 2008.
11. 米林 裕一郎, 中妻 啓, 西本 卓也, 嵯峨山 茂樹: “Orpheus: 歌詞の韻律を利用した Web ベース自動作曲システム,” 情報処理学会インタラクション 2008 論文集, IPSJ Symposium Series Vol.2008, No.4, pp.27-28, Mar. 2008.
12. 内山 裕貴, 宮本 賢一, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: “調波音を強調したクロマに基づく音楽音響信号からの自動和音認識,” 日本音響学会春季研究発表会講演集, pp.901-902, Mar. 2008.
13. Masahiro Araki, Tsuneo Nitta, Kouichi Katsurada, Takuya Nishimoto, Tetsuo Amakasu Shinnichi Kawamoto: “Proposal of a Hierarchical Architecture for Multimodal Interactive Systems,” Workshop on W3C’s Multimodal Architecture and Interactives, Nov. 2007.
14. 西本 卓也，狩谷 幸香，渡辺 隆行: “早口音声聴取における単語親密度とメンタルワークロードの検討,” 日本音響学会 2007 年秋季研究発表会講演論文集, 2-Q-23, pp.593-594 (in CD-ROM), Sep. 2007.
15. 西本 卓也，渡辺 隆行: “早口音声の聴取における学習効果と加齢の影響,” Human Interface 2007 論文集, 3134, pp.937-942, Sep. 2007.
16. 西本 卓也，酒向 慎司，嵯峨山 茂樹，小田 浩一，渡辺 隆行: “視覚障害者用早口合成音声に対する慣れの効果,” 日本音響学会 2007 年春季研究発表会講演論文集, 2-8-13, pp.357-360 (in CD-ROM), Mar 2007.
17. 西本 卓也: “マルチモーダル対話システムのためのアーキテクチャ階層化,” FIT2006 イベント企画「音声・マルチモーダル対話記述とその標準化」予稿集, Sep. 2006.
18. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: “Maximum Likelihood Based General Joint Adaptation to Noise and Long Reverberation,” in Proc. ASJ Autumn Conf., 1-11-23, pp.67-68 (in CD-ROM), Mar. 2006.
19. 陳 映融, 米田 隆一, 西本 卓也, 嵯峨山 茂樹: “マルコフ確率場モデルに基づく統計的な音楽情報の解析,” 日本音響学会 2006 年春季研究発表会 講演論文集, 2-2-10, pp.709-710 (in CD-ROM), Mar. 2006.
20. 米田 隆一, 西本 卓也, 嵯峨山 茂樹: “最大マージンのアプローチによる統計的な音楽情報の解析,” 日本音響学会 2006 年春季研究発表会 講演論文集, 2-2-11, pp.711-712 (in

CD-ROM), Mar. 2006.

21. 武田 晴登, 西本 卓也, 嵯峨山 茂樹: “HMM を用いたリズムとテンポの反復推定による多声 MIDI 演奏のリズム認識,” 日本音響学会 2006 年春季研究発表会 講演論文集, 3-2-3, pp.721-722 (in CD-ROM), Mar. 2006.
22. 武田 晴登, 西本 卓也, 嵯峨山 茂樹: “和音の発音順序交替を許容した動的計画法による多声音楽音 MIDI 演奏の楽譜追跡,” 日本音響学会 2006 年春季研究発表会 講演論文集, 3-2-4, pp.723-724 (in CD-ROM), Mar. 2006.
23. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: “Maximum Likelihood Based General Joint Adaptation to Noise and Long Reverberation,” in Proc. ASJ Autumn Conf., 1-11-23, pp.67-68 (in CD-ROM), Mar. 2006.
24. 西本 卓也，酒向 慎司，嵯峨山 茂樹，大島 一恵，小田 浩一，渡辺 隆行: “早口合成音声に対する聴取者の慣れの効果の検討,” 日本音響学会 2005 年秋季研究発表会講演論文集, 3-6-14, pp.355-356, Sep. 2005.
25. 酒向 慎司，西本 卓也，嵯峨山 茂樹: “HMM 音声合成手法による早口音声合成の検討,” 日本音響学会 2005 年秋季研究発表会, 3-6-15, pp.357-358, Sep. 2005.
26. 會田 卓也, 西本 卓也, 中沢 正幸, 大川 茂樹, 嵯峨山 茂樹: “頭部モーションセンサと音声を用いた対話インタフェースの提案,” ヒューマンインタフェースシンポジウム 2005 講演論文集, 2531, pp.601-604, Sep. 2005.
27. 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “視線制御モデルによる擬人化音声対話エージェントの制御,” 2005 年度人工知能学会全国大会 (第 19 回) 論文集, 3B2-07, Jun 2005.
28. 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “確定的アニーリング EM アルゴリズムを用いた調波時間構造化クラスタリングによる音楽信号分析,” 日本音響学会 2005 年秋季研究発表会講演論文集, 3-10-16, in CD-ROM, Sep. 2005.
29. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: “Maximum Likelihood Based Compensation of HMM Parameters for Channel Distortion,” in Proc. ASJ, in CD-ROM, Sep. 2005.
30. 齊藤 翔一郎, 西本 卓也, 嵯峨山 茂樹: “Specmurt 分析と HMM を用いた音楽音響信号の調認識,” 日本音響学会 2005 年秋期研究発表会講演論文集, 3-10-10, in CD-ROM, Sep. 2005.
31. 米田 隆一, 西本 卓也, 嵯峨山 茂樹: “最大エントロピーモデルに基づく統計的な音楽情報の解析,” FIT2005 第 4 回情報科学技術フォーラム講演論文集, pp.267-268, Sep. 2005.
32. 山本 遼, 山本 隼, 西本 卓也, 嵯峨山 茂樹: “ストロークをベースとした確率文脈自由文法による手書き数式の認識,” FIT2005 第 4 回情報科学技術フォーラム講演論文集, pp.43-44, Sep. 2005.
33. 槻 武也, 西本 卓也, 嵯峨山 茂樹: “残響音声の認識のための音響モデル変換,” 日本音響学

会 2005 年春季研究発表会講演論文集, 3-5-6, pp. 87-88, Mar. 2005.

34. 井上 和士, 西本 卓也, 嵯峨山 茂樹: “複素スペクトル円心 (CSCC) 法と雑音音源方向推定を組み合わせた雑音抑圧,” 日本音響学会 2005 年春季研究発表会講演論文集, 1-6-22, pp. 451-452, Mar. 2005.
35. 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “ガウス基底 2 次元分布モデルを用いた時空間クラスタリングによる音響ストリームの分離,” 日本音響学会 2005 年春季研究発表会講演論文集, 3-7-19, pp. 601-602, Mar. 2005.
36. 中潟 昌平, 西本 卓也, 嵯峨山 茂樹: “動的計画法に基づく自動対位法,” 日本音響学会 2005 年春季研究発表会講演論文集, 3-7-12, pp. 587-588, Mar. 2005.
37. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: “Model Convolution by State Splitting of HMM for Robust Speech Recognition in Presence of Convolutional Noise,” 日本音響学会 2005 年春季研究発表会講演論文集, 3-5-5, pp. 85-86, Mar. 2005.
38. 西本 卓也, 荒木 雅弘, 伊藤 克亘, 宇津呂 武仁, 甲斐 充彦, 河口 信夫, 河原 達也, 桂田 浩一, 小林 隆 夫, 嵯峨山 茂樹, 下平 博, 伝康晴, 徳田 恵一, 中村 哲, 新田 恒雄, 坂野 秀樹, 広瀬 啓吉, 峯松 信明, 三村 正人, 森島 繁生, 山下 洋一, 山田 篤, 四倉 達夫, 李 晃伸: “Galatea: 音声対話擬人化エージェント開発キット,” 第 12 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS), pp.125-126, Dec 2004.
39. 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “視線制御モデルを用いた擬人化音声対話エージェントの提案,” 2004 年度人工知能学会全国大会 (第 18 回) 論文集, 2E1-08, Jun 2004.
40. 西本 卓也, 中沢 正幸, 嵯峨山 茂樹: “音声対話における擬人化エージェントの身体動作表現の利用,” 2004 年度人工知能学会全国大会 (第 18 回) 論文集, 2C2-01, Jun 2004.
41. 西本 卓也, 塩谷 真, 高橋 寿平, 醍醐 英治: “テレマティックス音声対話ガイドライン作成に向けた検討,” シンポジウム「ケータイ・カーナビの利用性と人間工学」研究論文集, pp.125-128, Mar 2004.
42. 西本 卓也, 荒木 雅弘, 伊藤 克亘, 宇津呂 武仁, 甲斐 充彦, 河口 信夫, 河原 達也, 桂田 浩一, 小林 隆夫, 嵯峨山 茂樹, 下平 博, 伝 康晴, 徳田 恵一, 中村 哲, 新田 恒雄, 坂野 秀樹, 広瀬 啓吉, 峯松 信明, 三村 正人, 森島 繁生, 山下 洋一, 山田 篤, 四倉 達夫, 李 晃伸: “Galatea: 音声対話擬人化エージェント開発キット,” インタラクション 2004 論文集, pp.27-28, Mar 2004.
43. 嵯峨山 茂樹, 武田 晴登, 亀岡 弘和, 西本 卓也: “音楽情報処理と音声認識,” 日本音響学会 2004 年春季研究発表会講演論文集, 2-6-9, pp.785-788, Sep. 2004.
44. 井上 和士, 鎌本 優, 岡島 崇, 西本 卓也, 嵯峨山 茂樹: “複素スペクトル円心 (CSCC) 法によるマイクロホンアレーを用いた雑音除去,” 日本音響学会 2004 年春季研究発表会講演論文集, 2-3-7, pp.619-620, Sep. 2004.
45. Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama: “Noise-Driven Temporal

Trajectory Filtering of Spectral Parameters for Robust Speech Recognition,” 日本音響学会 2004 年春季研究発表会講演論文集, 1-1-14, pp.27-28, Sep. 2004.

46. 亀岡 弘和, 齊藤 翔一郎, 西本 卓也, 嵯峨山 茂樹: “Specmurt 法による音楽信号の音高可視化における共通調波構造パターンの自動決定,” 日本音響学会 2004 年春季研究発表会講演論文集, 2-6-15, pp. 803-804, Sep. 2004.
47. 鎌本 優, 守谷 健弘, 西本 卓也, 嵯峨山 茂樹: “チャンネル間相関を用いた多チャンネル信号の可逆圧縮符号化,” FIT2004 第 3 回情報科学技術フォーラム (平成 16 年 9 月 7 日（火）〜9 日（木）) 一般講演論文集第 4 分冊, pp.123-124, Sep. 2004.
48. 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “Harmonic-GMM の最尤推定と情報量規準に基づく多重音の基本周波数検出および調波構造分離,” 情報処理学会第 66 回全国大会予稿集, 3ZA-1, pp.2-427-2-428, Mar. 2004.
49. 菅原 啓太, 米田 隆一, 西本 卓也, 嵯峨山 茂樹: “HMM と音符連鎖確率を用いた旋律への自動和声付け,” 日本音響学会 2004 年春季研究発表会講演論文集, 1-9-2, pp. 665-666, Mar. 2004.
50. 中潟 昌平, 西本 卓也, 嵯峨山 茂樹: “DP 法に基づく対位法における複数の対旋律候補の自動生成,” 日本音響学会 2004 年春季研究発表会講演論文集, 1-9-3, pp. 667-668, Mar. 2004.
51. 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “拘束つき混合正規分布モデルの MAP 推定による同時発話音声の F0 追跡,” 日本音響学会 2004 年春季研究発表会講演論文集, 2-7-6, pp.275-276, Mar. 2004.
52. 織田 誠也, 亀岡 弘和, 西本 卓也, 嵯峨山 茂樹: “MAP 推定を用いた Harmonic Clustering による多重音中の非調和性音源の調波構造検出,” 日本音響学会 2004 年春季研究発表会講演論文集, 2-9-3, pp.689-690, Mar. 2004.
53. 高橋 佳吾, 西本 卓也, 嵯峨山 茂樹: “対数周波数スペクトルの逆畳み込みによる基本周波数解析 (Specmurt 法),” 日本音響学会 2004 年春季研究発表会講演論文集, 2-9-4, pp.691-692, Mar. 2004.
54. Chandra Kant Raut, 山本 仁, 西本 卓也, 嵯峨山 茂樹: “Polynomial-Approximation-Based Model Combination for Noisy Speech Recognition,” 日本音響学会 2004 年春季研究発表会講演論文集, 2-11-11, pp.121-122, Mar. 2004.
55. 岡嶋 崇, 鎌本 優, 西本 卓也, 嵯峨山 茂樹: “マイクロフォンアレイ入力の周波数領域での幾何学的処理による雑音中の音声認識,” 日本音響学会 2004 年春季研究発表会講演論文集, 3-10-1, pp.583-584, Mar. 2004.
56. 鎌本 優, 堀内 俊治, 水町 光徳, 中村 哲, 西本 卓也, 嵯峨山 茂樹: “最短 Golomb 定規間隔配置 Delay-and-Sum 型マイクロフォンアレイを用いた雑音環境下の音声認識,” 日本音響学会 2004 年春季研究発表会講演論文集, 3-10-6, pp.593-594, Mar. 2004.

57. 新田 恒雄, 西本 卓也, 川本 真一, 下平 博, 森島 繁生, 四倉 達夫, 山下 洋一, 小林 隆夫, 徳田 恵一, 広瀬 啓吉, 峯松 信明, 山田 篤, 伝 康晴, 宇津呂 武仁, 伊藤 克亘, 甲斐 充彦, 李 晃伸, 中村 哲, 嵯峨山 茂樹: “Galatea: 音声対話擬人化エージェント開発キット,” 第 8 回日本顔学会大会（フォーラム顔学 2003）, Vol.3, No.1, p.189, Sep 2003.
58. 西本 卓也, 荒木 雅弘, 小林 哲則: “ディストラクション評価に基づく車内音声対話コンテンツの比較,” FIT2003 イベント企画, 「車載情報システムにおける音声インタフェース」予稿集, Sep 2003.
59. 西本 卓也, 北脇 裕康, 高木 治夫: “非同期型音声会議システム VoiceCafe,” 情報技術レターズ (FIT2003 講演論文集), LK-005, pp.273-274, Sep 2003.
60. 西本 卓也, 嵯峨山 茂樹: “擬人化エージェント Galatea のための VoiceXML 処理系,” 第 17 回人工知能学会全国大会, 2C2-04, Jun. 2003.
61. 西本 卓也, 高山 元希, 荒木 雅弘: “音声インタフェースのための認知的負荷測定法の検討,” 日本音響学会講演論文集, 2-4-13, pp.83-84, Mar. 2003.
62. 高山 元希, 西本 卓也, 荒木 雅弘: “二重課題法による音声対話システムにおける認知負荷の測定, 情報処理学会全国大会講演論文集, Vol.65, No.2. pp.2.395-2.396, Mar. 2003.
63. 西本 卓也, 荒木 雅弘, 新美 康永: “擬人化音声対話エージェントのためのタスク管理機能,” 日本音響学会 2002 年春季研究発表会, 1-5-15, pp.29-30, Mar. 2002.
64. 西本 卓也, 高木 治夫: “視覚障害者向けタイピング練習ソフト「ウチコミくん」,” 情報処理学会インタラクション 2002, pp.191-192, Mar. 2002.
65. 住吉 悠希, 西本 卓也, 荒木 雅弘, 新美 康永: “インターネットラジオ番組制作支援ツール,” 情報処理学会インタラクション 2002, pp.153-154, Mar. 2002.
66. 川原 毅彦, 木田 智史, 西本 卓也, 高木 治夫: “非同期型音声会議におけるディクテーション機能,” 日本音響学会 2001 年春季研究発表会, 1-3-20, Mar. 2001.
67. 西本 卓也, 幸 英浩, 川原 毅彦, 荒木 雅弘, 新美 康永: “非同期型バーチャル会議システム AVM,” 電子情報通信学会総合大会, SD-4-9, Mar. 2000.
68. 西本 卓也, 新美康永: “音声認識の自己目的的な楽しさ,” 日本音響学会平成 11 年度春季講演論文集, 3-Q-31, pp.189-190, Mar. 1999.
69. 西本 卓也, 藤澤 正樹, 新美 康永: “音声ウェブブラウザ VOXplorer の評価,” 日本音響学会平成 10 年度秋季講演論文集, 1-R-26, pp.169-170, Sep. 1998.
70. 西本 卓也, 新美 康永: “ネットサーフィン支援のための音声対話システム,” 第 12 回人工知能学会全国大会, S7-03, pp.141-142, Jun. 1998.
71. 西本 卓也, 新美 康永: “ネットサーフィンにおける音声入力語彙とその役割,” 日本音響学会春季講演論文集, 2-Q-23, pp.165-166, Mar. 1998.
72. 西本 卓也, 小林 豊, 新美 康永: “WWW 上のデータベース検索のための汎用音声インタフェース,” 日本音響学会講演論文集, 2-Q-20, pp.179-180, Mar. 1997.

73. 会田 清, 西本 卓也, 李 圭建, 白井 克彦: “音声認識技術を利用した日本語発音学習システム,” 1996 年電子情報通信学会総合大会, D-692, Mar. 1996.
74. 西本 卓也, 小林 隆, 小林 哲則, 白井 克彦: “マルチモーダル作図システムの音声認識部における非コマンド発話のリジェクション,” 日本音響学会講演論文集, 3-P-28, pp.217-218, Mar. 1995.
75. 西本 卓也, 志田 修利, 山岡 紳介, 小林 哲則, 白井 克彦: “音声・マウス・キーボードを用いたマルチモーダル作図環境,” 日本音響学会講演論文集, 1-7-21, pp.41-42, Mar. 1994.

A.8 著書 (共著・寄稿)

1. Takuya Nishimoto, Shigeki Sagayama: “Galatea: Open-source software for developing anthropomorphic spoken dialog agents,” Computer Processing of Asian Spoken Languages, Shuichi Itahashi, Chiu-yu Tseng, Eds. Americas Group Publications, U.S. Apr. 2010.
2. Nobutaka Ono, Kenichi Miyamoto, Hirokazu Kameoka, Jonathan Le Roux, Yuuki Uchiyama, Emiru Tsunoo, Takuya Nishimoto, Shigeki Sagayama: “Harmonic and Percussive Sound Separation and its Application to MIR-related Tasks,” Advances in Music Information Retrieval, ser. Studies in Computational Intelligence, Z. W. Ras and A. Wieczorkowska, Eds. Springer, 274, pp.213-236, Feb., 2010.
3. 西本 卓也: “インタフェースシステムの導入原則に関する一考察,” 白井 克彦 監修: 情報システムとヒューマンインターフェース, 早稲田大学出版部, 2010.
4. 西本 卓也: “音声聴取におけるメンタルワークロードの測定,” 情報福祉の基礎知識—障害者・高齢者が使いやすいインタフェース, 情報福祉の基礎研究会 編著, ジアース教育新社, Apr. 2008.
5. 佐藤 知正 (著), 東京大学 21 世紀 COE 実世界情報プロジェクト (監修): 人と共存するコンピュータ・ロボット学—実世界情報システム, オーム社, Dec. 2004.（一部の執筆を担当）
6. Shin-ichi Kawamoto, Takuya Nishimoto, Shigeki Sagayama, et al.: “Galatea: Open-source Software for Developing Anthropomorphic Spoken Dialog Agents,” Life-Like Characters – Tools, Affective Functions and Applications, H. Prendinger, M. Ishizuka (Eds.), Springer, 2003.
7. 鯉江 英隆, 西本 卓也, 馬場 肇: バージョン管理システム (CVS) の導入と活用, ソフトバンクパブリッシング, Dec. 2000.

参考文献

[1] 海保 博之, 原田 悦子, 黒須 正明: “認知的インタフェース,” 新曜社, 1991.

[2] Rasmussen (海保 他 訳): インタフェースの認知工学, 啓学出版, 1986.

[3] 佐伯 胖:“機械と人間の情報処理 認知工学序説,” 竹内 啓 編, 意味と情報, 東京大学出版会, 1988.

[4] C. Rutkowski: An introduction to the human applications standard computer interface, Part I. : Theory and principles, BYTE, 7 (11), 291-310, 1982.

[5] S. K. Card, I. P. Moran, A. Newell: The Psychology of Human-Computer Interaction, Lawrence Erlbaum Associates, 1983.

[6] D. A. Norman: Cognitive Engineering, In D. A. Norman and S. W. Draper (Eds.), User Centered System Design, Lawrence Erlbaum Associates, 1986.

[7] Paul M. Fitts (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, volume 47, number 6, June 1954, pp. 381–391. (Reprinted in Journal of Experimental Psychology: General, 121(3):262–269, 1992).

[8] アラン・クーパー, 山形 浩生 (訳) : コンピュータは，むずかしすぎて使えない！, 翔泳社, Feb. 2000.

[9] D.J. Cochran, M.W.Riley, L.A.Stewart: “An evaluation of the strengths, weaknesses and uses of voice input devices,” Proc. Human Factors Society – 24th Annual Meeting, 1980.

[10] J.M. Nye: “Human factors analysis of speech recognition systems,” Speech Technology, I, pp.50–57, 1982.

[11] C. Schmandt, M. S. Ackerman, D. Hindus: “Augmenting a window system with speech input,” IEEE Computer, 23, 8, pp.50–56, Aug. 1990.

[12] G. L. Martin: “The utility of speech input in user-computer interface,” Int. J. Man-Machine Studies, 30, pp.355–375, 1989.

[13] D. A. Norman（野島久雄訳）: “誰のためのデザイン？ – 認知科学者のデザイン原論,”

新曜社, 1990.

[14] B. Shneiderman, “Designing the User Interface,” Second Edition, Addison-Wesley Publishing Company, 1992.

[15] 海保 博之, 加藤 隆: “人に優しいコンピュータ画面設計,” 日経 BP 社, 1993.

[16] 小林 哲則, 竹内 陽児, 白井 克彦: “音声・マウス・キーボードによる統合的入力環境,” 信学技報, HC92–68. Mar 1993.

[17] 西本 卓也, 小林 隆, 小林 哲則, 白井 克彦: “マルチモーダル作図システムの音声認識部における非コマンド発話のリジェクション,” 日本音響学会春季研究発表会講演論文集, pp.217–218, 3–P–28, Mar 1995.

[18] William Chia-Wei Cheng: “Tgif’s WWW Home Page,” http://bourbon.cs.usla.edu/tgif/.

[19] 志田 修利, 西本 卓也, 小林 哲則, 白井 克彦: “音声・マウス・キーボードを併用した作図システム S-tgif とその評価,” 信学技報, SP94–29, Jun 1994.

[20] M. J. Hirayama, T. Sugahara, Z. Peng, J. Yamazaki: “Interactive listening to structured speech content on the Internet,” Proceedings of ICSLP’98, pp.1627–1630, Dec. 1998.

[21] 岡田 美智男: 口ごもるコンピュータ, 共立出版, 東京, 1995.

[22] 西本 卓也, 新美 康永: “非同期音声メッセージシステムの設計,” 信学技報, MVE97-98, pp.39–46, Feb. 1998.

[23] 榎本 美香, 土屋 俊: “オーバーラップ発話の評定方法とその基礎統計〜日本語地図課題対話を通して〜,” 情処研報, 99-SLP-29-25, pp.145–150. Dec. 1999.

[24] 川口 由紀子, 土屋 俊: “ターン交替規則の破綻例の会話の含みによる説明の試み〜日本語地図課題対話を通して〜,” 情処研報, 99-SLP-29-26, pp.151–156, Dec. 1999.

[25] 西 宏之，五味 和洋, 小島 順治: “音声対話における確率的発声終了検出法,” 信学論 D, Vol.J70-D, No.11, Nov. 1987.

[26] 向後 千春，山西 潤一: “あいづち留守番電話の試作,” 日本認知科学会第 8 回大会発表論文集, pp.72–73, Jul. 1991.

[27] N. Ward: “Using prosodic clues to decide when to produce back-channel utterances,” Proceedings of ICSLP’96, pp.1728–1731, Oct. 1996.

[28] 菊池 英明, 杉田 洋介, 白井 克彦: “自由会話における時間的制約の影響の分析,” 人工知能学会研究会資料, SIG-SLUD-9702-5, pp.31–36, Oct. 1997.

[29] 堀内 靖雄, 小磯 花絵, 土屋 俊, 市川 薫: “自発的音声対話における話者交替の制御に関する発話末の統語的・韻律的特徴,” 情処研報, 96-SLP-10-9, pp.45–50, Feb. 1996.

[30] 岡登 洋平, 加藤 佳司, 山本 幹雄, 板橋 秀一: “韻律パターンの認識を用いた相槌挿入とその評価,” 情処研報, 96-SLP-10-7, pp.33–38, 1996.

[31] 岡登 洋平, 加藤 佳司, 山本 幹雄, 板橋 秀一: “相槌を打つ音声対話システムの評価,” 人工知能学会研究会資料, SIG-SLUD-9804-2, pp.7–12, Feb. 1999.

[32] J. Hirasawa, N. Miyazaki, M. Nakano, T. Kawabata: “Implementation of coordinative nodding behavior on spoken dialog systems,” Proceedings of ICSLP’98, pp.2347–2350, Dec. 1998.

[33] T. Nishimoto, H. Yuki, T. Kawahara, Y. Niimi: “An asynchronous virtual meeting system for bi-directional speech dialog,” Proceedings of Eurospeech’99, pp.2471–2474, Sep. 1999.

[34] Bruce Balentine, Devid P. Morgan: “How to build a speech recognition application,” 2nd Ed., EIG Press, 2001.

[35] 大本 浩司, 牛田 博英, 中嶋 宏, 石田 勉: “電話音声で情報提供を行うアプリケーションの UI 設計と評価の実践報告,” ヒューマンインタフェース学会研究報告集, Vol.3, No.4, pp.35-40, 2001.

[36] Byron Reeves, Clifford Nass: “The media equasion, ” CSLI, Cambridge, 1996.（細馬 宏通 訳: “人はなぜコンピューターを人間として扱うか「メディアの等式」の心理学,” 翔泳社, 東京, 2001.）

[37] Michael W. Eysenck, Mark Keane: “Cognitive psychology – A student’s handbook, ” 4th Ed., Psychology Press, East Sussex, 2000.

[38] Nick Lund: “Attention and pattern recognition,” Routledge, Philadelphia, 2001.

[39] 小島 真一, 本郷 武朗, 星野 博之, 内山 祐司: “音声対話の運転への影響評価法の開発,” 自動車技術会学術講演会前刷集, No. 91-99, pp. 17-20, 1999.

[40] 小島 真一, 本郷 武朗, 星野 博之, 内山 祐司: “音声対話の運転への影響評価法の開発,” 情処研報, 1999-MBL-010-010, 1999.

[41] 清水 司, 小島 真一, 脇田 敏裕, 本郷 武朗: “運転中における音声対話システムの評価,” 情処研報, 2000-SLP-32-17, pp.87-92, 2000.

[42] David L. Strayer, William A. Johnston: “Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone, ” Psychological Science, 12, pp. 462-466, 2001.

[43] 内山 祐司, 小島 真一, 本郷 武朗, 脇田 敏裕: “運転状況適応型音声情報提示システム,” シンポジウム「ケータイ・カーナビの利用性と人間工学」, pp.11-15, 名古屋, 2001.

[44] 西本 卓也, 高山 元希, 荒木 雅弘: “音声インタフェースにおける認知的負荷測定法とその評価,” 2002-SLP-45-5, pp.29-34, 2003.

[45] 西本 卓也, 高山 元希, 荒木 雅弘: “ 音声インタフェースにおける認知的負荷測定法の検討, ” 2003 年春季日本音響学会講演論文集, 2-4-13, pp.83-84 2003.

[46] 田中 敏: 実践心理データ解析, 新曜社, 東京, 1996.

[47] 田中 敏: 実践心理データ解析 改訂版, 新曜社, 東京, 2006.

[48] 渡辺 哲也: “スクリーンリーダの速度・ピッチ・性別の設定状況,” 電子情報通信学会論文誌 D-I, Vol. J88-D-I, No.8, pp.1257-1260, Aug. 2005.

[49] 浅川 智恵子, 高木 啓伸, 井野 秀一, 伊福部 達: “視覚障害者への音声提示における最適・最高速度,” ヒューマンインタフェース学会論文誌, Vol.7, No.1, pp.105-111, Feb. 2005.

[50] 西本 卓也, 酒向 慎司, 嵯峨山 茂樹, 小田 浩一, 渡辺 隆行: “早口合成音声の聴取実験によるテキスト音声合成の評価,” 電子情報通信学会技術報告, WIT2005-5, pp.23-28, May 2005.

[51] 西本 卓也, 渡辺 隆行: “早口音声の聴取における学習効果と加齢の影響,” Human Interface 2007 論文集, 3134, pp.937-942, Sep. 2007.

[52] 天野 成昭, 近藤 公久, 坂本 修一, 鈴木 陽一: 親密度別単語了解度試験用音声データセット (FW03), NII 音声資源コンソーシアム, 2006.

[53] 渡辺 俊朗: “単語了解度による規則合成音の評価法に関する検討,” 電子情報通信学会論文誌 (A), Vol.J72-A, no.10, pp.1503-1509, 1989.

[54] 淀川 英司, 中根 一成, 東倉 洋一: 視聴覚の認知科学, 電子情報通信学会, 1998.

[55] 天野 成昭, 近藤 公久, 日本語の語彙特性, 三省堂, 東京, 1999.

[56] 坂本 修一, 鈴木 陽一, 天野 成昭, 小澤 賢司, 近藤 公久, 曽根 敏夫: “親密度と音韻バランスを考慮した単語了解度試験用リストの構築,” 日本音響学会誌, 54 巻, pp. 842–849, 1998.

[57] 佐藤 逸人, 森本 政之, 佐藤 洋: “ “ 聞き取りにくさ ” による音声伝達性能の評価,” 日本音響学会誌, 63 巻, pp.275–280, May 2007.

[58] S. G. Hart, L. E. Staveland: “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload, Amsterdam, North Holland Press, 1998.

[59] 芳賀 繁: メンタルワークロードの理論と測定, 日本出版サービス, (2001).

[60] Christopher D. Wickens: “Processing resources in attention, ” in R. Parasuraman & David Roy Davies (Eds.), Varieties of attention, (pp. 63-102), New York, Academic Press, 1984.

[61] Sandra G. Hart: “NASA-Task Load Index (NASA-TLX); 20 Years Later,” Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 904-908, 2006.

[62] 三宅 晋司, 神代 雅晴: “メンタルワークロードの主観的評価法 — NASA-TLX と SWAT の紹介および簡便法の提案 —,” 人間工学, Vol. 29, pp. 399-408, 1993.

[63] 情報福祉の基礎研究会（代表: 市川 熹）編著: 情報福祉の基礎知識 — 障害者・高齢者が使いやすいインタフェース —, ジアース教育新社, 東京, (2008).

[64] 篠原 一光, 山田 尚子, 神田 幸治, 臼井 伸之介: “日常生活における注意経験と主観的メンタルワークロードの個人差,” 人間工学, Vol. 43, pp. 201–211, 2007.

[65] Diane McGuinness: “Hearing: individual differences in perceiving,” Perception 1 (4), pp.465–473, (1972).

[66] 西本 卓也, 北脇 裕康, 高木 治夫: “非同期型音声会議システム VoiceCafe,” 情報技術レターズ (FIT2003 講演論文集), LK-005, pp.273-274, Sep 2003.

[67] 西本 卓也, 塩谷 真, 高橋 寿平, 醍醐 英治: “テレマティックス音声対話ガイドライン作成に向けた検討,” シンポジウム「ケータイ・カーナビの利用性と人間工学」研究論文集, pp.125-128, Mar 2004.

[68] Takuya Nishimoto, Makoto Shioya, Juhei Takahashi, Hideharu Daigo: “A study of dialogue management principles corresponding to the driver’s workload,” Biennial Workshop on Digital Signal Processing for In-Vehicle and mobile systems, Sesimbra, Portugal, Sep 2005.

[69] Susan Weinschenk, Dean T. Barker: Designing Effective Speech Interfaces, Wieley, 2000.

[70] Bruce Balentine, Leslie Degler: It’s Better to Be a Good Machine Than a Bad Person: Speech Recognition and Other Exotic User Interfaces in the Twilight of the Jetsonian Age, ICMI Press, 2007.

[71] ヤコブ・ニールセン, 篠原 稔和, グエル (訳): ウェブ・ユーザビリティ 顧客を逃がさないサイトづくりの秘訣, エムディエヌコーポレーション, 2000.

[72] Donald A. Norman, 岡本 明, 安村 通晃, 伊賀 聡一郎, 上野 晶子 (翻訳) : エモーショナル・デザイン—-微笑を誘うモノたちのために, 新曜社, Oct. 2004.

[73] 西本 卓也, 高木 治夫: “視覚障害者向けタイピング練習ソフト「ウチコミくん」,” 情報処理学会インタラクション 2002, pp.191-192, Mar. 2002.

[74] 山本 吉伸，松井 孝雄，開 一夫，梅田 聡，安西 祐一郎: “計算システムとのインタラクション，楽しさを促進する要因に関する一考察,” 日本認知科学会，認知科学，1 巻 1 号, pp.107-120，1994.

[75] Mihaly Csikszentmihalyi: Beyond Boredom and Anxiety，Jossey Bass Publishers, 1975. M. チクセントミハイ，今村 浩明 訳: 楽しむということ，思索社, 1991.

[76] Mihaly Csikszentmihalyi: Flow，Harper & Row Publishers, 1990. M. チクセントミハイ，今村 浩明 訳: フロー体験—喜びの現象学，世界思想社, 1996.

[77] 佐藤 郁哉: 暴走族のエスノグラフィー– モードの叛乱と文化の呪縛 –, 新曜社, 1984.

[78] 羽尻 公一郎: “チャットにおけるフロー体験,” 社会言語科学 Vol.4 No.1 pp.17-23, 2001.

[79] Jane Webster，Linda K. Trevino，Lisa Ryan: “The Dimensionavity and Correlates of Flow in Human-Computer Interactions,” Computers in Human Behavior Vol.a,

pp.411-426, 1993.

[80] Susan Wiedenbeck, Sid Davis: “Intrinsic Motivation, Ease of Use and Usefulness Perceptions as Mediators in Computer Lerning,” Proceedings of HCI International 2001, Vol.1, pp.1553-1557, 2001.

[81] 西本 卓也, 新美 康永: “音声認識の自己目的的な楽しさ,” 人工知能学会研究会, SIG-SLUD-9804, pp.13-18, 1999.

[82] Donna L. Hoffman and Thomas P. Novak: “Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations,” Journal of Marketing, Vol. 60, pp. 50–68, 1996.

[83] 赤木 昭夫: インターネット・ビジネス論, 岩波書店, 1999.

[84] 西本 卓也, 櫻井 晴章, 荒木 雅弘, 新美 康永: “ディクテーション作業における楽しさの分析,” 人工知能学会 SIG-SLUD-A103-01(3/8), pp.01-06, Mar. 2002.

[85] 西本 卓也, 新美 康永: “ネットサーフィン支援のための音声対話システム,” 第 12 回人工知能学会全国大会, S7-03, pp.141-142, Jun 1998.

[86] 西本 卓也, 藤澤 正樹, 新美 康永: “音声ウェブブラウザ VOXplorer の評価,” 日本音響学会平成 10 年度秋季講演論文集 1-R-26, pp.169-170, Sep 1998.

[87] 高山 元希, 西本 卓也, 荒木 雅弘, 新美 康永: “電話音声応答システムにおける効果音の役割,” 電子情報通信学会技術研究報告, SP 2001-132, pp.55-62, Jan. 2002.

[88] 西本 卓也, 住吉 悠希, 荒木 雅弘, 新美 康永: “視覚障害者のための電子メール環境における操作性の検討,” ヒューマンインタフェース学会研究報告集, Vol.2 No.5, pp.33-38, Dec. 2000.

[89] 西本 卓也, 嵯峨山 茂樹, 藤原 扶美, 下永 知子, 渡辺 隆行: “対面朗読者と視覚障害者の対話の分析とその応用,” 情報処理学会研究報告, 2007-SLP-11, pp.55-60, Feb. 2007.

[90] Takuya Nishimoto, Takayuki Watanabe: “The Comparison Between the Deletion-Based Methods and the Mixing-Based Methods for Audio CAPTCHA Systems,” Proceedings of Interspeech 2010, pp.266-269, 2010-09-27, Makuhari, Chiba, Japan, 2010.

[91] 會田 卓也, 西本 卓也, 大川 茂樹, 嵯峨山 茂樹: “頭部モーションセンサと音声を用いた対話インタフェースの検討,” 電子情報通信学会技術報告, WIT2006-87, pp.85-90, Jan. 2007.

[92] Bolt, R. A. : “Put-that-there: Voice and gesture at the graphics interface, ” ACM Computer Graphics, Vol. 14, No. 3, pp. 262-270, 1980.

[93] 竹林 洋一: “音声自由対話システム TOSBURG II －ユーザ中心のマルチモーダルインタフェースの実現に向けて－,” 電子情報通信学会論文誌, D-II, Vol. J77-D-II, No. 8, pp.1417-1428, 1994.

[94] 神尾 広幸, 松浦 博, 正井 康之, 新田 恒雄: “マルチモーダル対話システム MultiksDial,” 電子情報通信学会論文誌, D-II, Vol. J77-D-II, No. 8, pp. 1429-1437, 1994.

[95] Miyahara, K., Inoue, H., Tsunesada, Y., Sugimoto, M. : “Intuitive Manipulation Techniques for Projected Displays of Mobile Devices,” In Proceedings of ACM CHI2005 Extended Abstract, Portland, Oregon, pp.1881-1884 (2005).

[96] 會田 卓也, 西本 卓也, 中沢 正幸, 大川 茂樹, 嵯峨山 茂樹: “頭部モーションセンサと音声を用いた対話インタフェースの提案,” ヒューマンインタフェースシンポジウム 2005 講演論文集, 2531, pp.601-604, Sep. 2005.

[97] 西本 卓也, 西村 雅史, 赤堀 一郎, 石川 泰, 磯谷 亮輔, 伊藤 克亘, 大淵 康成, 金澤 博史, 國枝 伸行, 外山 聡一, 新田 恒雄: “音声認識応用に関する学会試行標準,” 情報処理学会研究報告, 2005-SLP-55, pp.47-52, Feb. 2005.

[98] 西本 卓也, 岩田 英三郎, 櫻井 実, 廣瀬 治人: “探索的検索のための音声入力インタフェースの検討,” 情報処理学会研究報告 2008-HCI-127(2), pp.9-14, Jan. 2008.

[99] Gary Marchionini: “Exploratory search: from finding to understanding,” Communications of the ACM, Vol.49, Issue 4, pp.41-46, 2006.

[100] Jef Raskin, 村上 雅章 (訳): ヒューメイン・インタフェース, ピアソン・エデュケーション, 2001.

[101] 嵯峨山 茂樹, 西本 卓也, 中沢 正幸: “擬人化音声対話エージェント,” 情報処理学会誌, Vol.45, No.10, pp.1044-1049, Oct. 2004.

[102] 安藤 彰男: リアルタイム音声認識, 電子情報通信学会, 2003.

[103] M. Nakano, K. Dohsaka, N. Miyazaki, J. Hirasawa, M. Tamoto, M. Kawamori, A. Sugiyama, T. Kawabata：“Handling rich turn-taking in spoken dialogue systems,” Proc. of Eurospeech-99, pp. 1167–1170, 1999.

[104] 藤江 真也, 福島 健太, 三宅 梨帆, 小林 哲則: “相槌生成／認識機能を持つ音声対話システム,” 人工知能学会 SIG-SLUD, Vol.45, pp.41-46, Nov. 2005.

[105] 西村 良太, 北岡 教英, 中川 聖一: “応答タイミングを考慮した雑談音声対話システム,” 人工知能学会 SIG-SLUD 46, 21-26, Mar. 2006.

[106] G. Skantze, A. Hajalmarsson: “Towards Incremental Speech Generation in Dialogue Systems,” In Proceedings of SIGdial, Tokyo, Sep. 2010.

[107] Brenda K. Laurel: ”Interface as Mimesis,” User Centered System Design, edited by Donald A. Norman, Stephen W. Draper, pp.67-86, Lawrence Erlbaum Associates, New Jersey, 1986.

[108] 盧 迪, 中沢 正幸, 西本 卓也, 嵯峨山 茂樹: “擬人化エージェントとの円滑なマルチモーダル対話のための強化学習を用いた割り込み制御の検討,” HAI シンポジウム 2009, 2D-1, Tokyo, Dec. 2009.

[109] 盧 迪, 久保 伸太郎, 深山 覚, 中沢 正幸, 西本 卓也, 嵯峨山 茂樹, “マルチモーダル入力と強化学習による擬人化エージェントの対話制御の検討,” 2010 年度人工知能学会全国大会（第 24 回）論文集, 1J1-OS13-5, pp.1-4, Jun. 2010.

[110] Richard S. Sutton, Andrew G. Barto（三上 貞芳, 皆川 雅章 訳）: 強化学習, 森北出版, 2000.

[111] 盧 迪, 深山 覚, 西本 卓也, 嵯峨山 茂樹: “音声入力への応答タイミング決定のための強化学習の検討,” 電子情報通信学会技術報告 (音声研究会)(共催 日本音響学会 聴覚研究会), Mar. 2011.

[112] Takuya Nishimoto, Masahiro Araki, Yasuhisa Niimi: “RadioDoc : A Voice-Accessible Document System,” Proceedings ICSLP2002, pp.1485-1488, Denver, Sep. 2002.

[113] Takuya Nishimoto, Takayuki Kawasaki: “Support tools for broadcasting self-created radio programs for the visually impaired,” CSUN 17th Annual International Conference, Technology and Persons with Disabilities, Mar. 2002.

[114] 西本 卓也, 光部 杏里, 渡辺 隆行: “ラジオ放送番組におけるスポーツ実況中継の分析,” 電子情報通信学会技術報告, WIT2006-6, pp.27-32, May 2006.

[115] 西本 卓也, 宮川 祥子, 川崎 隆章: “インターネットラジオによる情報発信支援ツールの設計,” 電子情報通信学会技術研究報告, WIT 2001-7, pp.35-40, May 2001.

[116] 住吉 悠希, 西本 卓也, 荒木 雅弘, 新美 康永: “インターネットラジオ番組制作支援ツール,” 情報処理学会インタラクション 2002, pp.153-154, Mar. 2002.

[117] 西本 卓也, 川崎 隆章: “ラジオ放送支援システム「オラビー」の開発,” 電子情報通信学会技術報告, WIT2006-25, pp.49-54, Jul 2006.

[118] 西本 卓也, 志田 修利, 小林 哲則, 白井 克彦: “マルチモーダル入力環境下における音声の協調的利用 —音声作図システム S-tgif の設計と評価—,” 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.12, pp.2176-2183, Dec. 1996.

[119] 西本 卓也, 新美 康永: “音声認識の自己目的的な楽しさ,” 人工知能学会研究会資料, SIG-SLUD-9804, pp.13-18, Feb. 1999.

[120] 西本 卓也, 幸 英浩, 川原 毅彦, 荒木 雅弘, 新美 康永: “非同期型音声会議システム AVM の設計と評価,” 電子情報通信学会論文誌 D-II, Vol.J83-D-II, No.11, pp.2490-2497, Nov 2000.

[121] 西本 卓也, 荒木 雅弘, 小林 哲則: “ディストラクション評価に基づく車内音声対話コンテンツの比較,” FIT2003 イベント企画, 「車載情報システムにおける音声インタフェース」予稿集, Sep. 2003.

[122] 西本 卓也, 高山 元希, 櫻井 晴章, 荒木 雅弘: “音声インタフェースのための対話負荷測定法,” 電子情報通信学会論文誌 D-II, Vol.J87-D-II, No.2, pp.513-520, Feb. 2004.

[123] 西本 卓也, 狩谷 幸香, 渡辺 隆行: “早口音声聴取における単語親密度とメンタルワーク

ロードの検討,” 日本音響学会 2007 年秋季研究発表会講演論文集, 2-Q-23, pp.593-594 (in CD-ROM), Sep. 2007.

[124] 西本 卓也, 渡辺 隆行: “単語親密度を統制した超早口音声の聴取に対する慣れの検討,” 電子情報通信学会論文誌 D, Vol.J94-D, No.1, pp.209-220, Jan. 2011.

[125] 西本 卓也: “マルチモーダル対話システムのためのアーキテクチャ階層化,” FIT2006 イベント企画「音声・マルチモーダル対話記述とその標準化」予稿集, Sep 2006.