

音声スポット： “FOCUS” on Speech

北山 広治[†] 後藤 真孝^{††} 伊藤 克亘[‡] 小林 哲則[†]

[†] 早稲田大学 理工学部 〒 169-8555 東京都新宿区大久保 3-4-1

^{††} 産業技術総合研究所 〒 305-8568 茨城県つくば市梅園 1-1-1

[‡] 名古屋大学大学院 情報科学研究科 〒 464-8603 名古屋市千種区不老町 1

E-mail: †{kitayama, koba}@tk.elec.waseda.ac.jp, ††m.goto@aist.go.jp, ‡itou@is.nagoya-u.ac.jp

あらまし 音声に含まれる非言語情報である言い淀み（有声休止）と声の高さを活用した新しい音声インタフェース「音声スポット」の提案と評価を行う。会話中のユーザの音声、音声認識システムに対する発話か、会話相手の人に対する発話かを、音声だけから識別することが困難なため、従来、人間同士の会話中に音声認識システムは利用されていなかった。この問題を、音声スポットでは、通常の会話では現れない「有声休止の後に、高い声で話す」という特殊な発声方法を音声認識システムへの発話とすることで解決する。また、音声スポットを応用した「オンデマンド会話支援システム」「BGM付き電話システム」を構築し、音声スポットの有効性を確認した。

Speech Spotter: “FOCUS” on Speech

Koji Kitayama[†] Masataka Goto^{††} Katunobu Itou[‡] Tetsunori Kobayashi[†]

[†]School of Science and Engineering, Waseda University

^{††}National Institute of Advanced Industrial Science Technology (AIST)

[‡]Graduate School of Information Science, Nagoya University

Abstract In this paper, we propose a novel speech interface function called *Speech Spotter* that uses two kinds of non-verbal speech information: filled pause (lengthening of a vowel) and voice pitch. Automatic speech recognition systems have not been used in human-human conversation because it is difficult to judge whether a user talks to another person or a system. The *Speech Spotter* function solves this problem by making a system accept only a special voice-command utterance that is uttered with a high pitch just after a filled pause. By using this function, we built two application systems: “an on-demand human-human conversation support system” and “a telephone system with BGM-playback function”. The results of using these systems showed that the *Speech Spotter* function is robust and convenient enough to be used in daily human-human conversation at a site or over a cellular phone.

1 はじめに

本研究では、新たな音声入力インタフェース機能を導入することにより、人間同士が日常会話をおこないながら、音声認識によって計算機の支援が受けられるようにすることを目的とする。人間同士の会話中に、あたかもそこに第三者がいるかのように計算機の支援を受けられると便利である。例えば、人と会話中に今日が何日かを知りたくなったり、天気予報や、スポーツの試合結果を知りたくなった場合、通常は会話を中断して近くにいる第三者に尋ねるか、計算機に向かって調査をして解決することが多い。その代わりに、計算機が人間同士の会話を聞くことができ、ユーザの知りたい時に情報を提供してくれると、会話を中断せずに情報支援が受けられて有用である。

従来このような会話中の情報支援を実現することは

難しかった。音声情報のみで実現する場合、ワードスポッティング技術でキーワード検出して支援することが一般的である [1]。しかし、検出したキーワードが計算機に対する「処理対象音声」なのか、会話相手に対する「通常会話音声」なのかを話題に限定せず区別するとは困難であった。一方、音声以外の情報を併用する方法では、ボタンを押す（スイッチを入れる）ことによって区別する方法や、画像情報からユーザの顔や視線の方向によって識別する方法 [2, 3] が提案されている。また、画像情報を利用しつつ、会話内容を理解して情報支援を行う方法 [3, 4] が提案されている。しかし、いずれの方法でも、話題を限定せず音声情報のみの一般的な環境で、「処理対象音声」の区別を行うことは困難であった。

本研究では、音声情報だけで人間同士の会話中の処

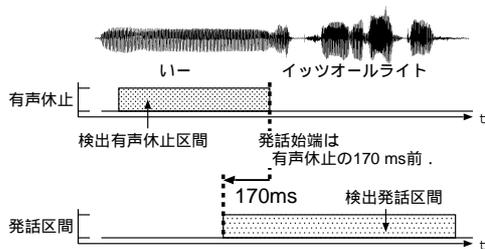


図 1: 発話区間の開始の決定

理対象音声と同定できる新たな音声入力インタフェース機能「音声スポット」を提案する。「音声スポット」では、通常の会話は無視し、母音を延ばして言い淀んだ後に故意に高い声で発声された、不自然で特殊な発話だけを計算機の処理対象とみなす。例えば、「えー、今日は何日」のように、言い淀んだ後に入力したい単語を高い声で発声（以下、下線は故意に高く発声することを示すものとする）すれば、計算機がその答えを教えてくれる。

2 音声スポット

音声スポットとは、ユーザが通常の会話では口にしていない不自然な発声を行うことによって、処理対象音声を指定することができる、新たな音声インタフェース機能である。従来の典型的な音声認識システムでは、処理対象音声と通常会話音声との区別ができなかったために、常にマイクから音声が入力されている状況下では、ユーザの発声はすべて処理対象音声として扱われた。そのため、ユーザとシステムが一对一で対話する形式が主だった。

音声スポットでは、ユーザが以下の二つの手順で発声することで、システムにその言葉が処理対象音声であることを伝えることができる。

1. 有声休止（「えー」や「あー」等の任意の母音の引き延ばし）によって言い淀む。
2. 言い淀んだ直後に、入力したい言葉を高い声（高いピッチ）で発話する。

日常会話中の通常の発声は、ほとんど上記に該当しない（言い淀んだ後に高い声で話すことがほとんどない）ため、システム側は的確に無視することができる。

このように音声スポットでは、ユーザが音声認識システムにコマンド入力したいという意図を、言い淀みと声の高さという二つの非言語情報で伝える点が新しい。音声スポットを使うことで以下の利点が得られる。

- 人間同士の会話中の音声入力が可能
人間同士で会話を続けながら、ユーザの望むときだけ音声認識システムを利用することができる。
- 省力化
ボタンやマウス、カメラといった他の入力装置を使わずに、音声だけで操作が可能である。
- 心的負担が少ない
システムの支援が不要な場合、存在を意識せず振

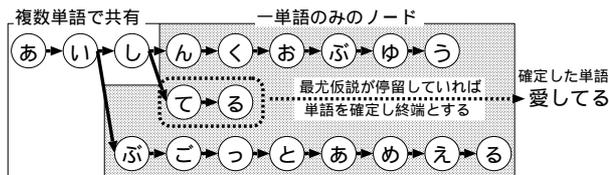


図 2: 発話区間終端の決定: 一定時間固有ノードに停留した場合に認識を終える。



図 3: 発話区間終端の決定: 一定時間無音ノードに停留した場合に認識を終える。

る舞うことができ、心的負担が少ない。

3 音声スポットの実現方法

音声スポットを実現するには、以下の四つの処理を順に実行する必要がある。

1. ユーザの音声から、言語情報に依存せずに言い淀み箇所を検出。
2. 検出した言い淀み箇所を目印に、発話区間¹始端を決定。
3. ユーザの発話内容から自動的に発話区間終端を決定。
4. 決定した発話区間中の音声が高声であるか通常の声であるかを識別。

3.1 言い淀みの検出

言い淀みの検出には、文献 [5] のリアルタイム有声休止検出手法を採用する。この手法は、有声休止（母音の引き延ばし）が持つ二つの音響的特徴（基本周波数の変動が小さい、スペクトル包絡の変形が小さい）をボトムアップな信号処理によってリアルタイムに検出する。任意の母音の引き延ばしを言語非依存に検出できるという特長を持つ。

3.2 発話区間始端の決定

発話区間の検出には、文献 [6] の発話区間決定手法を採用する。基本的には、検出した有声休止の終端を、発話区間始端とすればよい。しかし、有声休止と同じ音で続く発話（例えば、図 1「いー、イツオールライト」など）での誤認識を避けるために、有声休止の終端から 170ms 秒前の時刻を、発話区間始端とする。なお、有声休止の開始端を発話区間始端としない理由は、音声認識処理中に母音が不必要に長く続くと雑音が入る可能性が高くなり、誤認識を招きやすいためである。170ms という数値は、予備実験より決定した [6]。

3.3 発話区間終端の決定

発話区間の始端はユーザが明示的に指示するが、終端は、自動的に決定されるとユーザの負担は少ない。

¹ ここでいう発話区間とは、人が発話している区間を指すのではなく、音声認識器が音声認識処理を行なう区間を表す。

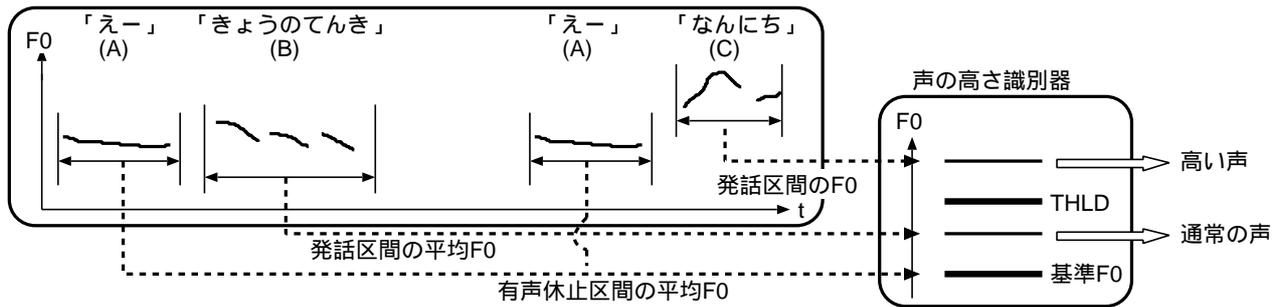


図 4: 声の高さの識別処理の概要

そこで、文献 [7, 8] に基づいて、音声認識器の各時刻の最尤仮説が一定の条件を満たしたときに、発話区間終端であると決定する [6]。音声認識器は、発話区間始端が決定された後、即座に認識処理を始める。音声認識器が各時刻において (10ms ごとに) 出力する最尤仮説を監視し、以下に挙げる二つのノードのいずれかに一定時間以上停留していたら、その時刻を発話区間終端とする。

1. 木構造辞書中で複数単語に共有されていない固有ノード (一単語のみのノード) [7]。
図 2 の 3 単語 (アイシクオブユウ, 愛してる, アイブゴットアメエル) を持つ木構造辞書をサーチ中とする。図 2 の網掛け部分は固有ノードに当たる。最尤仮説が図 2 波線で囲む部分である「てる」に到達しており、かつ数十ミリ秒連続して、最尤仮説が「てる」であれば、「愛してる」であると十分判断できるとし、発話の終端とする。
2. 文法上の文末の無音にあたる無音ノード。
ここでは、図 3 に例示したように、入力中の音声が無音になった可能性が連続して高いときに、入力音声を終了したと考える [8]。

3.4 声の高さの識別

文献 [9, 10] の手法を採用し、基本周波数 (以下, F_0) を元に、声の高さの識別を行なう。声の高低には個人差があり、個人ごとに相対的に判断する必要があるため、何らかの個人ごとの基準が必要となる。声の高さに関して、各個人の基準となる基準基本周波数 (以下, 基準 F_0) という考えを導入する。発話区間中の F_0 の平均が、基準 F_0 に対してある閾値より高ければ、高く発声されたと識別する。この基準 F_0 は有声休止から推定する。有声休止中は、調音器官の変化が小さく、 F_0 が安定する特徴を持ち、地声の F_0 (すなわち、基準 F_0) に近くなると仮定できる。そこで、有声休止区間中の F_0 の平均を話者固有の基準 F_0 とする [9, 10]。 F_0 推定と有声休止検出には、3.1 の言い淀み箇所の検出と同じく、文献 [5] の検出手法を用いる。

以上の処理の概要を図 4 に示す。図の左側は、「えー、今日の天気、えー、何日。」(「何日」は故意に高く発声) を発話したときの F_0 の変化を示している。(A) の F_0 が基準 F_0 推定に用いられ、(B) の平均 F_0 は閾値 (THLD: threshold) 以下であるため通常の声と識別さ

れ、(C) の平均 F_0 は閾値以上であるため高い声と識別されることを示している。こうして高い声と識別された場合の音声認識結果 (図 4 の (C) のみ) をシステムへの入力と解釈することで、音声スポットの機能は実現される。

4 音声スポット応用システム

音声インタフェース技術の新たな応用方法を提示するとともに、音声スポットの有効性を確認するために、「オンデマンド会話支援システム」「BGM 付き電話システム」という二つのシステムを提案する。

4.1 オンデマンド会話支援システム

「オンデマンド会話支援システム」は、人間 (ユーザ) 同士の会話中に、ユーザが望むときだけ、音声で必要な情報を検索できるシステムである。例えば、友人同士の会話で、ふと天気予報や今日のニュースを知りたくなったり、今日が何年、何日かを確認したくなるときにあたかも側に居る人に尋ねるように、システムに音声で問いかけることができ有効である。

図 5 に、オンデマンド会話支援システムの実行画面を示す。システムの起動直後に情報表示用のウィンドウが開く (図 5(a))。ユーザが情報検索を行いたいとき、まず、音声スポットの操作に従い、故意に母音を延す。すると、画面上部に緑色のバーが右に向かって延びてゆく。そのまま、システムが有声休止を一定時間検出するとバーが赤色になる (図 5(b))。その後、検索したい情報を高い声で発声すると、高い声と識別したことを示す効果音 (現実装では「ポン」という音) が鳴り、ユーザの求める情報が画面に表示される (図 5(c))。

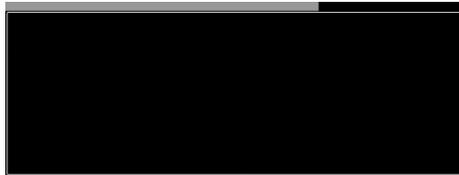
現在の実装では、情報検索結果は画面への出力と音声合成を使った音声出力の二つが可能となっている。ユーザの好みに応じて、画面または音声のどちらか一方、もしくは両方へ出力できる。特に、次の BGM 付き電話システムと組み合わせて、人間同士の通話中に会話支援を受けるときには、音声出力を使うことで画面の一切不要なシステムとなって便利である。

4.2 BGM 付き電話システム

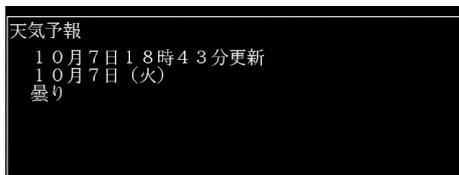
「BGM 付き電話システム」は、人間 (ユーザ) 同士が電話で通話をしながら、ユーザの望むときに、楽曲の選曲・再生が可能なシステムである。



(a) 起動直後の画面には何も表示されていない



(b) 有声休止を検出し始めると画面上部に表示される緑色のバーが画面左端から徐々に長くなっていき、さらに言い淀み続けてバーが画面右端に達して赤になると、音声入力可能となる



(c) 言い淀んだ直後に高い声で「今日の天気」と発声すると、システムからその回答が得られる

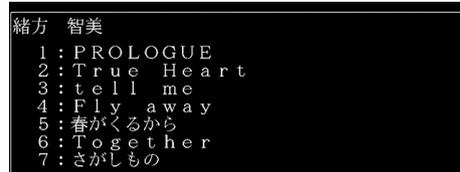
図 5: オンデマンド会話支援システムの画面表示例

図 6 に BGM 付き電話システムの実行画面を示す。システムの起動直後と言い淀み箇所の検出時の動作は、オンデマンド会話支援システム(図 5(a)(b))と同様である。また、画面出力だけでなく、音声合成出力も可能となっている点も同様である。現在の実装では、以下の二つの選曲方法に対応している。

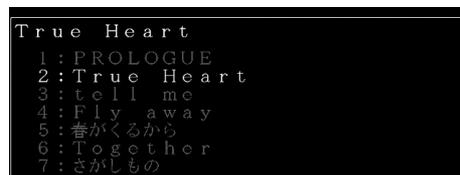
- 直接曲名を言って楽曲を再生する方法
音声スポットの操作に従い、有声休止後に曲名を高い声で発声すると、高い声と識別したことを示す効果音が鳴り、その楽曲が流れ始める(図 6(a))。画面には、音声認識結果の曲名が表示される。音声合成出力では、ユーザの入力した楽曲を確認する目的で「トゥルーハート」と読み上げ、楽曲が再生される。
- アーティスト名を言って楽曲一覧から選曲して再生する方法
上記同様、有声休止後にアーティスト名を高い声で発声すると、効果音が鳴り、画面に音声認識結果のアーティスト名と、そのアーティストによる曲目一覧が表示される(図 6(b-1))。ユーザがその中から希望楽曲の曲名または左側に書かれた番号を音声スポットの発声法で発話すると、その楽曲が流れ始める(図 6(b-2))。画面上では、選択された曲名の色がハイライトされ、その他は暗くなる。音声合成出力では、「緒方智美の曲は、1 番プロローグ、2 番トゥルーハート、3 番テルミー、4 番フライアウェイ、5 番春がくるから、6 番トゥギャザー、



(a) 曲名から音楽再生する例: 有声休止後に高い声で曲名「True Heart」を発話すると、その曲が再生される



(b-1) アーティスト名から選曲して音楽再生する例: 有声休止後に高い声でアーティスト名「緒方智美」を発話すると、緒方智美の曲目一覧が表示される



(b-2) アーティスト名から選曲して音楽再生する例: 曲目一覧を見た後に、再び有声休止後に高い声で選択番号「2番」を発話すると、選択された曲が再生される

図 6: BGM 付き電話システムの画面表示例

7 番さがしもの」のように、入力したアーティスト名に続き、曲名を順に読み上げるようになっている。ユーザは、システムの音声出力中に割り込んで音声入力することが可能である、

本システムの特長として、システムによる音楽再生中や音声合成中にも、システムへのコマンド入力や人間同士の自由な会話が可能なのが挙げられる、例えば「んー、ストップ」等の発声により再生を停止したり、別の曲名を言って楽曲を変更したりすることができる。また、会話の BGM として楽曲を再生するだけでなく、実際に楽曲を聞きながらその曲について議論したり、感想を述べ合うような使い方ができる。このような BGM の共有や再生楽曲についての意見交換は、従来遠隔地間では非常に困難だったことであり、特に音楽を日常的に自らの意志で聞く人々に対して有効である。

対象楽曲は、アーティスト名、楽曲名、楽曲音響信号のファイル名の一覧を、事前に楽曲データベースとして構築しておく必要がある。また、前のオンデマンド会話支援システムと組み合わせると、人間同士が実際に一緒にいる部屋で、その部屋の BGM を変える目的で楽曲選曲・再生機能を利用してもよい。

5 音声スポットを組み込んだ音声インタフェースの実装

「オンデマンド会話支援システム」と「BGM 付き電話システム」は、基本となる音声スポット機能を利

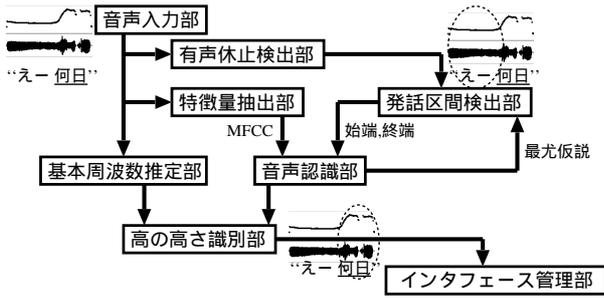


図 7: 全体の処理の流れ

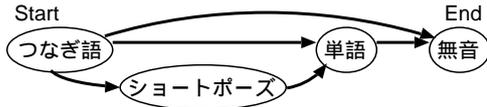


図 8: 音声スポットを組み込んだシステムで用いる文法

用し、共通のシステム構成要素で実現できる。以下、まず音声スポット機能の共通部分の実装について述べ、その後、各応用システムの実装の違いを述べる。

5.1 音声スポットの実装

図 7 に、音声スポット機能を構成する各システム構成要素(プロセス)と、全体の処理の流れを示す。システム構成要素は、図中の囲み字で示されており、ネットワーク(LAN)上の複数の計算機で分散して実行することが可能である。また、音声言語情報を分散環境で効率良く共有するために、ネットワークプロトコル RVCP [11, 12] (Remote Voice Control Protocol) を用いて実装した。

まず、音声入力装置(卓上マイクやピンマイク、ヘッドセットマイク等)から音声入力部に入力された音響信号は、ネットワーク上にパケットとして送信される。特徴量抽出部、有声休止検出部、基本周波数推定部がそのパケットを受けとり、それぞれの処理を行う。発話区間検出部では、有声休止検出部、音声認識部の結果を受信し、発話区間の始端と終端を決定する。声の高さ識別部では、基本周波数推定部、有声休止検出部、音声認識部の結果を受信し、発話が通常の声か高い声かを識別する。高い声の発話と識別された認識結果だけが、インタフェース管理部に通知され、認識結果中のコマンドが実行される。

音声認識部には、julian 3.3beta[13] を、毎フレームの単語仮説を発話区間検出部に送信することができるように拡張 [6] して用いた。音声特徴量は、MFCC + Δ MFCC + Δ power の 25 次元とし、CMN は行わない。音響モデルは、ASJ-JNAS, PB の男性話者 133 人分(約 2 万文)から学習し、混合数 16、状態数 2000 の Triphone とした。また、単語認識を基本とし、有声休止を含む代表的なつなぎ語(えー、えーっと、あー等の 12 種類に対応)から始まる、図 8 に示す文法とした。

5.2 オンデマンド会話支援システムの実装

図 7 のインタフェース管理部を変更することで、以下の機能を持つオンデマンド会話支援システムを実装した。

- 天気情報の検索機能
「あー、明日の天気」のような発声で、東京都の天気情報(今日、明日、明後日、1 週間)を提示する。その情報を記載している WWW ページ (Yahoo Japan “<http://www.yahoo.co.jp/>”) から、w3m[14] を使いテキストデータとして取得して実装した。
- 現在の日時の照会機能
「えー、今日は何日」のような発声で、現在の日時を提示する。システムが動作する OS (Linux) 上で得られる時刻 (date コマンド結果) を取得して実装した。
- ニュースの見出しの表示機能
「んー、経済ニュース」のような発声で、各種ニュース(経済、株・為替、国際、企業、IT、政治、社会、スポーツ)の見出し部分を表示する。天気と同様に、WWW ページ (日本経済新聞社 “<http://www.nikkei.co.jp/>”) から取得して実装した。

5.3 BGM 付き電話システムの実装

図 7 のインタフェース管理部を変更し、楽曲再生機能を組み込むことで、BGM 付き電話システムを実装した。楽曲データベースには、RWC 研究用音楽データベース: ポピュラー音楽 [15] の 100 曲 (RWC-MDB-P-2001 No. 1 ~ 100) を収め、楽曲名 100 語、アーティスト名 34 語を登録した。

現在の実装では、音声認識に利用する音声はヘッドホン付きヘッドセットマイクから収録する形態とした。

6 評価実験

音声スポットの基本性能を確認するために、通常会話音声の棄却性能と、音声スポットを組み込んだ二つの応用システムの運用結果を述べる。

6.1 通常会話音声の棄却性能評価実験

音声スポットでは、ユーザの故意に発する「言い淀み」と「声の高さ」の二つを同時に用いている。しかし、単純に「声の高さ」だけを用いて、ユーザが処理対象音声を指定できる可能性もある。そこで、通常会話音声の棄却性能を、(1)「声の高さ」だけで棄却、(2)音声スポットで棄却、の二つの方法で比較評価し、音声スポットの有効性を検証する。通常会話音声を処理対象音声と誤って識別した回数が少ないほど、人間同士の会話に不必要に干渉しない優れた手法と言える。

本評価では、(1)と(2)を的確に比較するために、声の高さの識別に用いる閾値は、双方共通の値とする。その閾値は、文献 [9] で示される、話者共通の閾値で

表 1: 実験結果：通常会話音声処理対象音声と誤った回数

	声の高さのみ	音声スポット
誤識別発話数	1338	60

ある 300cent とした。評価データは、PASD 音声コーパス [16] のうち、(1) の方法で有声休止区間から基準 F0 を求める必要があるため、有声休止区間の存在するデータ (81 データ：約 6 時間) を対象とした。その各評価データから、あらかじめ基準 F0 を求め (1) の実験で利用する。

表 1 の実験結果から、声の高さだけで識別しようとすると誤りが多いが、音声スポットによって有声休止が先行することも条件とすれば、誤りを大幅に (95.5%) 削減できることがわかる。この結果から、二つの非言語情報を組み合わせることが有効であることが分かる。

6.2 音声スポットの運用結果

音声スポットを組み込んだ二つの応用システムの運用結果を述べる。いずれの場合も、音声スポットを用いることで、人間同士の自由な会話中に音声認識システムを利用することができた。また、システムの支援が不要なときに、わざわざマイクのスイッチを切らなくてよく、必要なときだけ音声スポットの不自然な発声を行えば支援が得られるため、手間がかからず有用であった。

6.2.1 オンデマンド会話支援システムの運用結果

二人の人間がマイクの近くで会話中に、音声で、天気や日時、ニュースを調べることができ、本システムが的確に動作することが確認された。従来は会話中にそれらの情報を知りたくなると、WWW ブラウザ等を操作して調べるわずらわしさがあったが、本システムでは音声入力だけで手軽に確認ができ、便利であった。通常と異なる発声方法のときだけシステムが反応すること自体を面白いと感じるユーザもあり、普通の発声と音声スポットの発声とを、会話中に遊び感覚で使い分ける場面も見られた。

6.2.2 BGM 付き電話システムの運用結果

音声スポットを使うことで、市販の携帯電話や固定電話で通話を行ないながら、ユーザが好きなときに BGM を鳴らすことができた。携帯電話自体から音楽再生を行う着信メロディーや待ち受けメロディーは既に存在するが、通話中に BGM を聞くという感覚は従来経験したことのないものであり、試用したユーザに好評であった。なお、音楽は携帯電話の音声と同じ伝送路を通るため、運用実験に用いた様々な携帯電話のキャリアや機種によって、音質が大きく異なった。

7 まとめ

本稿では、故意に言い淀んでから故意に高い声で発声するという不自然な発声を取って採用することで、ユーザが処理対象音声と、そうでない音声とを使い分けることを可能にする「音声スポット」という音声イ

ンタフェース機能を提案した。さらに、音声スポットを有効に活用する二つの応用システムを提案した。一つ目の「オンデマンド会話支援システム」では、人間同士で会話を行いながら、計算機による情報支援を音声だけで受けることを可能にした。二つ目の「BGM 付き電話システム」では、自宅で BGM を流しながら友人と会話する行為を、携帯電話を用いて遠隔地間で行なうことを可能にした。これは、我々の調査した限り、人間同士の電話による会話中に、話者が音声だけで計算機の支援を意図的に得ることを初めて可能にした事例である。これらのシステムを運用した結果、日常会話を行ないながら実際に音声認識システムを利用できることを確認した。

なお、本研究は、音声補完 [11, 12, 17]、音声シフト [9, 10]、音声スタータ [6] の一連の研究に続く第四弾であり、音声の非言語情報を積極的に活用することで、より使いやすい音声インタフェース機能を実現したものである。

参考文献

- [1] 河原達也, 石塚健太郎, 堂下修司, “発話検証に基づく音声操作プロジェクトとそれによる講演の自動ハイパーテキスト化,” 情処学論, Vol.40, No.4, pp.1491-1498, 1999.
- [2] 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—,” 信学論 (D-II), Vol.J78-D-II, No.12, pp.2176-2183, 1996.
- [3] 松坂要佐, 東篠剛史, 小林哲則, “グループ会話に参加する対話ロボットの構築,” 信学論 (D-II), Vol.J84-D-II, No.6, pp.898-908, 2001.
- [4] K. Nagao and A. Takeuchi, “Social Interaction: Multimodal Conversation with Social Agents,” *Proc. AAAI-94*, Vol.1, pp.22-28, 1994.
- [5] 後藤真孝, 伊藤克亘, 速水悟, “自然発話中の有声休止箇所のリアルタイム検出システム,” 信学論 (D-II), Vol.J83-D-II, No.11, pp.2330-2340, 2000.
- [6] 北山広治, 後藤真孝, 伊藤克亘, 小林哲則, “音声スタータ: “SWITCH” on Speech,” 情処技報 2003-SLP-46, pp.67-72, 2003.
- [7] 井ノ上直己, 中村誠, 酒寄信一, 山本誠一, 谷戸文廣, “単語固有セルでのゆう度判定を用いた音声認識処理の高速化手法,” 信学論 (D-II), Vol.J79-D-II, No.12, pp.2110-2116, 1996.
- [8] 内藤正樹, 黒岩真吾, 山本誠一, 武田一哉, “部分文仮説のゆう度を用いた連続音声認識のための音声区間検出法,” 信学論 (D-II), Vol.J80-D-II, No.11, pp.2895-2903, 1997.
- [9] 尾本幸宏, 後藤真孝, 伊藤克亘, 小林哲則, “音声シフト: 音高を利用した新たな音声入力インタフェース,” WISS 2001, 近代科学社, pp.17-26, 2001.
- [10] 尾本幸宏, 後藤真孝, 伊藤克亘, 小林哲則, “音声シフト: “SHIFT” on Speech,” 情処研報, 2002-SLP-40, pp.13-18, 2002.
- [11] 後藤真孝, 伊藤克亘, 速水悟, “音声補完: “TAB” on Speech,” 情処研報, 2000-SLP-32, pp.81-86, 2000.
- [12] 後藤真孝, 伊藤克亘, 秋葉友良, 速水悟, “音声補完: 音声入力インタフェースへの新しいモダリティの導入,” コンピュータソフトウェア (日本ソフトウェア科学会論文誌), Vol.19, No.4, pp.10-21, 2002.
- [13] A. Lee, T. Kawahara and K. Shikano, “Julius — an open source real-time large vocabulary recognition engine,” *Proc. Eurospeech2001*, pp.1691-1694, 2001.
- [14] w3m project, “<http://w3m.sourceforge.net/>”
- [15] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一, “RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース,” 情処研報, 2001-MUS-42, pp.35-42, 2001.
- [16] 重点領域模擬対話音声コーパス, “<http://winnie.kuis.kyoto-u.ac.jp/taiwa-corpus/>”
- [17] 後藤真孝, 解説 “音声補完: 言い淀むと助けてくれる音声インタフェース,” 情報処理 (情報処理学会誌), Vol.43, No.11, pp.1210-1216, 2002.