

# ELIZA 的アプローチによって未知語を 疑似獲得する音声対話システム

高橋 勉<sup>†</sup> 川端 豪<sup>†</sup>

ユーザ発話中の未知語を疑似的に獲得しシステム発話に利用する機能を持った音声対話システムを構築する。ユーザ発話中には例えば未登録の選手名など未知語が多く含まれている。システムは未知語を検出しビタビアライメントを用いてその音声区間の始点・終点を決定する。また、検出された区間の音声を未知語として記憶する。システムの次の発話順番において合成する音声にこの記憶した単語を含めることによって、あたかも新しい単語を獲得したかのような振る舞いをさせることができる。このような ELIZA 的なアプローチによって未知語を疑似獲得する。

## Acquisition of unknown spoken words based on the ELIZA approach for the spoken dialog system

TSUTOMU TAKAHASHI<sup>†</sup> TAKESHI KAWABATA<sup>†</sup>

This paper describes the spoken dialog system which has the function of unknown spoken word acquisition. The user of spoken dialog system sometimes utters unknown words, for example unregistered player's name. When an unknown word is detected in the user's speech, the system precisely locates the start and end points of the unknown words using the Viterbi alignment, and memorizes the speech pattern of the unknown word. In the next turn, the system can synthesize its utterance including the memorized word. Such ELIZA-like approach makes a show that the system acquired the new vocabulary with only surface processing.

### 1. はじめに

現在、音声対話システムは音声認識技術の進歩に伴い、介護における話し相手や、車載情報サービスなど、多岐にわたり人間はコンピュータと対話するようになった。しかし、人間同士の対話と比較すると、コンピュータの応答や対話の進行が不自然になることも多い。

亀田はその原因の一つとして円滑性（発話者の意図に沿い、対話が速やかに行われること）の欠如を報告している。コンピュータの単語辞書内に未登録な単語、即ち未知語に遭遇した場合にその現象は著しい。従来の対話システムでは未知語に対してまったく対応できなかつたり、あるいはできたとしても人間への質問や話題転換が頻繁に起こってしまい、対話の円滑さが損なわれる場合がある[1]。

未知語処理の手法として、入力音声に対し形態素解析を行いデータベースの文例から類似性をもとに未知語処理す

る手法[2]や複数のサンプルの音素系列から未知語の音素を推定する手法などがある[3]。また、単語 N-gram に音韻連鎖モデルを統合し未知語処理する手法も用いられている[4]。

これらの正統的未知語処理とは全く異なるアプローチとして ELIZA という、対話システムがある[5]。テキストベースの対話において、未知の単語を抽出すると、それを記憶しておいて、何らかのタイミングでシステムの応答に組み込むことで、あたかも未知語を獲得したかのような振る舞いをする仕組みを持っている。本報告では、音声対話システムにおいて、この ELIZA と同じようにユーザの発声した未知単語を記憶しておいて、システムの発話を組み込むことで、疑似的に未知語を獲得するかのような動作を行わせることを考える。テキスト入力の ELIZA では単語は区切って入力されるのでその単語をそのまま記憶すればよいが、音声入力の場合は何らかの方法で未知語部分を切り出し記憶する必要がある。このため、話題に特有の文型を音声認識用の文法に書き込み、未知語を発見した後ビタビアライメントを行う。

<sup>†</sup> 関西学院大学

Kwansei Gakuin University

本研究の目的は、未知語彙を疑似獲得し、音声で応答を返す音声対話システムの可能性を検討することである。まず、ELIZA 的アプローチにより未知語を疑似獲得する。未知語部分を音節連鎖に展開した文法を作成し、未知語の発話に対し認識を可能にする。そして、ビタビアライメントによる未知語音声の切り出しを行い、記憶し、記憶された単語の声質を変更し応答音声に利用する。この手法を実装した音声対話システムにより主観評価実験を行い、その効果を評価する。音声認識には音声認識エンジン Julius を用いる。

## 2. ELIZA 的アプローチによって未知語を疑似獲得する音声対話システム

### 2.1 システム構成

図 1 にシステムの構成図を示す。音声入力に対し、文法・辞書部分から通常時と未知語検出のための文法を用いて、音声認識を行う。未知語検出のための文法が動作するとビタビアライメントによって未知語音声を記憶する。対話制御では応答事例ベースが用意されており、応答文には未知語が含まれる。発話音声には、人間が前もって音声を録音しておく。本研究ではサンプリング周波数 16kHz で収録した。再生時にはサンプリング周波数を 18kHz と高くした。また、記憶した未知語音声もサンプリング周波数 16kHz で録音し、サンプリング周波数 18kHz で再生した。これは音質の違いに気づかれないようにするためである。音声合成では、これらの音声を用いて、発生する。

図 2 に未知語疑似獲得処理の流れを示す。音声認識を行い、それに並行し、音声の録音を行う。録音された未知語を

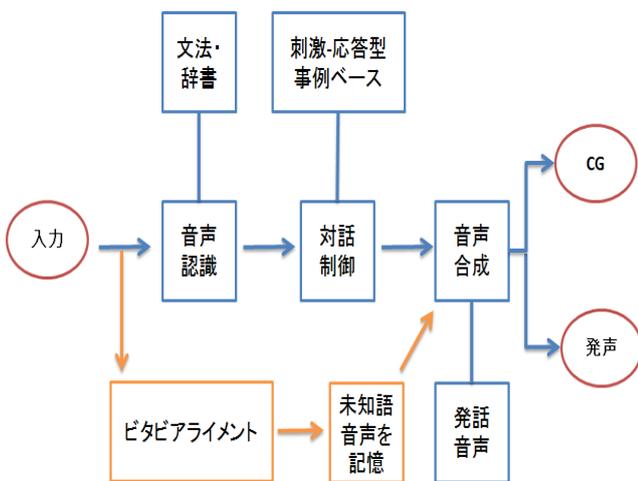


図 1 システムの構成図

含んだ音声ファイルに対し、ビタビアライメントによる未知語区間のフレーム値を用いて、未知語部分の切り出しを行う。結果、未知語のみの音声ファイルが抽出される。このような流れで処理を行う。

### 2.2 未知語区間判定のための音声認識文法

本研究では CFG に基づく認識を行う。CFG とは文脈自由文法のことであり、変数と規則がそれぞれ数多く設けられており、規則と変数を分離し自由に組み合わせることで、様々な文章を作ることが可能になる。

未知語部分に対し図 3 に示すように音節連鎖に展開した文法を作成する（ただし、CFG は図示しにくいのでネットワーク表現している）。音声の未知語区間は、任意個数の音節連鎖として認識される。例えば、「先発は高橋選手です」や、「先発は佐藤だ」など「高橋」や「佐藤」を未知語とする音声が発話された場合に「先発はたかはし選手です」、「先発はさとうだ」のように認識され、「たかはし」、「さとう」などの音節連鎖として認識される。このように、あらゆる音節の連鎖を可能にし、様々な未知語に対して認識を可能にする。

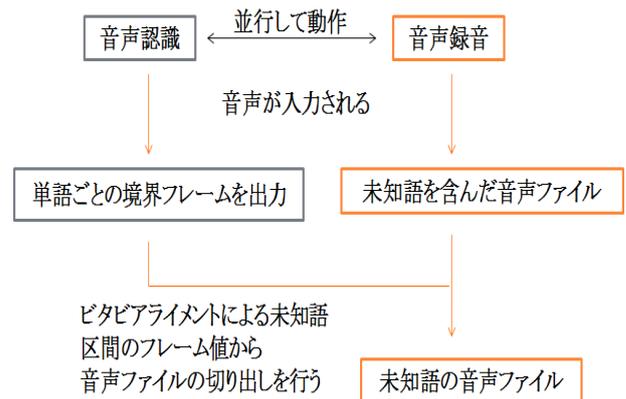


図 2 未知語疑似獲得処理の流れ

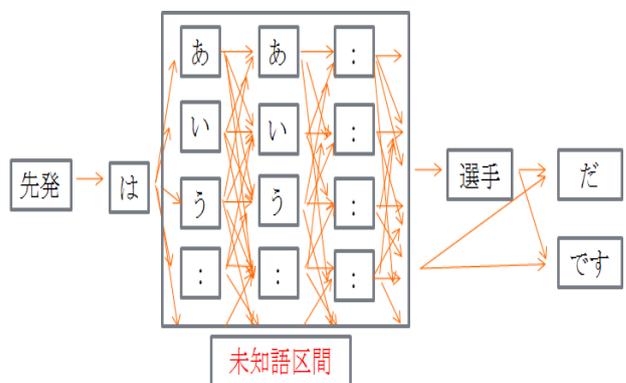


図 3 音声認識文法のネットワーク表現

### 2.3 ビタビライメントによる未知語区間獲得

ビタビライメントによる未知語区間の判定を行う。

ビタビライメントでは図4に示すように音素ごとの発声パターンの統計的モデルである音素 HMM に基づき入力音声の特徴量の最尤状態系列のみを算出する。ここで求められた最尤状態系列の結果から入力音声の未知語区間の決定を行う。決定された未知語区間はフレーム値で出力される。

本研究では、ビタビライメントにより出力された未知語区間のフレーム値を用いて、未知語部分の切り出しを行う。

## 3. 評価実験

### 3.1 実験準備

いくつかの対話の内容を仮定しておき、音声認識を可能にするため辞書ファイルを作成した。システムがそれらを認識したときの応答文を考え、サンプリング周波数 16kHz で収録しておき、サンプリング周波数 18 kHz 再生するように応答音声を作成した。また、対話をより親近感や円滑にするため、Microsoft Agent を用い、図5に示すようにエージェントを作成した。このエージェントにはうなずきや首をかしげるなどいくつかの反応パターンを用意しておき、認識内容によって、反応させた。これらの応答音声とエージェントの反応パターンを組み合わせ、対話実験を行った。

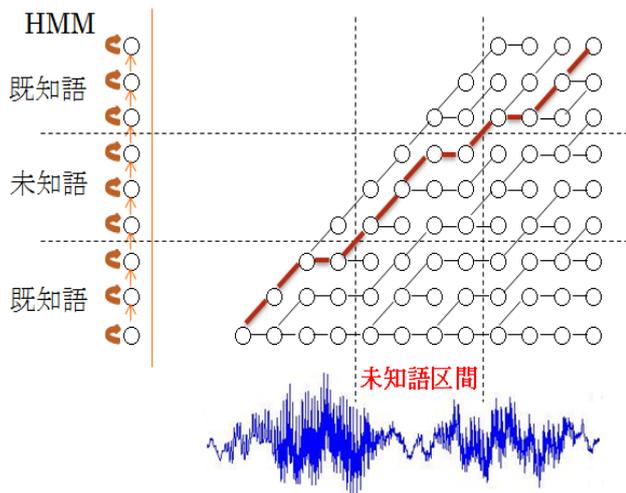


図4 ビタビライメントによる未知語区間決定

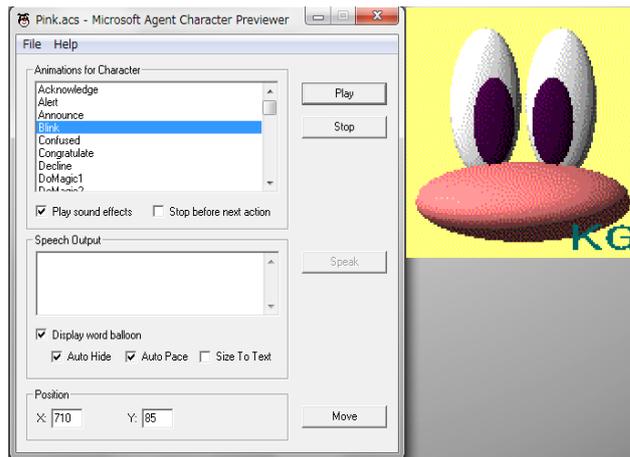


図5 エージェントの利用

### 3.2 実験方法

成人9名に被験者実験を行った。被験者には、初めに音声認識に慣れてもらうため、音声認識可能な発話リストを用意し、練習を行った。ある程度音声認識が可能になったところで実験を行った。実験は、表1に示すようにあらかじめ用意した発話リストに沿って、対話してもらい、対話システムの応答音声を聞いてもらう。表2はその発話リストに沿って、対話した場合のシステムの応答の例である。表中の○○○や△△△△の部分には未知語彙と設定し被験者に自由に発生してもらった。システムはこの○○○や△△△△などの未知語彙を疑似獲得し、次のシステムの応答音声として用いた。

音声録音はサンプリング周波数 16kHz で行い、未知語彙を獲得した。獲得した未知語彙はサンプリング周波数 18kHz で応答音声とした。これは獲得した未知語彙を被験者にシステムの音声のように感じさせ、自分の声だと思わせないためである。

発話リストの対話がすべて終わった後アンケートに答えてもらった。

表1 発話リスト

	発話リスト
1	やあ
2	野球の試合見た？
3	○○○はカーブがすごかったな
4	そういえば、今阪神何位？
5	でも、大阪桐蔭の○○○○を獲得したらいいよ
6	来年に期待しよ
7	巨人だろ
8	来年も巨人が優勝だろ
9	そうやな、じゃーね

### 3.3 評価アンケート

評価項目として表 3 に示すように質問項目は 25 個用意し、各々について SD 法を用い、「かなり (+2)」、「やや (+1)」、「どちらでもない (0)」、「やや (-1)」、「かなり (-2)」の 5 段階で評価を行った。

図 6 は本実験に対する主観評価実験の結果である。

「楽しい、愉快的」や「リラックスできる」、「興味深い」などに影響されており、全体的には良い印象を与えていることがわかる。しかし、「発音が不自然」や「音声の品質が悪い」などにも影響があり、未知語彙の獲得について悪い印象を与えている結果にもなった。

表 2 対話例

	対話例
ユーザ	やあ
システム	やあ
ユーザ	野球の試合見た？
システム	見た見た
ユーザ	〇〇〇はカーブがすごかったな
システム	〇〇〇調子よかったね
ユーザ	そういえば、今阪神何位？
システム	5位だよ。弱いね
ユーザ	でも、大阪桐蔭の△△△△を獲得したらしいよ
システム	△△△△いいね
ユーザ	来年に期待しよ
システム	そうだね。来年はどこが優勝するかな？
ユーザ	巨人だろ
システム	巨人いいね
ユーザ	来年も巨人が優勝だろ
システム	優勝しそうだね
ユーザ	そうやな、じゃーね
システム	さようなら

### 4. 結論・考察

人間の音声から未知語獲得の手法を提案し、この手法を用い、主観評価実験を行った。アンケート項目の「楽しい、愉快的」や「リラックスできる」、「興味深い」などの項目に影響を与えており、面白さや興味を記していることがわかる。しかし、「発音が不自然」や「音声の品質が悪い」などの項目にも特徴が見られた。この原因の一つは獲得した未知語彙をシステムが応答音声として用いたとき、あらかじめ収録していた音声との違いを感じたためである。

表 3 アンケート項目

1	良い	悪い
2	楽観的	悲観的
3	やりがいのある	やりがいのない
4	社交的な	内気な
5	楽しい、愉快的	苦しい、苦勞する
6	盛り上がる	盛り上がりりに欠く
7	眠たくない	眠い
8	意味のある	価値のない
9	重要な	ささいな
10	リラックスできる	緊張する
11	賢い	愚かな
12	評判の良い、立派な	不評な、みっともない
13	興味深い	退屈な
14	生き生きとした	生気のない
15	複雑な	単純な
16	発音が自然	発音が不自然
17	音声の品質が良い	音声の品質が悪い
18	耳に快い	調和しない
19	発声のタイミングが適度	時を誤った発声
20	意味が通じている	意味が通じていない
21	こっちのことが理解できている	理解できていない
22	活発的な	内気な
23	対話が自然に流れた	対話が不自然
24	対話が流暢	対話が訥弁
25	人間的	機械的

例えば、ユーザが「○○○はカーブがすごかったな」に対して、システムが「○○○調子いいね」と応答したとき、「○○○」と「調子いいね」で音質の違いを感じ、違和感を抱いたからである。また、完全に未知語彙を獲得できるとは限らず、後続する音声の一部も巻き込んでしまう場合も見られたためである。

このように未知語を疑似獲得する対話システムは音質などを考慮した改善法が必要であると考えられる。

### 参考文献

- 1) 福岡知隆, 服部峻, 久保村千秋, 亀田弘之: 「係り受け解析を用いた未知語意味カテゴリー推定法の有効性に関する諸検討」, DEIM Forum 2012 E8-2
- 2) 福岡和隆, 税田竜一, 久保村千秋, 服部峻, 亀田弘之: 「文の類似性を用いた未知語処理手法の提案とそれに基づく円滑な対話応答システムの作成」, 情報処理学会創立 50 周年記念 (第 72 回) 全国大会
- 3) 伊藤克亘, 速水悟, 田中和世: 「単語発声の複数サンプルを利用した未知語の音韻系列の推定」, 電子情報通信学会論文誌 D-II Vol. J83-D-II No.11 pp.2152-2159 (2000)
- 4) 谷垣宏一, 山本博史, 匂坂芳典: 「クラスに依存した語彙の確率的記述に基づく階層型言語モデル」, 電子情報学会通信学会論文誌 D-II Vol. J84-D-II No.11 pp.2371-2378 (2001)
- 5) Joseph Weizenbaum: A Computer Program For the Study of Natural Language Communication Between Man and Machine. Communications of the ACM Vol 9, No.1, pp.36-35 (1966)

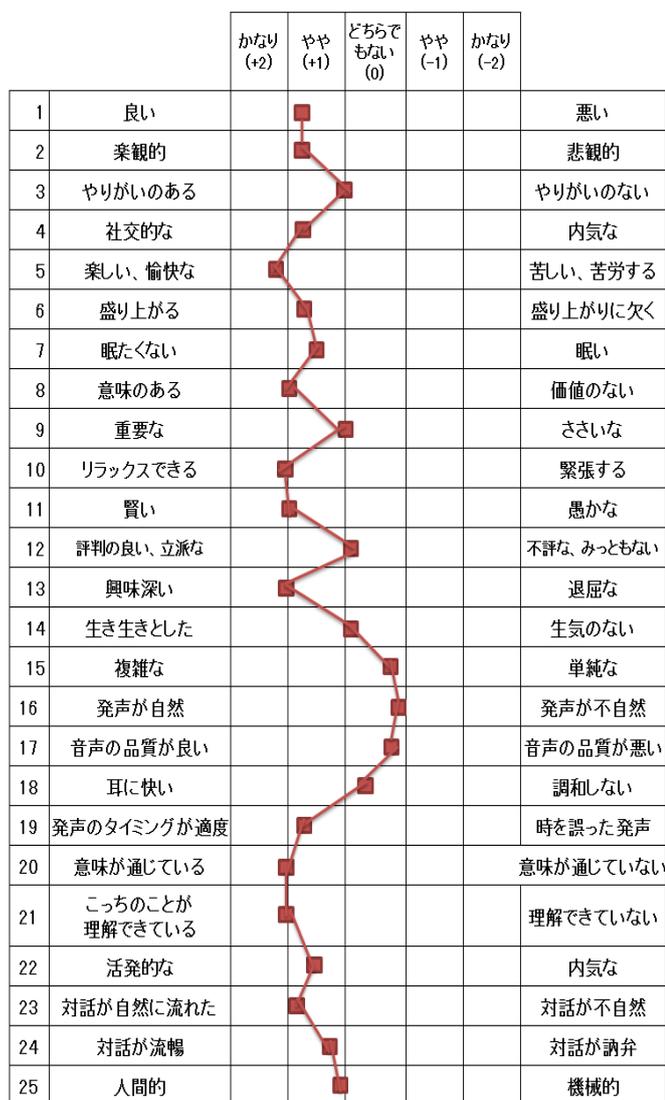


図 6 本実験に対する主観評価結果 (9 名平均)