

# 雑音環境下のための音声案内システム：ユーザーの聞き返しに基づく音量の自動調整

小暮 計貴<sup>†</sup>      鈴木 光<sup>‡</sup>      吉永 眞宏<sup>‡</sup>      北原 鉄朗<sup>‡</sup>

<sup>†</sup> 日本大学大学院総合基礎科学研究科地球情報数理科学専攻      <sup>‡</sup> 日本大学文理学部情報システム解析学科

## 1. はじめに

音声認識の技術が普及し、カーナビゲーションやインターネットでの情報検索など、多くの場面で音声対話が活用されている<sup>1)</sup>。このような音声対話システムが広まり、雑音に頑健な音声対話システムを構築する必要が生じた。明瞭に音声を聞き取るための研究としては伊積ら<sup>3)</sup>によるスピーカーの位置など物理的な要因を検証対象としたものや、水野ら<sup>4)</sup>による雑音に強い周波数を利用することで個人の声質を変化させずに音声合成を行う研究がなされてきた。しかし、伊積らの研究では周囲の雑音量が変化した時に対応することができない。このように、周囲の雑音量を考慮してリアルタイムに音声を聞き取りやすくする研究はあまり多くなされていない。

我々は、このような背景の下、周囲の雑音量を測定して自動的に合成音声の音量を制御したり、発話を遅延させるシステムを実現した<sup>6)</sup>。これにより、雑音下でも合成音声を聞けるようになったが、対話システムに応用した時の有効性の検証は行われていなかった。

本稿では、文献6)で提案した音量制御と発話延期を対話システムに組み込み、また、ユーザーによる聞き返しに基づく再発話や音量制御を導入する、これにより、ユーザーがシステムによる音声を聞き逃すのを削減できると期待される。

## 2. システム構成

### 2.1 システムの概要

駅のプラットフォームのように大きな騒音が継続的に発生する環境を想定し、そのような環境下でユーザーが聞き流す事なく対話を進める事ができる音声対話システムを構築する。ここでは、特定の駅のプラットフォームなどに設置され、そこから任意の駅に対する乗り換え方法(利用する路線と乗り換え駅)を案内するシステムを構築する。また、視覚障害を持つ人でも利用できるように、ユーザー、システムの両者とも音声による発話のみを用いるものとする。

システム構成図を図に示す。顔検出部が人間の顔を検出すると、ユーザーに対して呼びかけの発話を行い、ここから音声同士の対話が始まる。発話内容例を以下に示す。

システム 「こんにちは、乗り換え案内システムです」  
 システム 「どこ駅まで行きますか？」  
 ユーザー 「 駅まで」  
 システム 「 駅までで宜しいですか？」  
 ユーザー 「はい」  
 システム 「桜上水駅から 線××行きで 駅に乗り換

えてください」

システム 「 駅から 線××行きで 駅に到着です」  
 システム 「料金は 円、乗車時間は 分掛かります」  
 システム 「ご案内は以上です、ご利用有難う御座いました」

現在の実装では、本システムは京王線桜上水駅に設置されるとの前提から、同駅からユーザーが指定した駅までの乗り換え方法を案内する。乗り換え方法は「駅すばあと Web API」を用いて検索する。乗り換え案内情報が複数文にわたる場合は、一文発話した後にユーザーが「はい」と発言を行うか、3秒間たつと次の文を発話するようになっている。音声合成には OpenJTalk、音声認識には Julius を用いる。

### 2.2 雑音レベルに基づく音量制御と発話延期

文献6)で提案したように、雑音レベルを計測してそれに基づいて合成音声の音量制御・発話延期を行う。マイクロフォンアレイから取得した雑音に対して、適宜システムの発話音量を変更させたり、発話の延期を行う。

### 2.3 聞き返しなどに基づく再発話と音量制御

ユーザーの聞き返しに基づいて、以下のアクションを行う。

- (1) 「え？」や「何？」と発話された場合再度同じ、発話を行う。
- (2) 「聞こえない」や「もっと大きな声で」と発話された場合、音量を上げて再度発話。
- (3) 「うるさい」や「もっと小さな声で」と発話された場合、音量を下げて再度発話。

(1)に関しては、システムの発話内容をユーザーが理解できない時に「え？」と発話を行う事を考慮する。その反応に対し、再度発話を行えるようにする。(2)に関しては、ユーザーがシステムの発話内容が聞こえなかった場合、「聞こえない」と発話される事を想定する。その反応に対し、システム側は音量を大きくし、「これくらいで宜しいですか？」と発話を行い、ユーザーの要求する音量を調整する事が出来る。(3)は音量を下げる事以外は(2)と同様である。

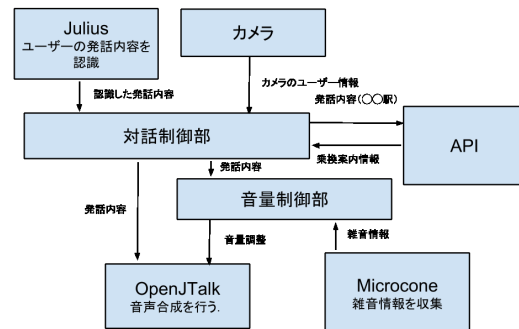


図1 システム構成図

Voice Guidance System for Noisy Environment: Automatic Adjustment of Volume When Reasked by User

by Kazuki Kogure, Hikaru Suzuki, Masahiro Yoshinaga and Tetsuro Kitahara (Nihon University)

### 3. 実験

#### 3.1 実験方法

雑音下で、本システムを用いてユーザーが聞き逃さずに必要な情報を得ることが出来るかを、実験によって検証した。駅のホームの雑音を再現した環境で、被験者は、指定された駅までの行き方(利用路線、乗換駅)、運賃、所要時間を本システムを用いて検索して、指定の解答用紙に記入する。この作業を30回繰り返す。ただし、30回の試行は、システム A (2.2節、2.3節の処理をどちらも導入しない) システム B (2.2節の処理のみを導入した) システム C (2.2節、2.3節の処理を両方導入した) を10回ずつ試行するものとし、各試行においてどのシステムが実行されているかは被験者には伝えないものとした。駅のホームの雑音は、7ch マイクロフォンアレイ Microcone を用いて東京都内の駅のホームで録音したものを、6台のスピーカーで再生することにより再現した。被験者は、21歳から24歳の正常な聴力を有する男性2名、女性1名である。

#### 3.2 実験結果

表1 左は正答数と総項目数を表し、右は正答率を表す。

被験者	システム		
	A	B	C
a	50/80 65%	72/84 86%	75/78 96%
b	64/80 80%	75/84 89%	88/92 96%
c	53/82 64%	76/88 86%	84/88 95%

表2 「え？」や「聞こえない」の回数、正答数、正答率

被験者	システム		
	発話した回数	正答数	正答率
a	7	6	85%
b	8	7	88%
c	2	1	50%
合計	17	14	82.3%

実験結果を表1に示す。これは、システム A~C の各10回の試行において記入すべき全項目(全利用路線、全乗換駅、運賃、所要時間)の個数と正答項目数、正答率である。被験者によって全項目数が異なるのは、APIが現在時刻に基づいて乗車方法を検索するため、実験を行うタイミングによって案内される乗車方法に差があるからである。

この結果から3人共、AとBの正答率を比べるとBが高く、BとCの正答率を比べるとCが高いことが分かる。Bの方が高いことから、レベルに基づく音量制御と発話延期は対話システムでも有効であることが分かった。また、Bに比べCの方が正答率が高いことから、聞き返しによってより確実にシステムの発話内容を聞き取れるようになった事がわかる。

表2は「え？」や「聞こえない」と発話をした回数、その時の正答数や正答率を表したものである。3人の合計の正答率は82.3%と高い正答率が出ており、この結果からユーザーの発話に基づく再度発話や音量調整は聞き取りやすさを向上している事が言える。この時の間違い17.7%の要因は3つ考

えられる。ユーザーが「え？」と聞き返すのが遅れ、システムによる次の発話が始まってしまったために、本来聞き直したい文の次の文が再発話されてしまった事である。2つ目はシステムが聞き返しの言葉を別の言葉として誤認識してしまったことである。3点目は再発話自体は上手くいったが、再発話のイントネーションがおかしいために、聞き取れなかったことである。

雑音が少なく発話内容が聞こえる環境であっても確認の為に「え？」と発話を行った場面もあった。例えば、どんどん音声が出て来ると、焦ってしまう為、「え？」と発話する事により、もう一度聞ける事によって冷静になれたという意見もあった。

発話延期に関しては、聞き取りやすさに関しては向上したと言えるが、スムーズさが失われたという意見もあった。その問題の解決案として、延期中にユーザーが「いいよ」などの発話を行ったら、延期をキャンセルするなどの工夫が必要かもしれない。また、システムがユーザーの発話と間違えて、発話を延期してしまう時もみられた。提案手法に関しては、「え？」や「聞こえない」という言葉以外にも聞き返しと考えられる言葉をピックアップし、認識できるようにしなければならない。

### 4. おわりに

本稿では、雑音環境下でも合成音声聞き取りやすい音声対話システムを目指し、雑音レベルに基づく音量制御や発話延期、ユーザーによる聞き返しに基づく再発話や音量調整を導入した。実験の結果、聞き取りやすさが向上する結果を得る事が出来たが、課題も多く存在する。今後はより雑音下での有効な対話システムを構築する為に、課題を解決していく。

謝辞 「駅すばあと web API」をご提供くださった(株)ヴァル研究所に感謝する。また、本研究の一部は SCAT 研究助成による助成を受けた。

### 参考文献

- 1) 角谷 直子 他：“カーナビの地名入力における誤認識時の訂正発話の分析と検出”，情報処理学会研究報告，SLP，pp.61-66，2001-7
- 2) 中村 哲 他：“実音響環境に頑健な音声認識を目指して”，電子情報通信学会，EA2002-12，pp.31-36，2002-4
- 3) 山辺 祐史 他：“駅コンコース案内放送の明瞭度向上手法”，RTRI REPORT，Vol.27，No.6，pp.49-54，2013
- 4) 水野 秀之 他：“多様性豊かな音声合成技術”，NTT 技術ジャーナル，pp.29-32，2013
- 5) 奥乃 博 他：“ロボット聴覚の現状と展望”，日本ロボット学会誌，Vol.28，No.1，pp.2-5，2010
- 6) 鈴木 光 他：“雑音環境下のための音声案内システム：周囲の雑音レベルに合わせた音量の自動調整”，情報処理学会第76回全国大会 6s-1，2014。