

## 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

鈴木 光<sup>†</sup> 吉永 眞宏<sup>†</sup> 小暮 計貴<sup>‡</sup> 北原 鉄朗<sup>†</sup><sup>†</sup> 日本大学文理学部情報システム解析学科<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

## 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るようになってきている<sup>1)</sup>。こうした音声対話システムが広まるにつれて、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。それにより、雑音に頑健な音声認識を行うための研究が数多くなされてきた<sup>2)</sup>。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>によるスピーカーク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。利用者が音声合成による発話を聞き取りづらく感じたときは手動で音量を調節する必要があるが、公共の場に設置されたシステムでは、音量を調節できないことも少なくない。特に視覚によって情報を得ることの難しい視覚障害者にとっては音量調節は必須になってくると考えられる。

本研究では、視覚障害者がシステムを利用することを想定して、周囲の雑音レベルに合わせて音量を自動的に調節するシステムを開発する。このシステムでは、7chのマイクロホンアレイから送られてきた音響信号から雑音レベルを抽出し、その雑音レベルに合わせてシステムの発話音量を自動調整する。しかし、発話の音量にも限界があり、電車の通過音などの突発的な雑音が発生した場合、音声案内は使用しているスピーカークが出力できる最大音量を超えてしまう事がある。そこで、雑音量があまりにも大きい場合は発話を停止し、一定時間後に再び発話をする。これにより、利用者は音声合成による発話を雑音下でも聞き取りやすくなると予想される。

## 2. システム構成

利用者は発話用マイクロホンの手前に位置して音声対話を行う。音声対話の内容は東京都内の乗り換え案内とし、利用者は「 駅から 駅まで行きたい」のように発話を行う。それに対してシステムは「駅すばあと Web API」を用いて最短経路を取得し、音声合成による発話を行うものとする。

音量調整は音声認識用とは別に用意したマイクロホンアレイを用いて行う。マイクロホンアレイから取得した雑音のレベルに対して、適宜システムの発話音量を変化させる。

## 2.1 雑音レベルの計測

雑音レベルは、マイクロホンアレイから得られる音響信号に基づいて推定する。現在の実装では、7chのマイクロホンアレイ「Microcone」からロボット聴覚オープンソフトウェア「HARK」<sup>5)</sup>を利用して約1秒間毎に音響信号を取得する。それに対してRMSを計算し、あらかじめ騒音計を用いて作成したRMSと騒音レベル(dB)の変換式に代入し、騒音レベル(dB)を算出する。

## 2.2 再生音量の変更

音量の変更は計算した雑音レベルを元に行う。システムの発話音量より周囲の雑音の方が大きい場合、雑音と同じ値まで発話音量を増幅する。また、周囲の雑音がシステムの発話音量より小さい場合は発話音量の縮小も行う。これにより環境に最適な発話音量の自動調節を実現する。

## 2.3 発話の延期

音量調節による雑音対策の他に発話の延期による対策も施す。これは、電車の警笛など最大音量を超える突発的な雑音に対応するためである。現在の実装では、100dBを超える雑音を感知した場合は発話を中断し、1秒の間隔をおいた後に発話を行うようになっている。

## 3. 実 験

提案手法によって利用者がシステムの発話を聞き取りやすくなったかどうかを実験する。

## 3.1 実験方法

実験は外からの騒音が入りにくい密室で行った。被験者は21歳から24歳の正常な聴力を有する男性3人、女性3人の計6人である。以下の流れで実験を行った。

- (1) 被験者の位置を中心に60度おきに6箇所にスピーカークを設置。被験者までの距離はそれぞれ4メートル。
- (2) スピーカークから雑音を再生する。
- (3) 雑音に慣れてもらう(30秒間)。
- (4) 音量調整を行わない音声案内を再生。
- (5) 提案手法の音声案内を再生。

## 3.2 実験内容

雑音はMicroconeを使って東京都内の駅のホームで録音したものを再生し、発話にはOpenJTalkで音声合成した女声を用いた。発話頻度は10秒に1回である。被験者は実験中に音声案内が聞こえたかどうかを6段階で評価する。それぞれの評価は以下のようにした。

- 1 まったく聞こえなかった
- 2 声は聞こえるが何を言っているのかわからない
- 3 聞き取れるが、大部分が聞き取りにくい
- 4 聞き取れるが、一部聞き取りにくい
- 5 聞き取れる
- 6 聞き取れるが、音量が大きすぎる

Voice Guidance System for Noisy Environment: Automatic Adjustment of Volume based on Noise Level  
by Hikaru Suzuki, Masahiro Yoshinaga, Kazuki Kogure (Nihon University) and Tetsuro Kitahara (Nihon University)

表 1 通常の音声案内．各セルの左は各回答番号を選んだ回数，右は其中で選択問題に正答した割合を表す．

被験者	回答番号					
	6	5	4	3	2	1
A	1 100%	10 90%	5 80%	3 67%	1 —	0 —
B	0 —	17 94%	2 100%	2 50%	0 —	0 —
C	0 —	12 100%	5 100%	3 100%	0 —	0 —
D	0 —	14 93%	1 100%	2 50%	4 —	0 —
E	0 —	16 94%	1 0%	0 —	0 —	3 —
F	0 —	15 93%	2 0%	2 0%	2 —	0 —
平均	0.2 100%	14 94%	2.7 75%	2 58%	1.2 —	0.5 —

表 2 提案手法による音声案内．各セルの左は各回答番号を選んだ回数，右は其中で選択問題に正答した割合を表す．

被験者	回答番号					
	6	5	4	3	2	1
A	1 100%	15 93%	5 0%	0 —	0 —	0 —
B	0 —	16 94%	5 100%	0 —	0 —	0 —
C	0 —	19 100%	2 100%	0 —	0 —	0 —
D	0 —	16 100%	4 100%	0 —	1 —	0 —
E	0 —	21 100%	0 —	0 —	0 —	0 —
F	0 —	20 95%	0 —	0 —	0 —	0 —
平均	0.2 100%	17.8 97%	2.7 69%	0 —	0.2 —	0 —

発話内容は、「 から までの料金は 円です。」とし，次のような 4 つの選択肢から聞き取ったものを選んでもらった。

- 「青砥から青井までの料金は 230 円です。」
- 「青砥から青井までの料金は 230 円です。」
- 「青井から青砥までの料金は 530 円です。」
- 「青井から青砥までの料金は 530 円です。」

音声案内を聞き取れるかどうかを確かめるのが目的なのでこの実験では「駅すばあと Web API」を用いず，また，元々料金を知っていることの効果を防ぐため，でたらめな料金を案内することとした．他に「音量は適切だったか?」，「発話の遅延は適切だったか?」など記述式の設問も設けた．

### 3.3 実験結果と考察

通常の音声案内を再生した結果を表 1，提案手法による音声案内の結果を表 2 に示す．

表は横軸が聞き取りやすさの違いによる 6 段階の評価，縦軸がそれぞれの被験者を表している．結果の左側がそれぞれの評価が記録された回数，右側が類似文章による選択問題の正答率となっている．それぞれの結果で，音声発話が正常に行われなかったデータは削除している．また，6 段階評価の [1]，[2] については聞き取れなかった評価のため，4 択問題による聞き取り判断は行わないものとする．

表 1 の結果から，音量調節を行わない場合は 6 人中 4 人が [1] または [2] の評価をしているため，聞き取りに困難を感じていると判断できる．平均に着目すると，すべての人が 1 回以上は聞き取りに困難を感じている結果となった．また，評価 5 を選択した回数は提案手法を用いた場合，通常再生より平均で 3.8 回増加し，6 人中 5 人で評価 5 を選択した回数が増加している．類似文章による選択問題の正答率も上昇していることから，提案手法によって聞き取りやすくなっていると言える．

表 3 再発話

被験者	発生発話数	評価 5	評価 4
A	6	5	1
B	6	5	1
C	6	6	0
D	7	6	1
E	7	7	0
F	6	5	1
平均	6.3	5.7	0.7

6 人の被験者のうち，被験者 B だけが提案手法を利用して適切な音量で聞き取れたと回答した回数が低下した．ここで被験者 B の類似文章による選択問題の正答率に注目すると聞き取り方が不安定な時の場合，通常の音声案内の正解率が 50 % なのに対して提案手法の正答率は 100 % となった．これは聞き取りやすく感じた回数は低下したが，実際に正しく聞き取れた回数は増加したことを意味している．

また，聞き取りが不安定な時の正答率の平均が低下した原因については以下のようなことが考えられる．提案手法が音量調節を行う際に周囲の環境音が静かなときに必要以上に再生音量を小さくしてしまう．実際に，記述回答では提案手法で周りが静かになったときに音量が音声案内の音量が小さくなりすぎたとの回答が複数得られた．

提案手法による延期が発生した発話の全てが評価 5 または評価 4 であった．その結果を表 3 に示す．左の列から「被験者」，「延期が発生した発話数」，「延期が発生した発話で評価 5 が記録された回数」，「延期が発生した発話で評価 4 が記録された回数」となっている．通常の音声案内と提案手法を比較すると，評価 5 の出現回数が平均 3.8 回程度増加し，評価 4 以下の出現回数が平均 3.5 回減少していることから，発話の延期が行われることによって聞き取りづらく感じていた発話が聞き取りやすくなったと考えられる．

## 4. おわりに

本稿では，合成音声の発話中に自動で音量調整や発話の延期を行うことでユーザーの聞き取りやすさに影響を与えることができるかを検証した．被験者実験を行った結果，提案手法によりユーザーの音声発話聞き取りを補助することが出来ることを確認した．しかし，提案手法では周囲が静かな時に音量が極端に小さくなりすぎるといった問題があるので，音量の最低値を引き上げることでこの問題を解決する予定である．

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: “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: “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: “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: “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: “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