

周囲の雑音やユーザーの聞き返しに基づいて音量調節を行う 音声対話システム

小暮 計貴^{1,a)} 吉永 眞宏^{2,b)} 鈴木 光^{2,c)} 北原 鉄朗^{2,d)}

概要: 実環境で音声対話システムを用いるには、雑音に対する対処が不可欠である。本研究では、システムの発話内容をより聞き取りやすくするために3つの手法を提案する。1つ目は、周囲の雑音に基づいて音量調節を行う手法。2つ目は電車の通過音など突発的な雑音に対処する為に、雑音に基づいて発話延期を行う手法。3つ目は、ユーザーの聞き返し（「え？」「何？」などの発話）に基づいて音量調節や再度発話を行う手法である。これらをシステムに導入する事により、システムの発話をユーザが聞き取れなかったために対話が成立しないなどの事態が避けられると期待される。それぞれの提案手法の有効性を確かめる為に二つの実験で評価を行い、共に聞き取りやすさが向上する結果を得られた。

キーワード: 対話システム, 雑音, 音量調節, 発話延期

Spoken Dialogue System with Volume Adjustment based on Ambient Noise Level and User's Response

Abstract: The robustness to ambient noise is indispensable in order to use a spoken dialogue system in real environments. In this paper, we propose three methods for allowing the user to hear the system's utterances content in noisy environments. The first method is volume adjustment based on the ambient noise level. The second method is utterance delay when a short but very loud noise occurs. The third method is volume adjustment and utterance repetition based on the user's response. These methods were evaluated by two experiments in order to confirm the effectiveness of the proposed methods.

Keywords: Dialogue System, Noise, Volume Adjustment, Utterance Delay

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 [1]。音声対話を円滑に行うためには、雑音環境下で音声認識を正確に行い、ユーザーに情報を正確に伝える必要がある。それを実現する為に、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 [2], [3]。

しかし、音声対話の研究では、システムが確実に音声対話を聞こえさせる為の研究は、あまり行われていない。明瞭に音声を取り取るための研究としては伊積ら [4] 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 [5] による、雑音下でも聞き取りやすい声がある事に着目し、雑音に強い周波数帯域の特徴を調べる事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、これらの研究では周囲の雑音レベルが変化した時に対応することができない。このように、周囲の雑音レベルを考慮してリアルタイムに音声を取り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

¹ 日本大学大学院 総合基礎科学研究科
Graduate School of Integrated Basic Sciences, Nihon University

² 日本大学文理学部
College of Humanities and Sciences, Nihon University

a) kogure@kthrlab.jp

b) yoshinaga@kthrlab.jp

c) hikaru@kthrlab.jp

d) kitahara@kthrlab.jp

なってくると考えられる。

本稿では、システムが確実に音声対話を聞こえさせる為の手法として、3つの手法を提案する。1つ目は、周囲の雑音レベルに合わせて音量を自動的に調節する手法である。この手法では、7chのマイクロホンアレー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節する。発話音量には限界がある為、自動調節のみでは電車の通過音など突発的な雑音に対処出来ない。そこで、2つ目の提案手法として、雑音レベルが大きい場合は発話を延期し、雑音レベルが小さくなった時に再び発話を行う手法を提案する。3つ目は、ユーザーの聞き返しに基づく再度発話である。雑音による音量調節を行ったとしても、ユーザーがシステムの発話を聞き逃してしまう可能性がある。この手法は、そのような場合に対する対処である。以上の3つの手法をシステムに導入する事で、雑音環境下での対話システムの発話内容の聞き取りやすさが向上すると考えられる。

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逃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。

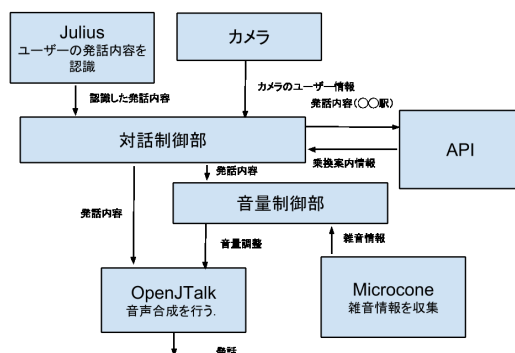


図1 システム構成図

システム構成図を図1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。音声認識には Julius[7]、音声合成には Open JTalk[8]を用いる。ユーザーが聞き返した時に使われる発話内容を Julius の辞書ファイルに登録する。

2.2 雑音レベルの計測

雑音レベルは、マイクロホンアレー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ー「Microcone」からロボット聴覚オープンソフトウェア「HARK」[6]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。

2.3 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。現在の実装では、64dB以下の雑音を感知した場合は発話音量の変更が適応される。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。

2.4 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、64dBを超える雑音を感知した場合は発話を中断し、周囲の雑音レベルが64dB以下になった時に、1秒の間隔をおいた後に発話を行うようになっている。また、発話の中断時は、「静かになるまでお待ちください。」と15秒毎に発話を繰り返す。

2.5 聞き返しなどに基づく再度発話と音量制御

ユーザーが聞き返した時に使われる発話内容を Julius の辞書ファイルに登録する。その発話内容をシステムが認識

した時に、システムは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。

3. 実験 1

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。ここでは、雑音レベルに基づく音量調節と発話延期の有効性のみ注目しており、聞き返しによる再度発話や音量調節は行っていない。

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 7 人、女性 7 人の計 14 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 箇所スピーカーを設置。被験者までの距離はそれぞれ 4 メートル。実験時の配置図を図 2 に示す。
- (2) スピーカーから雑音を再生する。本実験で用いた雑音の雑音レベルの変化を図 3 に示す。
- (3) 雑音に慣れてもらう (30 秒間)。
- (4) 音量調節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。

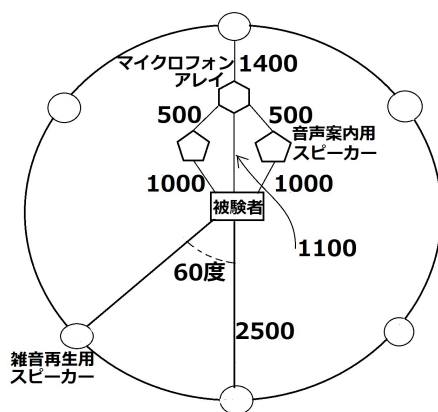


図 2 実験時の機材の配置 (単位は mm)

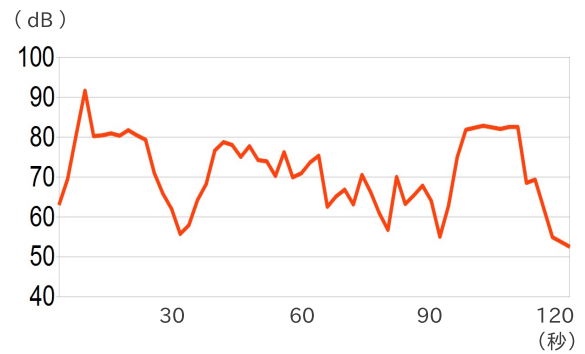


図 3 実験で用いた雑音の雑音レベルの変化

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 7 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
- 7 聞き取れるが、音量が非常に大きい

発話内容は、「 から までの料金は 円です。」や「 から 線に乗り、 に向かいます」など 5 種類用意し、毎回駅名や料金、手段を変えて案内を行い、次のような選択肢から聞き取ったものを選んでもらった。また、元々料金や行き方を知っていることの影響を防ぐため、でたらめな料金や行き方を案内することとした。

(例 1)

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

(例 2)

- 「亀有から西部多摩川線に乗り、亀戸へ向かいます」
- 「亀戸から西部多摩川線に乗り、亀有へ向かいます」
- 「亀有から西部多摩湖線に乗り、亀戸へ向かいます」

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影響を防ぐため、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

実験結果を図4と図5に示す。図4はユーザーの聞き取りやすさを示し、図5はユーザーの聞き取りの正答率を示す。図4の提案手法のユーザーの聞き取りやすさの割合に着目すると、ベースライン手法に比べ、評価5の割合が14%増加し、評価1と評価2の割合が6%減少する結果となった。この結果から、提案手法を導入する事により、雑音環境下での発話内容の聞き取りやすさが向上する事がわかる。図5に着目すると、提案手法の聞き取りの正答率はベースライン手法に比べ、正答率が8%増加している。この結果から、提案手法を導入する事で、発話内容をより聞き取れる事が分かる。

表1は発話延期発生回数とその時に選択された評価の数を被験者別に示している。この表に着目すると、提案手法による延期が行われた時の発話の聞き取りやすさの評価は3,4,5のみ選択されている事が分かる。また、全ての被験者が評価5を80%以上選んでおり、発話延期が行われた発話内容の聞き取りやすさは高い結果を得る事が出来た。評価3と評価4を選択された理由としては、延期後は雑音の大きさが静かになり、発話音量も極端に小さくなる為、多少聞き取りにくくなったと考えられる。

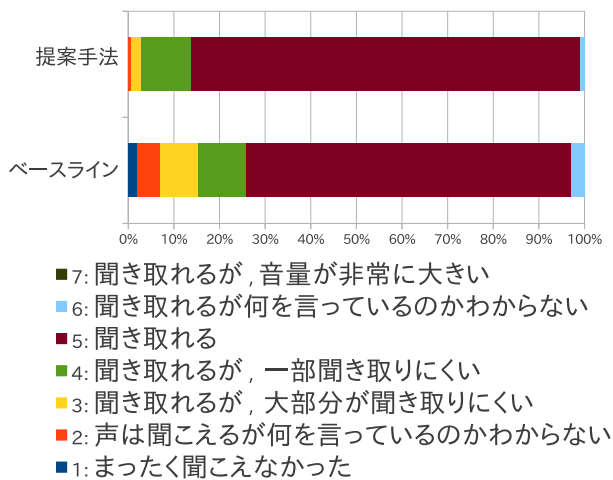


図4 実験結果1(ユーザーの聞き取りやすさ)

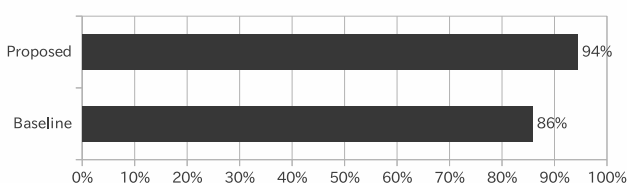


図5 実験結果2(ユーザーの聞き取りの正答率)

表1 発話延期発生回数とその評価

被験者	発生回数	評価7	評価6	評価5	評価4	評価3	評価2	評価1
A	6	0	0	5	1	0	0	0
B	6	0	0	5	1	0	0	0
C	6	0	0	6	0	0	0	0
D	7	0	0	6	1	0	0	0
E	7	0	0	7	0	0	0	0
F	6	0	0	5	1	0	0	0
G	6	0	0	6	0	0	0	0
H	6	0	0	6	0	0	0	0
I	5	0	1	4	0	0	0	0
J	5	0	0	5	0	0	0	0
K	5	0	0	5	0	0	0	0
L	5	0	0	4	0	1	0	0
M	5	0	0	4	1	0	0	0
N	5	0	0	4	1	0	0	0
平均	5.7	0	0.1	5.1	0.4	0.1	0	0

4. 実験2

4.1 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(2.3節2.5節のどちらの処理も導入しない)システムB(2.3節、2.4節の処理のみを導入した)システムC(全ての処理を導入した)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。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。実験環境に関しては、実験1と同様である。

4.2 実験結果

表2 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3 「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2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の方が高い事から, 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事から, 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3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節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, どんどん音声が流れて来ると, 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である。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。

5. おわりに

本稿では, 雑音環境下でも合成音声を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節を導入した。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。大きな課題としては2つある。1つ目は, ユーザーの発話音と雑音が区別出来ない為, ユーザーの発話で音量調節や発話延期が行われる。2つ目はユーザーの発話の聞き返しの意図が聞こ

えない時のみを想定しているシステムとなっている為, 聞き返しの意図を考慮しなければならない。例えば, 「え?」というユーザーの聞き返しでは, 聞こえないという意図で発した言葉とは限らない。対話内容が理解できなかったり, 目的地の駅名が正しく検索されていない為に発した可能性も考えられる。システムには, そのようなユーザーの意図に応えられるアクションを行う事が求められる。今後はこのような課題を解決し, より雑音下での有効な対話システムを構築する。

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, 北岡 競英, 中川 聖一: “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] T.Ishi, S.Matsuda, T.Kanda, T.Jitsuhiro, H.Ishiguro, S.Nakamura & N.Hagita (2008). A Robust Speech Recognition System for Communication Robots in Noisy Environments. *IEEE Transactions on Robotics*, 3, 759-763.
- [3] 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
- [4] 伊積 康彦, 石突 光隆: “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
- [5] 水野 秀之, 中嶋 秀治, 井島 勇佑, 神山 補相名, 村上 博子: “多様性豊かな音声合成技術”, NTT 技術ジャーナル, pp.29-32, 2013
- [6] 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
- [7] A. Lee & T. Kawahara (2009). Recent Development of Open-Source Speech Recognition Engine Julius. *APSIPA ASC 2009*, pp.131-137
- [8] <http://open-jtalk.sourceforge.net/>