5P-04

周囲の雑音に基づき音量調節を行う音声対話システム: セミブラインド音源分離の導入の検討

小暮 計貴† 北原 鉄朗‡

† 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 ‡ 日本大学文理学部情報科学科

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[1]. 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[3] 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[4] による雑音に強い周波数を利用し, 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.

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[5]. これにより, 雑音下でも合成音声を聞けるようになる. また, 実際の対話システムにこのシステムを組み込んで, 実験をする事により, 対話システムでも有効である事を確認した. しかし, このシステムは対話システムを利用するユーザーの音声やシステムの音声も雑音として処理してしまう問題がある. 本稿では, セミブラインド音源分離を利用し, この問題を解決する事を検討する.

2. システム構成

2.1 対話システムの概要

本システムの音声対話のシステム構成図を図 1 に示す. 本システムは, 特定の駅のプラットホ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.

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開始される. 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」

本システムは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「駅すぱあと Web API」を用いて検索する. 乗り換え案内情報が複数文の場合は, 一文発話した後にユーザーが「はい」と発言を行うか, 3 秒間経つと, 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.

本システムの特徴的な点は, 周囲の雑音の音量に合わせて合成音声の音量を自動的に調節し, 合成音声の音量を十分に上げても聞き取りにくいほどの大きな雑音が発生したときには, 雑音が解消されるまで発話を延期するところである. これを実現するため, ユーザ発話用とは別にマイクロフォンアレイ (Microcone) が用意されており, これにより時々刻々と測定した雑音の音量に基づいて「音量制御部」による音量調整または「対話制御部」による発話延期が行われる.

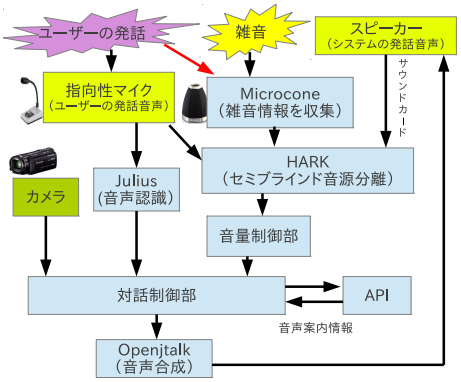

図 1 システム構成図

2.2 セミブラインド音源分離の導入

このシステムにおける重要な問題は, 雑音収集用のマイクロフォンアレイがユーザの発話やシステムの合成音声を拾い, 雑音の音量が大きいと判断してしまうことである. それにより, 周囲が静かであるにも関わらず, ユーザの発話の音量が大きいために発話延期が発生したり, 発話延期時に再生される「静かになるまでお待ちください」という合成音声が大きな音量で再生されるために発話延期が繰り返され, 対話が正常に進行しなくなるという問題がある.

本稿では, セミブラインド音源分離[6] を用いてこの問題の解決を図る. セミブラインド音源分離とは, 重畳されているいくつかの音源のうち, 1 つの原音源が既知であるときに使用できる音源分離手法である. 合成音声は原音源が既知であるし, ユーザ発話は指向性の強い発話用マイクで収録した音声を原音源とみなすことで, 雑音の音響信号から合成音声やユーザの音声を分離 (除去) する. その後, 残った雑音の音量を計測して合成音声の音量調節・発話延期を行うことで, ユーザの発話や合成音声の音量に影響されない処理を実現す

Spoken Dialogue System with Volume Adjustment based on Ambient Noise Level: A Consideration of Introduction of Semi-blind Sound Source Separation

by Kazuki Kogure and Tetsuro Kitahara (Nihon University)

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.

る．セミブラインド音源分離には，HARK[7] に搭載されているものを用いる．

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　(音量調節，発話延期システムを導入しない)
システム B　(音量調節，発話延期システムを導入する)
システム C　(音量調節，発話延期システムと 2．2 節を導入)

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．また，システム B とシステム C に関しては，発話延期回数を記録した．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，22 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　実験結果の正答率

	システム A	システム B	システム C
被験者 1	83%	95%	98%
被験者 2	85%	90%	96%
被験者 3	85%	96%	98%

表 2　発話延期回数

	システム B	システム C
被験者 1	4	0
被験者 2	3	0
被験者 3	5	1

実験結果を表 1 に示す．3 人の被験者共，システム A よりも B が高く，システム B よりも C が高い結果となった．システム A の正答率よりも B や C のシステムが高い事から，改めて音量調整システムや発話延期システムが雑音環境下において，有効である事が分かる．30 回の試行中におけるユーザー発話音声やシステム発話音声での発話延期回数を表 2 に示す．表 2 から，システム B のシステムに比べ，システム C の発話延期回数が大幅に減った事が分かる．また，システム C の発話延期回数が被験者 3 の 1 回のみという結果から，セミブラインド音源分離の手法は有効である事が分かる．

システム B と C の被験者実験中における雑音量 (dB) の変化をグラフにした結果を図 2 と図 3 に示す．この図の黒の四角で囲まれている値はユーザーの発話時の値を示し，赤で囲まれている値は雑音が流れている時の値を示している．黒の四角で囲まれている値に着目すると，システム B の発話時の音量が約 66dB から 69dB に対し，システム C の結果は約 61dB から 63dB の音量となっている．この事からユーザーの音声が抑圧されている事が分かる．この結果から，セミブラインド音源分離によって，ユーザーの発話音量を抑圧できている結果を得られた．赤の四角で囲まれている値に着目すると，システム B では，本システムの雑音測定の限界である 75dB も越えつづけた箇所があったのに対し，システム C ではこのような現象は見られなかった．これは，セミブラインド音源分離によって雑音も一緒に抑圧されたからと考える．

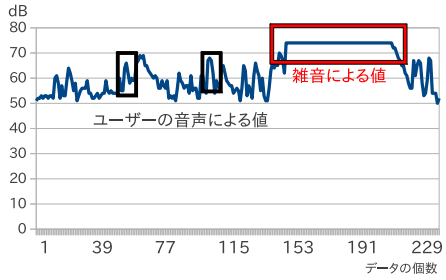

図 2　システム B の dB の変化

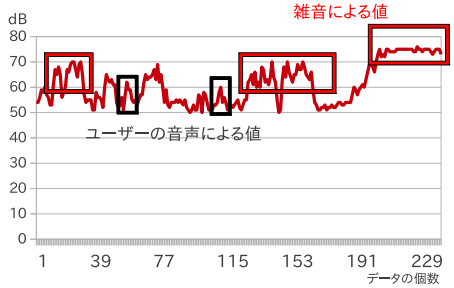

図 3　システム C の dB の変化

4. おわりに

本稿では，ユーザーの発話音声やシステムの発話音声が雑音として入力されてしまう問題を解決する為に，セミブラインド音源分離を導入する事を検討した．実験結果から，セミブラインド音源分離の導入は有効である事が分かった．しかし，本システムは決まった雑音環境下でその環境に適した閾値でしか，実験を行っていない．今後は，様々な環境で意図した通りに動作する事を確認し，どの雑音環境下においても，使用できる対話システムを構築する必要がある．

謝辞　「駅すぱあと 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) 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
6) 浅野 太:“一般化固有値展開を用いたセミブラインド音源分離”，電子情報通信学会，EA2002-6，pp.31-36，2002-4
7) 中臺 一博 他:“ロボット聴覚オープンソースソフトウェア HARK の紹介”，第 15 回システムインテグレーション部門講演会，SI2014 ,pp.1712-1716，2014-12

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.