

令和元年6月24日現在

機関番号：13101

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00352

研究課題名(和文) 音声・環境音を早期から識別できる手法の開発による音環境認知支援システムの実現

研究課題名(英文) Prototyping of environment sound recognition assist system by developing a sound identification method which can output recognition results as early as possible every moment with streaming processing

研究代表者

岩城 護 (Mamoru, Iwaki)

新潟大学・自然科学系・准教授

研究者番号：20262595

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：(1) 室内生活環境音調査を研究協力者一人あたり1週間実施した。行動に影響を与えたとされる生活環境音の種類を分類した。それらを収録し、データベースを構築した。各収録音に対して聴取実験を行い、音の印象に対する評価を付した。
(2) 音が到着した直後から音認識を開始し残りの部分が到着するごとに認識の確信の度合いを更新していくことによる音認識法を提案し、感情音声と環境音に対してそれぞれ適用した。コンピュータプログラムによるシミュレーションの結果、所望の漸次的な認識が確認できた。
(3) 到来した音情報の提示法として骨導ヘッドフォンの利用について検討し、到来方向知覚のズレの発生とその補正法を示した。

研究成果の学術的意義や社会的意義

自然な音声インターフェースのにとって、非言語情報の自動認識は重要な課題である。この認識は音の聴取と同時進行しなければならない。既存の手法は全収録後に認識といったバッチ処理的であるという欠点がある。本研究の特色は、この要件を満たす漸次的な認識手法を開発することである。その原型は、時々刻々と得られる短時間フレーム毎の認識をベイズ更新によって取り込みそれまでの認識結果を逐次更新していくことである。聴覚障がい者の行動支援機器の開発や、会話の進行に応じて適切な相槌を打ったり、話し手の感情を察して次の発話に備えたりすることができる、人間にとって自然な音声インターフェースの開発などに応用できる。

研究成果の概要(英文)：(1) We conducted a one-week-long investigation of indoor environmental sounds for each research subject. Sounds which had an influence on the behavior after the sounds were selected and recorded to construct a sound database. The database was labelled by sound impression based on hearing experiment.

(2) We developed a sound recognition method which can output recognition results every moment and update degree of belief about the result as the following part of the sounds arrives, which was examined in both emotional speech recognition and environment sound recognition experiments. According to computer simulations, it was confirmed that the method had the intended property.

(3) We examined bone-conduction headphones as a presentation device for information of occurring sounds. Because perceived sounds direction through bone-conduction headphones had a tendency to slip off, we developed a correction method.

研究分野：音声聴覚情報処理

キーワード：音認識 漸次的な認識法 ベイズ推定 ベイズ更新 環境音データベース 感情音声 行動支援 骨導ヘッドフォン

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

音声聴覚によるコミュニケーションや環境認識は日常生活にとって重要な役割を果たしており、聴覚障がいが生じると途端に生活が不自由になる。同様に、もし機械・コンピュータの音声・音環境認識能力が不十分だったなら、機械・コンピュータは人間にとって自然なパートナーや支援機器となることは難しい。従来、自動音声認識の研究には人間と機械との間で音声によるインターフェースを提供することが期待されてきた。音声信号から知覚される情報は言語情報(書き取り可能)と非言語情報(書き取りが困難)とに大別されるが、従来の自動音声認識の研究では発話の書き取りが主たる目的だった。一方、人間は会話音声から話者性や感情も同時に知覚している。このおかげで、(A)人間は状況に応じて音声を理解したり間髪を入れずに円滑にコミュニケーションしたりすることができる。さらに、(B)人間は環境音から周囲の出来事に対して素早く反応することができ、顔を向けたり危険を回避したりすることができる。今後の自動音声認識研究では、これらの機能を実現し統合していくことが重要な課題のひとつである。このような技術が実現すれば、会話者や環境に適応した自然な音コミュニケーション装置や、聴覚機能を補う支援機器など、有用な機器の開発に応用できる。これまでに我々は、(A)と(B)に関して基礎技術の検討してきた。

2. 研究の目的

本研究計画では、音声や環境音からコミュニケーションに有効な情報を時々刻々と検出する方法を開発し、受聴開始の早期からの情報提供を可能とするコンピュータ-人間インターフェースやコミュニケーション支援機器開発に応用する。

(1) 音声の話者性、感情の漸次的な識別： 音声の漸次的な識別手法を開発する。それは、(a)音声信号が刻々と到来するのに応じて短時間での話者・感情認識(原始的な認識)結果を得て、(b)その系列に基づいて時々刻々と話者・感情認識に関する確信の度合いを更新する、というアプローチをとる。

音声情報処理では話者性や感情などの非言語情報を抽出し有効に利用しようとする試みが活発である。例えば、話者識別では、話し方には話者の違いがよく表れることから、音声の基本周波数(F0)パターンを分析することが多い。感情識別では、声の大きさ、F0変化パターン、話速などが分析されている。しかし、最大値、最小値、平均値などの特徴を使用しているため、既存の話者・感情自動識別手法では音声信号を全て受信してからでないと識別を始められなかったり、遅延が大きかったりといった弱点がある。人間は音声を聴き続けているうちに判断を訂正したり確信の度合い高めたりすることができるのに対して、既存の手法は基本的に全体の音声信号に対して最後に一つの識別結果が得られるのみである。これまでの手法に加えて、時々刻々と話者・感情を認識しそれを逐次更新していくという特徴に焦点を当てることによって、自然な音声コミュニケーションや行動支援情報を提供するための自動音声認識手法が開発できると考えられる。

(2) 環境音の漸次的な識別： (1)と同様の方法に基づいて、音が鳴り始めた早期から時々刻々と環境音が識別できるという特長をもった手法を開発する。また、聴取者の行動に結びつく環境音を調査し、その対応関係がラベル付けされたデータベースを作成する。

環境音の認識は、周囲の状況を把握しそれに応じた行動を起こすといった、実環境内での人間の行動にとって重要な要素である。「耳の不自由な人たちが感じているアンケート調査報告書」((福)聴力障害者情報文化センター、E & Cプロジェクト(現(財)共用品推進機構)、1995年9月)で明らかにされたように、日常生活における周囲の音環境に関する情報支援の必要性は高い。

(3) 音源情報の提示、PCへの実装： 音源方位は音源の区別にとって重要であるが、さらに行動支援においてはその方向へ意識が向きそして顔が向けられるということは重要な要素である。音源毎に情報を提示するための手法を開発し、タブレットPCや携帯端末を用いて音を収集しながら時々刻々と識別情報を提示する聴覚支援機器を試作する。

3. 研究の方法

(1) 音声・環境音収録と知覚属性のラベリング： 話者分析、感情分析のために音声を収録しデータベースを作成する。音環境分析のために1週間の間に行動のきっかけとなった室内環境音をアンケート調査、収録しデータベースを作成する。各音データに対して聴取実験を行い、知覚された属性も付与する。

(2) 音声から感じられる感情の漸次的な認識法の検討： 音声を短時間のフレーム列に分解し、各フレームから得られる音声特徴を用いることによって、音声から感じられる感情を、フレーム毎に推定する方法を開発する。新しいフレームが得られる毎に推定された感情に対する確信は増していくと考える。事前確率を更新する機構をもったベイズ推定によってこれを計算する。また、感情の主観評価の手法であるSAM(Self-Assessment Manikin)評定を参考にし、感情音声をValence(快適度)、Arousal(覚醒度)、Dominance(支配度)で評価し、音声波形から得

られる特徴からこれらの値を推定することによる感情推定を試みる。

(3) 環境音の漸次的な認識法の適用範囲拡大： 日常の行動に影響を与える室内環境音を対象として、(1)で開発した漸次的な音認識法を適用する。音声とは異なる発音機構から生じているため、着目すべき特徴の種類は多岐にわたる。着目すべき特徴量が生じる時間スケールは短いものから比較的長いものまでである。漸次的な認識法で生じる問題点を検討し改良する。

(4) 音源に関する情報提示： 音の情報を提示する方法を検討する。携帯端末や眼鏡型端末などを利用した方法、骨導ヘッドフォンによる方法などを検討する。骨導ヘッドフォンによる提示法には、伝音性難聴に限られるものの、視覚を占有することなく音到来とほぼ同時に情報提示できるという利点がある。

4. 研究成果

(0) 音声・環境音収録と知覚属性のラベリング： 生活環境における音環境分析のために、研究協力者(16人)に対して、1週間にわたる生活環境音調査を実施した。研究協力者が影響を受けていたと思われる生活環境音の種類を分類した。同意が得られた研究協力者(6人)からこれらの生活環境音を収録し、データベースを構築した。また各収録音に対して聴取実験を行い、音の印象に対する評価を付した。

(1) 音声の話者性、感情の漸次的な認識法の適用範囲拡大： 感情認識の適用範囲を拡大するために、有声音からなる単語に対する感情音声を対象として特徴分析を行った。研究協力者(18人)に対して、知覚された感情と情動パラメータ(Arousal, Valence, Dominance)を調査した。得られた回答に基づいて研究協力者が知覚した発話者の感情と発話者が意図していた感情との間に生じたずれの分析を試みた。また、聴取者ごとにAVD情動パラメータの分布に生じる違いの分析を試みた。感情音声間でのAVD値の違いや話者間での違いをAffine変換で表すことができる可能性が示された。

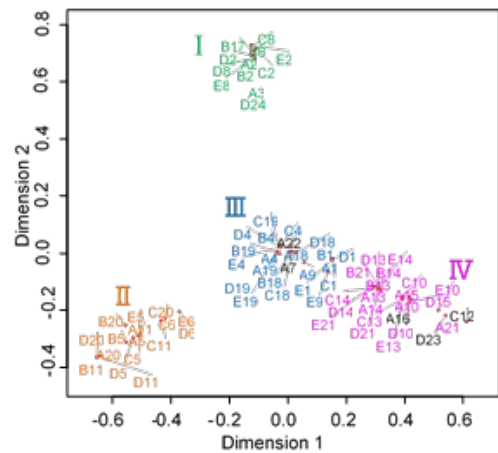


図1 行動に影響を与える室内環境音の散布図(多次元尺度構成法により2次元で表現した結果、4つの集団に分類されることが分かった。Mardia's fit measure 2 = 0.74)

(2) 環境音の漸次的な認識法の適用範囲拡大： RWCによる既存の環境音データベースのデータを対象として、漸次的な環境音認識法を提案しその認識性能を評価した。瞬時的に得られる特徴時系列を更新過程付きベイズ推定で識別した。音の開始から150ミリ秒後の認識結果は既にほぼ正解に到達しており、音源の変化に対しても追従できることが示された。7種類の環境音に対する平均認識率は約91%であった。

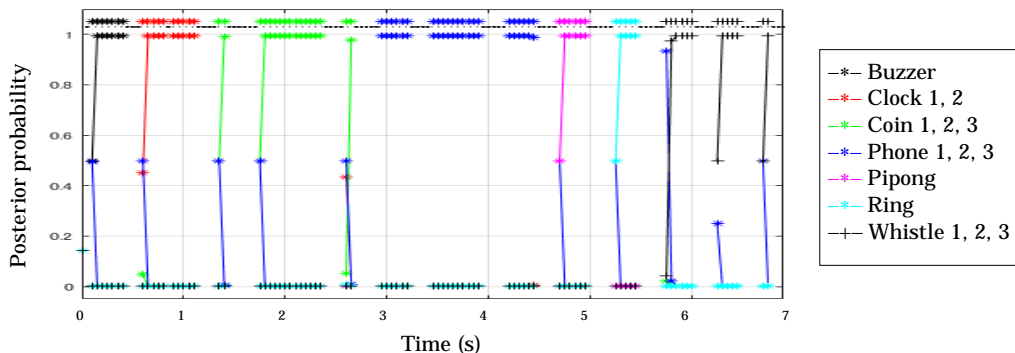


図2 音信号の入力に伴って時々刻々と変化する事後確率(新しい音が入力した際、暫くの間は事後確率が激しく変化するが、音が継続して到着する間に一つだけ大きな値の事後確率となり、音が漸次的に識別されることが確認できた。)

(3) 音源情報の提示方法の検討： 骨導ヘッドフォンを用いた音源方向提示方法を検討した。これまでは携帯端末による視覚的な提示方法を検討していたが、携帯端末を常に注視していなければならないという欠点があった。音源方向の提示に特化すれば、骨導ヘッドフォンによって音源方位を提示できる可能性がある。そこで、骨導ヘッドフォンによる音源方法知覚能力を

調査した。研究協力者（7人）に対して、気導ヘッドフォンで知覚される音源方向と骨導ヘッドフォンで知覚される音源方向には違いがあるものの、音源方向知覚が可能であることが分かった。さらには、骨導ヘッドフォンで知覚される音源方向は気導ヘッドフォンによるものよりも正面寄りになるため、ヘッドホンアンプにおいて補正することの必要性が示唆された。そこで、知覚方向を側方へ修正するために両耳間レベル差を強調する手法を提案しその性能を評価した。

また、骨導ヘッドフォンによる音源の順向性マスキングを調査した。研究協力者（6人）に対して、気導ヘッドフォンよりも順向性マスキングの影響時間が長いこと、周波数特性が異なることが分かった。これにより、骨導ヘッドフォンで報知音を鳴らす際には、断続間隔を広めに行なわなければならないという、指針が得られた。

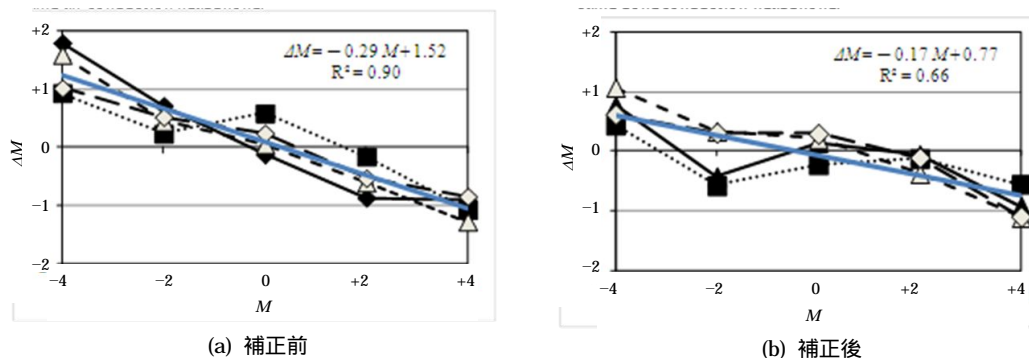


図3 骨導ヘッドフォンによる音源方向の提示における方向知覚特性のズレと補正効果（気導ヘッドフォンによる音の提示方向 M に対して骨導ヘッドフォンで知覚される方向にはズレ ΔM が生じており、骨導ヘッドフォンでは正面寄りに知覚方向がずれることが分かった。両耳間レベル差を強調することによって補正できることが分かった。）

(4) 音環境情報の提示による聴覚支援システムの開発：開発した手法をストリーミング処理によって実装した。特に、環境音の自動認識法では、構築した環境音データベースの音を4つのカテゴリーに分類して使用し、自動認識の平均認識率と音の到来に従って変化する推定音源の確率の様子を調査した。音の開始から750ミリ秒後の推定音源の確率はほぼ安定状態に到達しており、音源の変化に対しても追従できることが示された。一方、平均認識率は約56%であり、改善の余地が残った。特に、1秒に満たない短い音に対する認識率が悪いことから、改善すべき課題を明らかにすることができた。

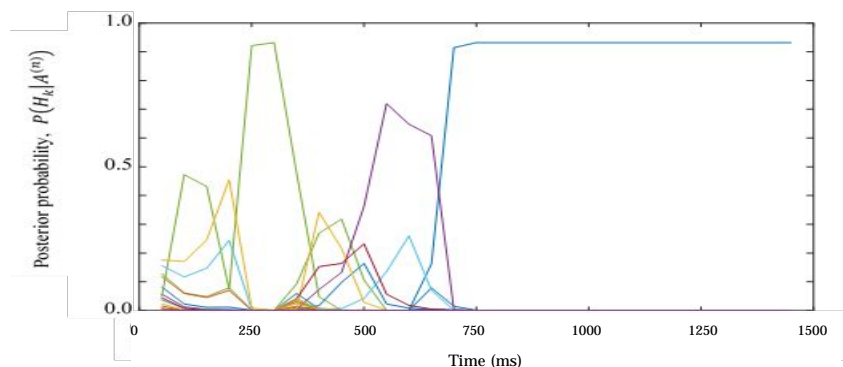


図4 音信号の入りに伴って時々刻々と変化する事後確率（行動に繋がる室内環境音に対する認識実験においても、音が漸次的に識別されることが確認できた。）

5. 主な発表論文等

〔雑誌論文〕(計1件)

Mamoru Iwaki, Yoshiki Chigira, "Emphasizing interaural level difference corrects the shift of perceived direction of sound for consumer-grade bone-conduction headphones," The Journal of the Acoustical Society of America, 査読有, Vol.140, Iss.4, p.3277, Nov. 2016.

DOI: <http://dx.doi.org/10.1121/1.4970412>

〔学会発表〕(計16件)

碓井峻, 岩城護, "骨導音の非同時マスキング効果における周波数依存性の計測", 日本音響学会 2019年春季研究発表会講演論文集, 3-P-25, pp.511-514, March 2019.

Mamoru Iwaki, Shun Nakayama, "Sound-Recognition Method for Helping Us Respond

Appropriately to Sounds in Daily Life," Proceedings of The 7th 2018 IEEE Global Conference on Consumer Electronics, 査読有, pp.680-682, Oct. 2018.

DOI:10.1109/GCCE.2018.8574822

碓井峻, 岩城護, "骨導音の非同時マスクング効果の計測 狭いノッチ幅の場合", 電子情報通信学会ソサイエティ大会講演論文集, A-5-4, p.24, Sept. 2018.

清水理博, 岩城護, "骨導ヘッドフォン聴取における純音の音源知覚方向のズレとILDとの関係(ILDをパラメータとした骨導ヘッドフォンにおける音源知覚方向のズレ計測)", 日本音響学会2018年春季研究発表会講演論文集, 2-P-38, pp.737-740, March 2018.

Mamoru Iwaki, Shun Nakayama, "Environmental Sound Database Focusing on Listener Impressions in Indoor Daily Life and Its Characteristics," Proceedings of The 6th 2017 IEEE Global Conference on Consumer Electronics, 査読有, pp.1-3, Oct. 2017.

DOI:10.1109/GCCE.2017.8229361

清水理博, 岩城護, "骨導ヘッドフォン聴取における純音の音源知覚方向のズレとITDとの関係", 電子情報通信学会信越支部大会講演論文集, P-11, p.156, Oct. 2017.

田村美友貴, 岩城護, "情動パラメータから成る感情空間における感情音声の分布について", 電子情報通信学会信越支部大会講演論文集, P-13, p.158, Oct. 2017.

荒木俊太郎, 岩城護, "ノッチ雑音に対する骨導音の傾向性マスクング効果の計測", 日本音響学会2017年春季研究発表会講演論文集, 2-Q-30, pp.707-710, March 2017.

中山俊, 岩城護, "人が自宅内で聴取する生活環境音の種類に関する調査", 日本音響学会2017年春季研究発表会講演論文集, 2-P-3, pp.143-144, March 2017.

吉田涼馬, 岩城護, "同時マスクングから推定される骨導音の知覚特性", 日本音響学会2017年春季研究発表会講演論文集, 2-Q-18, pp.677-680, March 2017.

清水理博, 岩城護, "ITDをパラメータとした骨導ヘッドフォンにおける音源方向のズレの計測", 日本音響学会2017年春季研究発表会講演論文集, 2-Q-22, pp.689-690, March 2017.

Mamoru Iwaki, Yoshiki Chigira, "Compensation of sound source direction perceived through consumer-grade bone-conduction headphones by modifying ILD and ITD," Proceedings of The 5th 2016 IEEE Global Conference on Consumer Electronics, pp.62-65, Oct. 2016.

DOI:10.1109/GCCE.2016.7800336

Mamoru Iwaki, Shun Nakayama, "A sound recognition method with incrementally taking sound stream into consideration," Proceedings of The 5th 2016 IEEE Global Conference on Consumer Electronics, pp.240-241, Oct. 2016.

DOI:10.1109/GCCE.2016.7800401

清水理博, 岩城護, "骨導ヘッドフォンにおける純音知覚方向のズレ特性の近似", 日本音響学会2016年秋季研究発表会講演論文集, 2-P-15, pp.587-590, Sept. 2016.

吉田涼馬, 岩城護, "ノッチ雑音同時マスクング法による気導音と骨導音に対する聴覚フィルタの推定", 日本音響学会2016年秋季研究発表会講演論文集, 2-P-16, pp.591-594, Sept. 2016.

中山俊, 岩城護, "聴取印象によるラベル付けに基づいた環境音データベース作成の試み", 日本音響学会2016年秋季研究発表会講演論文集, 2-Q-2, pp.41-42, Sept. 2016.

[その他]

ホームページ

<http://bsp.eng.niigata-u.ac.jp/personal/iwaki/16K00352/>

6. 研究組織

(1)研究分担者

なし

(2)研究協力者

なし

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。