

浴室空間でのスマートサービス提供に向けた音声認識の予備的評価

定光 郁斗[†] 陳 思楠[†] 中村 匡秀^{†,††} 佐伯 幸郎^{†††}

[†] 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 理化学研究所・革新知能統合研究センター 〒103-0027 東京都中央区日本橋 1-4-1

^{†††} 高知工科大学 〒782-8502 高知県香美市土佐山田町宮ノ口 185

E-mail: †ikuto.sadamitsu@gmail.com, ††chensinan@gold.kobe-u.ac.jp, †††masa-n@cmds.kobe-u.ac.jp,
††††saiki.sachio@kochi-tech.ac.jp

あらまし 近年、音声認識を利用したスマートサービスが増加しており、スマートフォンと連携して生活の質を向上させている。我々の研究グループでは、スマートフォンと給湯器を連携し、付加価値の高い音声認識を利用したサービスの検討を、給湯器メーカーであるノーリツと共に行っている。その中で、浴室環境での音声認識の実用性が不明であることが問題となった。本研究では浴室空間で様々な浴室環境を考慮した音声認識実験を行い、浴室での音声認識の実用性を評価した。環境の膨大な組み合わせ数、肉声での実験に要する過度な労力、音声認識精度の目視による評価の不十分さといった課題があったが、直交表を利用した組み合わせの数の削減、スピーカー音声での実験による効率化、分散分析での評価・検定といったアプローチで大学構内に再現した浴室環境で実験を行った。実験結果から、浴室環境を考慮すれば音声認識は実用可能であり、サービスとして検討可能であると考えられる。

キーワード 音声認識, 浴室環境, 直交表, 分散分析

Preliminary Evaluation of Speech Recognition for Smart Service Provision in Bathroom Space

Ikuto SADAMITSU[†], Sinan CHEN[†], Masahide NAKAMURA^{†,††}, and Sachio SAIKI^{†††}

[†] Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{††} Riken AIP 1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 Japan

^{†††} Kochi University of Technology, 185 Miyanokuchi, Tosayamada, Kami City, Kochi 782-8502, JAPAN

E-mail: †ikuto.sadamitsu@gmail.com, ††chensinan@gold.kobe-u.ac.jp, †††masa-n@cmds.kobe-u.ac.jp,
††††saiki.sachio@kochi-tech.ac.jp

Abstract In recent years, smart services using voice recognition are increasing, and they are improving the quality of life by linking with smartphones. Our research group is working with Noritz, a manufacturer of water heaters, to examine services that use voice recognition with high added value by linking smartphones and water heaters. Among them, it became a problem that the practicality of speech recognition in the bathroom environment was unclear. In this study, we conducted speech recognition experiments considering various bathroom environments in a bathroom and evaluated the practicality of speech recognition in bathrooms. There were problems with the enormous number of combinations of environments, the excessive effort required for experiments with human voice, and insufficient visual evaluation of speech recognition accuracy. Experiments were conducted in a bathroom environment reproduced in a university campus by approaches such as efficiency improvement by experiments and evaluation and testing by analysis of variance. From the experimental results, it is considered that speech recognition is practical and can be considered as a service if the bathroom environment is taken into account.

Key words Voice Recognition, Bathroom Environment, Orthogonal Table, Analysis of Variance

1. はじめに

近年、音声認識を利用したスマートサービスが増加している。それらはスマートフォンと連携し、利用者が音声で命令を出すと、命令に従い解決策を実行する機能で、人々の生活の質を向上させている。我々の研究グループでは、給湯器メーカーであるノーリツと共同研究を行っており、アプリ操作による命令や対話から浴室環境を個人に適したものにしてくれる、付加価値の高い音声認識を利用したサービスを検討している。研究開発の中で実際のユースケースを想定した**浴室環境での音声認識**は可能なのか不明であることが問題となった。このような背景のもと、浴室に依存した環境の中で、音声認識率を確かめる実験を行う。本研究では、浴室空間で様々な環境要因を考慮した音声認識実験を行い、浴室での**音声認識の実用性を評価**することを目的とする。また、どういった環境要因が認識率に影響を与えており、どのような環境下であれば音声認識が実用に耐え得るのかを明らかにする。上記の実験を行う過程で**環境の膨大な組み合わせ数**、**肉声での実験に要する過度な労力**、**音声認識精度の目視による評価の不十分さ**が課題として出てきた。これらの課題を受け、提案手法では3つのアプローチを行う。

A1: 直交表を活用した実験計画

浴室環境における音声認識に影響を与え得るデバイス、湯はり/状態(シャワー)、防水パック、換気扇、距離の5つの要因を洗い出したが、これら全ての組み合わせを考慮して実験を行うと、膨大な試行回数になってしまうため、実験の試行回数を効果的かつ最小限にする実験計画のツールである直交表を活用し、8種類の環境を設定した。

A2: 録音音声を活用した実験手法の設計

18種類の入力内容を設定し、1回ずつ発声する時間を測定したところ、1分強の時間がかかった。これを8種類の環境で試行回数を重ねるのは、被験者に身体的にも精神的にも過度な労力が伴うと判断し、今回は録音した音声をスピーカーで流すことによって効率化させる。

A3: 分散分析を用いた評価・検定

実験結果のデータの平均値を単なる目視で比較すると差があるかどうかの明確な基準がない。そこで評価する指標として分散分析を用いた検定を行う。

以上のアプローチのもと、大学構内で再現した浴室環境で実験を行った。まずは予備実験として同じ被験者で10回の再生を行ったときの認識率のばらつきを確かめ、本実験で被験者が増えた場合の再生回数の指標とする。実験結果より大きなばらつきはなかったため、本実験では再生回数を5回とする。

本実験では被験者10名(男性5名、女性5名)の声を5回ずつ再生し、得られた実験結果を分散分析で検定し、有意差を確かめる。実験結果よりデバイス、湯はり/状態(シャワー)、防水パック、距離には有意差があり、統計的に音声認識に影響を与えていると示すことができた。考察として**浴室環境を考慮すれば音声認識は実用可能であり、サービスとして検討可能**であると考えられる。

本論文の以降の構成は次のとおりである。2章では本論の準

備として音声認識を利用したスマートサービスと研究背景、技術的課題について述べる。3章では課題を受けての提案手法、アプローチを説明し、4章で予備実験と本実験の実験的評価を行う。5章では評価から考察を述べ、最語に6章でまとめとともに今後の課題を述べる。

2. 準備

2.1 近年の音声認識を利用したスマートサービスの増加

近年、**音声認識を利用したスマートサービス**が増加している[1]。例としてAppleのSiri、AmazonのAlexaなどが挙げられる[2][3]。これらはスマートフォンと連携したサービスで、スマートフォンから音声で命令を出すと、命令に従い解決策を実行してくれる。例えば、音声認識を通して「電気をつけて」と命令すると、部屋の電気を自動で点灯し、「7時に起こして」と命令すると、自動で7時にアラームを設定する。これらの音声認識サービスは人々の生活の質を向上させている。

2.2 浴室空間におけるスマートサービス

我々の研究グループでは給湯機メーカーであるノーリツと共同研究を行っており、**スマートフォンと連携可能な給湯器**の研究開発を行っている。音声での命令によるアプリ操作やデバイスとの対話を通じて浴室環境を個人に適した環境に変化させることは、浴室空間に高い付加価値を生み出すことに繋がる。そのため、浴室空間における音声認識を利用したサービス提供の実現が求められている。

研究開発の中で実際のユースケースを想定した**浴室環境での音声認識**は可能であるか不明であることが問題点として挙げられた。例として「iPhone利用者がシャワーを出しているとき」、「Android利用者が入浴中にデバイスを保護する防水パックを使用しているとき」などの場合、音声認識が浴室外と同様に可能かどうかは不明である。また、開発研究において、浴室環境を構成する様々な環境要因のうちどの要因の変化が認識率に影響を与えているのかのデータが必要である。これらの背景から、浴室に依存した環境の中で、音声認識率を確かめる実験を行う。

2.3 技術的課題

浴室環境での様々な環境要因を考慮した音声認識実験は容易ではない。本研究では以下の3つの課題に着目する。

課題 P1: 環境の膨大な組み合わせ数

浴室という場所に依存した環境は多数存在する。環境を設定する中で、環境要因とその水準の数に応じて組み合わせが指数的に増えてしまう。その膨大な環境の組み合わせを全て実験するのは過度な労力と時間を要する。組み合わせのパターンを考え、効率化することは必要不可欠である。

課題 P2: 肉声での実験に要する過度な労力

課題 P1 を解決し、組み合わせを減らせたとしても限界はある。音声入力するコマンドにも種類があることを考慮すると、試行回数と実験時間が被験者にとって肉体的・精神的に負担の大きい実験になってしまう。よって肉声で実験を行うのは困難と考えられる。

課題 P3: 音声認識精度の目視による評価の不十分さ

様々な組み合わせで実験を行い、データが取れたとしても、

適切な統計的手法を用いなければ音声認識精度に影響を与える環境要因を特定することは困難である。目視による比較ではなく、有意な差があるかどうかを統計的に示せる明確な基準が必要である。

3. 提案手法

3.1 研究の目的

本研究の目的は、浴室環境で様々な環境要因を考慮した音声認識実験を行い、浴室での音声認識の実用性を評価することである。実験を通して、浴室のどんな環境要因が認識率に影響を与えるか、どのような環境下で音声認識が実用に耐え得るのかを明らかにし、浴室環境での音声認識の実用性を評価する基礎データを創る。

3.2 アプローチ

3.1の目的を達成するために、実際の浴室環境においてスマートフォンの音声認識機能を活用した実験の評価を行う。

2.3の課題を受けて、アプローチを示す。

3.2.1 A1: 直交表を活用した実験計画

多様な利用者・ユースケースを想定し、浴室環境における音声認識に影響を与え得る5つの要因を洗い出した。

F1: スマートフォン機種の違い

ユースケースとしても多様なスマートフォンが使用されることが考えられるため、機種によって音声認識エンジンやマイク性能に違いがあるのか確かめる。

F2: 浴室の利用状態の違い

シャワー使用中と入浴中ではノイズによって認識率に違いがあるのか確かめる。

F3: スマートフォン防水パットの有無

実際のユースケースとして防水のためにスマートフォンにパットを被せるユーザーもいる。防水パットの遮音性によって認識率に違いがあるのか確かめる。

F4: 浴室換気扇によるノイズ

ユースケースとして浴室で換気扇をつけていることが考えられる。その微小なノイズは音声認識に影響するのか確かめる。

F5: 発話音源からスマートフォンまでの距離

スマートフォンと発話者の距離はどのくらいが適切なのか確かめる。

これら5つの環境の全ての組み合わせを試すと

$$F1 \times F2 \times F3 \times F4 \times F5(\text{通り})$$

の実験が必要となり、膨大な数になってしまう。

この課題を受けて、今回の実験では直交表を活用する。直交表とは実験の回数を効果的かつ最小限にする実験計画のツールである[4]。各要因(F1, F2など)の値(水準)が満遍なく実験に現れ、任意の2列の水準の組み合わせが網羅され、かつ均等に現れるように設定する。これによって実験回数を大幅に削減することが可能になる。今回の実験で作成した直交表を表1に示す。

表1 直交表

実験No.	F1:デバイス	F2:湯はり/状態	F3:防水パット	F4:換気扇	F5:距離
1	android1	あり/入浴中	あり	動作中	15cm
2	android1	なし/シャワー	なし	なし	30cm
3	android2	あり/入浴中	あり	なし	30cm
4	android2	なし/シャワー	なし	動作中	15cm
5	iphone1	あり/入浴中	なし	動作中	30cm
6	iphone1	なし/シャワー	あり	なし	15cm
7	iphone2	あり/入浴中	なし	なし	15cm
8	iphone2	なし/シャワー	あり	動作中	30cm

3.2.2 A2: 録音音声を活用した実験手法の設計

入力内容をコマンド(選択, 命令)と対話に分けて設定した。表2に示す。

表2 入力内容

選択
はい、いいえ
1番、2番
すすむ、もどる

命令
明日の天気教えて
温度を上げて
シャワーを出して
音楽を流して
入浴時間を測って
追い炊きして
風呂自動して
足し湯して
ハロー、メイちゃん(起動ワード)

対話
今日は仕事でとても疲れた
友達と食事をまたしたい
相変わらず右足がしびれる
スーパーに買い物に行かなきゃ
ラーメン行ったら、めっちゃめっちゃ待たされた
昨日は久しぶりによく寝た

今回の実験ではスマートサービス提供のため完全回答で認識とする。また、スマートフォンの変換エンジンによる誤変換は考慮しない。

2に示す入力内容を1回ずつ発声し、時間を測定したところ1回目は1分6秒、2回目は1分14秒の時間がかかった。これを表1の直交表の8種類の環境で試行するため、1度の実験に約10分弱の時間を要する。試行回数を重ねることを考慮すれば、被験者は長時間、浴室で繰り返し発声しなければならない。これは被験者にとって身体的にも精神的にも過度な労力を要すると考えられる。よって今回は録音した音声スピーカーで流すことによって効率化を図る。騒音計を活用し、スピーカーから流れる音声を日常会話と同じデシベル数に合わせる。浴室外でスピーカーから音声認識を行った場合、どのデバイスでも認識率は95%を超えていた。この数値なら音源を肉声からスピー

カーに変更することが可能であると考えられる。これによって被験者は1度声を録音するだけでよく、時間と労力が大幅に削減される。

3.2.3 A3: 分散分析を用いた評価・検定

水準の異なる2つのデータがあるとき、水準の違いによって認識率に差があるかどうかを調べたい。しかしながら、水準の平均値を単なる目視で比較すると差があるかどうかの明確な基準がない。そこで評価する指標として**分散分析を用いた検定**を行う。水準の違いによる認識率のばらつき（群間変動）と同じ水準での認識率のばらつき（群内変動）を比較する（図1参照）[5]。

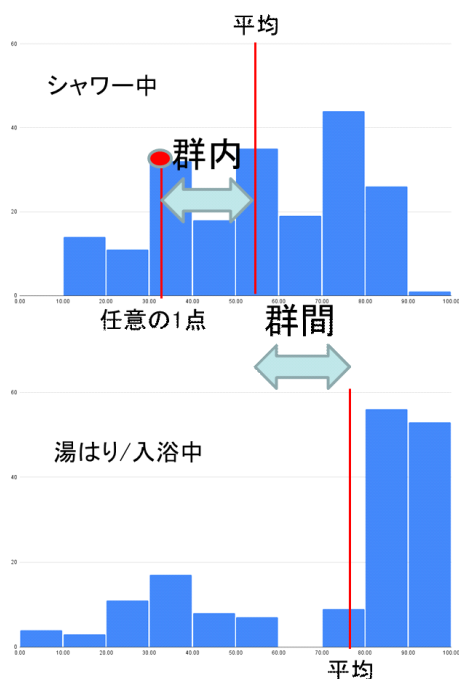


図1 群間変動と群内変動

要因の効果が大きければ、群間変動は群内変動に対して大きくなり、認識率に有意な差があると統計的に示すことができる。今回の実験ではエクセルのデータ分析ツールの一元配置分散分析を活用し、著者の定める有意水準と分析の結果得られたP値を比較することで統計的に有意な差があるかどうか確かめる[6]。

4. 実験的評価

4.1 実験環境

今回の実験の環境要因を示す。

- デバイス
- Android 端末 1: Google Pixel 7 (2022年10月13日発売)
- Android 端末 2: AQUOS SENSE6 (2021年11月4日発売)
- iPhone 端末 1: iPhone14 (2022年9月16日発売)
- iPhone 端末 2: iPhoneSE3 (2022年3月18日発売)

- 距離
- 15cm
- 30cm
- 防水パック
- 簡易防水ソフトケース (スマートフォン用)
- 騒音計
- サンワサプライ CHE-SD1
- 日常会話と同じ60デシベルに設定[7]
- スピーカー
- Anker SoundCore2

4.2 予備実験

まずは予備実験として同じ被験者で10回の再生を行ったときの認識率のばらつきを確かめ、本実験で被験者が増えた場合の再生回数の指標とする。実験手法として著者の声で3.2.2で設定した18種類の入力内容を1回録音する。3.2.1で設定した直交表の8種類の環境を大学構内で再現した浴室環境で構築し、各環境で録音した音声をスピーカーから再生し、スマートフォンで認識する実験を10回行う。実際に構築した浴室環境の写真を図2に示す。左がシャワー中の浴室であり、右が浴槽にお湯を張っている浴室である。

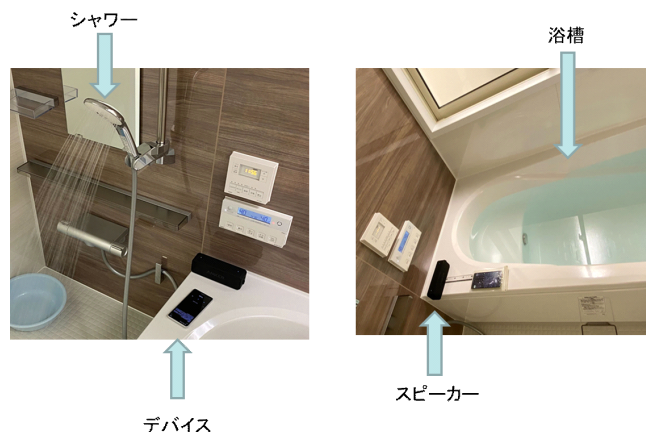


図2 浴室環境

予備実験で得られた結果を図3に示す。

図3より、同じ環境要因の中でも認識率に最大約25%の差があることが分かった（実験8）。また同じ実験環境の中で認識率が一定の実験もあった（実験7）。これらの結果より本実験では試行回数を1, 2回では不十分だが10回の必要はないと考えられるので5回とする。

4.3 本実験

本実験では被験者10名（男性5名、女性5名）の声で3.2.2で設定した18種類の入力内容を1回ずつ録音する。3.2.1で

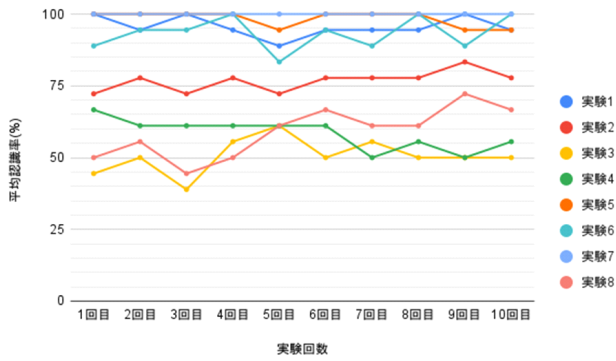


図3 予備実験結果

設定した直交表の8種類の環境を大学構内で再現した浴室環境で構築し、各環境で録音した音声スピーカーから再生し、スマートフォンで認識する実験を5回ずつ行う。得られた実験結果を分散分析で検定し、有意差を確かめる。本実験で得られた結果を箱ひげ図を図4に示す。

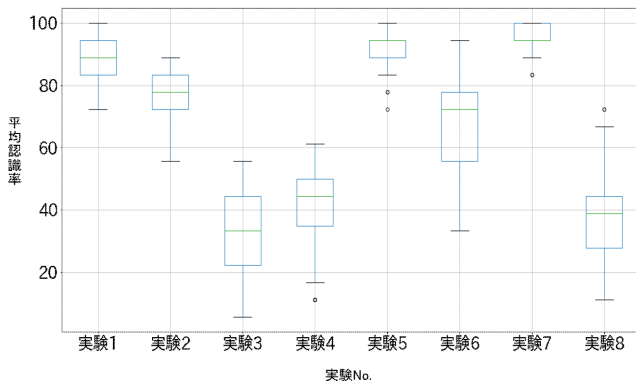


図4 本実験箱ひげ図

本実験の結果を分散分析で検定し得られたP値を表3示す。今回は有意水準を0.05に設定したので、P値が0.05を下回っていれば有意差がある。また大きく下回るほど有意差が大きく、その環境要因が音声認識に与える影響として大きいことになる。

表3 P 値

環境要因	P値
デバイス	1.31×10^{-54}
湯はり/状態	1.49×10^{-17}
防水パック	2.83×10^{-13}
換気扇	0.1042
距離	6.94×10^{-7}
性別	0.2998

得られたP値からデバイス、湯はり/状態、防水パック、距

離が有意水準0.05を下回っているの有意差があることが分かる。換気扇については有意水準を上回っているため有意差はないが、サンプルサイズをさらに大きくして、検出力を高めれば有意差が生じる可能性がある。しかしながら、スマートサービス提供のことを考えれば膨大なサンプルサイズで初めて有意差が出る程度であれば考慮しなくてよい。今回の実験は男性5名、女性5名で行ったので、性別でも分散分析を行ったが有意差はなく、性別は認識率に影響しないことが分かった。

iPhone 端末2において浴室環境が音声認識にとってプラスであるシャワーなし、防水パックなし、換気扇なし、距離が15cmの実験7の平均認識率が96%、浴室環境が全てマイナス要因の実験8の平均認識率が38.3%と多くの浴室環境が原因で認識率が低下していることが分かる。

最も大きく影響していたデバイスの個体差に着目し、4.1で発売日を見ると、モデルが新しいほど認識率が高いことが分かる。

5. 考 察

本実験から得られた結果から、浴室環境を考慮すれば音声認識は実用可能であり、サービスとして検討可能である。

最も影響の大きかったデバイスにおいて、最新モデルのAndroid 端末1の平均認識率が82.3%、型落ちモデルのAndroid 端末2の平均認識率が36.4%と大きな差はあったものの、Android 端末1と同時期に発売されたiPhone 端末1の平均認識率が79.4%と現在の最新モデルであれば、浴室環境での音声認識として十分な認識率があった。つまり現行の最新モデルであれば浴室環境での音声認識に耐えると考えられる。世間的な最新モデルの保有率は高くないが、年数が経てば現行の最新モデルも型落ちモデルとして扱われ、機種変更などにより世間的なスマートフォンのモデルも新しくなる。今後新たにリリースされるスマートフォンの音声認識機能が現行機種と比較して低下しないと仮定すると、将来的に、ユーザが所持するデバイスの音声認識精度の差は、浴室環境での音声認識サービスの実行に大きな影響を与えない程度に縮まると考えられる[8][9]。

次に影響の大きかった湯はり/状態において、シャワーなし/入浴中の平均認識率が77.1%、シャワー中の平均認識率が55.5%であった。浴室環境として大きな差があり、サービス検討の際はシャワーを出している間は音声認識を使用しないなど考慮することが望ましいが、ユースケースを考えた際にシャワーをしながら音声認識を使用するより、浴槽に浸かり入浴中に音声認識を使用することの方が多くと考えられるので、実用性としては十分な認識率といえる。

防水パックにおいて、被せていない場合の平均認識率が75.7%、被せている場合の平均認識率が56.9%であった。この浴室環境が最も考慮すべき点であると考えられる。防水機能のあるデバイスは増えてきているが、世間的にデバイスと水を接触させることに対する嫌悪感はある。実際のユースケースでも防水パックでデバイスを保護して浴室に持ち込むことも多い。

以上のことを踏まえて実験1に着目したい。防水パックを被せている場合の、有意差のない換気扇以外は全てプラス要因の

環境となっている実験1の平均認識率は88.1%であった。つまり、防水パックを被せていても、影響の大きいデバイスと湯はり/状態の環境がよければ、サービスとして申し分ない認識率であるということである。前述したようにデバイスと湯はり/状態はユースケースを考えれば問題ないので、浴室環境を考慮すれば音声認識は実用可能であり、サービスとして検討可能であると考えられる。

5.1 提案手法の利点

今回の実験の最も大きな利点は、**効率化**にあると考える。64種類の実験を直交表を活用し、8種類の実験で最小限かつ効果的にデータを得られた。音源を肉声からスピーカーに変えることで、被験者の肉体的・精神的負荷を抑えることに成功した。この2つの手法をとっても、浴室で予備実験と本実験を合わせて合計480回の音声を流し、データを得ることは容易ではなかった。どちらか1つでも欠けていたら、現実的に不可能な実験であった。分散分析を用いた評価・検定についても効率化したことによってとれた手法であるといえる。直交表を活用せずに、全ての組み合わせの実験を行っていけば、全ての環境要因の組み合わせのデータがあるがゆえに、数値を比較するだけで明確な差が分かるので、膨大な量のデータを愚直に比較するといった手法しかとれなかった。今回は効率化したことで、明確な基準を設けることが困難となり、分散分析を用いた評価・検定という手法をとることができた。スマートサービス提供に向けた音声認識の予備の評価としては十分なデータと評価方法から、統計的な結果が得られたといえる。

5.2 提案手法の限界

5.1で述べたように、今回の実験の最大の利点は効率化にあった。しかしながら、音声認識に影響を与え得る環境要因を増やしたとき、この手法には限界がある。例として新しく2水準の要因が5つ、4水準の要因が1つ出てきたとすると、直交表を活用しても今回の実験の倍の16種類の環境での実験が必要となる。今回の8種類の環境での実験でさえ、データを集めることは容易ではなかった。おそらく倍の量となると、音源から音声を流し続ける実験者の労力や、かかる時間のことも考えれば、新しい手法を取り入れることは必要不可欠だろう。また実際のユースケースとして、肉声でスピーカーのように毎回同じ音量・音程で話し続けることは困難であるので、肉声であることを考慮した新しい手法を取り入れることができれば、さらにユースケースに即した結果が得られる可能性がある。

6. まとめ

本研究では、浴室環境で様々な環境要因を考慮した音声認識実験を行い、浴室での音声認識の実用性を評価した。またその中で、浴室のどんな環境要因が認識率に影響を与えるか、どのような環境下で音声認識が実用に耐え得るのかを明らかにし、浴室空間での音声認識の実用性を評価する基礎データを創った。得られた結果の考察から、多くの環境要因が音声認識に影響を与えているが、浴室環境を考慮すれば音声認識は実用可能であると考えられる。

今後の課題は、今回の実験よりもユースケースに即した多く

の環境を評価できる手法を取り入れ、より精度の高い結果を得ることである。今回は予備の評価であったが、実際に浴室でのサービスを提供する際は、新しい手法を取り入れたさらに精度の高いデータが必要となる可能性がある。

謝辞 本研究の一部はJSPS科研費JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, 特別研究員奨励費22J13217, および、立石科学技術振興財団の研究助成を受けて行われている。

文 献

- [1] 辻野孝輔, 栄藤稔, 磯田佳徳, 飯塚真也, “実サービスにおける音声認識と自然言語インタフェース技術 (<特集> 音声対話システムの実用化に向けて),” 人工知能, vol.28, no.1, pp.75-81, 2013.
- [2] “Siri-Apple,” <https://www.apple.com/jp/siri/>. accessed January 23, 2022.
- [3] “Amazon-Alexa,” <https://developer.amazon.com/ja-JP/alexa>. accessed January 23, 2022.
- [4] 柏村孝義, 白鳥正樹, 于強, 国府田勲, “実験計画法, 数理計画法を用いた構造最適化,” 日本機械学会論文集 A 編, vol.62, no.601, pp.2180-2185, 1996.
- [5] “一元配置分散分析の計算方法,” <https://sigma-eye.com/2019/04/10/ichigen-bunsanbunseki/>. accessed January 23, 2023.
- [6] 小島隆矢, Excel で学ぶ共分散構造分析とグラフィカルモデリング, 株式会社 オーム社, 2003.
- [7] 白石君男, 神田幸彦, “日本語における会話音声の音圧レベル測定,” Audiology Japan, vol.53, no.3, pp.199-207, 2010.
- [8] 榎本慧太, “中古スマートフォンの価格形成要因分析,” Master’s thesis, 慶應義塾大学大学院経営管理研究科, 2017.
- [9] “[2022年] 日本のiphone・androidシェア率調査結果,” <https://simchange.jp/post-164095/>. accessed January 23, 2022.