

携帯端末の音声認識を利用した室内音響性能調査の試み

佐藤 彰治[※]，羽賀 葉^{※※}

Indoor Acoustic Performance Survey using Voice Recognition Software of Mobile Terminal

Shoji SATO , Konomi HAGA

Abstract- Evaluation of room acoustics requires measurement of monosyllabic intelligibility and word intelligibility by a large number of subjects. In addition, given the differences in hearing and word comprehension among individual subjects, standardization requires a significant amount of data. The purpose of this study is to conduct a word intelligibility test using voice recognition software of mobile terminals and examine its practicality. This report describes the reproducibility of the measured results.

Key Words : Acoustic Performance in Room, Mobile Terminals, Word Recognition Rate

1. はじめに

室内の音響特性（音声の聞き取りやすさ）は，一般に「単音節明瞭度試験」，「単語理解度試験」などの被験者実験により検討する。しかし，これらの実験では多数の被験者が必要になるため，その依頼や募集に時間や労力を要し，また聴力，単語理解力に個人差があるため標準化するのに多くのデータが必要となる。一方，昨今スマートフォン等の音声認識アプリの変換精度は長足の進歩がみられる。そこで，スマートフォンの音声認識機能の室内音響特性（聞き取りやすさ）調査への利用の可能性を検討するために，ここでは特に，音源からの距離などによる単語認識率の実測試験を行い，その「再現性」や「学習機能」の影響などについて考察する。

2. 学生を被験者とした単音節明瞭度試験

釧路高専（以下，本校）建築学分野では，毎年前期と後期に学生の工学実験（音環境）における演習で，「単音節明瞭度試験」を行っている。これは室の音響性能を調査するための実験の一つで，対象室各点における音声の聞き取りやすさを把握するためのものである。本演習では教壇部に設置したスピーカからの発生させた濁音，半濁音，拗音を含む単音節音声を学生が被験者となって聞き取り，記録する。学生は一回の試験で 100 音節聴取し，座席を移動してこれを 5 回繰り返す。各点（座席）5 回の聴取率

（正確に聞き取れた比率）のうち最高値と最低値を除く 3 回の平均値を各点の明瞭度とする。対象とする教室は，授業時間割や在籍人数によって変わるが，基本的には，音響的な配慮がなされた「大講義室」（階段教室）を使用している（図 1）。

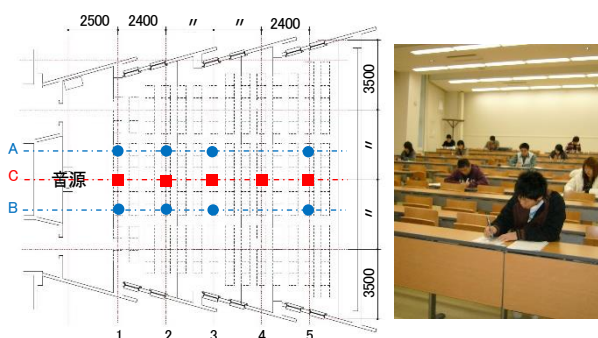


図 1 大講義室平面図と明瞭度試験風景

数年前より大講義室において行った試験結果を紹介する。図 1 の青丸の座席（A 列 4 点，B 列 4 点）で聴取した 5 シーズン分の結果を用いて，その平均明瞭度の分布（音源からの低下性）を図 2 左に示す。

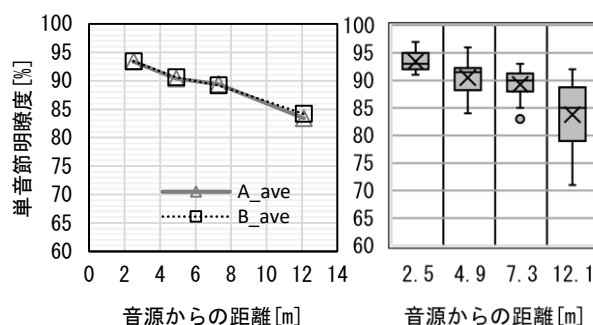


図 2 学生実験における平均単音節明瞭度

※ 釧路高専創造工学科

※※ 釧路高専専攻科

A 列と B 列は音源から見て対称の位置関係にあり、音源から後方部で若干の差異はあるものの、ほぼ同様の減衰分布となっている。AB 列のデータを一群（10 試験分）として扱った音源からの距離ごとの明瞭度分布（箱ひげ図）を図 2 右に示す。各点で若干のバラつきが見られ、特に音源から後方では四分位範囲で約 10%pt の明瞭度の差異がみられることから、強い再現性があるとはいえない結果となっている。

3. スマートフォンの音声認識アプリによる単語認識率（了解度）試験

3-1 発声音源の属性と使用単語（予備実験）

室内におけるスマートフォンの音声認識機能を使用した「音声の聞き取りやすさの調査」の試行に先立ち、実測方法、発声音源、音声認識アプリケーション（以下、スマホアプリ）などを検討するための予備実測を行った。まず、音源については国立情報科学研究所音声資源コンソーシアムより提供を受けた音声データベース¹⁴⁾を使用する。事前の確認よりスマホアプリによる単音節の認識は困難であることが分かっており、今回報告するすべての実験は上記データベースの中の「日本語単語」を使用した単語認識率（了解度）試験とした。また、今回使用する各機器は、特別に精度の高いものではなく廉価な市販品とし、スマホアプリは試験開始当初、比較的使用者の評価の高かったソフトを使用している。表 1 に各使用機器等を示す。

表 1 使用機器・音声認識アプリ

音源 PC	東芝 dynabook-R631
音源スピーカー	サンワサプライ USB スピーカー
スマホアプリ	Android: google 音声文字変換
	iPhone: Texter
精密騒音計	小野測器 LA-5560

実験用音声は、PowerPoint (Microsoft)により音声リストを再構成し、そのアニメーション機能によって外付けスピーカーを通して発生させることにした。この予備実験は本校建築製図室を使用することにした。図 3 に室平面、音源位置、測定位置（スマホ設置位置）を示す。

データベースに登録されている中の 100 単語を選定し音源の属性（発声者性別、発音間隔、音量）を変えてスマホアプリによる単語認識率の試験を繰り返し行った。その後、単語の親密度などを勘案して

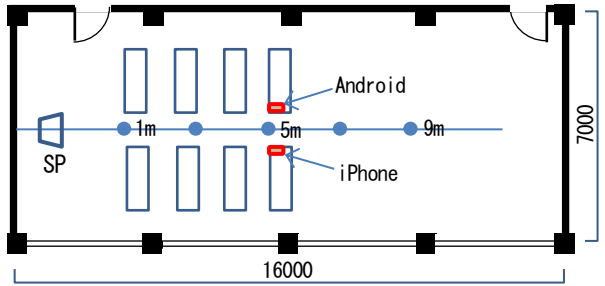


図 3 予備実験対象室（建築製図室）

20 単語に絞った。

この規模の室において暗騒音がない場合、被験者による実験ではどの場所でもほぼ 100%の単語了解度が得られるものと推察できる（被験者実験で確認済み）ため、スマホアプリの認識率ができるだけ高くなる基本条件について検討した。図 4 は音源の属性を変えて行った複数回の試験による 20 単語の平均認識率を示したものである。この結果から最終的に本実験では 10 単語を使用することとし、それらの発声順番を並べ替えた 5 つの音表を作成した（表 2）。この 50 単語リストを後述する大講義室における本実験に使用する。なお、音源の発声話者を認識率が高かった 40 代男性、発音間隔を 3sec とし、また講演時の話者の自然な大きさと判断される音量（SP100%, PC10%）を決めた。

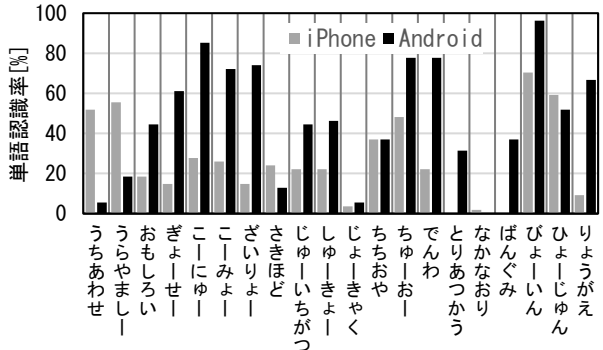


図 4 単語ごとの認識率

表 2 実験用単語リスト（50 単語）

	50 音昇順	50 音降順	50 音交互	ランダム 1	ランダム 2
1	ぎょーせー	ひょーじゅん	ぎょーせー	ぎょーせー	ひょーじゅん
2	こーにゅー	びょーいん	ひょーじゅん	こーみょー	でんわ
3	こーみょー	でんわ	こーにゅー	11 がつ	しゅーきょー
4	ざいりょー	ちゅーおー	びょーいん	ちゅーおー	ざいりょー
5	11 がつ	しゅーきょー	こーみょー	びょーいん	こーにゅー
6	しゅーきょー	11 がつ	でんわ	こーにゅー	びょーいん
7	ちゅーおー	ざいりょー	ざいりょー	ざいりょー	ちゅーおー
8	でんわ	こーみょー	ちゅーおー	しゅーきょー	11 がつ
9	びょーいん	こーにゅー	11 がつ	でんわ	こーみょー
10	ひょーじゅん	ぎょーせー	しゅーきょー	ひょーじゅん	ぎょーせー

3-2 大講義室における本試験の概要

室内での聞き取りやすさを定量的に、かつ標準化して評価する手法として確立するためには、同一条件における結果の「再現性」と繰り返し行ってもスマホアプリの認識率が向上しない、いわゆる「学習機能」の影響がないことが必要となる。ここで報告する本試験は主としてこれらの確認を目的として行っている。

認識率試験の対象室は前述した建築計画的に音響の配慮がなされている本校大講義室（前記載の図2）を使用する。認識率の測定点は同室の中央列（同図C列5点）とし、予備実験で決定した音源装置（属性）と音声リストを使用する。また、本試験の中で外付けマイクの接続について検討し、iPhoneアプリは接続した方が概ね高い認識率となり、反面Androidアプリはその認識率の変化は見られなかったため、その後の試験において、iPhone：外付けマイク接続、Android：同非接続で測定することとした。なお、これらの実測は2020年12月から翌年12月にかけて断続的に実施した。

3-3 認識率試験の結果と再現性

ここでは、設備、環境騒音を極力排除した中（室中央での等価騒音レベル30dB程度）での結果を示す。各測定点50単語のスマホによる聴取を、基本的に1日に2回繰り返している。なお、ここでの認識率とは全発声単語に対する正答数（正しく翻訳された個数）の割合を示すが、反応したが誤訳になった場合も正答の5割の加点として算出している。

図5に1日に2回連続して試験を行った際（4日分）の1回目と2回目の距離ごとの平均認識率を示す。前述した予備実験の結果を含めて、総じてiPhoneアプリはAndroidアプリの認識率を下回っており、特に音源からの距離が離れるにしたがってその差は顕著になるが、両アプリともに1回目と2回目の認識率に大きな差は見られなかった。また同様に経日変化も確認できなかったことから、繰り返し試験による認識率の向上傾向（学習機能）はなく、今後の試行において考慮する必要はないものと判断した。

都合6日間（10回分）の試験データを用いて音源距離ごとの認識率のバラつき（再現性）を箱ひげ図で検討する。図6-1はAndroidアプリによる結果であるが、音源から5m程度までの近距離においてはバラつき（分散）が僅少で高い再現性を示しているが、それより遠方の距離になると四分位範囲は大き

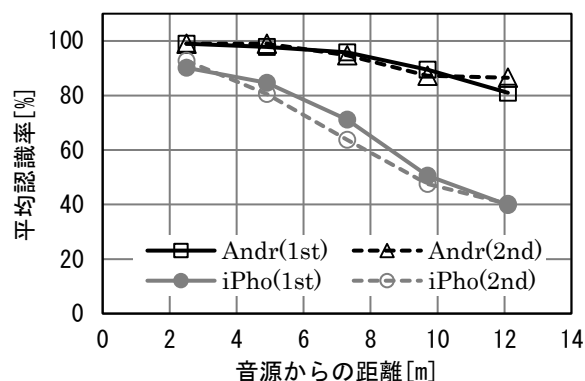


図5 平均単語認識率（1回目，2回目）

くなり、最遠方（12.1m）では同範囲が10%pt弱となっている。これは前述の学生被験者による単音節明瞭度試験と同様の傾向を示している。一方、図6-2はiPhoneアプリの結果であるが、距離にかかわらず大きなバラつきが見られ、四分位範囲で最大25%pt以上となる距離も見られた。また時折、何らかの不具合で認識アプリが反応しなくなる状況もあり、図中2.5mの外れ値はその影響によるものである。後日実施している会話文などの文節の認識率は両アプリともに高い認識率（正答率）を示しているが、「日本語単語」の認識（翻訳）では、アプリによる得手不得手があるものと推察される。

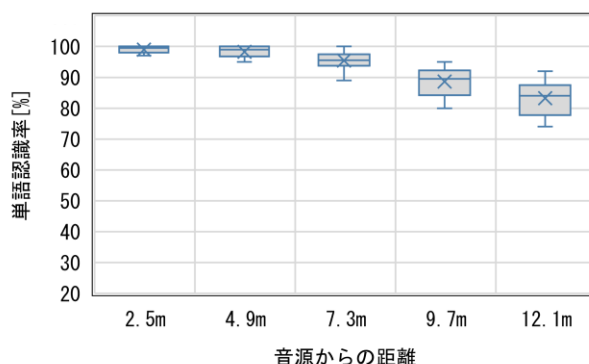


図6-1 認識率試験結果の分布（Android）

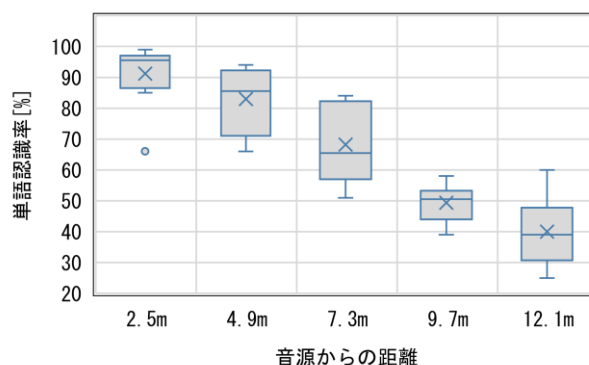


図6-2 認識率試験結果の分布（iPhone）

3-4 騒音下での認識率試験

実験過程で、試行的に設備騒音・環境騒音を加えた中での単語認識率試験を行った。設備騒音は対象室の天井に設置されている熱交換換気設備（8 台）による騒音を、環境騒音は側壁上部に設置されたスピーカーからバブルノイズ（パーティ騒音）を発生させた。なお対象室中央部で測定した等価騒音レベルは前者が 43dB、後者は 49dB であった。

図 7 に Android アプリによる認識率の結果を示す。音源からの距離と騒音の性質などからみて、比較的リーズナブルな結果と判断できる。

4. まとめと今後の展開

今回試行的に行った「スマホアプリによる音声認識試験」によって、以下のような知見が得られた。

- 1) スマートフォン機種やその音声認識アプリの種類によって、正しく変換（翻訳）されやすい単語が異なり、その認識率にも差異が発現する。
- 2) 同じく、機種によって試験ごとの認識率のバラつきに大きな差がみられる。
- 3) 繰り返し試験によるアプリの「学習機能」の影響は特にみられず、また今回は特に Android の音声認識アプリにおいて、ある程度高い再現性が確認された。
- 4) 騒音下における試験の結果、騒音レベルの増加による認識率の低下傾向からみて、この試験方法によって比較的妥当な結果が得られるものと推察された。

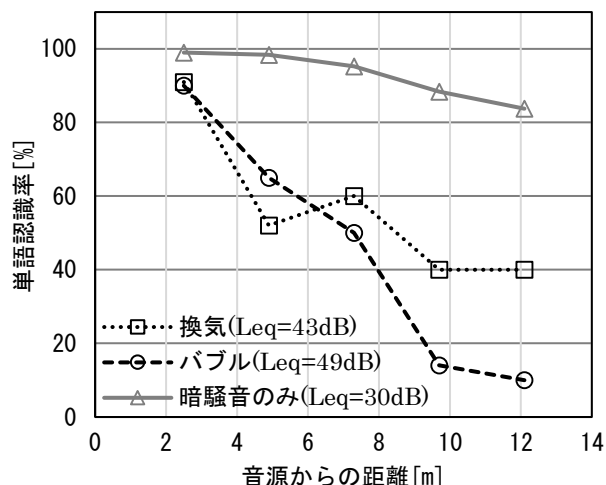


図 7 騒音下における認識率 (Android)

今回実施している試験件数は、まだ十分とは言えないものの、室の音響性能調査、特に「音声の聞き取りやすさ」の調査に利用し得る可能性が高いものと考えている。また、この手法は、室自体の音響性能調査のみならず、例えば昨今のコロナ禍の中で、様々な場面でみられるパーティーションを挟んだ会話などへの影響の検討などに応用できるものと考えており、現在その研究を進めている。

参考・引用文献

- [1] 国立情報科学研究所音声資源コンソーシアム，重点領域研究「音声言語」試験研究「音声 DB」連続音声データベース（改訂版），2015 年