博士論文

外乱環境における音声認識性能予測に関する研究
(Estimation of Speech Recognition Performance in Noisy-and-Reverberant Environments)

2015年3月

立命館大学大学院
情報理工学研究科
情報理工学専攻
博士課程後期課程

福森 隆寛

立命館大学審査博士論文

外乱環境における音声認識性能予測に関する研究

(Estimation of Speech Recognition Performance in Noisy-and-Reverberant Environments)

2015 年 3 月

March, 2015

立命館大学大学院情報理工学研究科

情報理工学専攻博士課程後期課程

Doctoral Program
in Advanced Information Science and Engineering
Graduate School of Information Science and Engineering
Ritsumeikan University

福森 隆寛

Takahiro Fukumori

研究指導教員：西浦 敬信 教授

Supervisor：Professor Takanobu Nishiura

本論文は立命館大学大学院情報理工学研究科に
博士 (工学) 授与の要件として提出した博士論文である.

福森 隆寛

審査委員：　主査　西浦 敬信 教授
副査　山下 洋一 教授
副査　福本 淳一 教授

外乱環境における音声認識性能予測に関する研究*

福森 隆寛

内容梗概

万人にとって使い勝手の良い理想的な情報機器の操作環境として，音声を利用して情報機器を操作するハンズフリー音声インタフェースが強く求められている．しかしながら，実環境において音声認識システムを利用すると，雑音や残響などの外乱が音声に混入することの影響を受けて音声認識性能が著しく劣化する．そのため，実際に音声インタフェースを利用する場合，事前に利用環境に存在する雑音や残響などの外乱の影響を音声認識システムに適応させる必要がある．

外乱の影響を音声認識システムに適応させるための関連手法として，外乱環境における音声認識性能の予測手法が注目されている．もし事前に音声認識性能を予測することができれば，予測結果に基づいて音声認識性能の改善手法を前処理等に反映させることが可能となり，常に最大限の音声認識性能を発揮させることが可能となる．

音声認識性能を予測することは，音声認識性能の改善に貢献できるだけでなく，これまでの音声認識評価に要するコストを大幅に削減できる側面も有する．従来の音声認識性能評価では，実際の利用環境において大量の音声データを収録することや，収録した音声データを認識評価することに膨大なコストが必要であった．そこで，雑音やインパルス応答などの外乱成分を少量収録するだけで音声認識性能を予測できれば，これまで必要だった大規模な音声収録や音声認識処理が省略されて，結果的に音声認識性能の評価コストを大幅に削減することができる．

本論文では，外乱環境においてコストをかけずに高精度に音声認識性能を予測する手法について検討する．具体的には，残響環境における音声認識性能の予測法の

*立命館大学大学院 情報理工学研究科 情報理工学専攻 博士論文.

確立した後，複数の外乱が混在する環境（雑音・残響環境）における音声認識性能の予測法の確立を目指す．

残響環境における音声認識性能の予測法では，これまで音声認識性能の予測としては不十分であった残響環境での音声認識性能の予測指標を提案する．本論文では，初期反射音と後続残響音の関係を表す室内音響指標の中でも特に Definition（D 値）に着目し，事前に様々な環境で計測した複数のインパルス応答を基に算出した D 値と音声認識性能の関係に基づいて残響指標を策定する．そして，策定した残響指標と音声認識性能の予測位置におけるインパルス応答に基づいて残響下における音声認識性能の予測を試みる．多種にわたる残響環境において，音声認識性能の予測評価実験を行った結果，提案手法の有効性を確認した．

複数の外乱が混在する環境における音声認識性能の予測法では，雑音と残響が音声認識性能に与える影響を予測するための指標を提案する．本論文では，雑音環境下における音声認識性能の予測指標の Perceputual Evaluation of Speech Quality（PESQ）と残響環境下における音声認識性能の予測指標の室内音響指標の D 値を組み合わせて，雑音・残響指標 Noisy-and-Reverberant Speech Recognition criteria with PESQ and Acoustic parameters（NRSR-PA）を策定する．そして，NRSR-PA を用いて音声認識性能の予測位置におけるインパルス応答，雑音，発話音声から音声認識性能の予測を試みる．評価実験の結果，従来の雑音指標・残響指標を個別に用いて音声認識性能を予測する手法よりも，NRSR-PA を用いることで頑健に音声認識性能を予測できることを確認した．

キーワード

音声認識，雑音，残響，音声認識性能の予測，PESQ，室内音響指標

Estimation of Speech Recognition Performance in Noisy-and-Reverberant Environments*

Takahiro Fukumori

Abstract

Hands-free speech interfaces are expected for an ideal environment that is easy for all users to operate information devices. The speech recognition performance of hands-free speech interfaces is, however, degraded due to noise and reverberation. To solve this problem, it is necessary to take the effects of noise and reverberation in the usage environment into consideration for developing a speech recognition system.

Performance estimation of speech recognition is one of the methods related to adapting noise and reverberation to the system. It is possible to always achieve a higher speech recognition performance by utilizing a suitable improved method based on the estimation results.

The estimation method contributes to not only improving the speech recognition performance but also reducing a lot of cost for large-scale recording and speech recognition. The conventional evaluation methods of the speech recognition performance require a huge cost for recording and recognizing a large amount of speech signals. If the performance can be estimated using an impulse response and noise measured in an evaluation environment, data volume and computation time will be significantly reduced because it is no longer necessary to record and recognize the speech signals.

In this doctoral dissertation, we proposed a method to accurately estimate the speech recognition performance in noisy and reverberant environment at lower cost.

*Doctoral Dissertation, Advanced Information Science and Engineering, Graduate School of Information Science and Engineering, Ritsumeikan University.

In particular, a method was first proposed so that the performance can be accurately estimated in reverberant environments. The method was then improved to estimate the performance in an environment that includes both noise and reverberation.

In order to estimate the speech recognition performance in reverberant environments, it is necessary to design a suitable measure for evaluating reverberant speech. We therefore proposed a method to estimate the performance in reverberant environments using the reverberant measure. Our method focused on early and late reflections on distant-talking speech recognition to determine the suitable measure. The measure was designed based on the relationship between the speech recognition performance and the ISO3382 acoustic parameters that expresses early and late reflections. The speech recognition performance was then obtained by using the designed measure and an impulse response in a position for the performance estimation. Evaluation experiments confirmed that the performance can be accurately and robustly estimated with the proposed measure.

It is indispensable to newly design a noisy and reverberant measure for estimating the speech recognition performance in both noisy and reverberant environments. We thus proposed the noisy and reverberant measure, which is referred to as "Noisy and reverberant speech recognition with perceptual evaluation of speech quality (PESQ) and acoustic parameters (NRSR-PA)". The NRSR-PA was designed using the relationships among the ISO3382 acoustic parameters which is a reverberant measure, the PESQ score which is a noisy measure, and the speech recognition performance. The performance was then estimated with the designed measure NRSR-PA in our evaluation experiments. Experimental evaluations demonstrated that the proposed measure is well suited for robustly estimating the performance in noisy and reverberant environments.

Keywords:

Automatic speech recognition, Noise, Reverberation, Estimation of speech recognition performance, Perceptual evaluation of speech quality, Acoustic parameters

目 次

図 目 次

表 目 次

第1章　序論

1.1. 研究背景と目的

情報機器の急速な発展に伴い，機器操作が大幅に複雑化しており，万人にとって使い勝手の良い操作環境が強く求められている．これまではキーボードとマウスが機器操作の基本であったが，近年のスマートホンの爆発的な普及によりタッチパネルを利用して操作する機会が急増してきた．ところが，情報機器に不慣れな高齢者や手足が不自由な身体障害者には，このようなタッチパネル操作が非常に困難であるのが現状である．

万人がタッチパネル操作を必要とせず，使い勝手の良い理想的な操作環境を実現するために，音声認識技術 [1, 2, 3, 4, 5, 6] を利用した情報機器の操作に多くの関心や注目が集まっている [7, 8, 9]．音声認識は音声に含まれている情報を機械的な手段で抽出する技術であり，ビデオや講義音声などから必要な情報を抽出する音声ドキュメント検索 [10, 11]，異なる言語を話す人々の円滑な会話を支援する音声翻訳 [12, 13]，音声を介して人と対話をしながら目的を遂行する音声対話システム [14, 15] などをはじめとする様々な利用シーンでの応用 [16, 17, 18, 19, 20] が期待されている．特に最近では，利用者がスマートホンなどの携帯端末に話しかけることで，音声認識技術によりタッチパネルを介さずに端末の基本機能（メール編集，アラーム設定，音楽再生など）を利用できるパーソナルアシスタント機能が音声インタフェースの飛躍的な発展を示すひとつの起爆剤となった．

現在，音声認識技術を用いたサービスが次々と普及しているが，マイクロホンを装着しない音声インタフェースは，図 1.1 に示す外乱要因によって音声認識性能が著しく低下するという問題がある．音声認識性能の低下原因として，実環境下で使用者がマイクロホンから離れて発話することで目的音声に雑音や残響等が混入するこ

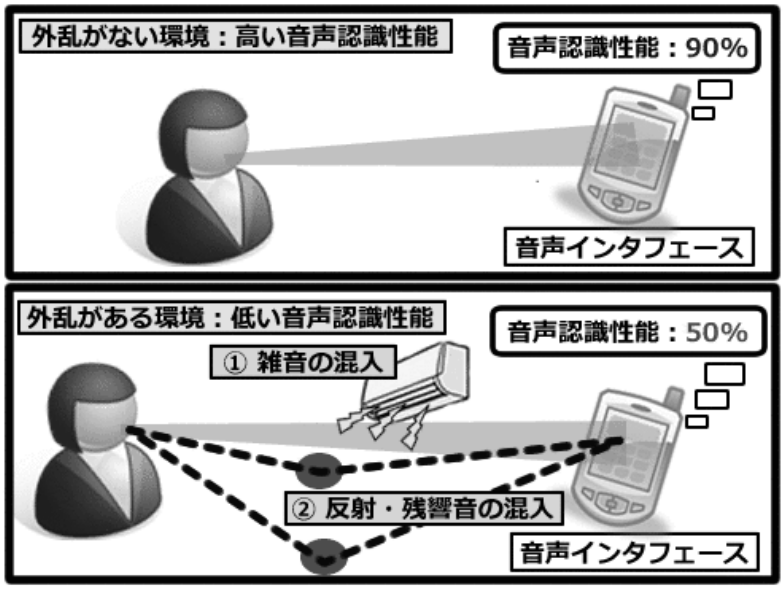


図 1.1 雑音と残響の混入による音声認識性能の低下

とが挙げられる [21, 22, 23, 24]．これまでに実環境下で音声認識性能を向上させるために数多くの雑音対策 [25, 26, 27, 28, 29, 30, 31] や残響対策 [32, 33, 34, 35, 36, 37] が提案されている．実際に音声インタフェースを利用する場合，事前にこれらの性能改善手法を適切に講じて，利用環境に存在する外乱の影響を音声認識システムに適応させる必要がある．

外乱の影響を音声認識システムに適応させるための関連手法として，外乱環境における音声認識性能の予測手法が注目されている．図 1.2 に音声認識性能の予測から改善までの流れを示す．もし事前に音声認識性能を予測することができれば，予測結果に基づいて外乱対策を音声認識システムの前処理等に適切に反映させることで，音声認識性能の劣化を未然に防ぐことができ，結果的に利用環境で音声認識性能を最大限に発揮できるようになる．たとえば，複数の外乱対策に対する音声認識性能を予測・比較することで，利用環境に適切な外乱対策を利用者に推奨することができる．

音声認識性能を予測することは，音声認識性能の改善に貢献できるだけでなく，

これまでの音声認識評価に要するコストを大幅に削減できると考えられる．これまで音声認識システムを導入する環境において，音声認識性能を評価するには，事前にその環境で収録した音声データを用いて音声認識実験を行うことが多かった [38]．しかしながら，実際の利用環境において大量の音声データを収録することや，収録した音声データを認識評価することは膨大なコストが必要となる上に，収録従事者や被験者の負担も大きくなる．そこで雑音やインパルス応答などの外乱成分を少量収録するだけで音声認識性能を予測することができれば，これまで必要だった大規模な音声収録や音声認識処理が省略されて，結果的に音声認識性能の評価コストを大幅に削減できる．

本論文では，外乱環境においてコストをかけずに音声認識性能を高精度に予測する手法について検討する．具体的には，残響環境における音声認識性能の予測法の確立した後，複数の外乱が混在する環境（雑音・残響環境）における音声認識性能の予測法の確立を目指す．

残響環境における音声認識性能の予測法では，これまで音声認識性能の予測としては不十分であった残響環境での音声認識性能の予測指標を提案する．本論文では，初期反射音と後続残響音の関係を表す室内音響指標の D 値に着目し，事前に様々な環境で複数箇所計測したインパルス応答を基に算出した D 値と音声認識性能の関係に基づいて残響指標を策定する．そして，策定した残響指標と音声認識性能の予測位置におけるインパルス応答に基づいて残響下における音声認識性能の予測を試みる．

複数の外乱が混在する環境における音声認識性能の予測法では，雑音と残響が音声認識性能に与える影響を予測するための指標を提案する．本論文では，雑音環境下における音声認識性能の予測指標の Perceputual Evaluation of Speech Quality（PESQ）と残響環境下における音声認識性能の予測指標の室内音響指標の D 値を組み合わせて，雑音・残響指標 Noisy-and-Reverberant Speech Recognition criteria with PESQ and Acoustic parameters（NRSR-PA）を策定する．そして，NRSR-PA を用いて音声認識性能の予測位置におけるインパルス応答，雑音，発話音声から音声認識性能の予測を試みる．

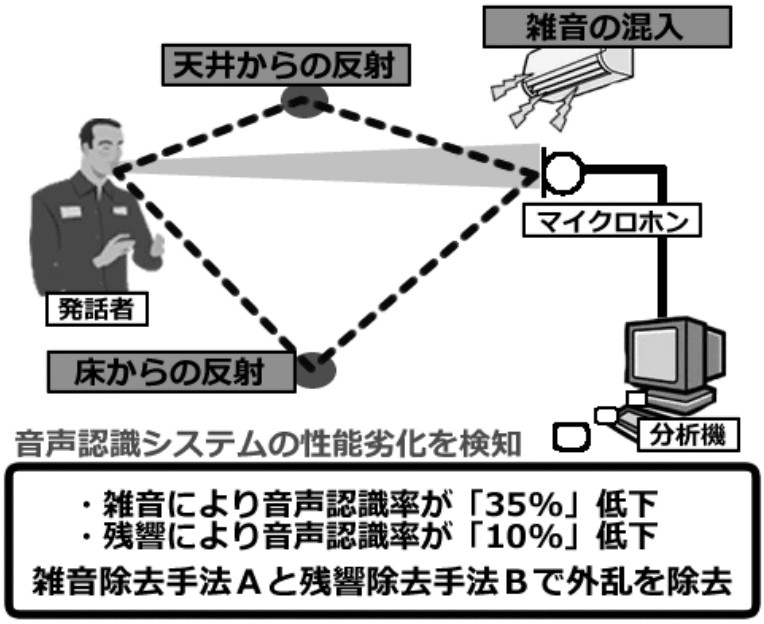


図 1.2 雑音・残響環境における音声認識性能の予測

1.2. 本論文の構成

本論文は，以下の全 5 章から構成される．2 章では，音声認識の原理，音声認識性能の評価手順と課題，そして雑音環境と残響環境のそれぞれに対する既存の音声認識性能予測手法の原理と課題について述べる．そして 3 章では，残響環境下における高精度かつ簡便な音声認識性能の予測手法について述べる．4 章では，雑音と残響が混在する環境における音声認識性能の予測手法について述べる．最後に 5 章で結論と今後の課題について述べる.

第2章　外乱環境における音声認識性能予測の基礎

2.1. はじめに

外乱成分が音声認識システムに与える影響を予測することで，その予測結果に基づいて外乱対策を音声認識システムの前処理等に適切に反映させることで，音声認識性能の劣化を未然に防ぐことができる．更に簡便な音声認識性能の予測手法を確立することで，実際の音声認識性能を評価するための音声収録や音声認識処理に必要なコストを大幅に削減できることが期待される．

本章は，音声認識性能の評価方法，そして雑音環境と残響環境のそれぞれに対する既存の音声認識性能予測手法の原理と課題について述べる．2.2 節では，音声認識の仕組みについて述べる．2.3 節では，一般的に用いられる音声認識性能を評価する手順について説明する．2.4 節では，雑音や残響の影響を受けることによる音声認識性能の劣化について述べる．2.5 節では，雑音環境，あるいは残響環境における従来の音声認識性能の予測手法の原理と課題について述べる．

2.2. 音声認識

音声認識は，人間の音声を機械的に自動認識する処理 [39] であり，一般的には入力音声をテキストとして出力することが多い．音声認識を行うには，大量の発話音声を記録した学習用データから音声を表現する特徴を学習し，入力された音声信号とそれらの特徴を照らし合わせながら，最も尤度の高い言語系列を認識結果として出力する統計的手法が良く用いられている．

音声認識では，音声を音響的な特徴と言語的な特徴に分けて処理する．音響的な特徴は，主に認識対象の音素の周波数特性をモデル (音響モデル) として表現する．音響モデルを構築する方法として，混合正規分布を出力確率とした隠れマルコフモデルが広く用いられている．一方，言語的な特徴は，音素の並び方に関する制約をモデル (言語モデル) として表現する．言語モデルの構築する方法として，認識対象の言語表現が多様な場合は n-gram が良く用いられ，認識対象の言語表現が人手で網羅出来る程度に小さい場合は文脈自由文法が良く用いられる．

ここで音声信号を分析して得られるパターン列を Y，単語列の集合を W とする．音声認識システムへの入力を y ($y \in Y$)，認識結果としての単語列の候補を w ($w \in W$) とするとき，認識結果の単語列 $\hat{w}$ を出力する音声認識システムは，ベイズの識別規則に従う．

$$\hat{w} = \underset{w \in W}{\operatorname{argmax}} P(w|y), \tag{2.1}$$

通常，$P(w|y)$ を直接算出すことは困難である．そこで，条件付き確率の定義より，

$$P(w|y) = \frac{P(y|w) \cdot P(w)}{P(y)}, \tag{2.2}$$

が成り立つため，式 (2.1) の $P(w|y)$ を最大化する代わりに，式 (2.2) の右辺を最大化 [40] する．式 (2.2) の $P(y)$ は，最適化する単語列 w とは無関係であるため，考慮する必要はない．したがって，ベイズの識別規則に基づく音声認識システムは，

$$\hat{w} = \underset{w \in W}{\operatorname{argmax}} P(y|w) \cdot P(w), \tag{2.3}$$

を算出する．なお，最大化すべき $P(y|w) \cdot P(w)$ のうち，$P(y|w)$ は音響モデルを用いて計算し，$P(w)$ は言語モデルから算出する．ここで音響モデルを残響や雑音を考慮しないクリーンな学習データから作成すると，残響や雑音を含む音声が入力された場合，特徴量に差異が生じるために音声認識性能が低下するという問題がある．

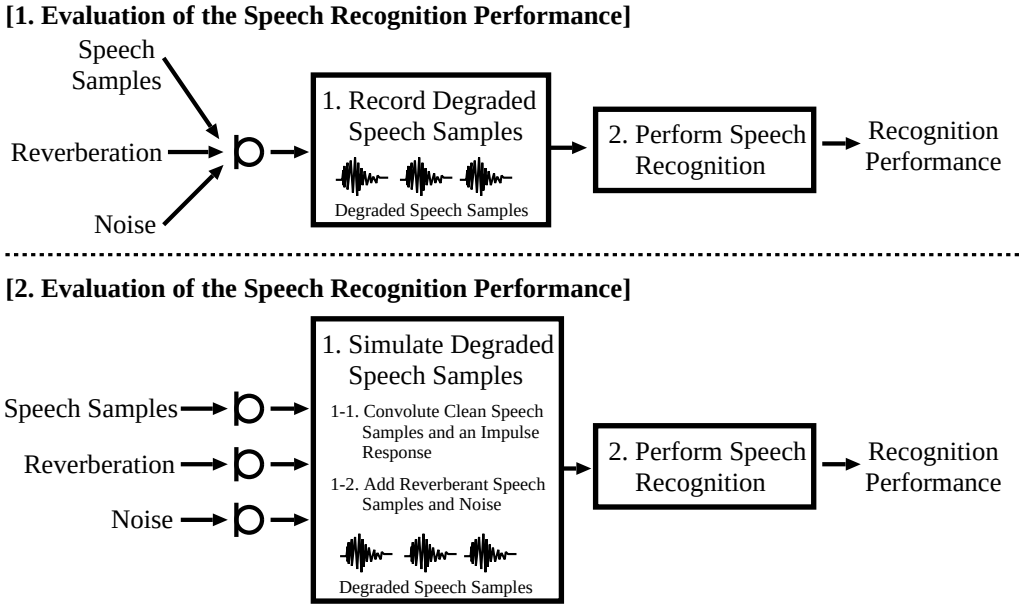


図 2.1　音声認識性能の評価手順

2.3. 音声認識性能の評価方法

ここでは，外乱環境における音声認識性能を評価するための手順について述べる．図 2.1 に評価音声収集から音声認識処理までの手順を示す．

音声認識処理では，大別して「(1) 実際に評価環境において音声データを収録（図 2.1 の上段）」と「(2) 評価環境を模擬した音声データを生成（図 2.1 の下段）」のいずれかを用いて評価音声を用意する．しかし，(1) については実際の利用環境において多くの被験者に発話させて大量の音声データを収録しなければならず，特に評価する発話位置が多いほど収録にかかる時間やコストが増大するという問題がある．また (2) については，クリーン音声，インパルス応答（残響），雑音を別々に収録した後に，クリーン音声と残響を畳み込みんだ残響音声に雑音を加算して評価音声を模擬する．そのため，被験者の音声収録回数が発話位置に関係なく 1 度だけで良いため，(2) は (1) と比べて被験者の音声収録の負担が大きく軽減できるが，一方で音声と残響の畳み込み処理に膨大な計算時間を要するという問題がある．

大量の評価用音声を収録した後に，その収録音声を用いて音声認識処理を行うが，

表 2.1 外乱と音声認識性能の関係調査のための実験条件

環境	和室 (T_{60}=450 ms，72ヶ所) ※ 壁からの距離：25 cm, 132 cm 会議室 (T_{60}=600 ms，120ヶ所) ※ 壁からの距離：25 cm, 335 cm エレベータホール (T_{60}=850 ms，120ヶ所) ※ 壁からの距離：25 cm, 300 cm
入出力間距離	100〜5,000 mm
音声	ATR 音素バランス 216 単語 [42, 43, 44] 女性：7 話者, 男性：7 話者
雑音	白色ガウス雑音 ピンク雑音 ヒューマンスピーチライク雑音 [41]
SNR	-5, 0, 5, 10, 15, 20, 30, 40, 50 dB
デコーダー	Julius rev. 4.2.1 [45, 46, 47]
HMM	IPA モノフォンモデル（性別依存）
特徴量	MFCC（12 次元）+ΔMFCC（12 次元）+ΔPower（1 次元）
分析長	25 ms (ハミング窓)
シフト長	10 ms

こちらも音声認識に用いる評価音声のデータ量に比例して計算量が増加する問題がある．

2.4. 外乱環境における音声認識性能

音声に雑音や残響などの外乱成分が混入することで，目的音声が歪み音声認識性能が低下する問題がある．本節では，具体的に外乱成分が音声認識性能に与える影響を分析するために表 2.1 に示す実験条件において音声認識実験を行った．

2.4.1 雑音環境における音声認識実験

雑音環境における音声認識実験では，雑音の種類や雑音量が異なる条件において音声認識性能を評価する．本実験では，周波数特性の異なる 3 種類の雑音（白色ガウス雑音，ピンク雑音，ヒューマンスピーチライク雑音（複数話者の音声を加算した信号）[41]）を用いた．そして評価音声として，クリーン音声に 3 種類の雑音を 9 種類の SNR で加算した評価音声を用いて音声認識率を算出した．SNR（Signal-to-Noise Ratio）は信号対雑音比を表し，SNR が低いほど雑音がクリーン音声よりも支配的であることを示す．

図 2.2 に雑音と音声認識性能の関係を示す．図 2.2 の結果より，全ての雑音に対して低 SNR 環境であるほど音声認識性能が低下することから，雑音量の影響を大きく受けると音声認識性能が低下することがわかる．また，一般的な生活環境を想定した SNR=5∼20 dB の条件では，雑音の種類によって音声認識性能に約 10∼20 %のばらつきがあることが確認できる．このように，雑音の影響の受け方によって，音声認識性能の劣化量が大きく異なることがわかる．

2.4.2 残響環境における音声認識実験

本項では室内や発話位置が異なる残響環境において音声認識評価実験を行った．まず残響時間が異なる 3 種類の環境（和室 : T_{60}=450 ms，会議室 : T_{60}=600 ms，エレベータホール : T_{60}=850 ms）において，図 2.3∼2.5 のように発話位置や発話方位などを変えて数十ヶ所∼百数十ヶ所のインパルス応答を計測した．なお，残響時間 T_{60} は，音の響きの長さを表し，残響時間が長いほど残響量が多いことを示す．そして，それぞれのインパルス応答とクリーン音声を畳み込んで，各発話位置における音声認識性能を算出した．

図 2.6 に残響と音声認識性能の関係を示す．図 2.6 中の線は各残響環境の音声認識性能の平均を表す．実験結果より，残響環境では長い残響時間ほど音声認識性能の平均が低下し，分散も上昇していることが確認できた．

次に壁・話者間距離差による音声認識性能の変化を表現するために，式 (2.4) に示す音声認識性能の相対変化量 $P_{diff}(d)$ [%] に基づいて音声認識性能の発話位置依存

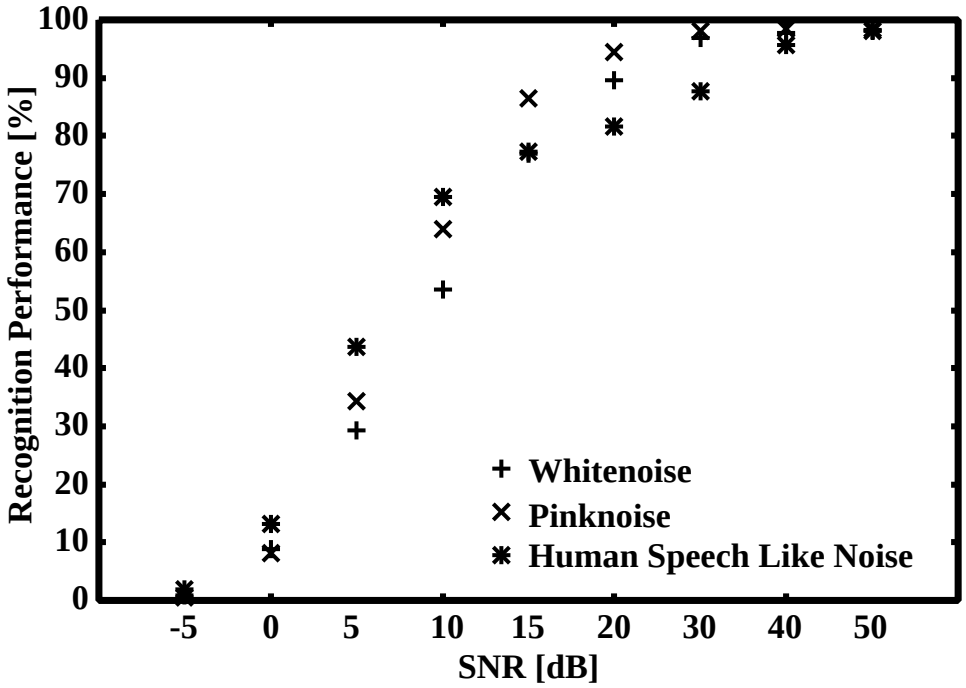


図 2.2 雑音環境における SNR と音声認識性能の関係

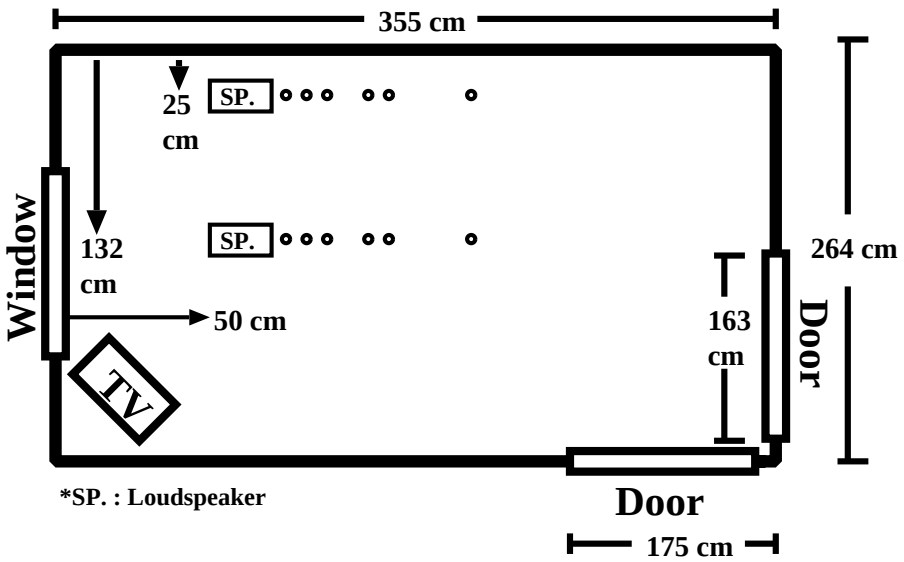


図 2.3 収録環境 (和室：T_{60}= 450 ms)

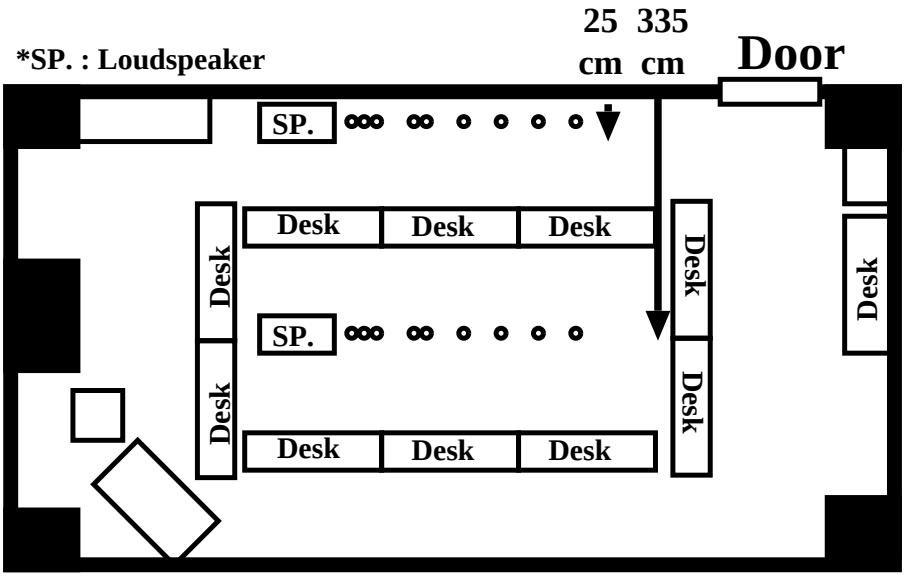


図 2.4 収録環境 (会議室：T_{60}= 600 ms)

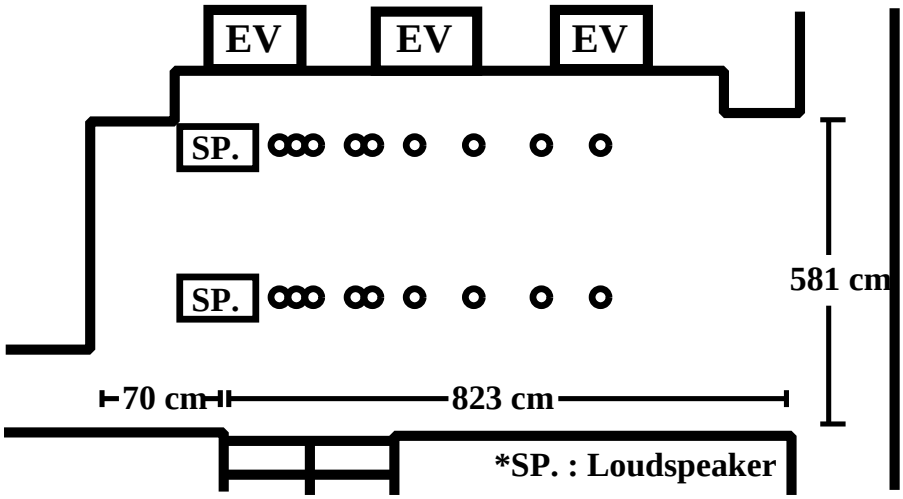


図 2.5 収録環境 (エレベータホール：T_{60}= 850 ms)

性の検証を行った．$P_{diff}(d)$ は壁に接近して発話した場合における音声認識性能に対して，壁から離反させて発話した場合における音声認識性能の変化量を相対的に表現した尺度である．

$$P_{diff}(d) = \frac{P_{close}(d) - P_{dist}(d)}{P_{close}(d)} \times 100. \tag{2.4}$$

なお d は入出力間距離，$P_{close}(d)$ は壁に接近させた場合の音声認識の正答数，$P_{dist}(d)$ は壁から離反させた場合の音声認識の正答数を示す．ここで $P_{diff}(d)$ が正値であれば壁に接近させた場合の音声認識性能が，$P_{diff}(d)$ が負値であれば壁から離反させた場合の音声認識性能が向上することを表す．図 2.7∼2.10 に評価実験結果を示す．実験結果より，和室 (T_{60}=400 ms) のような低残響環境においては，壁から離れて発話することで音声認識性能が向上したのに対して，エレベータホール (T_{60}=850 ms) のような高残響環境においては壁に接近して発話することで音声認識性能が向上した．また会議室 (T_{60}=600 ms) のように計測箇所付近に机などの障害物がある場合，壁以外の反射成分の影響により発話位置と音声認識性能の関係について顕著な傾向を確認することができなかった．そして発話方位に着目すると，スピーカの向きがマイクロホンに対して背面や右面では他方位と比較して音声認識性能の変化量 $P_{diff}(d)$ が大きいことがわかった．これはスピーカの向きが背面や右面の場合，直接音や極めて初期の反射音を受音することが難しく，その一方で音声認識性能を低下させる原因である後続残響を多く受音しているためだと考えられる．

2.5. 音声認識性能予測のための外乱指標

本節では，雑音環境，あるいは残響環境における音声認識性能を予測するための従来の外乱指標を述べる．

2.5.1 SNR（Signal to Distortion Ratio）

信号対雑音比 SNR（Signal to Distortion Ratio）は，信号成分と雑音成分のエネルギーを表現した指標であり，式 (2.5) のように表現される．

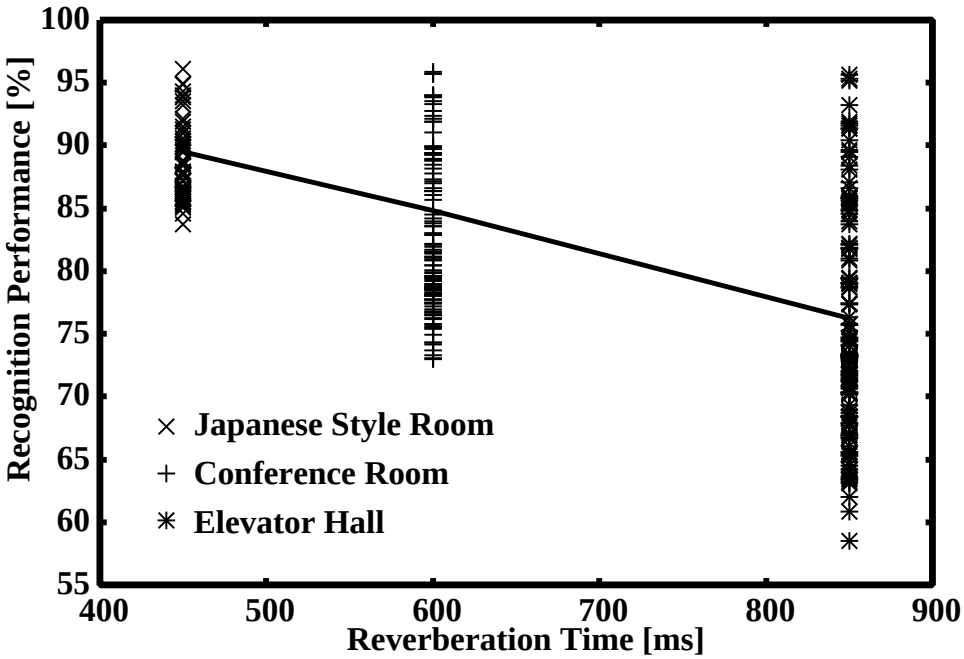
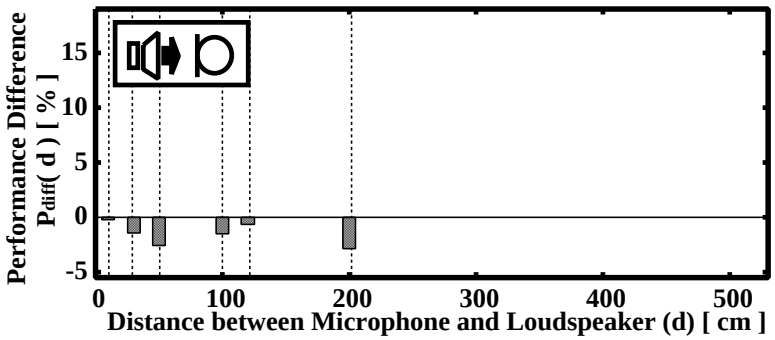
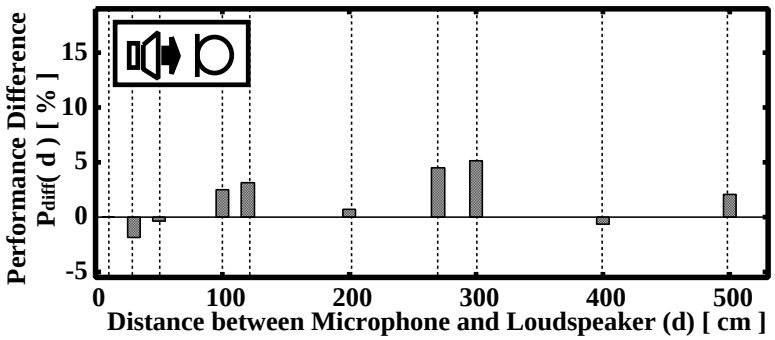


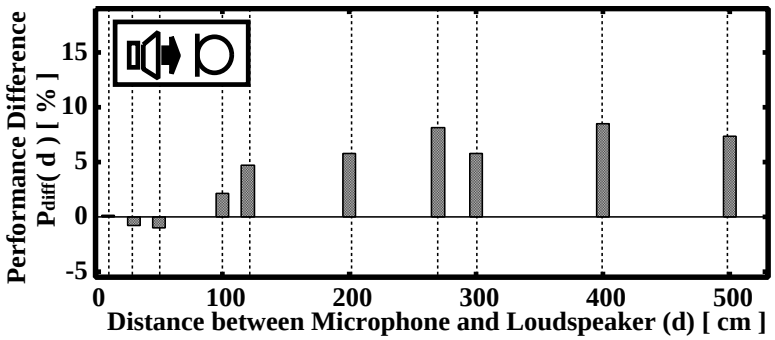
図 2.6 残響環境における残響時間と音声認識性能の関係



(a) 研究室 (T_{60}=400 ms)

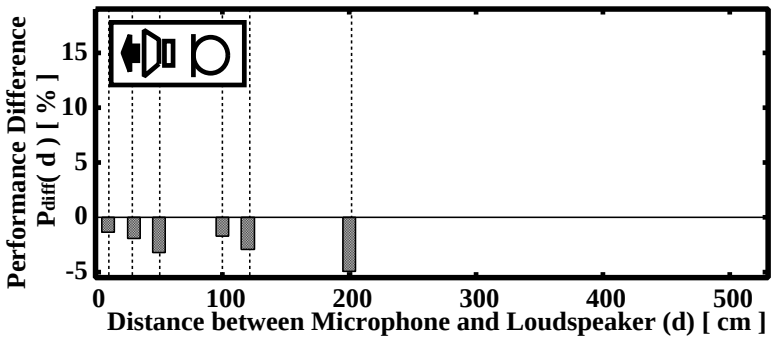


(b) 会議室 (T_{60}=650 ms)

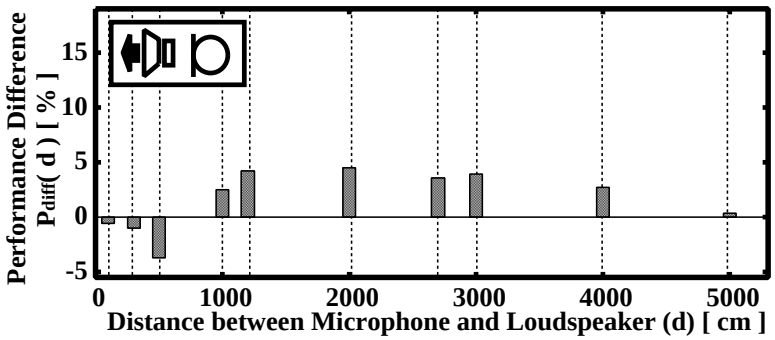


(c) エレベータホール (T_{60}=850 ms)

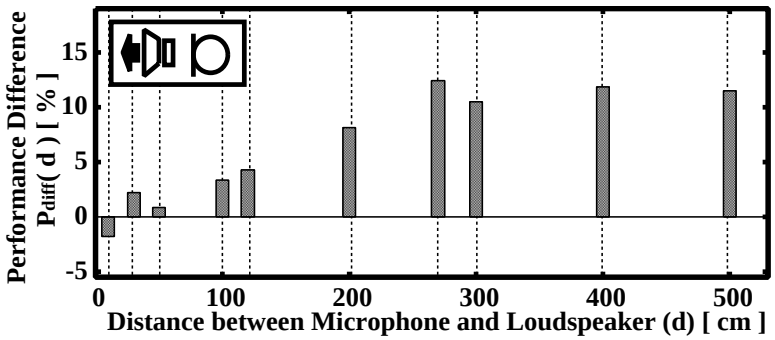
図 2.7 音声認識性能の変化量（正面から放射）



(a) 研究室 (T_{60}=400 ms)

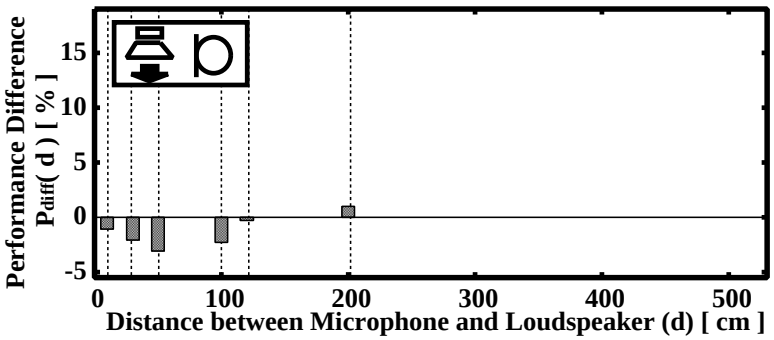


(b) 会議室 (T_{60}=650 ms)

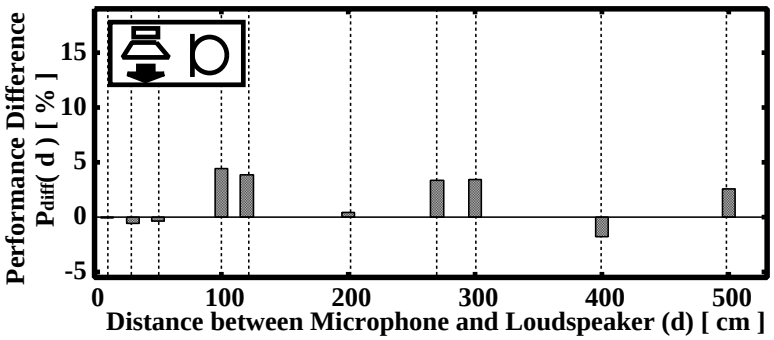


(c) エレベータホール (T_{60}=850 ms)

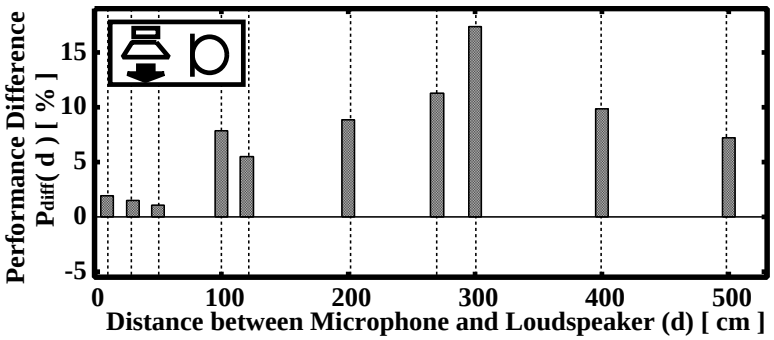
図 2.8 音声認識性能の変化量（背面から放射）



(a) 研究室 (T_{60}=400 ms)

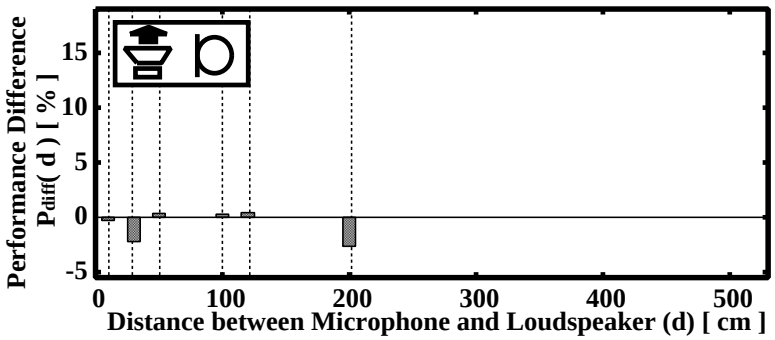


(b) 会議室 (T_{60}=650 ms)

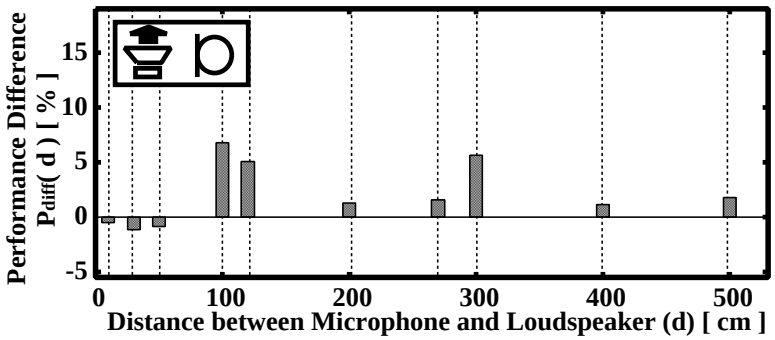


(c) エレベータホール (T_{60}=850 ms)

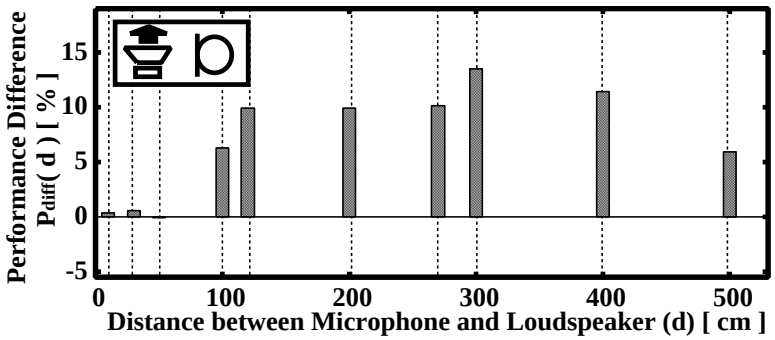
図 2.9 音声認識性能の変化量（左側方から放射）



(a) 研究室 (T_{60}=400 ms)



(b) 会議室 (T_{60}=650 ms)



(c) エレベータホール (T_{60}=850 ms)

図 2.10 音声認識性能の変化量（右側方から放射）

$$SNR \;=\; 10\log_{10}\left(\frac{\sum_{t=0}^{T-1} s^2(t)}{\sum_{t=0}^{T-1} n^2(t)}\right), \tag{2.5}$$

ここで，$s(t)$，$n(t)$ は時刻 t におけるクリーン信号および雑音信号の振幅を表し，T は分析区間を表す．一般的に SNR が高いほど，クリーン信号のエネルギーが雑音信号よりも支配的であり，雑音信号がクリーン信号に与える影響が小さいことを表す．SNR は現在の音声認識の雑音指標として積極的に利用されているが，非定常雑音を取り扱う場合，高精度な SNR 推定に長い分析区間と計算時間を用いるため，これに伴って音声認識性能予測に必要な計算コストが増加する問題があった．また図 2.2 においても，雑音の種類によって音声認識性能にばらつき（例えば，SNR=10 dB で 15%以上のばらつき）が確認できることから，SNR のみで音声認識性能を予測することに限界があると考えられる．

2.5.2 残響時間（T_{60}）

残響時間 (T_{60})[51] は室内音場を評価する基本的な概念であり響きの長さを表す．室内に放射した音が平衡状態に達した後，音を停止し，その後の残響エネルギー密度が音源停止直前のエネルギー密度に比べて 100 万分の 1（-60 dB）になるまでの時間を表したものである．残響理論では室内で拡散音場を仮定しているため，吸音材料をどの位置に配置してもその効果は変化せず，音源位置によって残響時間が変わらないと定義されている．また残響時間は M. R. Schroeder によって 2 乗積分法に基づく残響測定法 [49] が提案され，系の残響曲線はインパルス応答 $h(\lambda)$ を用いて式 (2.6) に基づき容易に算出できるようになった．

$$< Sd^2(t) > \;=\; N\int_t^{\infty} h^2(\lambda)d\lambda, \tag{2.6}$$

ここで N は単位周波数あたりのパワー，$< Sd^2(t) >$ は残響曲線を表す．これまで残響曲線は入力信号をランダム雑音として長時間かつ複数回観測した信号から集

合平均を利用して算出したのに対して M.R. Schroeder はインパルス応答 $h(\lambda)$ のみから集合平均を利用せずに残響曲線を算出する手法を提案した．残響時間は算出した残響曲線に基づき 60 dB 減衰するまでの時間となるが，計測したインパルス応答の後続部分は暗騒音に埋没し，実際に残響エネルギー密度が 60 dB 減衰する時間を算出することは困難である．この問題に対して，通常は初期部分を回帰した直線が 60 dB 減衰するまでの時間を残響時間とすることが一般的である．

残響時間は現在の音声認識の残響指標として積極的に利用されているが，仮定する拡散音場と実際の環境との差異から他の残響特性が変化し，同一環境でも計測箇所によって音声認識性能が変動する．そのため固有の値をとる残響時間のみで音声認識の難しさを表現することに限界があると考えられる．

2.6. まとめ

本章では，雑音環境と残響環境のそれぞれに対する既存の音声認識性能予測手法の原理と課題について述べた．2.2 節では，音声認識処理の構成について説明した．2.3 節では，音声認識性能の評価方法について述べた．2.4 節では，雑音や残響の影響を受けることによって音声認識性能が劣化することを示した．2.5 節では，雑音環境，あるいは残響環境における従来の音声認識性能の予測手法の原理を述べ，これらの手法では高精度かつ簡便な予測が難しいことを示した．

第3章　室内音響指標を用いた残響下における頑健な音声認識性能予測

3.1. はじめに

外乱環境において音声認識性能を予測することは，音声認識性能の改善につながるだけでなく，音声認識評価に関わるコスト削減にも貢献することができる．特にテレビ会議システムのような屋内での音声インタフェース利用を想定すると，外乱環境の中でも残響環境下における頑健な音声認識性能の予測が必要となるが，過去に有力な残響指標が提案されていない．これまでは 2.5 節でも述べた通り，残響下音声認識性能の優劣を判別する残響指標として同一室内では同じ値となる残響時間が提案されているが，仮定する拡散音場と実環境との差異から他の残響特性が変化することにより同一環境でも計測箇所によって音声認識性能が変動する．そのため残響時間は音声認識の難しさを表す指標として不十分であることが問題視されている．そこで本章では，ISO3382 Annex A で提案されている室内音響指標を用いた残響下における頑健な音声認識性能の予測法を検討する．

本章の構成を以下に示す．3.2 節で，提案手法に用いる室内音響指標について述べる．3.3 節で提案手法の詳細について述べる．3.4~3.8 節で，提案手法を用いて残響環境における音声認識性能の予測に関する実験を行い，その結果について述べる．3.9 節で，本章のまとめを述べる．

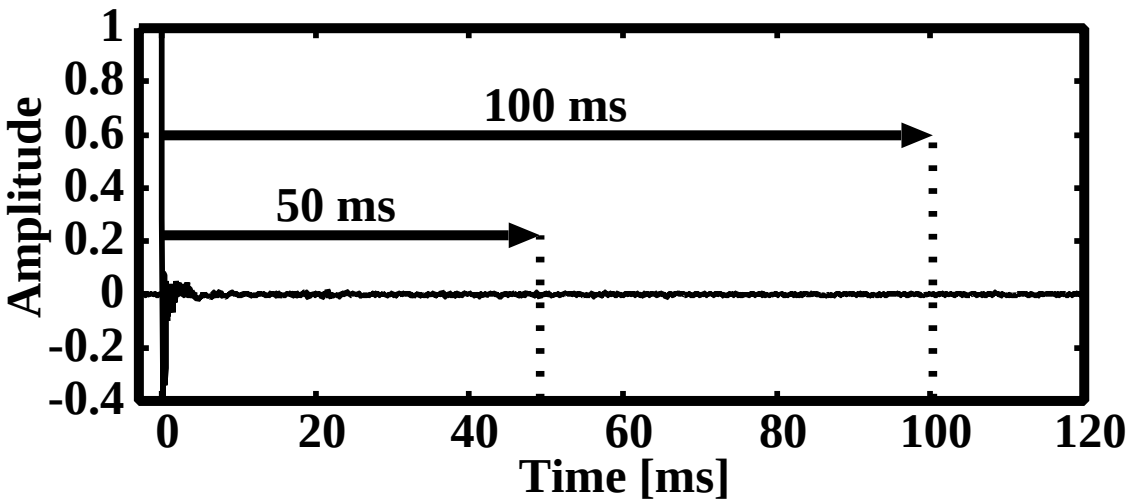


図 3.1 直接音からのインパルス応答長

3.2. 室内音響指標

3.2.1 音声認識における初期・後続反射音の影響

前章において同一環境でも計測箇所によって音声認識性能が変動することから，同一室内で固有の値となる残響時間では音声認識性能の予測が困難であることを述べた．そこで本節では音声認識に影響を与える残響特性を明らかにするために，音声認識性能の著しい低下が顕著に確認できる反射継続時間と音声認識性能の関係について調査する．

音声認識性能と反射音の関係を調査する方法として，TSP（Time Stretched Pulse）信号 [52, 53, 54, 55] を用いて系のインパルス応答を計測 [56, 57] し，図 3.1 および表 3.1 の実験条件に示す範囲に基づいて初期反射時間分だけインパルス応答を切り出した上で音声ドライソースと畳み込むことで，初期反射音の継続時間と音声認識性能との関係を調査する．なおハース効果 [51] に基づき本実験では直接音から最長 100 ms までの反射音を調査する．

図 3.2∼3.4 に初期反射音の継続時間と音声認識性能の関係を示す．音声認識性能は，マイクロホンとスピーカ間の距離が 500〜1,000 mm となる系を境界として低下

表 3.1 反射音と音声認識性能の関係調査のための実験条件

環境	研究室 (T_{60}=450 ms，6ヶ所) 廊下 (T_{60}=600 ms，6ヶ所) エレベータホール (T_{60}=850 ms，6ヶ所)
入出力間距離	100, 300, 500, 1,000, 2,000, and 3,000 mm
音声	ATR 音素バランス 216 単語 [42, 43, 44] 女性：2 話者, 男性：2 話者
デコーダー	Julius rev. 4.2.1 [45, 46, 47]
HMM	IPA モノフォンモデル（性別依存）
特徴量	MFCC（12 次元）+ΔMFCC（12 次元）+ΔPower（1 次元）
分析長	25 ms (ハミング窓)
シフト長	10 ms
インパルス応答長	5 ms, 10〜100 ms（10 ms 間隔）

する傾向が確認できた．さらに，同一残響時間でも音声認識性能に差異があることや，20〜30 ms 程度より後続の反射音，特に 60 ms 程度より後続の反射音は音声認識性能を大きく低下させる要因であることが確認できた．また図 3.4 におけるマイクロホンとスピーカ間の距離が 300 mm の結果では，直接音からのインパルス応答長が 10〜80 ms において音声認識性能はほぼ同程度であるため，本実験において最長 80 ms までの反射音を含むインパルス応答を用いても音声認識性能は低下せず，直接音から 60 ms 以降の後続の反射音が音声認識性能の劣化原因とならない環境が存在することも確認できた．この結果から音声認識性能の予測指標として，従来の残響時間では高精度な音声認識性能の予測が困難であることを再確認した．

そこで本章では，音声認識が著しく低下するまでの初期反射音の継続時間に基づき初期部分の反射音エネルギーと後続部分の反射音エネルギーの割合に着目する．この着目点に対して室内音響指標（ISO3382）[60] の導入を念頭に残響下音声認識のための残響指標の策定を試みる．

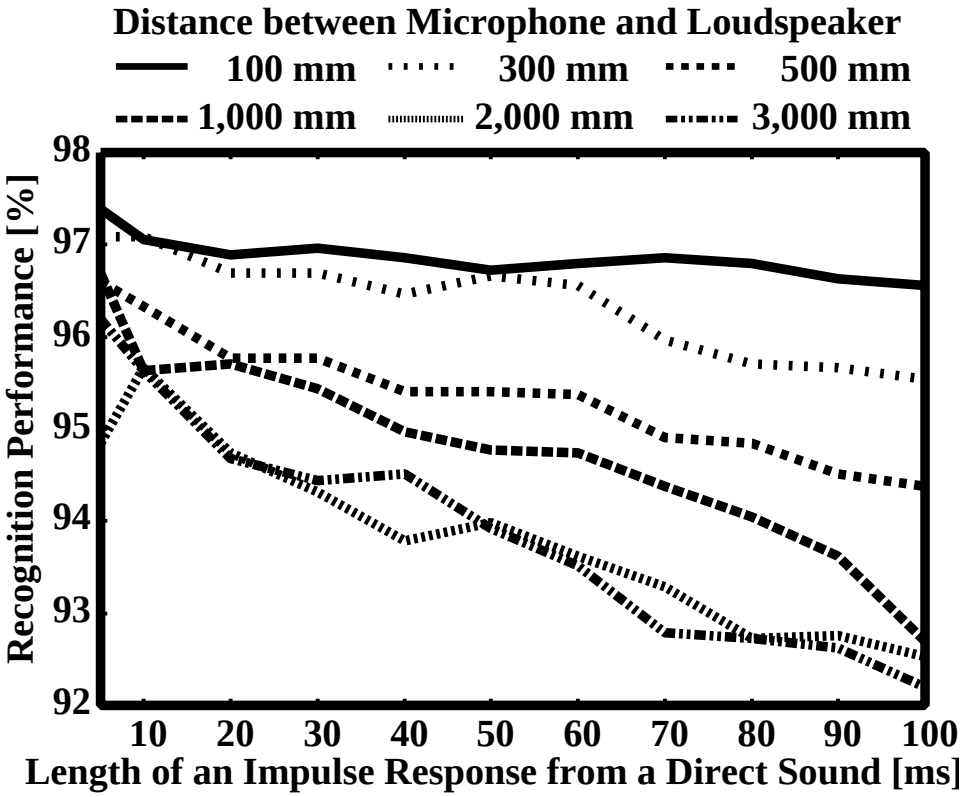


図 3.2 音声認識性能と初期反射音の関係（(a) 研究室, マイクと壁の距離 : 250 mm）

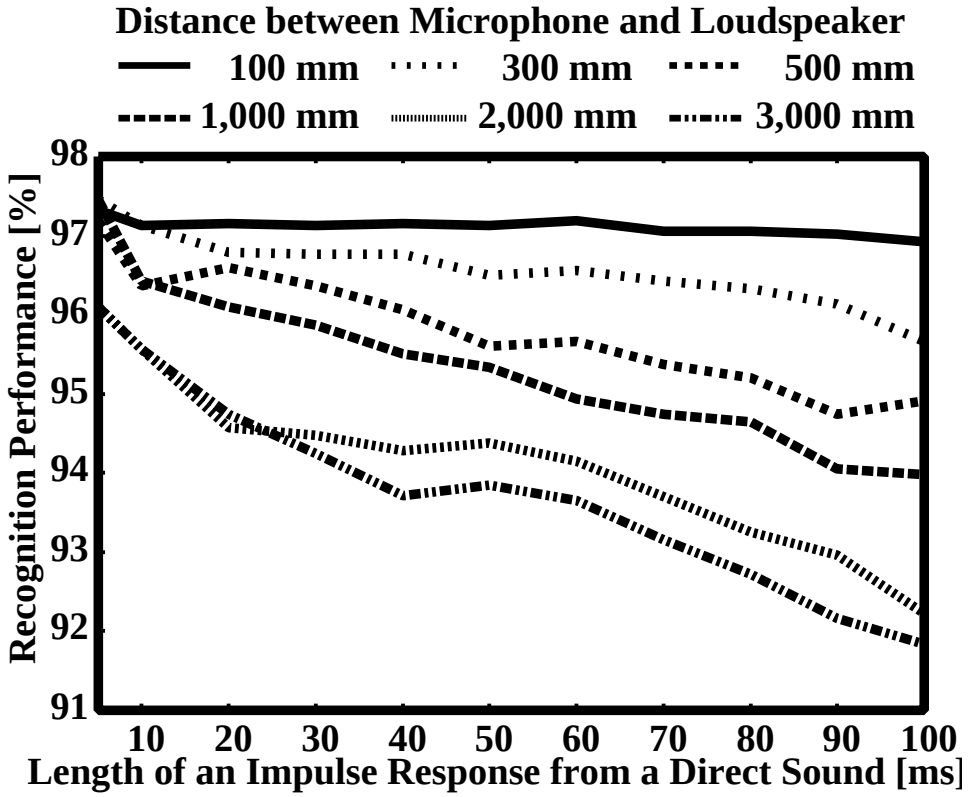


図 3.3 音声認識性能と初期反射音の関係（(b) 廊下, マイクと壁の距離：250 mm）

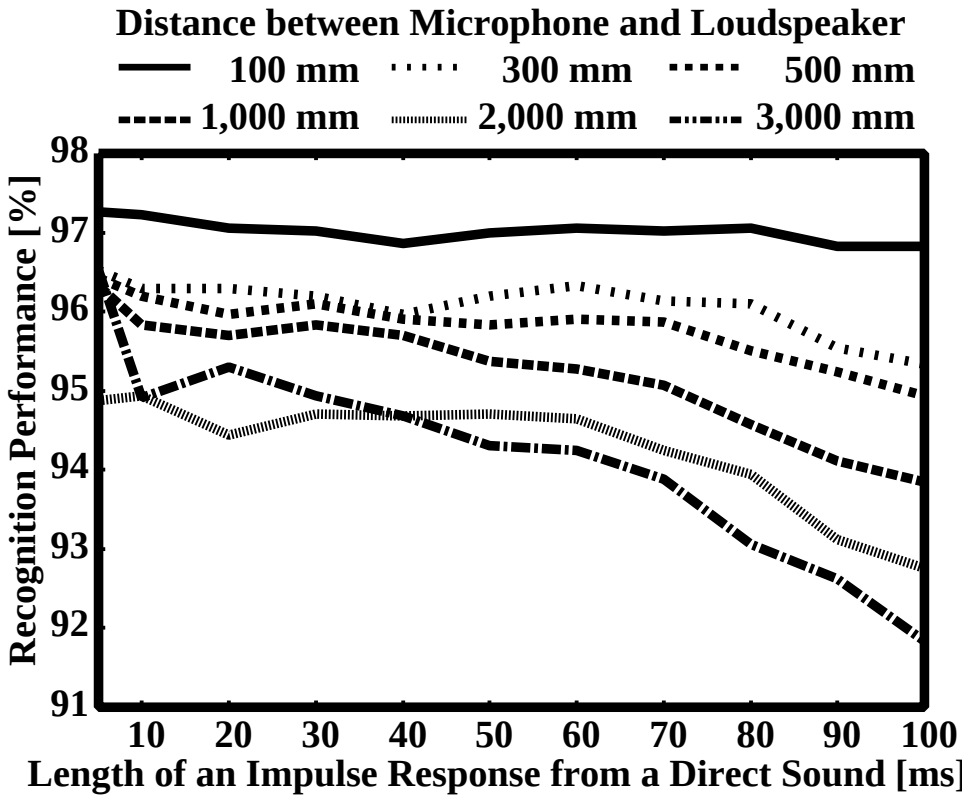


図 3.4 音声認識性能と初期反射音の関係（(c) エレベータホール, マイクと壁の距離：250 mm）

3.2.2 A 値（反射音の総合振幅）

計測したインパルス応答の反射エネルギーを表現する尺度としてよく利用されるのが直接音に対する反射音の総合振幅を表す A 値 [59] である．A 値は式 (3.1) のように定義される．

$$A = \sqrt{\int_{\epsilon}^{\infty} h^2(t)dt / \int_{0}^{\epsilon} h^2(t)dt}, \tag{3.1}$$

ここで $h(t)$ はインパルス応答の振幅を表す．また ϵ は直接音の持続時間を示し，インパルス応答の場合 3〜5 ms となる．A 値は受音信号における反射音エネルギーに対する直接音エネルギー比であり，同一室内でも各受音点により大きく異なる．音源に近接して受聴すると反射音に比べて直接音のエネルギーが高くなるため，A 値が低下するのに対して，遠方から受聴すると反射音のエネルギーが大きくなり，A 値は上昇する．しかし A 値では系の初期反射音と後続残響のどちらのエネルギーが大きいのかを判断できないため音声認識性能を著しく低下させる後続残響エネルギーを表現することが困難である．したがって反射エネルギーの中で音声認識性能に影響する成分を明確に示すことができず，A 値に基づいて音声認識性能を予測することは困難であると考えられる．

3.2.3 Definition（D 値）

ISO3382 Annex A で提案されている室内音響指標 [60] は残響時間を補う残響尺度として，音の初期部分の減衰状態を表現するために 1997 年に提案され，建築音響学の分野ではよく用いられている指標の 1 つである [61, 62]．この室内音響指標は以下の 4 つから構成される．

1. 音圧レベル
2. 残響時間
3. 初期反射音と後続残響音のバランス
4. 両耳パラメータ

この中で音の了解性に最も関連性がある「3．初期反射音と後続残響音のバランス」に着目し，音声認識システムの整合性を検証する．

初期反射音と後続残響音のバランスを構成する要素として，C 値 (Clarity)[63], D 値 (Definition)[64] と Ts (Centre time)[65] の 3 つが存在する．C 値は式 (3.2) より算出され，直接音と初期反射音のエネルギーに対する後続残響のエネルギー比を示す．D 値は式 (3.3) より算出され，直接音と初期反射音のエネルギーに対する直接音と全ての反射音のエネルギー比を示す．そして，Ts は式 (3.4) より算出され，2 乗インパルス応答の時間重心を示す．

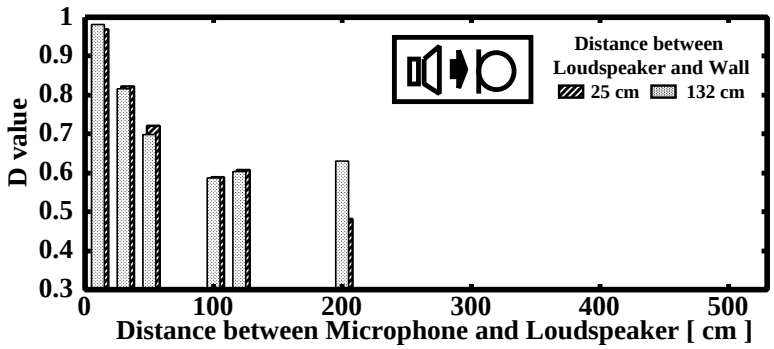
$$C_n = 10\ \log 10(\int_0^n h^2(t)dt / \int_n^\infty h^2(t)dt). \tag{3.2}$$

$$D_n = \int_0^n h^2(t)dt / \int_0^\infty h^2(t)dt, \tag{3.3}$$

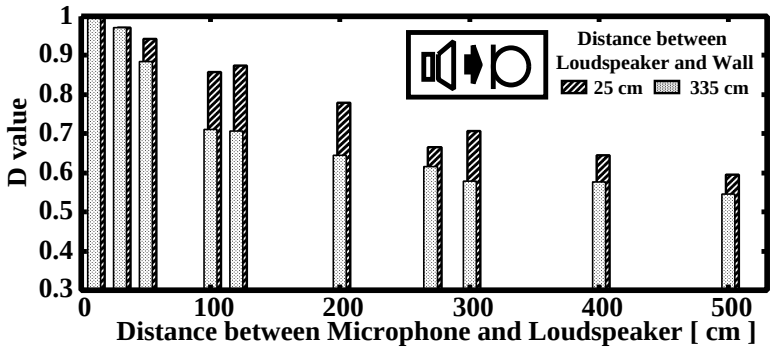
$$T_s = \int_0^\infty t h^2(t)dt / \int_0^\infty h^2(t)dt, \tag{3.4}$$

ここで $h(t)$ はインパルス応答を，n は初期反射音と後続残響音の境界時間を示す．C 値は n=80 ms（C_{80}）のとき音楽の明瞭性と高い相関があり，さらに D 値は n=50 ms（D_{50}）のとき音声の明瞭性と高い相関を有する．また Ts が高いほど後続残響音が大きいことを示し，残響感が増幅されて明瞭度が低くなる．直接音と初期反射音のエネルギーが大きいほど D 値は向上を示し，後続残響のエネルギーが大きいほど低下する．D 値は計測したインパルス応答から音声認識性能に影響を与える初期反射音と後続残響音の割合を表現できることから，音声認識性能に与える劣化の度合いを表現するパラメータとなる可能性がある．これまでの先行研究 [58] により，C 値・D 値と音声認識性能については強い相関があることがわかっている．C 値と D 値は可逆変換可能な指標であり，かつ D 値は音声の明瞭性を表現可能な指標として提案されていることから，本研究では D 値に注目する．

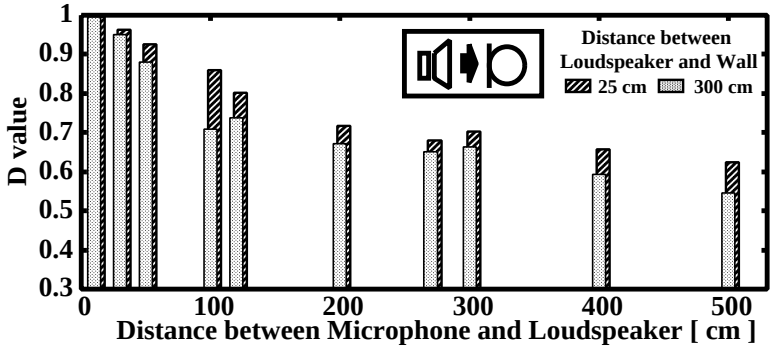
ここで 2.4 節における外乱環境における音声認識実験と同じ条件に基づいて，残響時間が異なる 3 種類の環境（研究室：T_{60}=450 ms，会議室：T_{60}=600 ms，エレベータホール：T_{60}=850 ms）において，発話位置，D 値，音声認識性能の関係を分析した．図 3.5∼3.8 に各残響環境における D 値の結果を発話方位別に示す．各図の横軸



(a) 研究室 (T_{60}=400 ms)

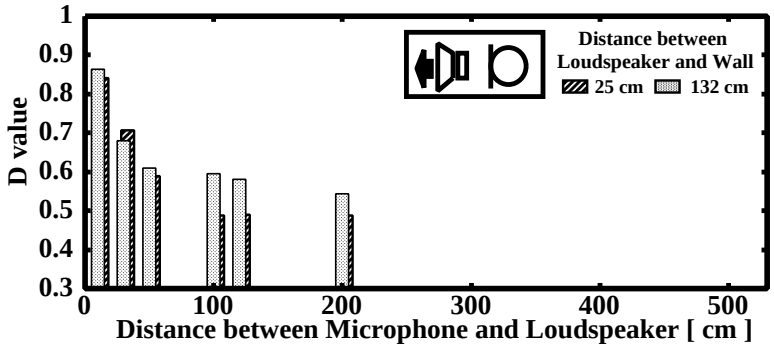


(b) 会議室 (T_{60}=650 ms)

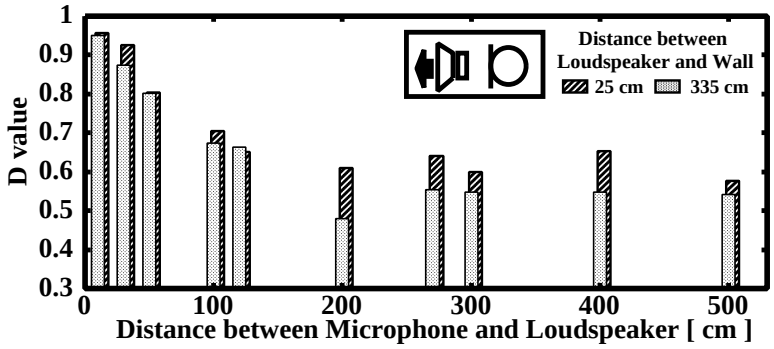


(c) エレベータホール (T_{60}=850 ms)

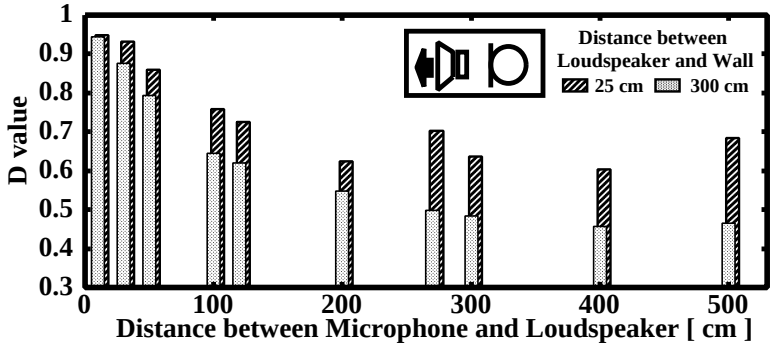
図 3.5 各残響環境の D 値（正面から放射）



(a) 研究室 (T_{60}=400 ms)

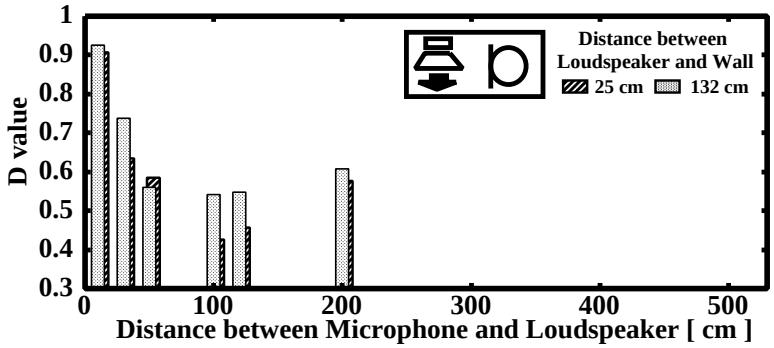


(b) 会議室 (T_{60}=650 ms)

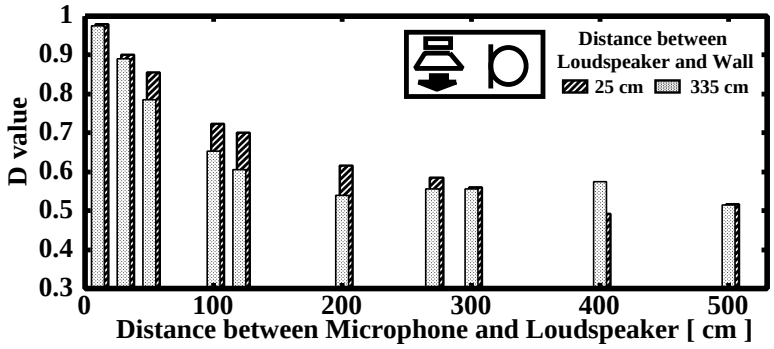


(c) エレベータホール (T_{60}=850 ms)

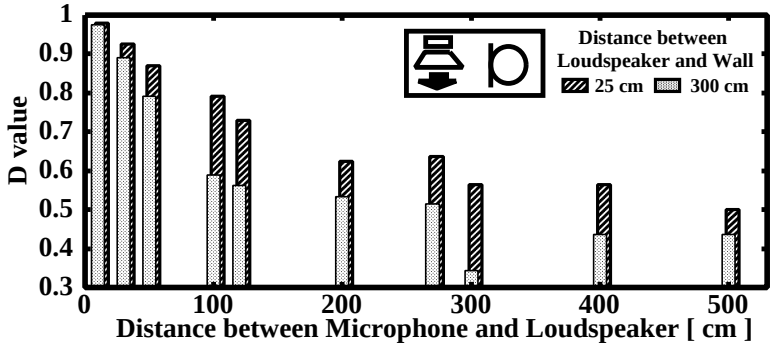
図 3.6 各残響環境の D 値（背面から放射）



(a) 研究室 (T_{60}=400 ms)

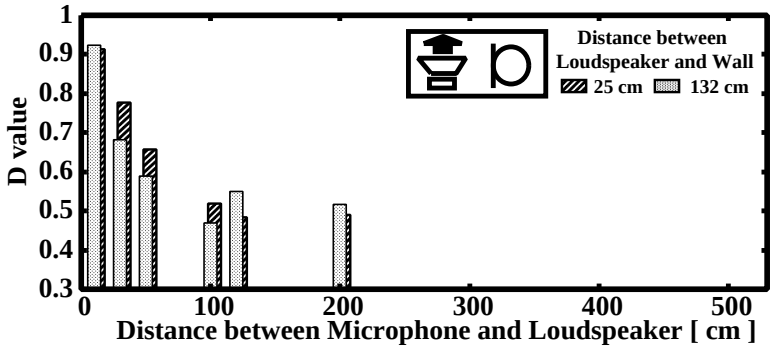


(b) 会議室 (T_{60}=650 ms)

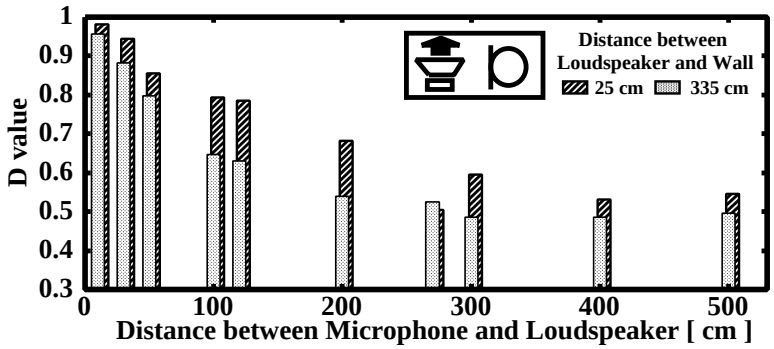


(c) エレベータホール (T_{60}=850 ms)

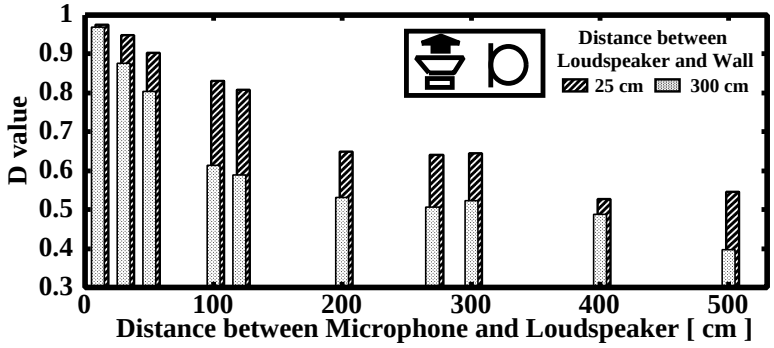
図 3.7 各残響環境の D 値（左側方から放射）



(a) 研究室 (T_{60}=400 ms)



(b) 会議室 (T_{60}=650 ms)



(c) エレベータホール (T_{60}=850 ms)

図 3.8 各残響環境の D 値（右側方から放射）

は入出力間距離を，縦軸はD値を表す．実験結果から，全体的に入出力間距離が長いほどD値が低下していることから，受音点から離れた発話は直接音や初期反射のエネルギーが小さく，後続残響のエネルギーが増加していることがわかる．また低残響環境では壁から離れて発話することでD値が向上するのに対して，高残響環境では壁に接近して発話することでD値が向上することを確認した．つまり，低残響環境では壁から離れた位置において，そして高残響環境では壁に接近した位置において，直接音や初期反射音を多く収音できることがわかる．

また，図2.7~2.10の壁と話者間の距離による音声認識性能の相対変化量と図3.5~3.8のD値の分析結果より，低残響環境では壁から離れて発話し，高残響環境では壁に接近して発話することで音声認識性能とD値がそれぞれ向上した．このことより，D値は発話位置に依存する音声認識性能を的確に表現できる指標の一つであることがわかった．

3.3. 音声認識性能予測アルゴリズム

前述のD値と残響下音声認識性能の関係を明らかにした上で，それぞれの相関関係について曲線近似し，残響下音声認識性能予測のための残響指標RSR-D_n (Reverberant Speech Recognition criteria with D_n) の策定を試みる．

3.3.1 残響指標RSR-D_nの策定

音声認識性能を予測するための残響指標RSR-D_nの策定アルゴリズムを図3.9に示す．

Step.1 インパルス応答計測

各環境でインパルス応答を数10～数100箇所にて計測する．その際，式(2.6)に基づいて算出した残響曲線から残響時間を算出する．残響時間は同一室内では理論上，固有の値をもつため，計測したインパルス応答の全てから残響時間を算出する必要は無く，数箇所のインパルス応答から算出した残響時間の平均

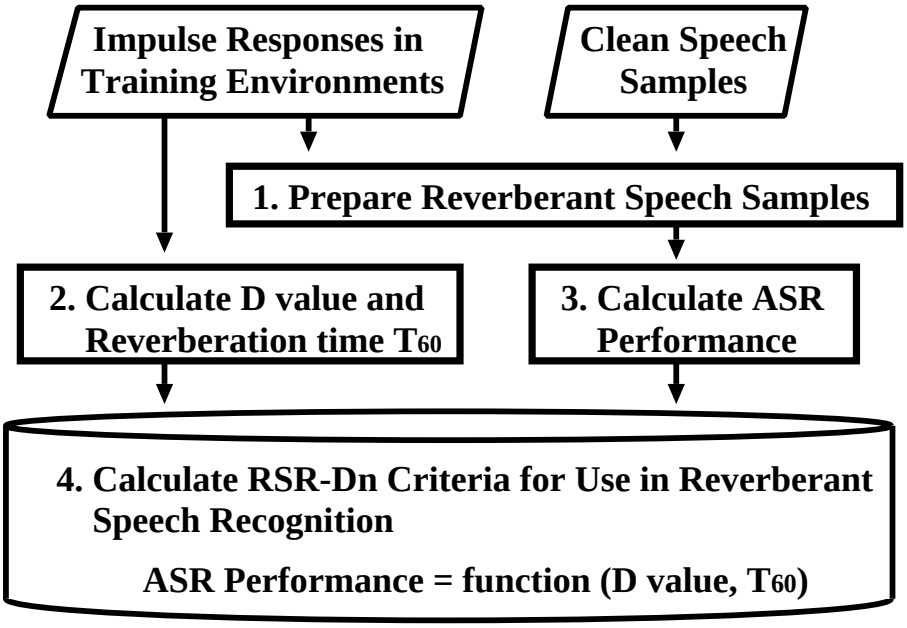


図 3.9 提案手法の概要（残響指標 RSR-D_n の策定）

を各環境の残響時間とすることが一般的である．

Step.2 D 値の算出

Step.1 で計測した各インパルス応答に対して式 (3.3) に基づいて D 値を算出する．また初期反射音と後続残響の境界時間を表す n は，音声認識性能と D 値の最大相関値を示すように設定する必要がある．そこで最適な境界時間 n を **3.4** 節で実験的に検討し，その結果を基に D_n を算出する．

Step.3 音声認識性能の算出

Step.1 で計測した各インパルス応答と学習データとして予め用意した音声ソースを畳み込み，音声認識エンジンを用いて音声認識性能を算出する．

Step.4 近似曲線の算出

表 3.2 近似曲線と音声認識性能予測値

近似曲線	$y = ax + b$	$y = ax^2 + b$
音声認識性能予測値 ($\hat{x}$)	$\hat{x} = \frac{y-b}{a}$	$\hat{x} = \sqrt{\frac{y-b}{a}}$

※ a，b：近似係数，x：音声認識性能，y：D 値

Step.2 と **Step.3** で各インパルス応答から算出した D 値と音声認識性能を基に近似曲線を算出する．算出する近似曲線は 1 次直線，2 次曲線とする．各近似曲線の定義式を表 3.2 に示す．なお本論文では 1 つの D 値から音声認識性能が一意に定まるように 2 次曲線の 1 次項を省略している．1 次直線，2 次曲線で分析を行う際に用いる係数予測方法は，最小二乗法 [67] を用いる．最小二乗法は，予測値と測定値の残差の二乗和が最小となるようにモデルパラメータを決定する方法である．

3.3.2 残響指標 RSR-D_n を用いた音声認識性能予測

策定した残響指標 RSR-D_n に基づく音声認識性能の予測アルゴリズムを図 3.10 に示す．音声認識性能を予測する系で計測したインパルス応答に基づいて残響時間と D 値を算出する．ここで同一残響時間の指標が存在しない場合，近接の残響時間の指標を線形補間する．そして同一残響時間における残響指標 RSR-D_n と D 値から音声認識性能の予測を試みる．

3.4. 評価実験 1 -残響指標 RSR-D_n のための最適境界時間の検討-

式 (3.3) における n は，初期反射音と後続残響音の境界時間を示し，D 値の算出において適切な値を設定する必要がある．そこで音声認識性能と D 値の間で高い相関を示す境界時間 n を検討するために，表 3.3 **(B)** に示す残響時間が異なる 3 環境で評価実験を行った．

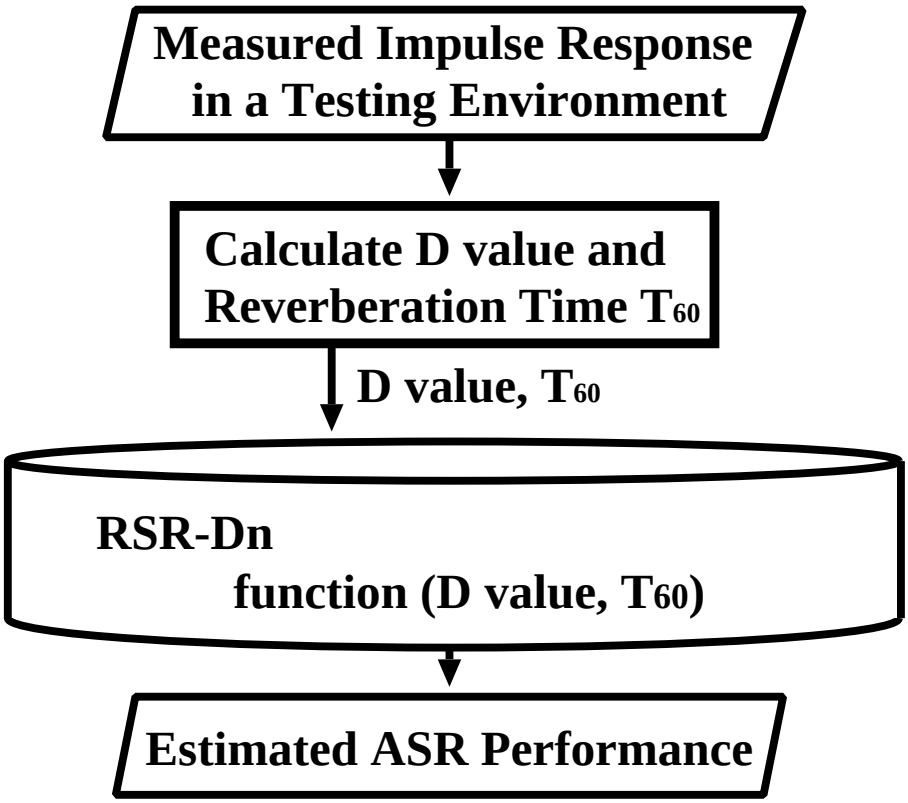


図 3.10 提案手法の概要（残響指標 RSR-D_n を用いた音声認識性能の予測）

3.4.1 実験条件

実験方法は **3.3** 節の分析アルゴリズムと同様である．また D 値は境界時間 n を 10〜90 ms の 10 ms 間隔に設定して算出する．そして境界時間 n ごとに算出した D 値と音声認識性能との関係を曲線近似し，3 環境の相関係数の平均を各近似曲線ごとに算出した．

3.4.2 実験結果

初期反射音と後続残響音の境界時間 n と各近似曲線の相関係数の関係を図 3.11 に示す．1，2 次曲線共に境界時間 n が 20 ms で最も高い相関係数を示し，以降は減少傾向にあることを確認した．したがって，今回の表 3.3 **(B)** に示す 3 環境における評価実験結果では残響指標 RSR-D_n のための境界時間 n は 20 ms が最適であることがわかった．本提案手法では，最も高い相関係数を確認した n=20 ms を採用して D 値 (D_{20}) および RSR-D_{20} を算出する．

3.5. 評価実験 2 -残響指標 RSR-D_{20} の策定実験-

3.5.1 実験条件

室内音響指標を用いた残響指標 RSR-D_{20} を策定するために，表 3.3 **(A)** に示す 9 つの学習環境にて計 732 箇所のインパルス応答を計測した．なお表 3.3 に示す環境は，様々な残響環境を想定するために，残響時間が異なる環境でインパルス応答を計測した．また各残響環境の中でも各系の D 値の分散が大きくなるように 10〜500 cm の入出力間距離および正背左右の放射面の条件で計測を行った．そして計測したインパルス応答を基に，室内音響指標と音声認識性能の関係について曲線近似して残響指標 RSR-D_{20} を策定した．

表 3.3 実験条件

(A) 学習環境	防音室 (T_{60}=100 ms，72ヶ所) 和室 (T_{60}=400 ms，72ヶ所) 研究室 (T_{60}=450 ms，72ヶ所) 会議室 (T_{60}=600 ms，120ヶ所) リビング (T_{60}=600 ms，72ヶ所) 廊下 (T_{60}=600 ms，120ヶ所) 浴室 (T_{60}=650 ms，28ヶ所) エレベータホール (T_{60}=850 ms，120ヶ所) 階段 (T_{60}=850 ms，56ヶ所)
(B) 境界時間 n を決定するための評価環境	和室 (T_{60}=400 ms，72ヶ所) 会議室 (T_{60}=600 ms，120ヶ所) 階段 (T_{60}=850 ms，56ヶ所)
(C) RSR-D_n 策定のための評価環境	和室 (T_{60}=400 ms，72ヶ所) 会議室 (T_{60}=600 ms，120ヶ所) 階段 (T_{60}=850 ms，56ヶ所)
(D) 性能予測環境 （オープン環境）	研究室 (T_{60}=450 ms，72ヶ所) 浴室 (T_{60}=650 ms，28ヶ所) エレベータホール (T_{60}=850 ms，120ヶ所)
入出力間距離	100〜5,000 mm

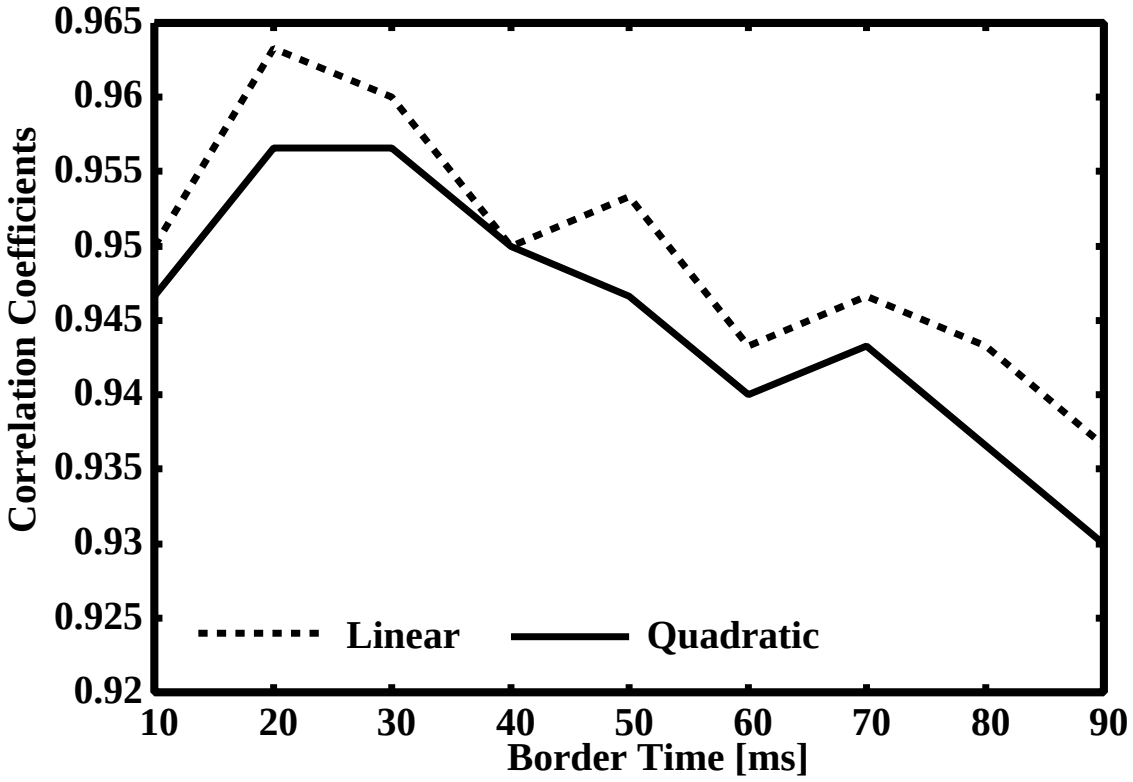


図 3.11　各近似曲線の相関係数と境界時間 n の関係

表 3.4 相関係数

	RSR-D_{20}L (Linear)	RSR-D_{20}Q (Quadratic)
T_{60}=400 ms	0.937	**0.939**
T_{60}=600 ms	**0.966**	0.963
T_{60}=850 ms	**0.977**	0.972

3.5.2 実験結果

表 3.3 **(A)** に示す 9 つの学習環境における D_{20} と音声認識性能の関係を図 3.12 に，拡大図を図 3.13 に示す．そして，この 9 環境の中から表 3.3 **(C)** に示す残響時間が異なる 3 環境について曲線近似を行った結果を図 3.14〜3.16 に，3 環境に対する各近似曲線の相関係数を表 3.4 に示す．また音声認識性能と D_{20} の関係を 1 次曲線で近似した結果を RSR-D_{20}L (Linear)，2 次曲線で近似した結果を RSR-D_{20}Q (Quadratic) と表している．

結果より，会議室 (T_{60}=600 ms) と階段 (T_{60}=850 ms) における両曲線の相関係数が 0.96 を上回り，高精度に近似可能であった．また和室 (T_{60}=400 ms) における両曲線の相関係数も 0.93 を上回っており，全体的に高精度な曲線近似が可能であった．この結果から D_{20} と音声認識性能の関係を 1 次，2 次曲線で近似した RSR-D_{20}L，RSR-D_{20}Q ともに有力な残響指標であることを確認した．

ここで策定した RSR-D_{20} の環境変化に対する頑健性について考察する．表 3.3 **(A)** に示す 9 つの学習環境における D_{20} と音声認識性能の関係を示した図 3.13 の残響時間が 600 ms の環境 (会議室，リビング，廊下) より，同一残響時間または近傍の残響時間をもつ環境における計測値の分布が類似していることがわかる．残響時間が 400〜450 ms の和室と研究室，850 ms のエレベータホールと階段においても同様の傾向が確認できる．このことから近傍の残響時間であれば異なる環境の RSR-D_{20} を用いても音声認識性能を頑健に予測できると考えられる．

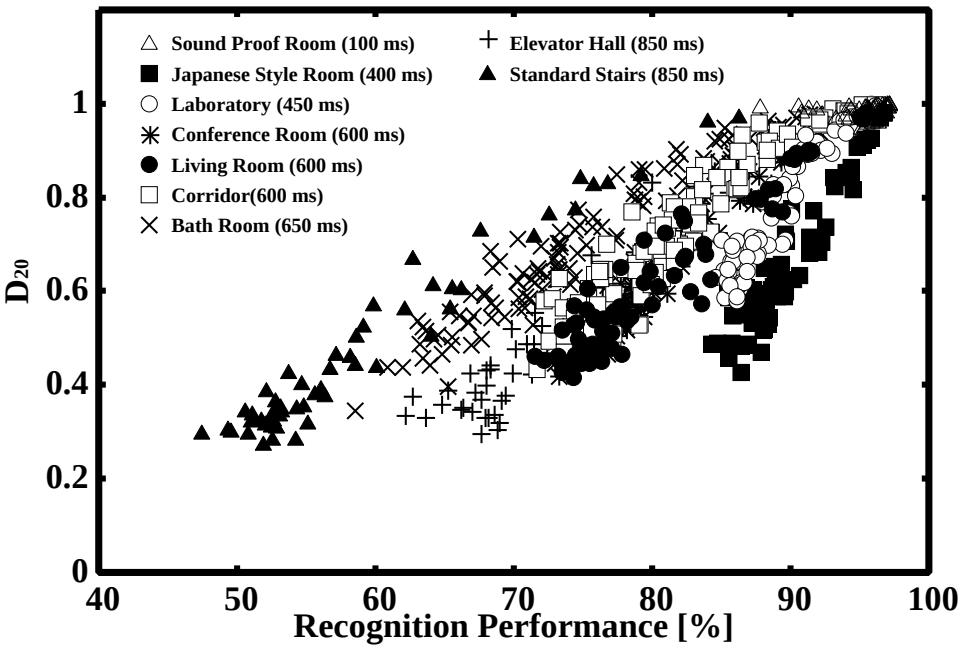


図 3.12 D_{20} と音声認識性能の関係（全体図）

3.6. 評価実験3 -残響下音声認識性能の予測-

3.6.1 実験条件

策定した音声認識指標の有効性を検証するために音声認識性能予測実験を行う．各環境の予測精度を比較するために，環境クローズテストおよび環境オープンテストを行う．環境クローズテストでは，環境が既知という条件で，学習時と同一環境の RSR-D_{20} から音声認識性能を予測する．本論文では表 3.3 (C) に示す 3 環境において策定した RSR-D_{20} を用いて同一環境の音声認識性能の予測を試みる．一方，環境オープンテストでは，環境が未知という条件で，学習時と残響時間は近いが環境が異なる RSR-D_{20} から音声認識性能を予測する．本論文では表 3.3 (C) に示す 3 環境のインパルス応答に基いて策定した RSR-D_{20} を用いて，表 3.3 (D) に示す 3 環境の音声認識性能の予測を試みる．予測精度評価には RSR-D_{20} から算出した音声認識性能の予測値とテストデータの真値との差を示す平均予測誤差を用いた．

なお提案手法との比較のために残響時間のみを用いた従来の音声認識性能予測も

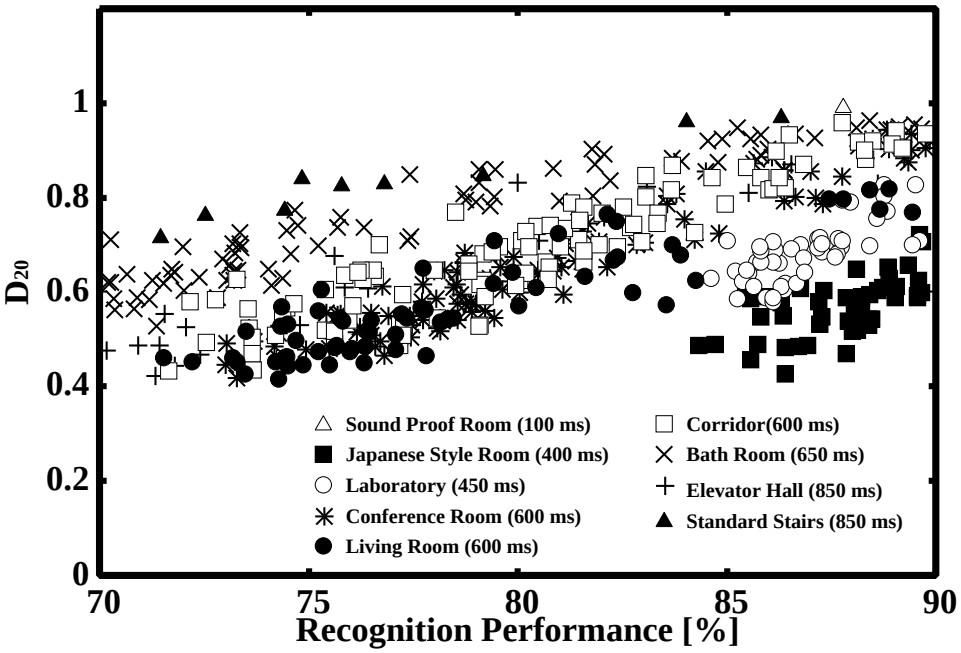


図 3.13 D_{20} と音声認識性能の関係（拡大図）

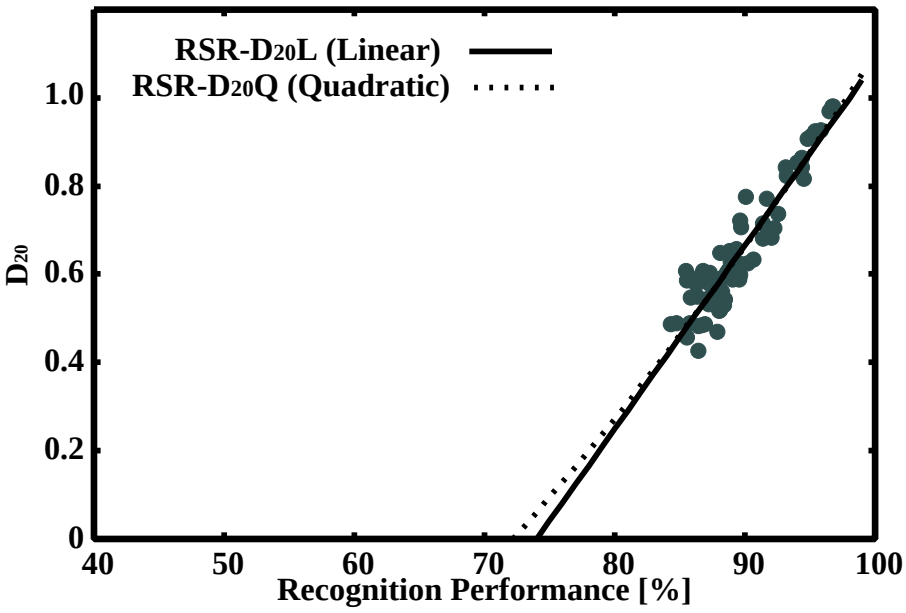


図 3.14 RSR-D_{20} と音声認識性能の関係（和室 (T_{60}=400 ms)）

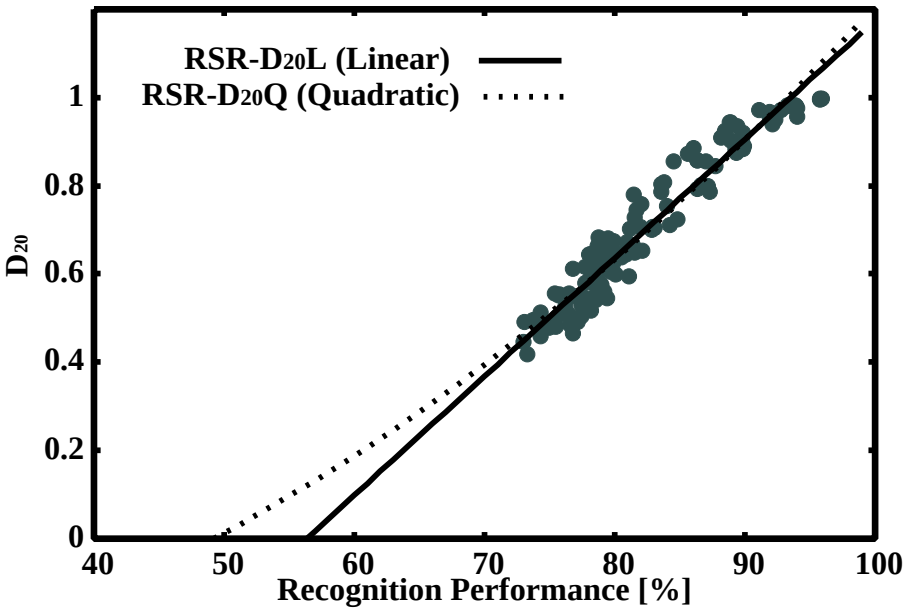


図 3.15 RSR-D_{20} と音声認識性能の関係（会議室 (T_{60}=600 ms)）

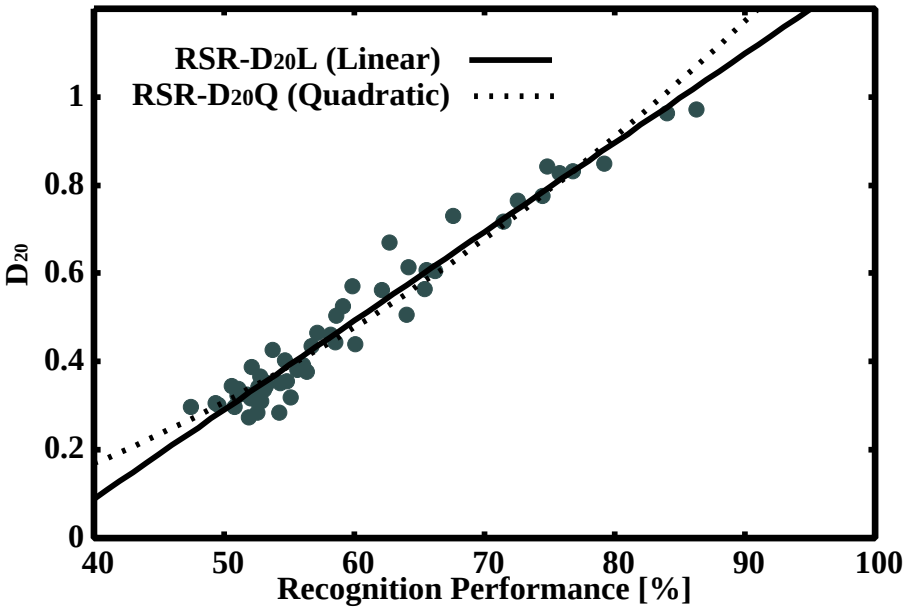


図 3.16 RSR-D_{20} と音声認識性能の関係（階段 (T_{60}=600 ms)）

併せて行った．従来法は表 3.3 (C) に示す 3 つのテスト環境の残響時間を基に，各環境に対する音声認識性能の平均に基づいて音声認識性能を予測した．

3.6.2 実験結果

図 3.17∼3.19 に各環境の環境クローズテストおよび環境オープンテスト結果を，表 3.5 に各テストの標準偏差を示す．高残響環境では RSR-D_{20} を用いた場合，平均性能予測誤差と標準偏差が従来法と比較して全体的に改善し，高精度に音声認識性能を予測できた．また残響時間のみを用いても十分に予測可能な低残響環境についても，同程度の予測精度を確認できた．そして環境オープンテストにおいて RSR-D_{20}Q の平均性能予測誤差と標準偏差ともに RSR-D_{20}L の結果よりも改善でき，高精度な音声認識性能の予測ができた．したがって音声認識性能と D_{20} の関係を 2 次曲線で近似した残響指標 RSR-D_{20}Q が残響下音声認識性能の予測指標として最適であると考えられる．

また本論文では音声認識性能と D 値の関係を 1 次，2 次曲線に基づき近似することで RSR-D_{20} を策定したが，さらに 3 次曲線 ($y = ax^3 + b$，x：音声認識性能，y：D 値，a, b：係数) を利用した近似も検討した．表 3.3 (C) に示す残響時間が異なる環境で RSR-D_{20} を策定した結果，各環境の相関係数が和室では 0.941，会議室では 0.959，階段では 0.960 となり，RSR-D_{20}L と RSR-D_{20}Q とほぼ同等の性能を達成した．これにより残響指標策定において高次数の曲線で近似する必要はなく RSR-D_{20}L や RSR-D_{20}Q を用いることで十分な性能が期待できると考えられる．

3.7. 評価実験 4 -CENSREC-4 を用いた音声認識性能予測-

前節では音声認識性能を推定するために提案した残響尺度の頑健性を独自収録したインパルス応答を用いて評価した．そして本節では策定した残響尺度の信頼性向上を目指して，現在公開されている音声コーパスリストの中から CENSREC（Corpus and Environments for Noisy Speech RECognition）[68] に着目し提案した残響指標

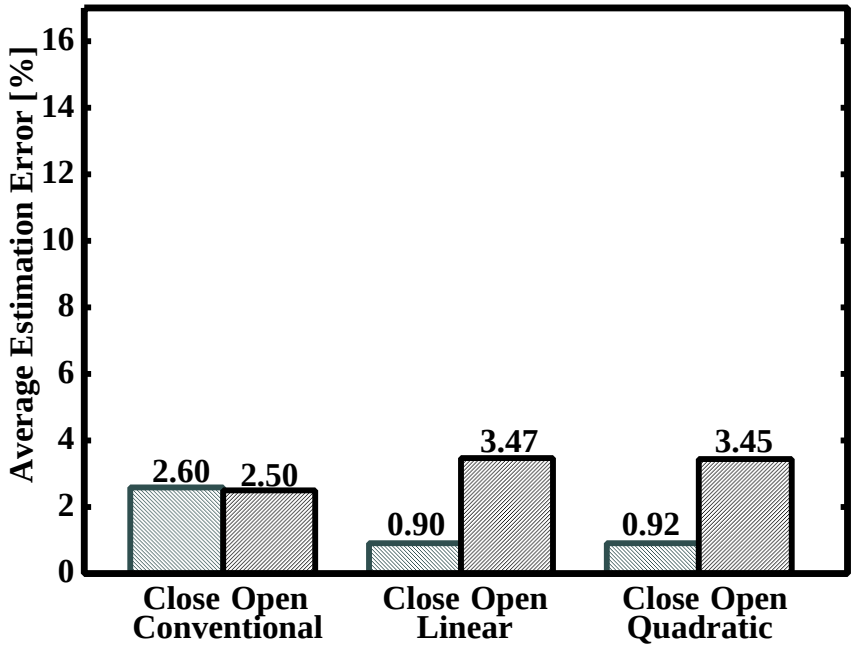


図 3.17 平均予測誤差（(a) 和室 (T_{60}=400 ms）)

表 3.5 標準偏差

	従来手法		RSR-D_{20}L (Linear)		RSR-D_{20}Q (Quadratic)	
	Close	Open	Close	Open	Close	Open
T_{60}=400 ms	3.10	3.26	1.10	3.62	1.13	3.60
T_{60}=650 ms	6.92	7.18	2.46	3.49	2.59	3.14
T_{60}=850 ms	8.80	17.64	2.41	5.35	2.81	5.23

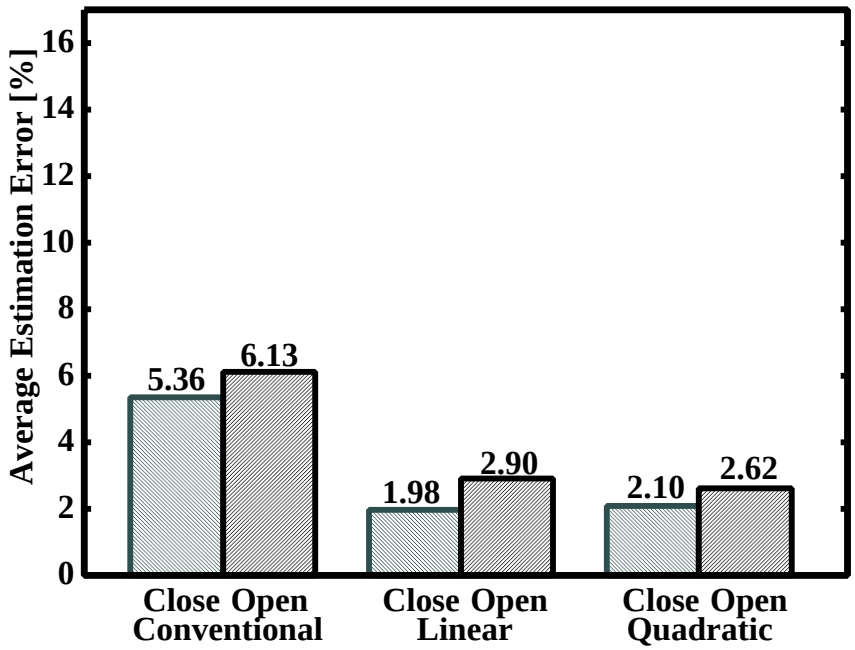


図 3.18 平均予測誤差（(b) 会議室 (T_{60}=600 ms)）

の頑健性を検証する．CENSREC とは，雑音・残響環境下の音声認識タスクの共通評価フレームであり，様々な実環境データや評価ツールが用意されている．ここでは残響環境下音声認識の評価環境として構築された CENSREC-4[69] を用いて策定した残響尺度の頑健性を検証する．CENSREC-4 は残響時間 (T_{60}) が異なる環境 (全 8 環境) のインパルス応答が 1 系分収録されている．

3.7.1 実験条件

表 3.6 に示す残響時間が異なる 3 環境において，音声認識性能推定用の残響指標 RSR-D_{20}L を策定した．なお指標策定に用いた評価音声として，CENSREC に収録されている 4,004 発話の連続数字音声を用いた．そして，策定した音声認識性能を予測するための残響指標の頑健性を表 3.7 に示す CENSREC-4 内の 5 環境のインパルス応答を用いて検証した．予測精度の評価尺度として RSR-D_{20}L から算出した音声認識性能の推定値とテストデータの真値との差分平均を表す平均推定誤差値を用

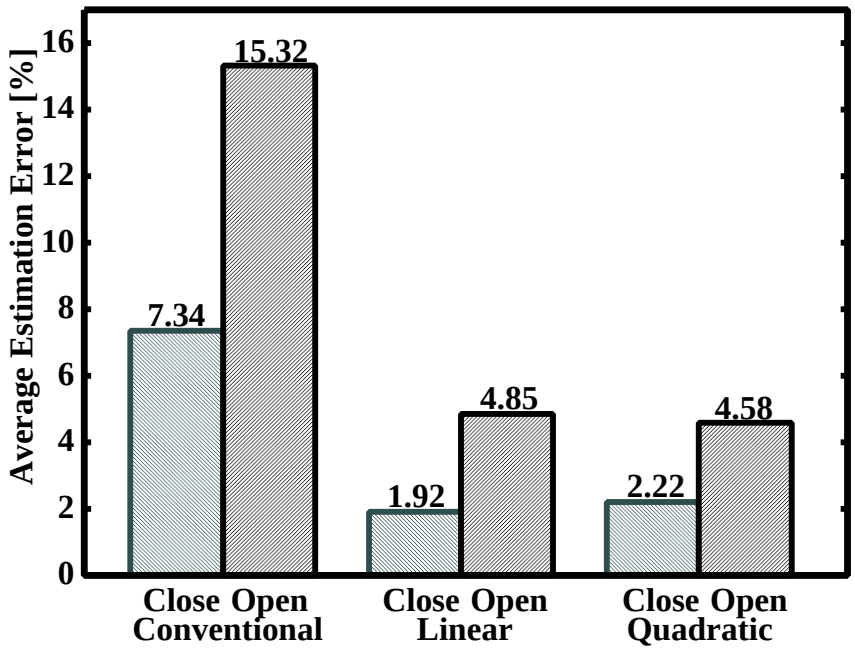


図 3.19 平均予測誤差（(c) 階段 (T_{60}=850 ms)）

いた．また従来手法として，前節の評価実験と同様に残響時間のみを用いた音声認識性能の予測法とした．なお，予測環境と同一残響時間の残響指標が学習セットに存在しない場合は，最近傍残響時間の残響指標より音声認識性能の予測を行った．

3.7.2 実験結果

残響指標 RSR-D_{20}L の策定結果を図 3.20 に，そして策定した残響指標を用いた音声認識性能の予測結果を表 3.8 に示す．表 3.8 には，各推定環境のインパルス応答の D_{20} と音声認識性能の真値も示す．これらの残響指標を用いて性能推定した結果，提案手法の平均推定誤差値が従来手法より小さかったことより，音声認識性能の予測における提案指標の有効性を確認した．これは性能推定に残響時間のみを用いる従来手法では性能推定値が系に関係なく一意に決定する問題点を各系のインパルス応答によって変動する D 値を用いることで解消できたためであると考えられる．

しかし，エレベータホールでの予測誤差が約 20 %であることから高残響環境下の

表 3.6 残響指標 RSR-D_{20}L の策定条件

環境	和室 (T_{60}=400 ms，72ヶ所) 会議室 (T_{60}=600 ms，120ヶ所) 階段 (T_{60}=850 ms，120ヶ所)
音声	連続数字音声 (4,004 発話)
計測距離	100〜5,000 mm
話者数	104 話者 (女性：52 名, 男性：52 名)
デコーダー	HTK
特徴量	MFCC（12 次元）+ΔMFCC（12 次元）+ΔΔMFCC（12 次元）+ log Power（1 次元）+Δlog Power（1 次元）+ΔΔlog Power（1 次元）
分析長	25 ms (ハミング窓)
シフト長	10 ms

表 3.7 音声認識性能推定実験条件

音声認識 推定環境	オフィス (T_{60}=250 ms，1ヶ所) 和室 (T_{60}=400 ms，1ヶ所) 会議室 (T_{60}=650 ms，1ヶ所) リビング (T_{60}=650 ms，1ヶ所) EV ホール (T_{60}=850 ms，1ヶ所)
音声	連続数字音声 (4,004 発話)
話者数	104 話者 (女性：52 名, 男性：52 名)
デコーダー	HTK
特徴量	MFCC（12 次元）+ΔMFCC（12 次元）+ΔΔMFCC（12 次元）+ log Power（1 次元）+Δlog Power（1 次元）+ΔΔlog Power（1 次元）
分析長	25 ms (ハミング窓)
シフト長	10 ms

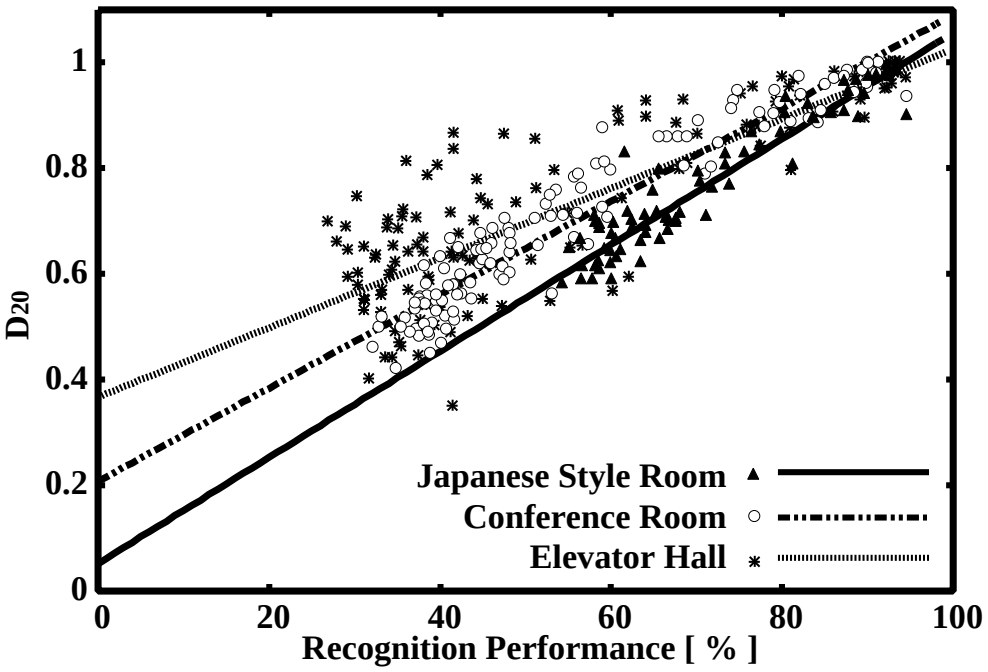


図 3.20 RSR-D_{20}L の策定結果

性能推定が困難であった．これは同じ D 値でも音声認識性能が異なる系が多数存在するためであると考えられる．したがって，提案手法は従来手法よりも高精度に推定できたが，更なる予測精度向上のために同じ D 値でも正確な予測ができる補正指標の検討が今後の研究課題であると考えられる．

3.8. 評価実験 5 -音声認識性能予測のコスト評価-

3.8.1 実験条件

ここでは，従来･提案手法による音声認識性能予測に必要なデータ量および計算時間を評価する．ここでの従来の音声認識性能評価とは，クリーン音声にインパルス応答を畳み込んだ評価音声データを大量に用意して音声認識性能を予測する手法を指す．そして提案手法は，インパルス応答から室内音響指標を算出して音声認識性

表 3.8 音声認識性能の予測結果

予測環境	オフィス		和室		会議室		リビング		EV ホール	
予測指標	T_{60}	D_{20}	T_{60}	D_{20}	T_{60}	D_{20}	T_{60}	D_{20}	T_{60}	D_{20}
認識率 [%]	93.1		54.3		74.1		65.3		30.7	
予測値 [%]	70.5	92.6	70.5	56.9	56.0	85.2	56.0	60.7	52.3	50.9
予測誤差 [%]	**22.6**	**0.5**	**16.2**	**2.9**	**18.1**	**11.1**	**9.3**	**4.6**	**21.6**	**20.2**

能を予測する．また本実験では計算機サーバ（Debian Linux 6.0.7，CPU:Intel Xeon 3.60 GHz，メモリ: 16 GB）を用いて評価した．

なお，従来手法を用いて正確に音声認識性能を評価するには，大量の音声データを用いて統計的な処理を行う必要がある．そのため，本実験ではクリーン音声として 12 話者分の ATR216 音素バランス単語を用い，その他は表 3.1 に示す条件で評価を行った．

3.8.2 実験結果

音声認識性能予測に必要なデータ量を表 3.9 に，そして計算時間を表 3.10 に示す．表 3.9 に示すデータ量の結果より，従来の音声認識性能評価では合計で約 20 GB の評価音声データを必要としていたのに対して，提案手法を用いることでデータ量を約 9.6 MB まで大幅削減することができた．また，表 3.10 に示す計算時間においても，従来手法 (1 環境あたり約 4 分) と比較して，提案手法 (1 環境あたり 1 ミリ秒) を用いることで，実時間で音声認識性能を予測することができた．これらの評価結果より，提案手法を用いることで，音声認識性能予測に要するコストを大幅に削減できることが明らかとなった．

3.9. まとめ

これまでに残響環境下に対する頑健な音声認識のための残響指標は存在せず，残響下の音声認識性能の予測が困難であるという問題があった．従来，音声認識の難

表 3.9 音声認識性能予測に必要なデータ量

従来手法	
音声データ	216 単語 × 12 話者 = 81 MB
合計	81 MB × 248ヶ所= **20.088 GB**

提案手法	
D 値	
合計	研究室 (72ヶ所 = 2.1 MB) + 廊下 (120ヶ所 = 4.5 MB) + 階段 (56ヶ所 = 3.0 MB) = **9.6 MB**

表 3.10 音声認識性能予測の計算時間

従来手法	
1. インパルス応答の畳込み: 214.9 秒 2. 音声認識: 120.1 秒	
合計	214.9 秒+120.1 秒= **335 秒**

提案手法	
1. **D 値の計算**: 1 ミリ秒	
合計	**1 ミリ秒**

しさを判別する残響尺度として同一室内で固有の値をとる残響時間 (T_{60}) が利用されていたが，同一環境でも計測箇所によって音声認識性能が変動することから，残響時間のみで音声認識性能を予測することは困難であった．そこで本章では，音声認識性能を残響に対して頑健かつ簡便に予測できる残響指標 RSR-D_n を提案し，音声認識性能の高精度な予測を試みた．はじめに 3.2 節で，室内音響指標が高精度に音声認識性能を予測できる残響指標である可能性を示した．そして，3.3 節で提案手法の詳細について述べた．最後に 3.4〜3.8 節で，提案手法を用いて残響環境における音声認識性能の予測実験を行い，その有効性を示した．今後は MTF (Modulation Transfer Function)[66] などの周波数指標も含めた音声認識に適した残響指標の確立を目指す．

第4章　室内音響指標とPESQを用いた雑音・残響下における頑健な音声認識性能予測

4.1. はじめに

実環境において音声認識システムを利用すると，雑音や残響などの外乱の影響を受けて音声認識性能が著しく劣化する．ここで外乱による性能劣化を事前に予測できれば，その結果に基づいて性能改善手法を前処理等に反映できる．特にスマートフォンなどの携帯を許容する音声インタフェースを想定すると，複数の外乱成分が存在する環境（例えば，残響と雑音が混在する環境など）における頑健な音声認識性能の予測手法の確立が必要となる．これまでに雑音下ではPerceputual Evaluation of Speech Quality（PESQ）を，残響下では室内音響指標（D値，残響時間（T_{60}））を用いて音声認識性能を予測する手法が提案されていた．しかし，これらの手法には予測指標が表現できない外乱が混入すると音声認識性能の予測精度が低下する問題がある．そこで本章では，雑音・残響下における音声認識性能の予測精度を向上させるために，PESQ，D値，T_{60} を用いた雑音・残響指標 Noisy-and-Reverberant Speech Recognition criteria with PESQ and Acoustic parameters（NRSR-PA）の策定を検討する．

本章の構成を以下に示す．4.2節で，提案手法に用いる室内音響指標について述べる．4.3節で提案手法の詳細について述べる．4.4~4.6節で，提案手法を用いて外乱環境（雑音・残響環境）における音声認識性能の予測実験を行い，その結果について述べる．4.7節で，本章のまとめを述べる．

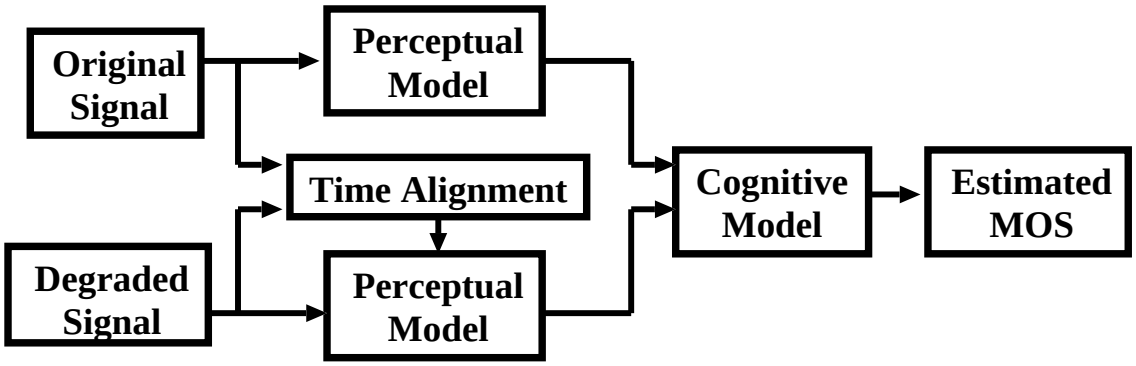


図 4.1 PESQ スコアの計測方法

4.2. 室内音響指標と PESQ

雑音に対する音声認識の難しさを表現するために，これまでに山田らが PESQ[70] を用いて音声認識性能を予測する手法を確立している [79]．ここ数年の間に，評価信号の主観的な品質を高精度かつ客観的に推定できるようになってきており [71, 72]，その中でも特に PESQ は代表的な客観品質評価法であり，音声データの品質を評価するために積極的に用いられている [73, 74, 75, 76, 77, 78]．ここで，図 4.1 に PESQ スコアの算出アルゴリズムを示す．はじめに，クリーン信号と劣化信号を知覚モデルを用いてセルと呼ばれる時間・バークスペクトル領域に射影する．そして，セル間のひずみから認知モデルを用いて主観 MOS の推定値（PESQ 値）を計測する．山田らは，PESQ と雑音下での音声認識性能に強い相関が保たれていることを確認し，雑音下で音声認識性能が予測できることを明らかにした．

筆者は前章で明らかにした通り，残響による音声認識性能の低下を予測するために，室内音響指標 [60] の D 値に基づいて，残響時間ごとに策定した残響指標 Reverberant Speech Recognition with D_n (RSR-D_n) を用いて音声認識性能を予測する手法を提案している [80]．これはインパルス応答の初期と後続の反射エネルギー比と音声認識性能の間に強い相関関係があることを明らかにし，このエネルギー比を表現できる室内音響指標の D 値を用いて高精度に音声認識性能を予測できることを実証している．D 値とは系のインパルス応答を基に式 (3.3) より算出され，直接音と初期反射音のエネルギーに対する直接音と全ての反射音のエネルギー比を示す．特に直接音

表 4.1 実験条件（従来指標と音声認識性能の関係分析）

音声認識環境	和室 (T_{60}=400 ms，72ヶ所) 会議室 (T_{60}=600 ms，120ヶ所) エレベータホール (T_{60}=850 ms，120ヶ所)
音声	ATR 音素バランス 216 単語 [42, 43, 44] 女性：2 話者, 男性：2 話者
デコーダ	Julius rev. 4.2.1 [45, 46, 47]
HMM	IPA モノフォンモデル（性別依存）
音声特徴量	MFCC（12 次元）+ΔMFCC（12 次元）+ ΔPower（1 次元）
雑音	白色雑音
SNR	-5, 0, 5, 10, 15 and 20 dB
分析長	25 ms (ハミング窓)
シフト長	10 ms

と初期反射音のエネルギーが大きいほど D 値は向上を示し，後続残響のエネルギーが大きいほど低下する．D 値は音声認識性能に影響を与える初期反射音と後続残響音の割合を表現できることから，音声認識性能に与える劣化の度合いを表現するパラメータとして有効であることが明らかとなっている [80]．

従来指標の問題点として，それぞれの指標が表現できる外乱成分とは異なる外乱成分が混入することで音声認識性能の予測精度が劣化することが挙げられる．ここで実際に雑音と残響が混在する環境において，従来指標と音声認識性能の関係を評価した．この実験では，表 4.1 に示す条件において，クリーン音声に残響を畳み込んだ信号に白色雑音を所望の SNR で加算した評価音声を用いて音声認識を行った．図 4.2 に D 値と音声認識性能の関係（会議室：T_{60}=600 ms，SNR：-5∼20 dB）を，図 4.3 に PESQ と音声認識性能の関係（和室：T_{60}=400 ms，会議室：T_{60}=600 ms，エレベータホール：T_{60}=850 ms，SNR：10, 20 dB）を示す．まず，図 4.2 の残響指標と音声認識性能の関係より，雑音（特に SNR）の影響を受けたことによって，同じ D 値に対して音声認識性能のばらつきが確認できる．また図 4.3 の雑音指標と音声認識性能の関係においても，残響（残響時間や発話位置）の影響を受けたことに

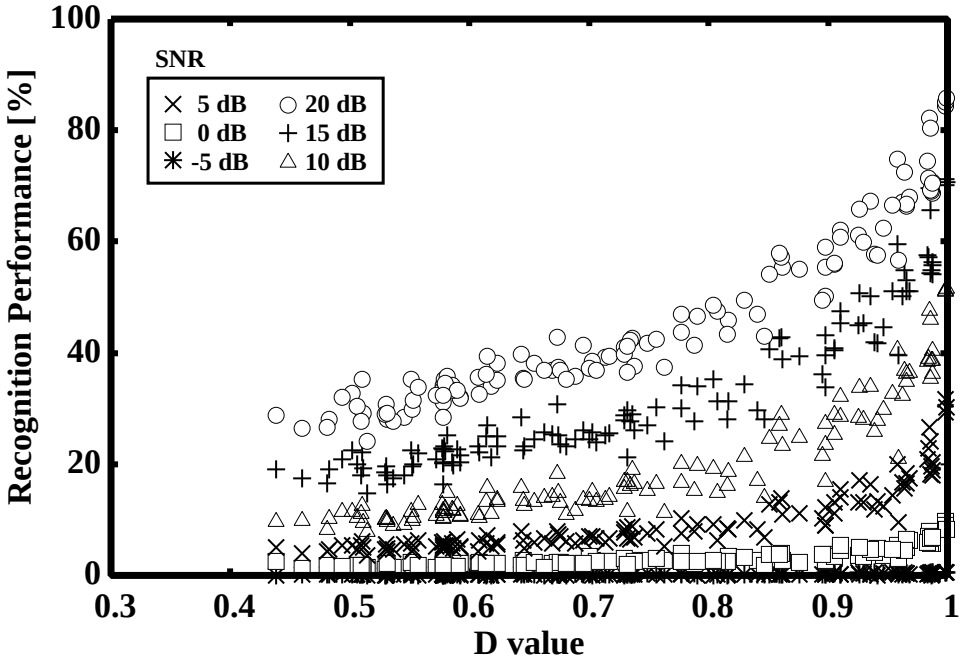


図 4.2 D_{20} と音声認識性能の関係（会議室，SNR：-5∼20 dB）

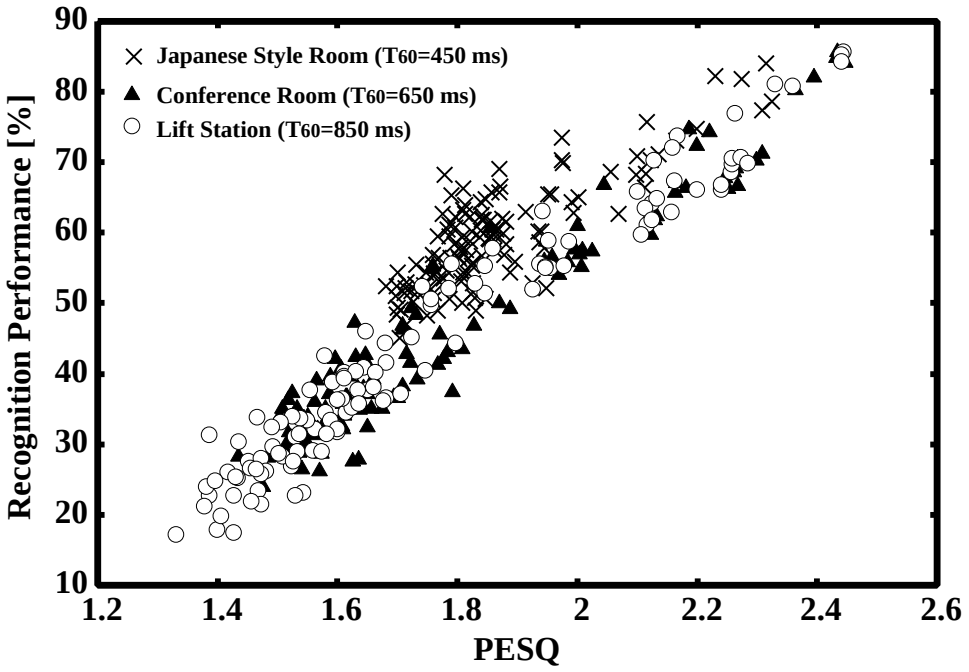


図 4.3 PESQ と音声認識性能の関係（和室，会議室，エレベータホール，SNR : 10, 20 dB）

よって，同じ PESQ に対して音声認識性能のばらつき（特に PESQ が 1.8 のときの和室の音声認識性能に 20 %以上のばらつき）が確認できる．これらの結果より，1 種類の指標だけで複数の外乱成分（雑音と残響）が音声認識性能に与える影響を表現することに限界があると予想される．

ただし，図 4.2 に着目すると，D 値と音声認識性能の関係が SNR に依存する傾向が確認できることから，D 値と SNR を組み合わせることで雑音・残響下において高精度な音声認識性能予測が期待できる．しかしながら，SNR を雑音（特に非定常雑音）と音声が混在する観測信号から正確に推定することは容易ではない上に計算コストの増大にも繋がるため，本研究においては SNR に代わって雑音成分が音声認識システムに与える影響を表現できる別の雑音指標の検討を考える．そこで，本研究では PESQ を用いることで，雑音成分が音声認識性能に与える影響を SNR や定常・非定常性に依存することなく正確かつ簡便に表現できるという従来研究の知見 [79] に着目し，雑音と残響成分が音声認識性能に与える影響を同時に表現できる新しい外乱指標の策定を試みる．具体的には，残響指標の D 値や残響時間では表現しきれない雑音成分の影響を雑音指標の PESQ で表現できるような雑音・残響指標を策定して，雑音と残響が混在する環境における頑健な音声認識性能の予測に取り組む．

4.3. 音声認識性能予測アルゴリズム

本研究では，前節で指摘した雑音･残響指標の問題点を解決するために，雑音･残響に対して頑健な音声認識性能予測指標を提案する．具体的には，雑音指標（PESQ），残響指標（室内音響指標）と音声認識性能の関係を重回帰分析して算出された予測式を予測指標とし，その指標を使って音声認識性能の予測を試みる．

4.3.1 雑音・残響指標 NRSR-PA の策定

音声認識性能を予測するための雑音・残響指標 NRSR-PA の策定アルゴリズムを図 4.4 に示す．

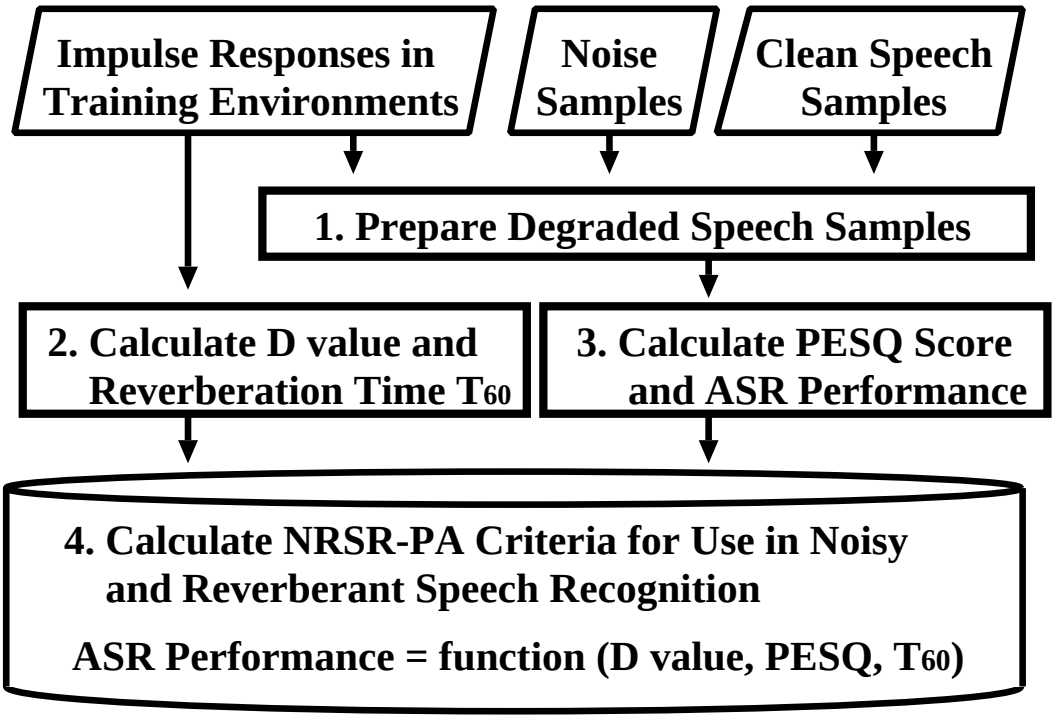


図 4.4 雑音・残響下音声認識における性能予測指標の策定手順

[事前準備] インパルス応答，雑音，クリーン音声の計測

雑音・残響下音声認識性能の予測指標を策定するために，事前にインパルス応答，雑音，クリーン音声を計測する．インパルス応答については，残響時間が異なる環境において，入出力間距離や発話方位が異なる伝達特性を数十〜数百箇所で計測する．また雑音は音声認識システムの利用環境に存在する雑音を計測し，クリーン音声は大規模音声データベースを利用したり，ヘッドセットマイクなどで話者の音声を近接収録する．

[**Step 1**] 音声認識評価と **PESQ** 計測用の劣化音声の作成

事前準備で計測した信号から音声認識評価と PESQ 計測に用いる劣化音声を作成する．具体的には，インパルス応答とクリーン音声を畳み込んだ残響信号に対して雑音を所望の SNR で加算する．

[**Step 2**] **D** 値，残響時間の算出

事前準備で計測した各インパルス応答に対して式 (3.3) に基づいて D 値を算出する．また初期反射音と後続残響の境界時間を表す n は，音声認識性能と D 値の最大相関値を示すように設定する必要がある．なお先行研究 [80] より n=20 ms が適切な境界時間であることが明らかとなっている．またインパルス応答から D 値と併せて残響時間を式 (2.6) に基づいて算出した残響曲線から算出する．なお残響時間は同一室内では同じ値をもつため，計測したインパルス応答の全てから残響時間を算出する必要は無く，数箇所のインパルス応答から算出した残響時間の平均を各環境の残響時間とすることが一般的である．

[**Step 3**] **PESQ** と音声認識性能の計測

Step 1 で作成した劣化音声を用いて，PESQ と音声認識性能を計測する．なお PESQ の計測には，劣化音声と併せてクリーン音声を用いる必要がある．そし

て音声認識性能は Julius[45] などの音声認識エンジンを用いて算出する．

[Step 4] 音声認識性能の予測式の算出

雑音・残響下における音声認識性能を予測するために，Step 2 と Step 3 で計測した D 値，PESQ，音声認識性能に対して残響時間ごとに重回帰分析を行い，雑音・残響指標 NRSR-PA の評価関数を策定する．策定した雑音・残響指標 NRSR-PA を示す $\mathrm{R}_{\mathrm{Est}}(x_{\mathrm{d}}, x_{\mathrm{p}}, T)$ は，式 (4.1) で表現される．

$$\mathrm{R}_{\mathrm{Est}}(x_{\mathrm{d}}, x_{\mathrm{p}}, T) \quad = \quad A_T \cdot x_{\mathrm{d}} + B_T \cdot x_{\mathrm{p}} + C_T, \tag{4.1}$$

x_{d} は D 値を，x_{p} は PESQ を，T は残響時間を，A_T, B_T, C_T は重回帰分析によって得られた回帰係数を表す．式 (4.1) は，D 値と PESQ の線形和で表現される音声認識性能の予測式が残響時間ごとに構成されることを表している．なお回帰係数の予測方法は，最小二乗法 [67] を用いる．

4.3.2 雑音・残響指標 NRSR-PA を用いた音声認識性能予測

4.3.1 で策定した雑音・残響指標 NRSR-PA を用いた音声認識性能の予測アルゴリズムを図 4.5 に示す．

[事前準備] インパルス応答，雑音，クリーン音声の計測

雑音・残響下音声認識性能を予測するために，事前に発話者と音声認識システム間のインパルス応答と劣化音声を事前に計測する．なお，音声認識性能の予測にはクリーン音声も併せて必要であるが，本研究では大規模音声データベースや事前にヘッドセットマイクなどで近接収録した話者音声を利用する．

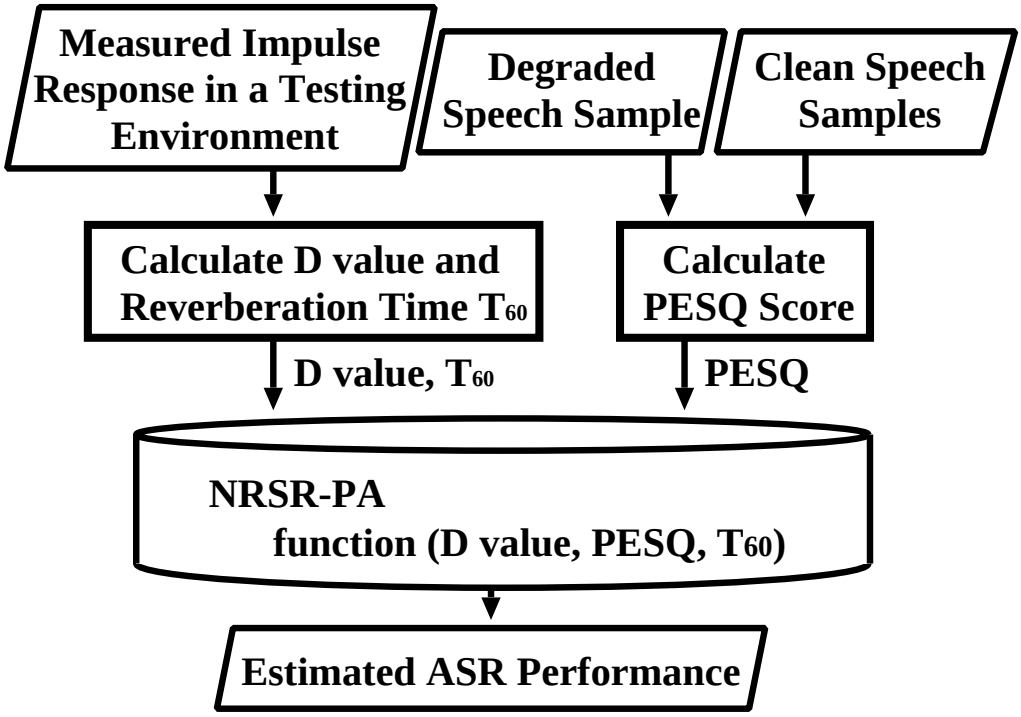


図 4.5 雑音・残響下音声認識における性能予測手順

[Step 1] D 値，残響時間，PESQ の計測

事前準備で計測したインパルス応答から式 (3.3) に基づいて D 値を，式 (2.6) に基づいて残響時間を計測する．また計測したクリーン音声と劣化音声から PESQ を併せて計測する．

[Step 2] 音声認識性能の予測

Step 1 で計測した D 値，PESQ と残響時間を式 (4.1) の雑音・残響指標 NRSR-PA の評価関数に適用することで音声認識性能の予測を試みる．

4.4. 評価実験1 -雑音・残響指標NRSR-PAの策定-

4.4.1 実験条件

D 値，PESQ，音声認識性能の関係を分析するために表 4.2 に示す 6 つの学習環境にて計 560 箇所のインパルス応答を計測した．なお表 4.2 に示す環境は，様々な残響環境を想定するために，残響時間が異なる環境でインパルス応答を計測した．また各残響環境の中でも，近距離発声だけでなく遠隔発声も考慮したハンズフリー発話環境を想定して 10〜500 cm の入出力間距離および正背左右の放射面の条件で計測を行った．本評価実験では，残響音声に対して白色雑音と電子協騒音データベース [81] の工場騒音を 8 種類の SNR で加算した．音響モデルは，IPA の日本語ディクテーション基本ソフトウェア [2] に収録されている性別依存モノフォンモデルを使用した．なお音声認識性能は特徴量や言語・音響モデルなどに依存するため，雑音・残響尺度策定と音声認識性能予測における認識条件を統一させた．

4.4.2 実験結果

残響時間が異なる 3 環境 (和室 (T_{60}=400 ms)，会議室 (T_{60}=600 ms)，エレベータホール (T_{60}=850 ms)) における D_{20}，PESQ，音声認識性能の関係を図 4.6〜4.11 に

表 4.2 実験条件

NRSR-PA 策定環境	和室 (T_{60}=400 ms，72ヶ所) 会議室 (T_{60}=600 ms，120ヶ所) エレベータホール (T_{60}=850 ms，120ヶ所)
音声認識性能 予測環境 (オープン環境)	研究室 (T_{60}=450 ms，72ヶ所) 廊下 (T_{60}=650 ms，120ヶ所) 階段 (T_{60}=850 ms，56ヶ所)
音声	ATR 音素バランス 216 単語 [42, 43, 44] 女性：2 話者, 男性：2 話者
デコーダ	Julius rev. 4.2.1 [45, 46, 47]
HMM	IPA モノフォンモデル（性別依存）
特徴量	MFCC（12 次元）+ΔMFCC（12 次元）+ΔPower（1 次元）
雑音	白色雑音, 工場騒音
SNR	-5, 0, 5, 10, 20, 30, 40, 50 dB
分析長	25 ms (ハミング窓)
シフト長	10 ms

示す．そして，図 4.6∼4.11 には，重回帰分析により得られた式 (4.1) の係数値（表 4.3）を用いて近似平面を描画している．また，そのときの相関係数を表 4.4 に示す．

まず，表 4.3 の NRSR-PA の係数値より，和室 (T_{60}=400 ms) における係数値 B_T，C_T を除くと，環境に依らずに同等の係数値が得られた．このことから，和室 (T_{60}=400 ms) のような低残響環境については，環境別に予測指標 NRSR-PA を策定する必要があるものの，それ以上の高残響環境であれば雑音や残響の環境に依存せずに音声認識性能を予測できると考えられる．

表 4.4 に示す相関係数より，NRSR-PA の相関係数が全ての雑音・残響環境において 0.93 を上回り，D 値，PESQ と音声認識性能の関係を高精度に近似できた．一方，D 値単体の相関係数は最大で 0.32 であり，雑音・残響下における音声認識性能と D 値の関係を高精度に近似することが難しかった．なお PESQ 単体の相関係数は最大

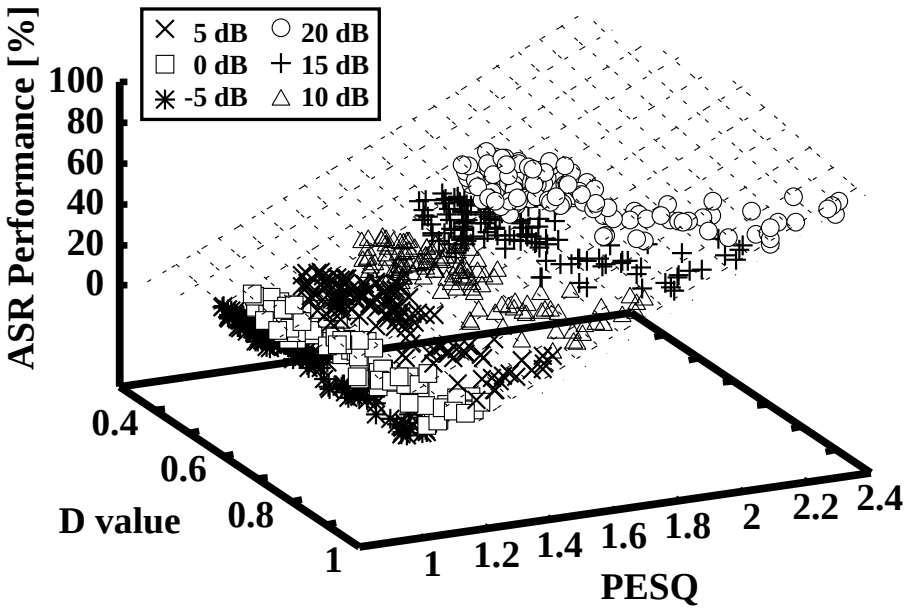


図 4.6 D 値，PESQ，音声認識性能の関係 (白色雑音，和室 (T_{60}=400 ms))

で 0.91 であったが，さらに D 値を組み合わせることで相関係数が最大で 0.96 に向上したことから，NRSR-PA を用いることで高精度な音声認識性能の予測が期待できると考えられる．この結果から音声認識性能の予測値を PESQ と D 値の線形結合で表現した NRSR-PA は有力な雑音・残響指標であることがわかった．

表 4.3 重回帰分析で得られた NRSR-PA の係数値

	白色雑音			工場騒音		
	A_T	B_T	C_T	A_T	B_T	C_T
T_{60}=400 ms	-35.0	74.0	-54.8	-33.6	68.1	-41.0
T_{60}=600 ms	-33.5	58.4	-28.5	-35.4	57.9	-23.4
T_{60}=850 ms	-26.0	57.4	-33.0	-29.5	58.0	-26.9

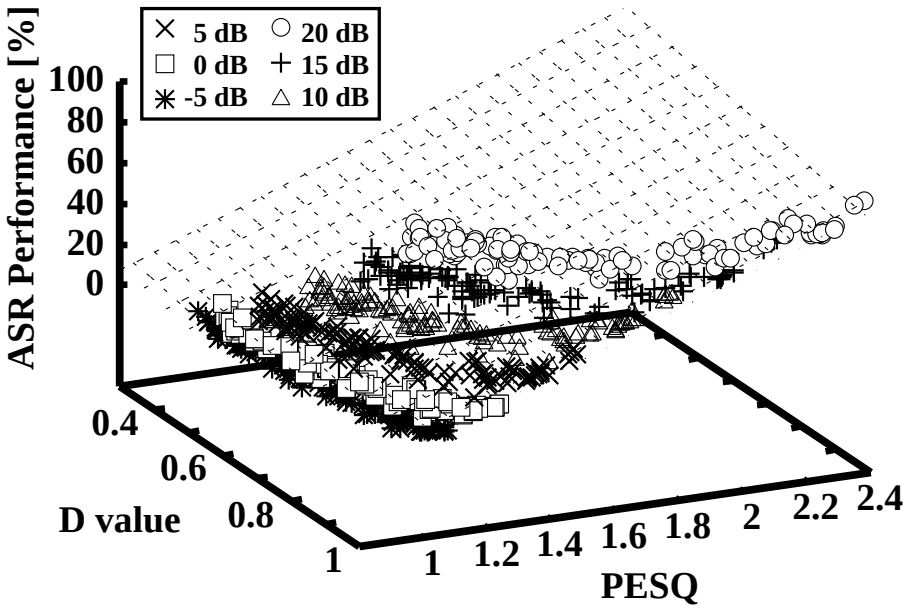


図 4.7 D 値，PESQ，音声認識性能の関係 (白色雑音，会議室 (T_{60}=600 ms))

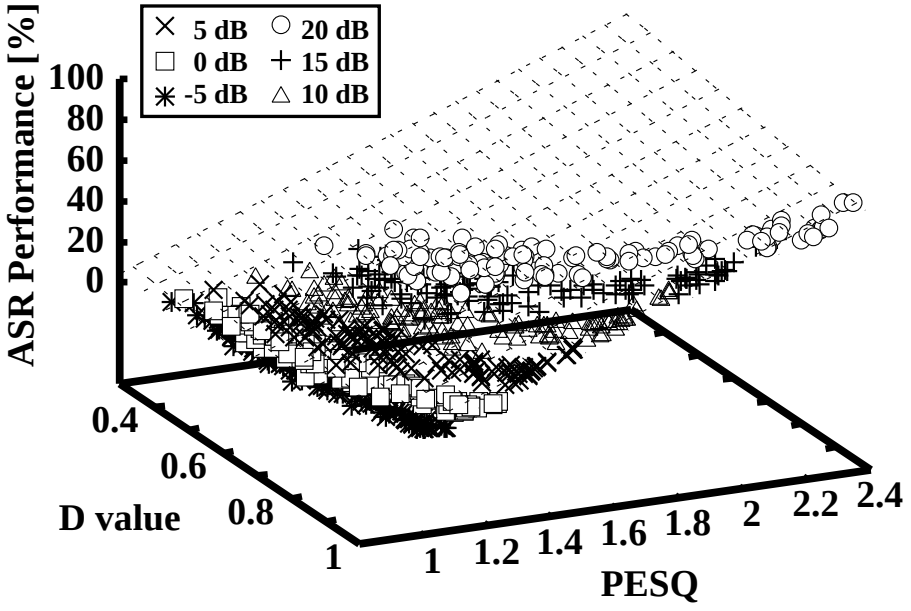


図 4.8 D 値，PESQ，音声認識性能の関係 (白色雑音，階段 (T_{60}=850 ms))

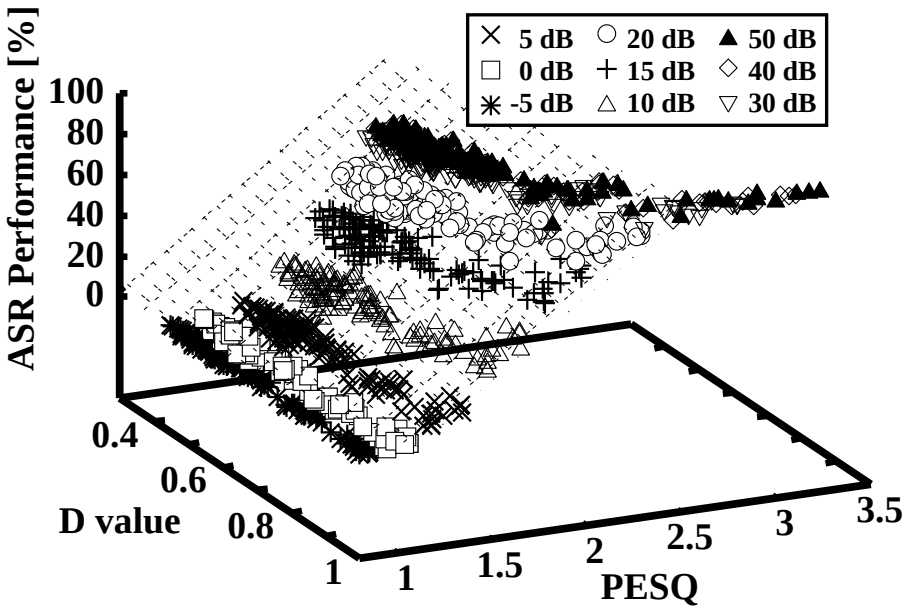


図 4.9 D 値，PESQ，音声認識性能の関係 (工場騒音，和室 (T_{60}=400 ms))

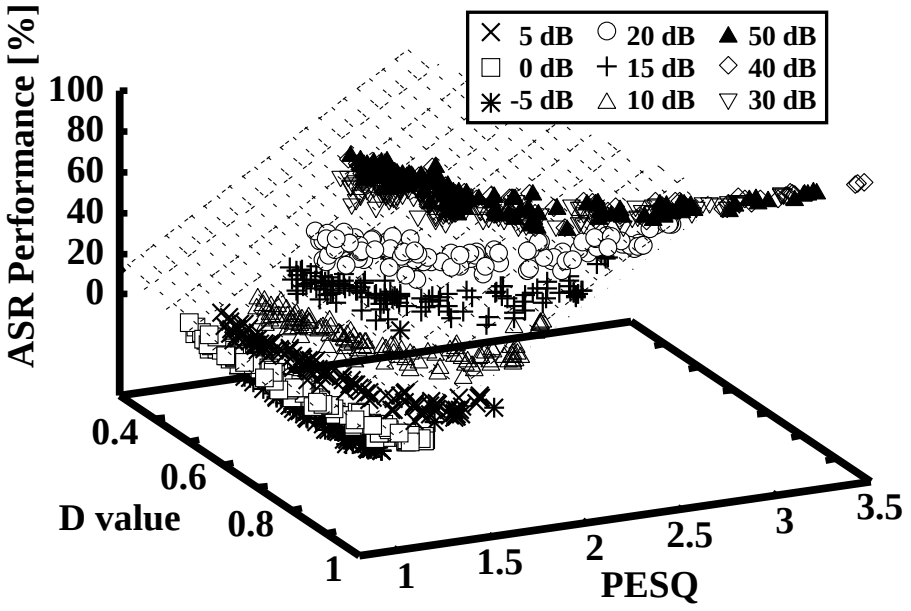


図 4.10 D 値，PESQ，音声認識性能の関係 (工場騒音，会議室 (T_{60}=600 ms))

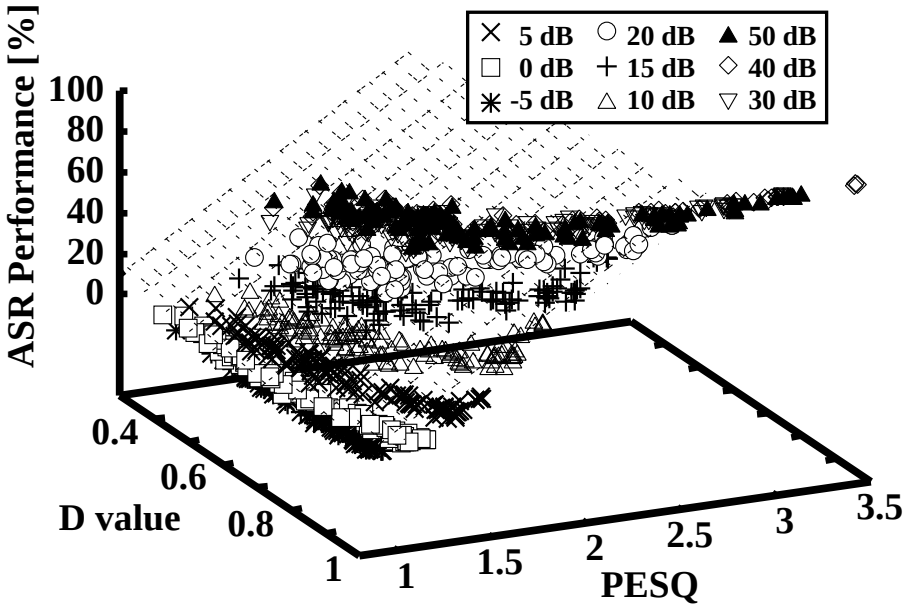


図 4.11 D 値，PESQ，音声認識性能の関係 (工場騒音，階段 (T_{60}=850 ms))

4.5. 評価実験2 -雑音・残響下音声認識性能の予測-

4.5.1 実験条件

4.4 節で策定した雑音・残響指標 NRSR-PA の有効性を検証するために音声認識性能予測実験を行う．各環境の予測精度を比較するために，環境クローズテストおよび環境オープンテストを行う．環境クローズテストでは，残響環境が既知という条件で，学習時と同一環境の NRSR-PA から音声認識性能を予測する．本研究では表 4.2 に示す 3 環境 (和室 (T_{60}=400 ms)，会議室 (T_{60}=600 ms)，エレベータホール (T_{60}=850 ms)) において策定した NRSR-PA を用いて同一環境の音声認識性能の予測を試みる．一方，環境オープンテストでは，残響環境が未知という条件で，学習時と残響時間は近いが環境が異なる NRSR-PA から音声認識性能を予測する．本研究では表 4.2 に示す 3 環境 (和室 (T_{60}=400 ms)，会議室 (T_{60}=600 ms)，エレベータホール (T_{60}=850 ms)) において策定した NRSR-PA を用いて，3 つのオープン環境 (研究室 (T_{60}=400 ms)，廊下 (T_{60}=600 ms)，階段 (T_{60}=850 ms)) の音声認識性能の

表 4.4 重回帰分析で得られた相関係数

白色雑音	T_{60}=400 ms	T_{60}=600 ms	T_{60}=850 ms
D 値	0.11	0.24	0.32
PESQ	0.91	0.90	0.90
NRSR-PA	**0.96**	**0.94**	**0.94**

工場騒音	T_{60}=400 ms	T_{60}=600 ms	T_{60}=850 ms
D 値	0.08	0.20	0.29
PESQ	0.90	0.89	0.90
NRSR-PA	**0.94**	**0.93**	**0.93**

予測を試みる．なお，音声認識性能予測では，雑音・残響指標 NRSR-PA の策定とは異なる雑音区間を用いて評価を行った．予測精度評価には NRSR-PA から算出した音声認識性能の予測値とテストデータの真値との差を示す平均予測誤差を用いた．なお本研究では，従来手法として D 値と PESQ を個別を用いて音声認識性能予測も併せて行った．

4.5.2 実験結果

図 4.12∼4.17 に各環境の環境クローズテストおよび環境オープンテスト結果を示す．また図中のエラーバーは，音声認識性能の予測誤差に対する標準偏差を表す．評価実験より，提案手法は，全ての残響環境や SNR に対して，D 値単体や PESQ 単体と同程度あるいはそれ以上の予測性能 (全環境で 10%以下の平均性能予測誤差) を達成できていることを確認した．なお，D 値単体では SNR が 10∼20 dB の音声に対しては D 値では表現が難しい雑音の影響を受けているのに対して，雑音と残響の影響が考慮された提案手法では予測精度の向上が確認できる (例えば，図 5 の左上の環境オープンテスト (SNR=20 dB) における予測誤差が D 値単体では 8.1 %であるのに対して，提案手法では 3.4 %であった)．また，PESQ 単体でも高残響環境の音声に

対してはPESQでは表現が難しい残響の影響を受けているのに対して，雑音と残響の影響が考慮された提案手法では予測精度の向上が確認できる (例えば，図6の右中の環境オープンテスト (SNR=30 dB) における予測誤差がD値単体では12.1 %であるのに対して，提案手法では5.4 %であった)．またSNRが-5〜0 dBのとき，全ての予測指標に対して1 %以内の平均性能予測誤差を達成した．これは，これらの音声に対する認識性能が最大約7 %であり，ダイナミックレンジも小さいために，顕著な差異を確認できなかったと考えられる．

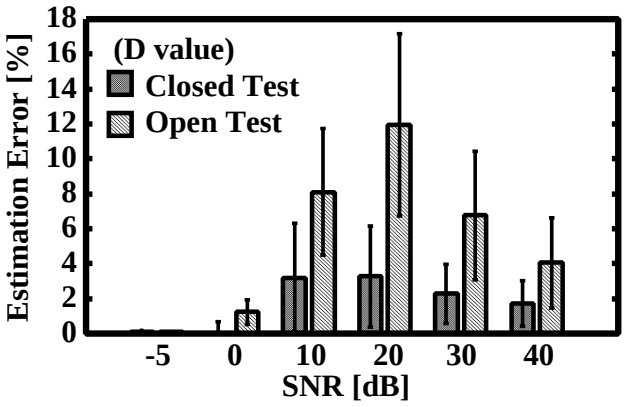
4.6. 評価実験3 -音声認識性能予測のコスト評価-

4.6.1 実験条件

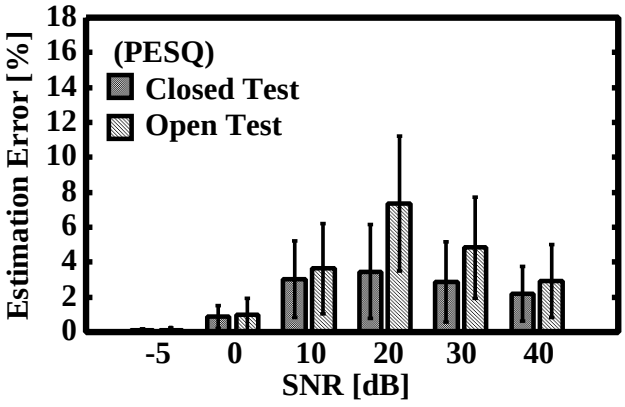
ここでは，従来・提案手法による音声認識性能予測に必要なデータ量および計算時間を評価する．ここでの従来の音声認識性能評価とは，クリーン音声にインパルス応答を畳み込んだ信号に雑音を付加した評価音声データを大量に用意して音声認識性能を予測する手法を指す．そして提案手法は，インパルス応答，クリーン音声，および雑音から室内音響指標とPESQを算出して音声認識性能を予測する．また本実験では計算機サーバ（Debian Linux 6.0.7，CPU:Intel Xeon 3.60 GHz，メモリ: 16 GB）を用いて評価した．

なお，従来手法を用いて正確に音声認識性能を評価するには，大量の音声データを用いて統計的な処理を行う必要がある．そのため，本実験では表4.2の実験条件に基づいて評価を行うが，クリーン音声のみ12話者分のATR216音素バランス単語を用いた．

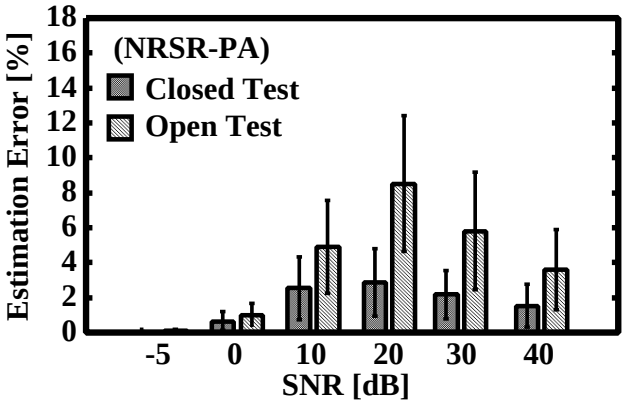
一方，提案手法で音声認識性能を予測するときは，複数の評価音声を用いて算出したPESQの平均値を用いるが，このPESQの平均値を算出するのに十分な音声データ数を明らかにする必要がある．そこで予備実験として，表4.2に示すクリーン音声（合計864発話），エレベータホールのインパルス応答 (1ヶ所)，そして白色雑音（SNR=10 dB）を用いて評価音声データを用意してPESQの平均値と分散値を評価した．その結果，50発話以上の評価音声データを用いることで，全てのデー



(a) 予測指標：D 値

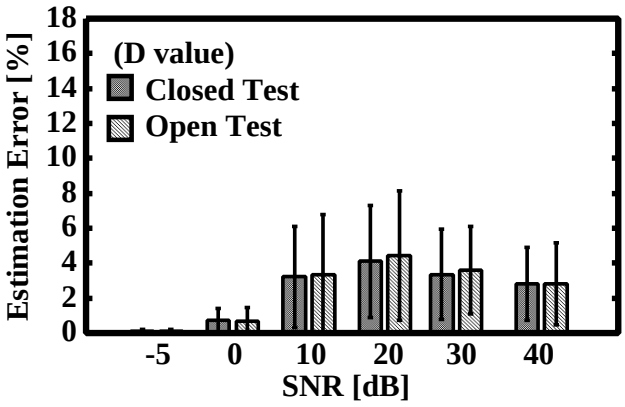


(b) 予測指標：PESQ

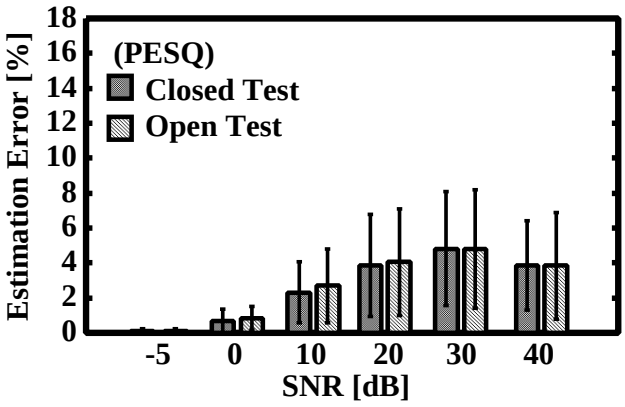


(c) 予測指標：NRSR-PA

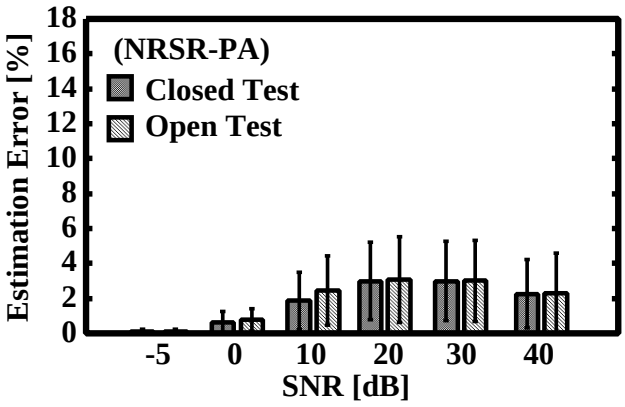
図 4.12 平均性能予測誤差（雑音：白色雑音，残響時間：450 ms）



(a) 予測指標：D 値

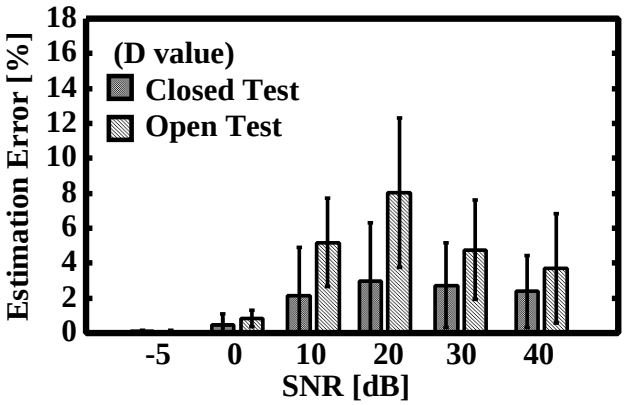


(b) 予測指標：PESQ

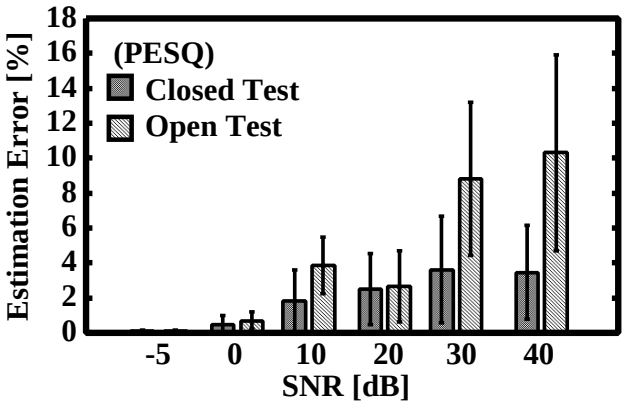


(c) 予測指標：NRSR-PA

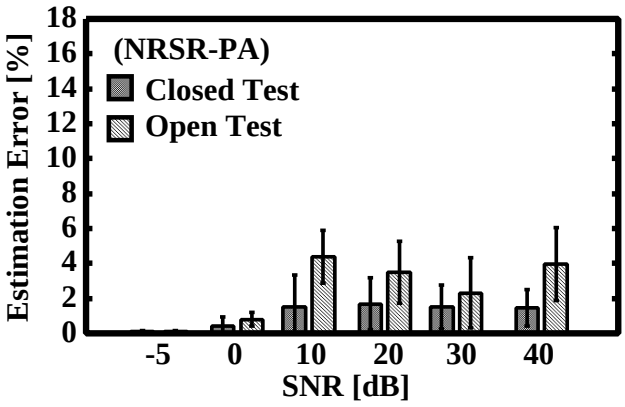
図 4.13 平均性能予測誤差（雑音：白色雑音，残響時間：600 ms）



(a) 予測指標：D 値

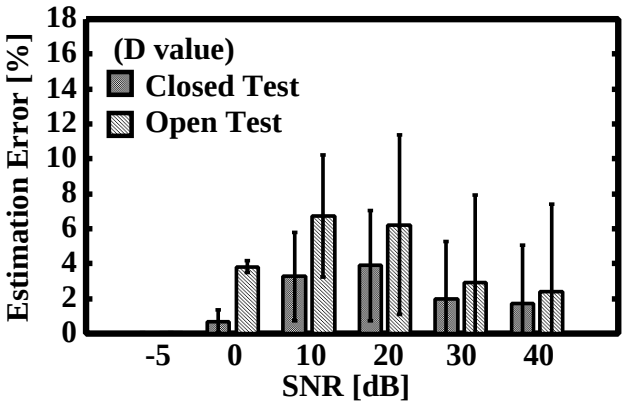


(b) 予測指標：PESQ

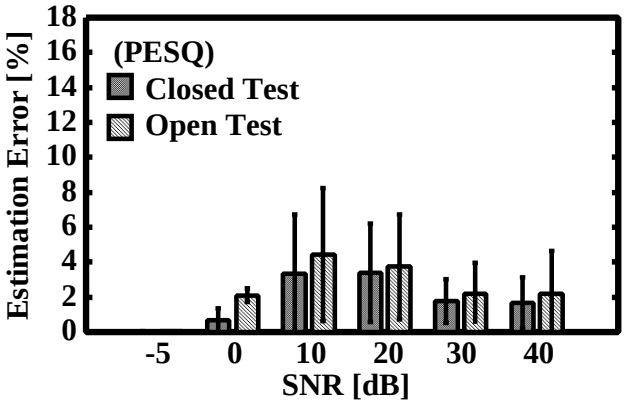


(c) 予測指標：NRSR-PA

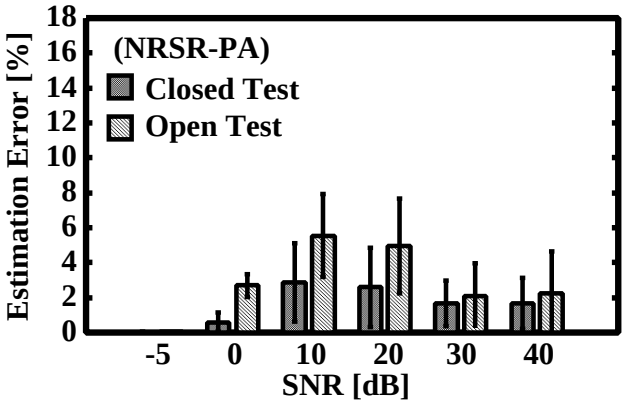
図 4.14 平均性能予測誤差（雑音：白色雑音，残響時間：850 ms）



(a) 予測指標：D 値

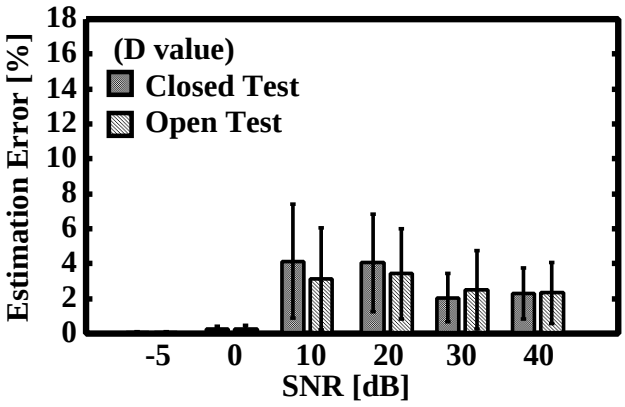


(b) 予測指標：PESQ

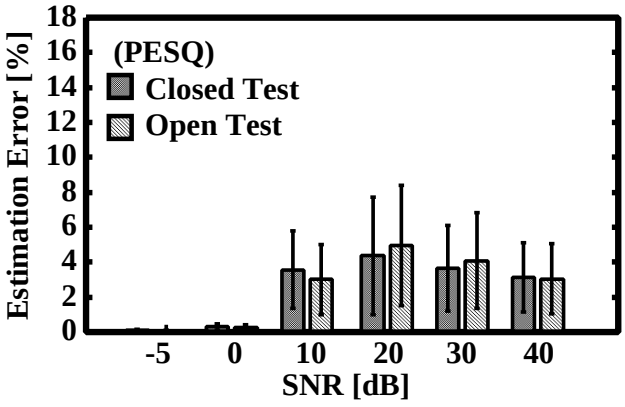


(c) 予測指標：NRSR-PA

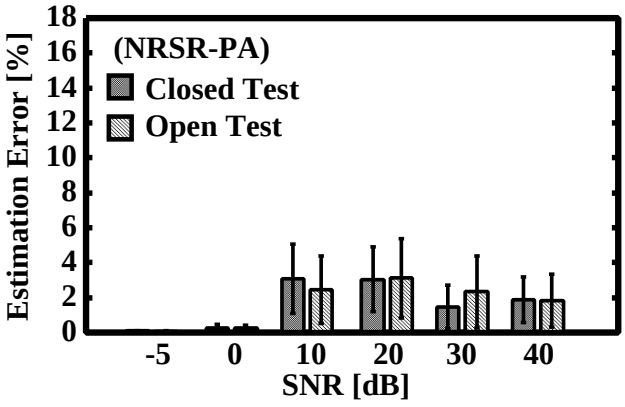
図 4.15 平均性能予測誤差（雑音：工場騒音，残響時間：450 ms）



(a) 予測指標：D 値

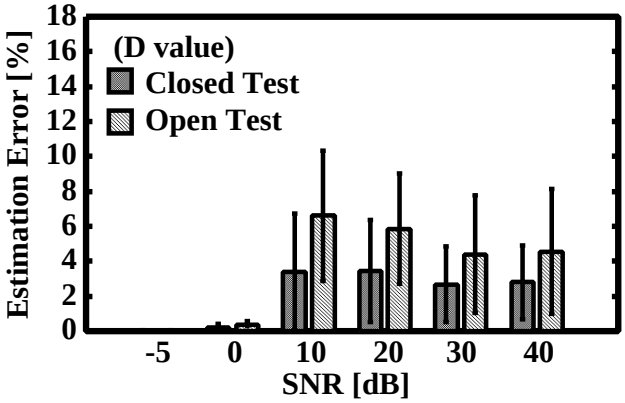


(b) 予測指標：PESQ

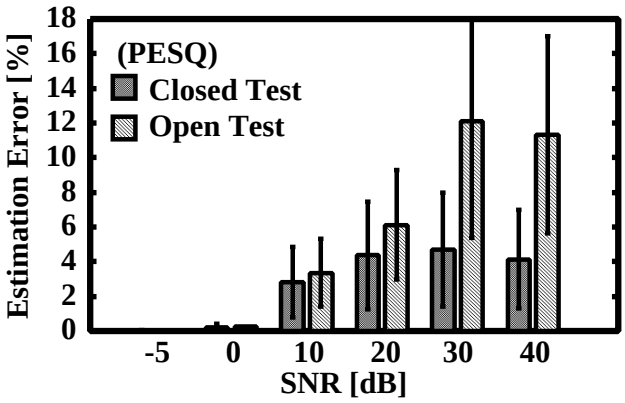


(c) 予測指標：NRSR-PA

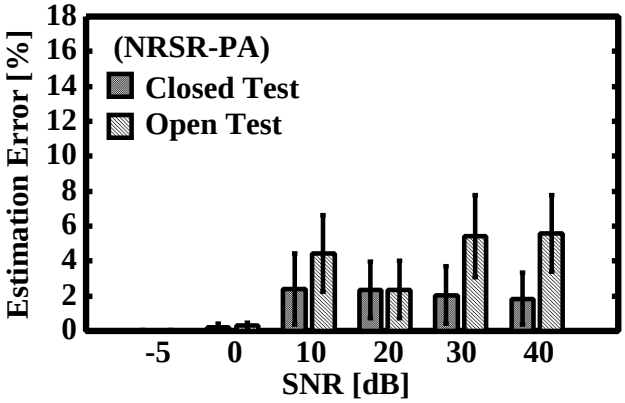
図 4.16 平均性能予測誤差（雑音：工場騒音，残響時間：600 ms）



(a) 予測指標：D 値



(b) 予測指標：PESQ



(c) 予測指標：NRSR-PA

図 4.17 平均性能予測誤差（雑音：工場騒音，残響時間：850 ms）

タを用いた場合と同等の PESQ の平均値と分散値を達成したことから，本実験では PESQ 算出に用いる評価音声データ数を各環境につき 50 発話とした．

4.6.2 実験結果

音声認識性能予測に必要なデータ量を表 4.5 に，そして計算時間を表 4.6 に示す．表 4.5 に示すデータ量の結果より，従来の音声認識性能評価では合計で約 321 GB の評価音声データを必要としていたのに対して，提案手法を用いることでデータ量を約 6.3 GB まで大幅削減することができた．また，表 4.6 に示す計算時間においても，従来手法 (1 環境あたり約 4 分) と比較して，提案手法 (1 環境あたり約 46 秒) を用いることで，およそ 8 倍の速さで音声認識性能を予測することができた．しかしながら，提案手法では 50 発話の PESQ を計算するのに 41.4 秒を必要としているため，更なる計算コスト削減のためには PESQ 算出の高速化が重要であると考えられる．これらの評価結果より，提案手法を用いることで，音声認識性能予測に要するコストを大幅に削減できることが明らかとなった．

4.7. まとめ

外乱による音声認識性能の劣化を事前に予測できれば，その結果に基づいて性能改善手法を前処理等に反映できる．これまでに雑音下では Perceputual Evaluation of Speech Quality（PESQ）を，残響下では室内音響指標（D 値，残響時間（T_{60}））を用いて音声認識性能を予測する手法が提案されていた．しかし，これらの手法には予測指標が表現できない外乱が混入すると音声認識性能の予測精度が低下する問題があった．そこで本章では，音声認識性能を残響に対して頑健かつ簡便に予測できる雑音・残響指標 NRSR-PA を提案し，音声認識性能の高精度な予測を試みた．はじめに 4.2 節で，提案手法に用いる室内音響指標と PESQ について述べ，個々だけで複数の外乱が存在する環境において音声認識性能を予測することの難しさを示した．そして，4.3 節で提案手法の詳細について述べた．最後に 4.4~4.6 節で，提案手法を用いて外乱環境（雑音・残響環境）における音声認識性能の予測実験を行い，その有効性を示した．今後は提案手法の実用化に向けて，D 値を算出するためのイン

表 4.5 音声認識性能予測に必要なデータ量

従来手法	
音声データ	216 単語 × 12 話者 = 81 MB
合計	81 MB × 248ヶ所×雑音（ 2 種類）× SNR（8 種類） = **321.408 GB**

提案手法	
1. D 値	
計	研究室 (72ヶ所 = 2.1 MB) + 廊下 (120ヶ所 = 4.5 MB) + 階段 (56ヶ所 = 3.0 MB) = 9.6 MB
2. PESQ	
音声	50 単語 = 1.575 MB
計	1.575 MB × 248ヶ所×雑音（ 2 種類）× SNR（8 種類） $\simeq$ 6.250 GB
合計	9.6 MB + 6.250 GB $\simeq$ **6.260 GB**

パルス応答や事前にクリーン音声を必要とする PESQ を簡便に推定できる手法の検討に取り組む予定である.

表 4.6 音声認識性能予測の計算時間

従来手法	
1. インパルス応答の畳込み: 214.9 秒 2. 雑音の加算: 21.2 秒 3. 音声認識: 120.1 秒	
合計	214.9 秒+21.2 秒+120.1 秒= **356.2 秒**

提案手法	
1. D 値 D 値の計算: 1 ミリ秒 **2. PESQ** 1. インパルス応答の畳込み: 4.2 秒 2. 雑音加算: 0.4 秒 3. PESQ の計算: 41.4 秒	
合計	1 ミリ秒+4.2 秒+0.4 秒+41.4 秒 $\simeq$ **46.0 秒**

第5章　結論

5.1. 本論文のまとめ

情報機器の急速な発展に伴い入力インタフェースが複雑化している中で，現在の基本となる入力インタフェースはキーボードとマウスである．しかし，情報機器に不慣れな高齢者や手足が不自由な身体障害者には，このような入力インタフェースを使用することが困難である．

万人にとって使い勝手の良い入力インタフェースを実現するために，マイクロホンを意識せずに音声入力ができるハンズフリー音声インタフェースの実現に高い注目が集まっている．しかしながら，実環境下における音声認識ではマイクロホンから離れた地点で発話すると壁や床からの反射音が混入することの影響で目的音声が歪むために音声認識性能が低下するという問題があった．この問題を解決するには，事前に利用環境に存在する雑音や残響などの外乱の影響を音声認識システムに適応させる必要がある．

外乱の影響を音声認識システムに適応させるための関連手法として，外乱環境における音声認識性能の予測手法が注目されている．利用環境に存在する外乱が音声認識性能に与える影響を事前に予測することができれば，その予測結果に基づいて外乱対策を音声認識システムの前処理等に反映させることが可能となり，結果的に音声認識性能の劣化を防ぐことができる．このようなに利用環境で最大限の音声認識性能を発揮させることを目指して，本論文では，雑音や残響が存在する外乱環境における音声認識性能を簡便かつ高精度に予測する手法について検討した．

2 章では，雑音環境と残響環境のそれぞれに対する既存の音声認識性能予測手法の原理と課題について述べた．まず外乱環境において音声認識実験を行い，雑音や残響の影響を受けることによって音声認識性能が劣化することを示した．次に既存

の音声認識性能の予測手法として，雑音環境においてはSNR(Signal to noise ratio)を用いて，残響環境においては残響時間を用いて音声認識性能を予測する手法について述べた．そして，これらの手法では雑音や残響が目的信号に与える影響を詳細に表現することができず，高精度な音声認識性能予測が難しいことを示した．

3章では，残響環境下においてコストをかけない高精度な音声認識性能の予測手法を提案した．提案する残響環境下音声認識性能の予測手法では，初期反射音と後続残響音の関係を表す室内音響指標の中でも特にDefinition（D値）に着目し，様々な環境で複数箇所計測したインパルス応答を基に算出したD値と音声認識性能の関係に基づいて残響指標RSR-D_nを策定した．そして，策定した残響指標RSR-D_nと音声認識性能の予測位置におけるインパルス応答を基に実環境を想定した残響下音声認識性能の予測を試みた．その結果，残響時間や発話位置が異なる様々な残響環境において，音声認識性能の予測実験を行った結果，提案手法の有効性を確認した．

4章では，複数の外乱要因（本論文では残響と雑音の2種類）が混在する環境において，高精度かつ簡便に音声認識性能を予測するための手法を提案した．具体的には，雑音・残響下における音声認識性能の予測精度を向上させるために，PESQ，室内音響指標のD値と残響時間を用いた音声認識性能の予測式を提案した．ここでは事前に計測した発話音声やインパルス応答を用いて算出したPESQ，D値，残響時間，音声認識性能から雑音・残響指標Noisy-and-Reverberant Speech Recognition criteria with PESQ and Acoustic parameters（NRSR-PA）を策定した．そしてNRSR-PAを用いて性能予測を行う発話位置におけるインパルス応答と発話音声から音声認識性能の予測を試みた．実環境を想定した雑音・残響環境において音声認識性能の予測実験，従来の雑音指標・残響指標を個別に用いて性能予測する手法よりもNRSR-PAは頑健に雑音・残響下音声認識性能を予測できることを確認した．

5.2. 今後の課題

本論文で提案した雑音・残響指標NRSR-PAを用いることで，外乱環境において高精度かつ簡便に音声認識性能を予測できることを確認した．しかし，実際の利用環境において，さらに高精度に音声認識性能を予測するためには，以下の問題点が

残されている．

1. 話者の個人性による音声認識性能の予測精度の劣化．

2. 雑音環境（SNR=10∼20 dB 程度）における音声認識性能の予測精度の劣化．

1. を解決するためには，話者の個人性を推定する技術を用いて，提案手法の性能を改善する必要がある．例えば，話者識別技術 [82, 83, 84, 85, 86, 87] や発話様式（平静音声，叫び声，滑舌の優劣，方言など）の推定技術 [88, 89, 90, 91] などを提案手法と併用することで音声認識性能の予測精度の向上が大いに期待できると考えられる．

2. を解決するためには，ITU-T 勧告 P.863 で規定された次世代のモバイル音声品質試験標準である POLQA（Perceptual Objective Listening Quality Assessment）[92] を併用することで高い予測精度が期待できる．特に，POLQA は PESQ と比べて，背景雑音が高い状況においても正確な音質評価が可能である上に，評価周波数帯域も 50∼14,000 Hz（ちなみに，PESQ は 100∼7,000 Hz）の広帯域にも対応していることから，近年では POLQA を用いた音声評価が主流になりつつある [93, 94]．また ITU-T 勧告 P. 863 では音響インタフェースを通じて録音された信号の評価に POLOA の利用を推奨しているため，今後はこのような指標と本論文で提案した音声認識性能の予測指標を組み合わせて評価を進めていく必要がある．

上記の問題を解決するために，今後も引き続き研究を行い，利用環境に存在する外乱成分を推定しながら，常に高い音声認識性能を発揮できるハンズフリー音声インタフェースを実現することで，より快適な社会の実現に僅かながらでも貢献できれば幸いである．

謝辞

本博士論文は, 立命館大学大学院情報理工学研究科博士後期課程において筆者が行った研究の成果をまとめたものです. 本研究を遂行するにあたり, 学内, 学外を問わず多くの方にお世話になりました. ここに深厚なる感謝の意を表します.

立命館大学情報理工学部西浦敬信教授には，筆者の本学在学中における研究活動を通じて多大なご指導を頂きました．西浦先生には指導教員として研究方法の初歩から, 研究の内容, 展開, 論文の執筆に至るまで丁寧にご指導頂きました．また研究活動のみならず, 各種活動の機会を与えて下さったことで, 音情報処理研究室に配属された学部 3 回生から 6 年半を有意義に過ごすことができました．ここに心から感謝の意を表します.

同学部山下洋一教授には，筆者が本学在学の間，終始懇切丁寧なご指導を頂きました．山下先生の厳しくも温かいご指導なくしては，筆者が本学における研究活動を成し遂げ，博士論文執筆にいたる道を見出すことはできませんでした．ここに深甚なる感謝の意を表します.

同学部福本淳一教授には，本論文審査委員として本論文の執筆におけるご指導を頂きました．福本先生から頂いた的確かつ有益な御助言によって，本論文をより良い方向へ進歩させることができました．心より深く御礼申し上げます.

同学部中山雅人助教には，毎週の研究進捗ミーティングや, 実際の研究の遂行や実際のプログラミング, 論文の執筆に至るまで, 常日頃から懇切なる御指導, 御助言を頂きました. ここに厚く御礼申し上げます.

同学部 Jeremy Stewart White 准教授には，博士論文の英文執筆にあたり，有益なご助言を頂きました．心より感謝申し上げます.

同学部森勢将雅助教（現在，山梨大学特任助教）には, 筆者が本大学院在学中に，計算機の使い方やデータ収録の方法など日頃から熱心な御指導，御討論を頂きまし

た．心より深く御礼申し上げます．

本研究の遂行にあたり，数々の有益な御助言を頂いた情報処理学会音声言語情報処理研究会雑音下音声認識評価ワーキンググループの皆様に心より感謝いたします．

個々には御名前を申し上げられませんが，筆者の研究上の議論に付き合っていただき，また筆者の至らない点を御援助頂きました立命館大学情報理工学部音情報処理研究室の多くの先輩，同期，後輩，秘書の皆様，そして多くの励ましを頂いた学内外の友人に心より御礼申し上げます．

最後になりましたが, 深い愛情と広い心で今日まで筆者を支えて頂いた家族と友人に心から感謝いたします．

参考文献

[1] 石井 健一郎，上田 修功，前田 英作，村瀬 洋，“わかりやすいパターン認識，” オーム社，2001.

[2] 鹿野 清宏，伊藤 克亘，河原 達也，武田 一哉，山本 幹雄，“IT Text 音声認識システム，” オーム社，2001.

[3] R.O. Dura, P.E. Hart and D.G. Stork, “パターン識別，” 新技術コミュニケーションズ，2001.

[4] 中川 聖一, “確率モデルによる音声認識，” 電子情報通信学会, 1998.

[5] 北 研二，中村 哲，永田 昌明，“音声言語情報処理 -コーパスにもとづくアプローチ-，” 森北出版, 1996.

[6] D.L. Ramon and M. Araki，“Spoken Multilingual and Multimodal Dialogue Systems,” *Wiley*, 2005.

[7] X. Huang, A. Acero, F. Alleva, M.Y. Hwang, L. Jiang, and M. Mahajan, “Microsoft Windows Highly Intelligent Speech Recognizer: Whisper,” *Proc. 1995 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004)*, vol. 1, pp. 93-96, 1995.

[8] 藤本 雅清，武田 一哉，中村 哲，“自動車内における連続数字音声コーパス CENSREC 2 の設計と評価，” 電子情報通信学会技術研究報告, vol. 105, no. 494, pp. 55-56, 2005.

[9] C.E. Mokbel and G.F.A. Chollet, “Automatic Word Recognition in Cars，” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 5, 1995.

[10] B. Chen, “Word Topic Models for Spoken Document Retrieval and Transcription,” *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, no. 2, pp. 1-27, 2009.

[11] 杉本 樹世貴，前沢 慎吾，西崎 博光，関口 芳廣，“検索対象と類似性の高い Web ページを利用した音声ドキュメント検索の検討，” 情報処理学会 音声ドキュメント処理ワークショップ，pp. 33-38, 2009.

[12] 中村 哲，“音声翻訳システムの研究開発，” 電子情報通信学会技術研究報告，vol. SP-108, no. 422, pp. 31-36, 2009.

[13] 山端 潔，磯谷 亮輔，安藤 真一，花沢 健，石川 晋也，江森 正，磯 健一，服部 浩明，奥村 明俊，渡辺 隆夫，“PDA で動作する旅行会話向け日英双方向音声翻訳システム，” 電子情報通信学会技術研究報告，vol. NLC-102, no. 199, pp. 55-62, 2002.

[14] E. Levin, R. Pieraccini, and W. Eckert, “A Stochastic Model of Human-Machine Interaction for Learning Dialog Strategies，” *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, pp. 11-23, 2000.

[15] S. Young, J. Schatzmann, K. Weilhammer, and Y. Hui, “The Hidden Information State Approach to Dialog Management,” *Proc. 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, vol. 4, pp. 149-152, 2007.

[16] 吉岡 理，荒井 和博，管村 昇，嵯峨山 茂樹，“音声認識機能を含むマルチモーダルインタフェースを持つ住所入力システムの開発評価，” 電子情報通信学会論文誌 D, vol. J80-D-II, no. 5, pp. 1007-1015, 1997.

[17] 伊田 政樹，森 弘之，中村 哲，鹿野 清宏，“据置き型情報提供端末向き雑音処理を用いた音声入力インタフェース，” 電子情報通信学会論文誌 D, vol. J84-D2, no. 6, pp. 868-876, 2001.

[18] 中川 聖一，富樫 慎吾，山口 優，藤井 康寿，北岡 教英，“講義音声ドキュメントのコンテンツ化と視聴システム，” 電子情報通信学会論文誌 D, vol. J91-D, no. 2, pp. 238-249, 2008.

[19] 三村 正人, 河原 達也，“会議音声認識における BIC に基づく高速な話者正規化と話者適応，” 電子情報通信学会論文誌 D, vol. J95-D, no. 7, pp. 1467-1475, 2012.

[20] 大村 絵梨，南條 浩輝，“多言語音声の同時認識システムにおける翻訳モデルとスコア計算の高速化，” 情報処理学会論文誌, vol. 53, no. 10, pp. 2349-2358, 2012.

[21] 滝口 哲也，中村 哲，鹿野 清宏 “雑音と残響のある環境下での HMM 合成によるハンズフリー音声認識法,” 電子情報通信学会論文誌 D，vol. J79-D-2, no. 12, pp. 2047-2053, 1996.

[22] B. kingsbury and N. Morgan, “Recognizing Reverberant Speech with RASTA-PLP,” *Proc. 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1997)*, vol. 2, pp. 1259-1262, 1997.

[23] 飛田 瑞広，菅原 昇，中津 良平，“音響伝送歪みの単語認識率への影響とその改善,” 電子情報通信学会論文誌 D，vol. J73-D-II, no. 6, pp. 781-787, 1990.

[24] T. Nishiura, M. Nakayama, Y. Denda, N. Kitaoka, K. Yamamoto, T. Yamada, S. Tsuge, C. Miyajima, M. Fujimoto, T. Takiguchi, S. Tamura, S. Kuroiwa, K. Takeda, and S. Nakamura, “Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments: Newest Part of the CENSREC Series -,” *Proc. Louisiana Real Estate Commission 2008 (LREC2008)*, pp. 968-971, 2008.

[25] J.L. Flanagan, J.D. Johnston, R. Zahn, and G. W. Elko, “Computer-Steered Microphone Arrays for Sound Transduction in Large Rooms,” *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1508-1518, 1985.

[26] O.L. Frost, “An Algorithm for Linearly Constrained Adaptive Array Processing,” *Proc. IEEE*, vol. 60, no. 8, pp. 926-935, 1972.

[27] M.J.F. Gales and S.J. Young, “An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise,” *Proc. 1992 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1992)*, vol. 1, pp. 233-236, 1992.

[28] H.M. Cung and Y. Normandin, “Noise Adaptation Algorithms for Robust Speech Recognition,” *Speech Communication*, vol. 12, no. 3, pp. 267-276, 1993.

[29] Y. Ephraim and D. Malah, “Speech Enhancement Sing a Minimum Mean Square Error Log-Spectral Amplitude Resonator,” *IEEE Transactions on Speech and Audio Processing*, vol. 33, no. 2, pp. 443-445, 1985.

[30] 堀井 圭祐, 福森 隆寛, 森勢 将雅, 中山 雅人, 西浦 敬信, 山下 洋一, 南條 浩輝, “雑音下音声受音における Weighted 反復スペクトル減算法を用いたミュージカルノイズの低減,” 電子情報通信学会論文誌 D, vol. J96-D, no. 3, pp. 664-674, 2013.

[31] M. Fujimoto and Y. Araki, “Combination of Temporal Domain SVD Based Speech Enhancement and GMM Based Speech Estimation for ASR in Noise -Evaluation on the AURORA2 Task-,” *Proc. European Conference on Speech Communication and Technology (EUROSPEECH)*, pp. 1781-1784, 2003.

[32] S. Furui, “Cepstral Analysis Technique for Automatic Speaker Verification,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254-272, 1981.

[33] M. Miyoshi and Y. Kaneda, “Inverse Filtering of Room Acoustics,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145-152, 1988.

[34] 清水 秦博, 梶田 将司, 武田 一哉, 板倉 文忠, “空間音響特性を考慮したスペースダイバシチ型音声認識,” 電子情報通信学会論文誌 D, vol. J83-DII, no. 11, pp. 2448-2456, 2000.

[35] T. Takiguchi, M. Nishimura, and Y. Ariki, “Acoustic Model Adaptation Using First-Order Linear Prediction for Reverberant Speech,” *IEICE Transactions on Information and Systems*, vol. E89-D, no. 3, pp. 908-914, 2006.

[36] H. Kameoka, T. Nakatani, and T. Yoshioka, “Robust Speech Dereverberation Based on Non-Negativity and Sparse Nature of Speech Spectrograms,” *Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, pp. 45-48, 2009.

[37] J.A.N. Flows and S.J. Young, “Continuous Speech Recognition in Noise Using Spectral Subtraction and HMM Adaptation,” *Proc. 1994 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1994)*, vol. 1, pp. 409-412, 1994.

[38] 荒木 雅弘, “フリーソフトでつくる音声認識システム,” 森北出版株式会社, 2007.

[39] 安藤 彰男，“リアルタイム音声認識，” 電子情報通信学会, 2005.

[40] 中川 聖一，“確率モデルによる音声認識，” 電子情報通信学会, 1998.

[41] 梶田 将司，小林 大祐，武田 一哉、板倉 文忠，“ヒューマンスピーチライク雑音に含まれる音声特徴の分析,” 日本音響学会論文誌，vol. 53, no. 5, pp.337-345, 1997.

[42] K. Takeda, Y. Sagisaka, and S. Katagiri, “Acoustic-Phonetic Labels in a Japanese Speech Database,” *Proc. European Conference on Speech Technology*, pp. 2013-2016, 1987.

[43] Y. Sagisaka, K. Takeda, M. Abe, S. Katagiri, T. Umeda, and H. Kuwabara, “A Large-Scale Japanese Speech Database,” *Proc. International Conference on Spoken Language Processing*, pp. 1089-1092, 1990.

[44] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, "ATR Japanese Speech Database as a Tool of Speech Recognition and Synthesis," *ELSEVIER Speech Communication*, vol. 9, no. 4, pp. 357-363, 1990.

[45] A. Lee, T. Kawahara, and K. Shikano, "Julius — an Open Source Real-Time Large Vocabulary Recognition Engine," *Proc. European Conference on Speech Communication and Technology*, pp. 1691-1694, 2001.

[46] A. Lee and T. Kawahara, "Recent Development of Open-Source Speech Recognition Engine Julius," *Proc. Asia Pacific Signal and Information Processing Association (APSIPA)*, pp. 131-137, 2009.

[47] 河原 達也，李 晃伸, "連続音声認識ソフトウェア Julius," 人工知能学会論文誌, vol. 20, no. 1, pp. 41-49, 2005.

[48] T. Yamada, M. Kumakura, and N. Kitawaki, "Performance Estimation of Speech Recognition System under Noise Conditions Using Objective Quality Measures and Artificial Voice," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 14, no. 6, pp. 2006-2013, 2006.

[49] M.R. Schroeder, "New Method of Measuring Reverberation Time," *Journal of the Acoustical Society of America*, vol. 37, pp. 409-412, 1965.

[50] R. Petrick, X. Lu, M. Unoki, M. Akagi, and R. Hoffmann, "Robust Front End Processing for Speech Recognition in Reverberant Environments: Utilization of Speech Characteristics," *Proc. International Speech Communication Association (INTERSPEECH)*, pp. 658-661, 2008.

[51] 日本音響学会, "新版音響用語辞典," コロナ社, 2003.

[52] Y. Suzuki, F. Asano, H.Y. Kim, and T. Sone, "An Optimum Computer-Generated Pulse Signal Suitable for the Measurement of Very Long Impulse

Responses," *Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1119-1123, 1995.

[53] N. Aoshima, "Computer-Generated Pulse Signal Applied for Sound Measurement," *Journal of the Acoustical Society of America*, vol. 69, no. 5, pp. 1484-1488, 1981.

[54] 鈴木 陽一，浅野 太，曽根 敏夫，"音響系の伝達関数の模擬をめぐって（その２）," 日本音響学会誌, vol. 45, no. 1, pp. 44-50, 1989.

[55] 大賀 寿郎，山崎 芳男，金田 豊，"音響システムとディジタル信号処理," 電子情報通信学会, 1995.

[56] 佐野 史明，"はじめてのインパルス応答計測," 日本音響学会誌，vol. 67, pp. 155-162, 2011.

[57] 金田 豊，"インパルス応答測定の際の留意点," 日本音響学会誌，vol. 55, pp. 364-369, 1999.

[58] 西浦 敬信，傳田 遊亀，"音声認識における初期反射音の影響についての検討"，日本音響学会 2006 年春季研究発表会講演論文集，pp. 141-142, 2006.

[59] H. Kuttruff, "Room Acoustics," *Spon Press*, 2000.

[60] ISO3382: Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters. Internatinal Organization for Standardization, 1997.

[61] 五十嵐 冬人 , 佐久間 哲哉，"室内音響インパルス応答の聴感的類似度に関する研究," 日本建築学会学術講演梗概集 (D-1)，pp. 31-32, 2001.

[62] 福山 忠雄 , 土屋 裕造 , 山崎 芳男，"教室等を対象とした音声明瞭性関連の物理指標の測定例 :学校の音環境に関する研究," 日本建築学会学術講演梗概集 (D-1)，pp. 159-160, 2003.

[63] V.R. Thiele, “Richtungsverteilung und Zeitforge der Schallrückwurfe in Räumen,” *Acustica*, vol. 3, pp. 195-200, 1961.

[64] W. Reichardt, O. Abdel Alim and W. Schmidt, “Definition und Messgrundlage Eines Objektiven Masses zur Ermittlung der Grenze Zwischen Brauchbarbietung,” *Acustica*, vol. 32, pp. 126-137, 1975.

[65] L. Cremer and H.A. Müler, “Principles and Applications of Room Acoustics, vol. 1,” *Applied Science*, 1982.

[66] T. Houtgast, H.J.M. Steeneken and R. Plomp, “Predicting Speech Intelligibility in Room Acoustics,” *Acustica*, vol. 46, pp. 60-72, 1980.

[67] 田中 敏幸，“数値計算法基礎,” コロナ社, 2006.

[68] AURORA-J/CENSREC: http://sp.shinshu-u.ac.jp/CENSREC/index.html.ja

[69] M. Nakayama, T. Nishiura, Y. Denda, N. Kitaoka, K. Yamamoto, T. Yamada, S. Tsuge, C. Miyajima, M. Fujimoto, T. Takiguchi, S. Tamura, T. Ogawa, S. Matsuda, S. Kuroiwa, K. Takeda, and S. Nakamura, “CENSREC-4: Development of Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments,” *Proc. International Speech Communication Association (INTERSPEECH)*, pp. 968-971, 2008.

[70] “Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codes,” ITU-T Recommendation P. 862, 2001.

[71] N. Egi, H. Aoki, and A. Takahashi, “Objective Quality Evaluation Method for Noise-Reduced Speech,” *IEICE Transactions on Communications*, vol. E91-B, pp. 1279-1286, 2008.

[72] 篠原 佑基，山田 武志，北脇 信彦，牧野 昭二，“雑音抑圧音声の総合品質モデルを用いたフルリファレンス客観品質評価法の検討,” 第 7 回 QoS ワークショップ, pp. 40-41, 2009.

[73] T. Yamada, Y. Kasuya, Y. Shinohara, and N. Kitawaki, “Non-Reference Objective Quality Evaluation for Noise-Reduced Speech Using Overall Quality Estimation Model,” *IEICE Transactions on Communications*, vol. E93-B, pp. 1367-1372, 2010.

[74] ETSI EG 202 369-3 V1.3.1, “Speech and multimedia Transmission Quality (STQ); Speech Quality Performance in the Presence of Background Noise Part 3: Background Noise Transmission –Objective Test Methods,” 2011.

[75] T. Yamada, M. Kumakura, and N. Kitawaki, “Objective Estimation of Word Intelligibility for Noise-Reduced Speech,” *IEICE Transactions on Communications*, vol. E91-B, pp. 4075-4077, 2008.

[76] K. Kondo and Y. Takano, “Estimation of Two-to-One Forced Selection Intelligibility Scores by Speech Recognizers Using Noise-Adapted Models,” *Proc. International Speech Communication Association (INTERSPEECH)*, pp. 302-305, 2010.

[77] 押田 賢浩, 大和田 昇, 陶山 健仁，“客観的品質評価尺度による移動話者追尾手法の性能評価，” 電子情報通信学会論文誌 A, vol. J93-A, no. 9, pp. 583-593, 2010.

[78] 青木 直史，“ピッチ波形複製法に基づくステガノグラフィを用いた VoIP におけるパケット損失の一隠蔽法，” 電子情報通信学会論文誌 B, vol. J86-B, no. 12 , pp. 2551-2560, 2003.

[79] T. Yamada, M. Kumakura, and N. Kitawaki, “Performance Estimation of Speech Recognition System under Noise Conditions Using Objective Quality Measures and Artificial Voice,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2006-2013, 2006.

[80] 福森 隆寛, 森勢 将雅, 西浦 敬信, 山下 洋一, “室内音響指標を用いた残響指標 RSR-Dn に基づく残響下音声認識性能の予測, ” 電子情報通信学会論文誌 D, vol. J94-D, no. 4, pp. 712-720, 2011.

[81] 電子協騒音データベース http://research.nii.ac.jp/src/JEIDA-NOISE.html

[82] S. Nakagawa, W. Zhang, and M. Takahashi, "Text-Independent/Text-Prompted Speaker Recognition by Combining Speaker-Specific GMM With Speaker Adapted Syllable-Based HMM," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 3, pp. 1058-1065, 2006.

[83] T. Matsui and K. Tanabe, "Comparative Study of Talker Identification Methods: dPLRM, SVM and GMM," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 3, pp. 1066-1073, 2006.

[84] 風間 道子，東山 三樹夫，山崎 芳男，"狭帯域音声波形包絡線の帯域間相関行列に現れる話者情報," 電子情報通信学会論文誌 A，vol. J92-A, no. 4, pp. 205-215, 2009.

[85] D.G. Romero and C.Y.E. Wilson, "Analysis of i-Vector Length Normalization in Speaker Recognition System," *Proc. International Speech Communication Association (INTERSPEECH)*, pp. 249-252, 2011.

[86] 小坂 哲夫，赤津 達也，加藤 正治，好田 正紀，"音素モデルを用いた話者ベクトルに基づく話者識別," 電子情報通信学会論文誌 D，vol. J90-D, no. 12, pp. 3201-3209, 2007.

[87] 森 幹男，竹田 勉，荻原 慎洋，谷口 秀次，高橋 謙三，"気導音声と骨導音声のスペクトル比を用いた話者識別の検討," 電気学会論文誌 C, vol. 127, no. 3, pp. 456-457, 2007.

[88] 平山 直樹，吉野 幸一郎，糸山 克寿，森 信介，奥乃 博，"擬似生成した複数方言言語モデル混合による混合方言音声認識," 情報処理学会論文誌, vol. 55, no. 7, pp. 1681-1694, 2014.

[89] 鈴木 雅之，黒岩 龍，印南 圭祐，小林 俊平，清水 信哉，峯松 信明，広瀬 啓吉，"条件付き確率場を用いた日本語東京方言のアクセント結合自動推定," 電子情報通信学会論文誌 D, vol. 96, no. 3, pp. 644-654, 2013.

[90] J. Pohjalainen, P. Alku, and T. Kinnunen, “Shout Detection in Noise,” *Proc. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, pp. 4968-4971, 2011.

[91] W. Huang, T.K. Chiew, H. Li, T.S. Kok, and J. Biswas, “Scream Detection for Home Applications,” *Proc. Industrial Electronics and Applications (ICIEA)*, pp. 2115-2120, 2010.

[92] “Perceptual Objective Listening Quality Analysis (POLQA),” ITU-T Recommendation P. 863, 2014.

[93] A. Hines, J. Skoglund, A. Kokaram, and N. Harte, “Robustness of Speech Quality Metrics to Background Noise and Network Degradations: Comparing ViSQOL, PESQ and POLQA,” *Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, pp. 3697-3701, 2013.

[94] 近藤 和弘, “客観的評価値を用いたデータを埋め込んだ音響信号品質の推定の基礎検討,” 電子情報通信学会技術研究報告, vol. 112, no. 292, pp. 69-74, 2012.

研究業績

学術論文

1. **福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一, “PESQ と室内音響指標を用いた雑音・残響指標 NRSR-PA に基づく雑音・残響下音声認識性能の予測,” 電子情報通信学会論文誌 D, vol. J98-D, no. 3, Mar. 2015.（※ 採録決定）

2. 堀井 圭祐, **福森 隆寛**, 森勢 将雅, 中山 雅人, 西浦 敬信, 山下 洋一, 南條 浩輝, “雑音下音声受音における Weighted 反復スペクトル減算法を用いたミュージカルノイズの低減,” 電子情報通信学会論文誌 D, vol. J96-D, no. 3, pp. 664-674, Mar. 2013.

3. **Takahiro Fukumori**, Takanobu Nishiura, Masato Nakayama, Yuki Denda, Norihide Kitaoka, Takeshi Yamada, Kazumasa Yamamoto, Satoru Tsuge, Masakiyo Fujimoto, Takiguchi Tetsuya, Chiyomi Miyajima, Satoshi Tamura, Tetsuji Ogawa, Shigeki Matsuda, Shingo Kuroiwa, Kazuya Takeda, and Satoshi Nakamura, ” CENSREC-4:An Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments, ” *Acoustical Science and Technology*, vol. 32, no. 5, pp. 201-210, Sep. 2011.

4. **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, ”室内音響指標を用いた残響指標 RSR-Dn に基づく残響下音声認識性能の予測, ” 電子情報通信学会論文誌 D, vol. J94-D, no. 4, pp. 712-720, Apr. 2011.

国際会議

1. **Takahiro Fukumori**, Makoto Hayakawa, Masato Nakayama, Takanobu Nishiura,

and Yoichi Yamashita, "Evaluation of Clipping-Noise Suppression of Stationary-Noisy Speech Based on Spectral Compensation," *Proc. the 43rd International Congress on Noise Control Engineering (inter.noise)*, Paper ID: p422, Nov. 2014.

2. Takayuki Furoh, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "A Study of Degraded-Speech Identification Based on Speech Features Including Spectral Centroid," *Proc. the 43rd International Congress on Noise Control Engineering (inter.noise)*, Paper ID: p424, Nov. 2014.

3. Yuki Nagano, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Efficiency Evaluation of Subspace-Based Spectral Subtraction Based on Iterative Eigenvalue Analysis in Real Environments," *Proc. the 43rd International Congress on Noise Control Engineering (inter.noise)*, Paper ID: p417, Nov. 2014.

4. **Takahiro Fukumori**, Naoto Kakino, Masato Nakayama, Takanobu Nishiura, Yoichi Yamashita, and Hiroaki Nanjo, "Shouted Speech Detection in Noisy Environments Based on Rahmonic and Mel-Frequency Cepstrum Coefficients," *Proc. the 7th Forum Acusticum 2014 (FA2014)*, Paper ID: R20C_2, Sep. 2014.

5. Liang Li, Kyoko Hasegawa, **Takahiro Fukumori**, Wataru Wakita, Satoshi Tanaka, Takanobu Nishiura, Kozaburo Hachimura, and Hiromi Tanaka, "Digital Museums of Cultural Heritages in Kyoto: The Gion Festival in a Virtual Space," *Proc. the 16th International Conference on Human-Computer Interaction*, pp. 523-534, Jun. 2014.

6. **Takahiro Fukumori**, Masato Nakayama, Takanobu Nishiura, and Yoichi Yamashita, "Estimation of Speech Recognition Performance in Noisy and Reverberant Environments Using PESQ Score and Acoustic Parameters," *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Paper ID: 144, Oct. 2013.

7. **Takahiro Fukumori**, Masato Nakayama, Takanobu Nishiura, and Yoichi Yamashita, "Interactive Acoustic Sound Field Reproduction with Web System for Gion Festival," *Proc. Culture and Computing*, pp. 135-136, Sep. 2013.

8. Naoki Yoshimoto, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Evaluation of High-Realistic Acoustic Sound Field Reproduction Method for Gion Festival Music," *Proc. Culture and Computing*, pp. 133-134, Sep. 2013.

9. **Takahiro Fukumori**, Masato Nakayama, Takanobu Nishiura, and Yoichi Yamashita, "Performance Estimation of Speech Recognition Based on Perceptual Evaluation of Speech Quality (PESQ) and Acoustic Parameters under Noisy and Reverberant Environments with Corpus and Environment for Noisy Speech RECognition 4 (CENSREC-4)," *Proc. the 21st International Congress on Acoustics (ICA)*, Paper ID: 1aSCb21, Jun. 2013.

10. Naoto Kakino, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Experimental Study of Shout Detection with the Rahmonic Structure," *Proc. the 21st International Congress on Acoustics (ICA)*, Paper ID: 1aSCb6, Jun. 2013.

11. Takayuki Furoh, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Detection for Lombard Speech with Second-Order Mel-Frequency Cepstral Coefficient and Spectral Envelope in Beginning of Talking-Speech," *Proc. the 21st International Congress on Acoustics (ICA)*, Paper ID: 1aSCb8, Jun. 2013.

12. Makoto Hayakawa, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Suppression of Clipping Noise in Observed Speech Based on Spectral Compensation with Gaussian Mixture Models and Reference of Clean Speech," *Proc. the 21st International Congress on Acoustics (ICA)*, Paper ID: 1aSCb7, Jun. 2013.

13. Naoki Yoshimoto, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Evaluation of Reproducing High-Realistic Acoustic Sound Field Based on the Radiation Characteristics of Musical Accompaniment for Gion Festival," *Proc. 2013 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2013)*, pp. 169-172, Mar. 2013.

14. Keisuke Horii, **Takahiro Fukumori**, Masato Nakayama, Takanobu Nishiura, and Yoichi Yamashita, "Musical Tone Reduction for Sound-Quality Improvement by Weighted Iterative Spectral Subtraction in Real Noisy-Environments," *Proc. 2013 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2013)*, pp. 13-16, Mar. 2013.

15. Makoto Hayakawa, **Takahiro Fukumori**, Masato Nakayama, and Takanobu Nishiura, "Clipping Ratio Estimation for the Clipping Noise Suppression," *Proc. 2013 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2013)*, pp. 241-244, Mar. 2013.

16. **Takahiro Fukumori**, Takanobu Nishiura, and Yoichi Yamashita, "Digital Archive for Japanese Intangible Cultural Heritage Based on Reproduction of High-Fidelity Sound Field in Yamahoko Parade of Gion Festival, " *Proc. 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2012)*, pp. 549-554, Aug. 2012.

17. **Takahiro Fukumori**, Masato Nakayama, Masanori Morise, Takanobu Nishiura, and Yoichi Yamashita, "An Identification of Speaker-Dependence in Reverberant-Robust Speech Recognition, " *ACOUSTICS 2012*, Paper Number: 4pSP4, May 2012.

18. Keisuke Horii, Masato Nakayama, **Takahiro Fukumori**, Masanori Morise, Takanobu Nishiura, and Yoichi Yamashita, "The Determination of Dynamic

Subtraction for Spectral Subtraction towards Musical Tone Reduction, " *ACOUSTICS 2012*, Paper Number: 4aSP24, May 2012.

19. Naoto Kakino, **Takahiro Fukumori**, Yasuhiro Kuratani, Masato Nakayama, Masanori Morise, and Takanobu Nishiura, "Distant-Talking Speech Enhancement Based on Spectrum Restoring with Phoneme Labels, " *ACOUSTICS 2012*, Paper Number: 4aSP23, May 2012.

20. **Takahiro Fukumori**, Masanori Morise, Takanobu Nishiura, and Yoichi Yamashita, "A Study of Speaker-Dependence Criteria in Reverberant Speech Recognition, " *Proc. 2012 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2012)*, pp. 241-244, Mar. 2012.

21. Keisuke Horii, **Takahiro Fukumori**, Masanori Morise, Takanobu Nishiura, and Yoichi Yamashita, "Musical Tone Reduction Based on Auditory Sense for Spectral Subtraction," *Proc. the 40th International Congress on Noise Control Engineering (inter.noise)* Paper ID: Mon-P-22, Sep. 2011.

22. **Takahiro Fukumori**, Masanori Morise, and Takanobu Nishiura, "Interactive Acoustic Sound Field Reproduction with Binaural Recording of Yamahoko Parade of Gion Festival," *Proc. the 40th International Congress on Noise Control Engineering (inter.noise)*, Paper ID: Tue-P-35, Sep. 2011.

23. **Takahiro Fukumori**, Masanori Morise, Takanobu Nishiura, Yoichi Yamashita, and Hiroaki Nanjo, "The Estimation of Optimum Subtraction Parameters for Iterative Spectral Subtraction towards Musical Tone Reduction," *Proc. the 40th International Congress on Noise Control Engineering (inter.noise)*, Paper ID: Mon-P-21, Sep. 2011.

24. Woong Choi, **Takahiro Fukumori**, Kohei Furukawa, Kozaburo Hachimura, Takanobu Nishiura, and Keiji Yano, "Virtual Yamahoko Parade in Virtual

Space," *Proc. the 6th Joint Workshop on Machine Perception and Robotics (MPR2010)*, Oct. 2010

25. Keiji Yano, Hiromi Tanaka, Kozaburo Hachimura, Takanobu Nishiura, Woong Choi, **Takahiro Fukumori**, Kohei Furukawa, Wataru Wakita, and Masaru Tsuchida, "Digital Museum of the World Intangible Cultural heritage "Kyoto Gion Festival" within Virtual Kyoto" *Proc. the 3rd International Symposium on Sentinel Earth − Advance in Satellite Imagery Data and GIS and Their Application −*, Nov. 2010.

26. Woong Choi, **Takahiro Fukumori**, Kohei Furukawa, Kozaburo Hachimura, Takanobu Nishiura, and Keiji Yano, "Virtual Yamahoko Parade in Virtual Environment, " *Proc. NICOGRAPH INTERNATIONAL 2010*, pp. 70-71, Jun. 2010.

27. Hiromi Tanaka, Keiji Yano, Kozaburo Hachimura, Takanobu Nishiura, Woong Choi, **Takahiro Fukumori**, Kohei Furukawa, Wataru Wakita, Masaru Tsuchida, and Naoki Saiwaki, "Digital Archiving of the World Intangible Cultural Heritage "Gion Festival in Kyoto": Reproduction of "Fune-boko" Float of the Gion Festival Parade in "Virtual Kyoto" —, " *Proc. ASIAGRAPH 2010*, Jun. 2010.

28. Woong Choi, **Takahiro Fukumori**, Kohei Furukawa, Kozaburo Hachimura, Takanobu Nishiura, and Keiji Yano, "Virtual Yamahoko Parade in Virtual Kyoto, " *Proc. SIGGRAPH 2010 (The 37th International Conference and Exhibition on Computer Graphics and Interactive Techniques)*, ISBN 978-1-4503-0210-4/10/0007, Jul. 2010.

29. **Takahiro Fukumori**, Masanori Morise, and Takanobu Nishiura, "Performance Estimation of Speech Recognition Based on Acoustic Parameters under Reverberation Environments with CENSREC-4, " *Proc. the 21st International Congress on Acoustics (ICA)*, Paper ID: 166, Aug. 2010.

30. **Takahiro Fukumori**, Masanori Morise, and Takanobu Nishiura, "Performance Estimation of Reverberant Speech Recognition Based on Reverberant Criteria RSR-Dn with Acoustic Parameters, " *Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH 2010)*, pp. 562-565, Sep. 2010.

31. **Takahiro Fukumori**, Yoshiki Hirano, Masanori Morise, and Takanobu Nishiura, "Performance Estimation of Speech Recognition Based on Acoustic Parameters under Reverberation Environments, " *Proc. the 10th Western Pacific Acoustics Conference (WESPAC 2009)*, Paper ID: 128, Sep. 2009.

研究会

1. 三宅 亮太，**福森 隆寛**，中山 雅人，西浦 敬信，"反復スペクトル減算のための連検定に基づく雑音環境識別手法の検討"，電子情報通信学会技術研究報告，vol. 114, no. 191, SIP2014-73, pp. 7-12, Aug. 2014.

2. 長野 優貴, **福森 隆寛**, 中山 雅人, 西浦 敬信, "反復固有値解析に基づくサブスペース型スペクトルサブトラクションの検討," 第 27 回 回路とシステムワークショップ, pp. 363-368, Aug. 2014.

3. **福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一，"室内音響指標を用いた最適音響モデル選択による残響下音声認識評価"，電子情報通信学会技術研究報告，vol. 113, no. 452, SP2013-108, pp. 7-12, Feb, 2014.

4. 柿野 直人，**福森 隆寛**，中山 雅人，西浦 敬信，南條 浩輝，"雑音環境下における叫び声検出のための特徴量次元数の削減"，電子情報通信学会技術研究報告，vol. 113, no. 452, SP2013-111, pp. 21-22, Feb, 2014.

5. 三宅 亮太，**福森 隆寛**，中山 雅人，西浦 敬信，"スペクトル減算法における高次統計量に基づく最適反復回数指標の検討"，電子情報通信学会技術研究報告，vol. 113, no. 452, SP2013-110, pp. 19-20, Feb, 2014.

6. 早川 惇, **福森 隆寛**, 中山 雅人, 西浦 敬信，“周波数領域のパワー包絡復元に基づく観測音声のクリッピングノイズ抑圧”，電子情報通信学会信号処理シンポジウム, pp. 180-181, Nov. 2013.

7. **福森 隆寛**，堀井 圭祐，中山 雅人，西浦 敬信，山下 洋一，“聴覚特性に基づく重み付け反復スペクトル減算法による音質改善の検討”，情報処理学会音声言語情報処理研究会，vol. 2013-SLP-96, no.6. pp. 1-8, May, 2013.

8. **福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一，“雑音・残響指標 NRSR-PDn に基づく雑音・残響下音声認識の予測性能評価”，電子情報通信学会技術研究報告，vol. 113, no. 29, SP2013-6, pp. 31-36, May, 2013.

9. 早川 惇, **福森 隆寛**, 中山 雅人, 西浦 敬信, “周波数帯域別スペクトル包絡補正による音声のクリッピングノイズ抑圧法の検討,” 電子情報通信学会技術研究報告, vol. 112, no. 423, SIP2012-101, pp. 121-126, Feb. 2013.

10. 吉元 直輝, **福森 隆寛**, 中山 雅人, 西浦 敬信, “祇園囃子の放射特性を考慮した高臨場音場の構築, ” 電子情報通信学会技術研究報告, vol. 112, no. 386, MVE2012-64, pp. 163-168, Jan. 2013.

11. **福森 隆寛**, 吉元 直輝, 中山 雅人, 西浦 敬信, 山下 洋一, “祇園祭音場の高臨場感再生に基づく日本無形文化財のデジタルアーカイブ, ” 日本音響学会聴覚研究会, vol. 42, no. 7, pp. 579-584, Oct. 2012.

12. **福森 隆寛**, 中山 雅人, 森勢 将雅, 西浦 敬信, 山下 洋一, “残響下音声認識における発話様式の実験的分析と評価, ” 電子情報通信学会技術研究報告, vol. 112, no. 49, SP2012-31, pp. 179-184, May 2012.

13. 堀井 圭祐，**福森 隆寛**，森勢 将雅，西浦 敬信，山下 洋一, “減算係数重み付けスペクトル減算によるミュージカルノイズ低減の検討,” 日本音響学会聴覚研究会，vol. 41, no. 7, pp. 577-582, Oct. 2011.

14. **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, “残響環境下音声認識における発話位置・話者依存性の分析評価,” 電子情報通信学会技術研究報告, vol. 111, no.

28, SP2011-10, pp. 55-60, May 2011.

15. 堀井 圭祐, **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, “帯域分割型スペクトル減算に基づくミュージカルノイズ低減手法の検討,” 電子情報通信学会技術研究報告, vol. 111, no. 28, SP2011-1, pp. 1-5, May 2011.

16. 堀井 圭祐, **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, “スペクトル減算を用いた音質改善のための減算係数最適化の検討, ” 電子情報通信学会技術研究報告, vol. 110, no. 401, SP2010-113, pp. 59-64, Jan. 2011.

17. **福森 隆寛**, 森勢 将雅, 西浦 敬信, “サラウンド収録に基づく祇園祭山鉾巡行の高忠実度音場再現, ” 電子情報通信学会技術研究報告, vol. 110, no. 382, MVE2010-122, pp. 371-376, Jan. 2011.

18. **福森 隆寛**, 森勢 将雅, 西浦 敬信, 南條 浩輝, “最適フロアリング係数を用いた反復スペクトルサブトラクションによるミュージカルノイズの低減, ” 電子情報通信学会技術研究報告, vol. 110, no. 56, SP2010-8, pp. 43-48, May 2010.

19. **福森 隆寛**, 森勢 将雅, 西浦 敬信, “残響指標 RSR-Dn に基づく残響環境下音声認識の予測性能評価, ” 電子情報通信学会技術研究報告, vol. 110, no. 56, SP2010-3, pp. 13-18, May 2010.

大会発表

1. 不老 孝之，**福森 隆寛**，中山 雅人，西浦 敬信，“スペクトル重心を用いた劣化音声識別の検討 ,” 2014 年日本音響学会秋季研究発表会, pp. 81-82, Sep. 2014.

2. 三宅 亮太，**福森 隆寛**，中山 雅人，西浦 敬信，“反復スペクトル減算のための連検定に基づく雑音環境識別 ,” 2014 年日本音響学会秋季研究発表会, pp. 1553-1554, Sep. 2014.

3. **福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一, “残響下音声認識における室内音響指標に基づく最適音響モデル選択法の検討,” 2014 年日本音響学会春季研究発表会, pp. 15-18, Mar. 2014.

4. 長野 優貴，**福森 隆寛**，中山 雅人，西浦 敬信, “サブスペース型スペクトルサブトラクションに基づく雑音抑圧の検討,” 2014 年日本音響学会春季研究発表会, pp. 245-246, Mar. 2014.

5. 柿野 直人，**福森 隆寛**，中山 雅人，西浦 敬信，南條 浩輝, “Rahmonic とメルケプストラムを用いた叫び声検出における特徴ベクトルの次元数削減,” 2014 年日本音響学会春季研究発表会, pp. 233-234, Mar. 2014.

6. 三宅 亮太，**福森 隆寛**，中山 雅人，西浦 敬信, “高次統計量を用いた雑音環境推定に基づくスペクトル減算法の最適パラメータ決定法の検討,” 2014 年日本音響学会春季研究発表会, pp. 247-248, Mar. 2014.

7. 早川 惇，**福森 隆寛**，中山 雅人，西浦 敬信, “スペクトル包絡の補償に基づく不特定話者音声のクリッピングノイズ抑圧,” 2014 年日本音響学会春季研究発表会, pp. 19-20, Mar. 2014.

8. 不老 孝之，**福森 隆寛**，中山 雅人，西浦 敬信, “Rahmonic とメルケプストラムに基づく劣化音声識別の検討,” 2014 年日本音響学会春季研究発表会, pp. 235-236, Mar. 2014.

9. 竹下 洋正, **福森 隆寛**, 中山 雅人, 西浦 敬信, “言語・音響尤度を併用した叫び声検出,” 電子情報通信学会関西支部学生会 第 19 回学生会研究発表講演会, p. 65, Mar. 2014.

10. 長野 優貴, **福森 隆寛**, 中山 雅人, 西浦 敬信, “サブスペース型スペクトルサブトラクションに基づく音声強調,” 電子情報通信学会関西支部学生会 第 19 回学生会研究発表講演会, p. 64, Mar. 2014.

11. 不老 孝之，**福森 隆寛**，中山 雅人，西浦 敬信，“劣化音声識別のための音響特徴量の分析・評価,” 第 16 回日本音響学会関西支部若手研究者交流研究発表会, p. 9, Dec. 2013.

12. 三宅 亮太，**福森 隆寛**，中山 雅人，西浦 敬信，“スペクトル減算法のための高次統計量を用いた雑音環境推定法の検討,” 第 16 回日本音響学会関西支部若手

研究者交流研究発表会, p. 8, Dec. 2013.

13. 不老 孝之，**福森 隆寛**，中山 雅人，西浦 敬信, “統計的確率モデルに基づく発話頭におけるロンバード音声検出の検討,” 平成 25 年電気関係学会関西連合大会, pp. 337-338, Nov. 2013.

14. 柿野 直人，**福森 隆寛**，中山 雅人，西浦 敬信，南條 浩輝, “Rahmonic とメルケプストラムを用いた叫び声検出の検討,” 2013 年日本音響学会秋季研究発表会, pp. 169-170, Sep. 2013.

15. 安福 元貴，**福森 隆寛**，中山 雅人，西浦 敬信, “残響曲線を利用した CMN による残響抑圧法の検討,” 2013 年日本音響学会秋季研究発表会, pp. 141-142, Sep. 2013.

16. 三宅 亮太，**福森 隆寛**，中山 雅人，西浦 敬信, “実騒音環境下における重み付き反復スペクトル減算法の最適パラメータの客観評価,” 2013 年日本音響学会秋季研究発表会, pp. 139-140, Sep. 2013.

17. 早川 惇，**福森 隆寛**，中山 雅人，西浦 敬信, “時間・周波数領域のパワー包絡補償に基づく連続音声のクリッピングノイズ抑圧,” 2013 年日本音響学会秋季研究発表会, pp. 135-136, Sep. 2013.

18. **福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一, “残響指標 RSR-Dn に基づく残響抑圧音声の認識性能予測,” 2013 年日本音響学会春季研究発表会, pp. 31-34, Mar. 2013.

19. 小川 純平，林田 亘平，**福森 隆寛**，中山 雅人，西浦 敬信, “主観的危険度に基づく環境音識別手法の評価,” 2013 年日本音響学会春季研究発表会, pp. 23-26, Mar. 2013.

20. 堀井 圭祐，**福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一, “Weighted 反復スペクトル減算法における反復回数の最適化,” 2013 年日本音響学会春季研究発表会, pp. 9-12, Mar. 2013.

21. 早川 惇，**福森 隆寛**，中山 雅人，西浦 敬信, “帯域別スペクトル補正による不特定話者音声のクリッピングノイズ抑圧,” 2013 年日本音響学会春季研究発表会, pp. 15-18, Mar. 2013.

22. 柿野 直人，**福森 隆寛**，中山 雅人，西浦 敬信, “Rahmonic を用いた叫び声の特徴量抽出,” 2013 年日本音響学会春季研究発表会, pp. 137-138, Mar. 2013.

23. 不老 孝之，**福森 隆寛**，中山 雅人，西浦 敬信, “発話頭における重み付き 2 次メル周波数ケプストラム係数とスペクトル傾斜を用いたロンバード音声検出の研究,” 電子情報通信学会関西支部学生会 第 18 回学生会研究発表講演会, p. 69, Mar. 2013.

24. 早川 惇，**福森 隆寛**，中山 雅人，西浦 敬信，“スペクトル包絡の周波数帯域別補正に基づく音声のクリッピングノイズ抑圧—クリッピング比の推定に関する一検討—,” 日本音響学会関西支部第 15 回若手研究者交流研究発表会, p. 8, Dec. 2012.

25. 柿野 直人，**福森 隆寛**，中山 雅人，西浦 敬信，“危機的状況下における発声音声の特徴量分析,” 日本音響学会関西支部第 15 回若手研究者交流研究発表会, p. 8, Dec. 2012.

26. 吉元 直輝，**福森 隆寛**，中山 雅人，西浦 敬信，“祇園囃子の放射特性を考慮した高臨場感再生の検討，” 日本音響学会関西支部第 15 回若手研究者交流研究発表会, p. 16, Dec. 2012.

27. **福森 隆寛**，中山 雅人，西浦 敬信，山下 洋一, “PESQ と室内音響指標を用いた雑音・残響下における音声認識性能の予測, ” 2012 年日本音響学会秋季研究発表会, pp. 37-38, Sep. 2012.

28. 堀井 圭祐，**福森 隆寛**，森勢 将雅，中山 雅人，西浦 敬信，山下 洋一, “実騒音環境下における Weighted 反復スペクトル減算法のパラメータ最適化に関する実験的検討, ” 2012 年日本音響学会秋季研究発表会, pp. 9-10, Sep. 2012.

29. 早川 惇，**福森 隆寛**，中山 雅人，西浦 敬信, “帯域分割スペクトル包絡の制御・置換による音声のクリッピングノイズ抑圧, ” 2012 年日本音響学会秋季研究発表会, pp. 11-12, Sep. 2012.

30. 小川 純平，林田 亘平，**福森 隆寛**，中山 雅人，西浦 敬信, “環境音モデル構築のための音響特徴量に基づく環境音分類に関する基礎的検討, ” 2012 年日本音響学会秋季研究発表会, pp. 3-4, Sep. 2012.

31. **福森 隆寛**, 吉元 直輝, 柿野 直人, 西浦 敬信, “Web システムによる祇園祭山鉾巡行の高臨場音場体験, ” 第 17 回計算工学講演会, 発表番号 : A-5-2, May 2012.

32. **福森 隆寛**，森勢 将雅，西浦 敬信，山下 洋一，“発話速度を用いた残響下音声認識の話者依存度推定法の検討，” 2012 年日本音響学会春季研究発表会, pp. 29-30, Mar. 2012.

33. 堀井 圭祐，**福森 隆寛**，森勢 将雅，西浦 敬信，山下 洋一，“オクターブバンド分割型スペクトル減算によるミュージカルノイズ低減の検討，” 2012 年日本音響学会春季研究発表会, pp. 31-32, Mar. 2012.

34. 柿野 直人，堀井 圭祐，**福森 隆寛**，森勢 将雅，西浦 敬信，“遠方音声受音のためのスペクトル復元に基づく音声強調法，” 電子情報通信学会関西支部学生会 第 17 回学生会研究発表講演会, p. 95, 2012.

35. **福森 隆寛**，森勢 将雅，西浦 敬信，“京都祇園祭の時空散歩への挑戦 ～船鉾の高臨場バーチャル再現～，” 日本音響学会関西支部第 14 回若手研究者交流研究発表会, p. 26, Dec. 2011.

36. 堀井 圭祐，**福森 隆寛**，森勢 将雅，西浦 敬信，山下 洋一，“音声の時間変動を考慮したスペクトル減算によるミュージカルノイズ低減の検討，” 日本音響学会関西支部第 14 回若手研究者交流研究発表会, p. 21, Dec. 2011.

37. **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, “発話者の音声特徴量に基づく残響下音声認識の話者依存評価, ” 平成 23 年電気関係学会関西連合大会, pp. 606-607, Oct. 2011.

38. 堀井 圭祐，**福森 隆寛**，森勢 将雅，西浦 敬信，山下 洋一，“帯域分割型スペクトル減算によるミュージカルノイズ低減のための減算係数最適化の検討，” 2011 年日本音響学会秋季研究発表会, pp. 125-126, Sep. 2011.

39. **福森 隆寛**，森勢 将雅，西浦 敬信，山下 洋一，“残響下音声認識における話者依存尺度の検討，” 2011 年日本音響学会秋季研究発表会, pp. 7-8, Sep. 2011.

40. 堀井 圭祐, **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, “聴覚特性重み付けスペクトル減算によるミュージカルノイズの低減, ” 電子情報通信学会関西支部学生会 第 16 回学生会研究発表講演会, p. 76, 2011.

41. **福森 隆寛**, 森勢 将雅, 西浦 敬信, “残響下音声認識性能の発話位置・話者依存性に対する評価, ” 2011 年日本音響学会春季研究発表会, pp. 153-154, Mar. 2011.

42. **福森 隆寛**, 森勢 将雅, 西浦 敬信, 山下 洋一, “残響下音声認識性能の発話位置依存性に対する評価, ” 平成 22 年電気関係学会関西連合大会, pp. 606-607, Nov. 2010.

43. 堀井 圭祐, **福森 隆寛**, 森勢 将雅, 西浦 敬信, 南條 浩輝, “聴覚特性に基づくスペクトル減算におけるミュージカルノイズの低減, ” 平成 22 年電気関係学会関西連合大会, p. 602, Nov. 2010.

44. **福森 隆寛**, 森勢 将雅, 西浦 敬信, 南條 浩輝, “高フロアリング係数を用いた反復スペクトルサブトラクションによるミュージカルノイズの低減の検討, ” 第 9 回情報科学技術フォーラム (FIT2010), pp. 203-204, Sep. 2010.

45. **福森 隆寛**, 林田 亘平, 森勢 将雅, 西浦 敬信, “祇園祭山鉾巡行における高忠実度音場再現に基づく日本無形文化財のデジタルアーカイブ化, ” 2010 年日本音響学会秋季研究発表会, pp. 461-462, Sep. 2010.

46. **福森 隆寛**, 森勢 将雅, 西浦 敬信, “室内音響指標（Clarity・Definition）を用いた残響下音声認識性能の予測精度評価, ” 2010 年日本音響学会秋季研究発表会, pp. 123-124, Sep. 2010.

47. 堀井 圭祐, **福森 隆寛**, 森勢 将雅, 西浦 敬信, 南條 浩輝, “スペクトル減算のための等ラウドネス曲線に基づく減算係数最適化の検討, ” 2010 年日本音響学会秋季研究発表会, pp. 711-712, Sep. 2010.

48. **福森 隆寛**，森勢 将雅, 西浦 敬信, “CENSREC-4 を用いた室内音響指標に基づく残響下音声認識性能の推定, ” 2010 年日本音響学会春季研究発表会, pp. 245-246, Mar. 2010.

49. **福森 隆寛**, 森勢 将雅, 西浦 敬信, “室内音響指標に基づく残響下音声認識性能の推定と評価, ” 平成 21 年度情報処理学会全国大会, pp. 2-255-2-256, Mar. 2010.

50. **福森 隆寛**，森勢 将雅, 西浦 敬信, “室内音響指標と回帰分析を用いた残響下音声認識性能の推定,” 2009 年日本音響学会秋季研究発表会, pp.151-152, Sep. 2009.

著書

1. 八村 広三郎, 田中 弘美 (編), **福森 隆寛**, 森勢 将雅, 西浦 敬信 (著), “シリーズ日本文化デジタル・ヒューマニティーズ 06 「デジタル・アーカイブの新展開」, ” ナカニシヤ出版, ISBN 978-4-7795-0585-0, Mar. 2012. (※ 福森：第 5 章, 第 6 章, 第 14 章, 第 15 章担当)

2. Takanobu NISHIURA, and **Takahiro FUKUMORI**, “Suitable Reverberation Criteria for Distant-Talking Speech Recognition, ” *INTECH Speech Technologies / Book 1*, ISBN 978-953-307-152-7, Jun. 2011.

受賞

1. **福森 隆寛**, （財）電気通信普及財団 第 27 回テレコムシステム技術学生賞 入賞受賞発表（”室内音響指標を用いた残響指標 RSR-Dn に基づく残響下音声認識性能の予測,” 電子情報通信学会論文誌 D, vol. J94-D, no. 4, Apr. 2011.）

2. **福森 隆寛**, 電子情報通信学会関西支部長賞 功労賞 受賞理由（”電子情報通信学会関西支部学生会に学生幹事長として貢献, Mar. 2012.）

3. **福森 隆寛**，日本音響学会関西支部 若手奨励賞 受賞発表 (”京都祇園祭の時空散歩への挑戦 〜船鉾の高臨場バーチャル再現〜，” 日本音響学会関西支部第 14 回若手研究者交流研究発表会, p. 26, Dec. 2011.)

4. **福森 隆寛**, 電気関係学会関西連合大会 奨励賞 受賞発表 (”残響下音声認識性能の発話位置依存性に対する評価, ” 平成 22 年電気関係学会関西連合大会, pp. 606-607, Nov. 2010.)