学 位 論 文

機械学習を用いた騒音付加音声の了解度推定に関する研究

山形大学大学院理工学研究科

小林 洋介

学 位 論 文

機械学習を用いた騒音付加音声の了解度推定に関する研究

2013年3月

山形大学大学院理工学研究科

小林 洋介

Doctoral Thesis

A study on speech intelligibility estimation using machine learning in noisy conditions

March 2013

Graduate School of Science and Engineering
Yamagata University

Yôsuke Kobayashi

概 要

本論文では，主として屋外を対象とした騒音に暴露される環境で，音声システムを用いた際の了解度の推定を検討した．了解度は人間による音声の知覚実験であり，システムの設計評価に用いるには金銭的，時間的コストがかかる．このため，高精度な了解度推定法が望まれる．既往の研究では，主として残響環境下での音声了解度推定の目安となるインデックス値を予測するSTI（Speech Transmission Index）や，定常騒音下での音声了解度インデックス値を予測するSII（Speech Intelligibility Index）がある．これらはISOやANSIで標準化されており，広く利用されている．しかし，インデックス値ではシステム設計の目安となるものの，主観評価値に対応する推定了解度を求めることはできない．また，STI，SII共に騒音のパワースペクトルの時間変動が定常であることを想定している．このため，パワースペクトルの時間変動が大きい騒音の予測は困難である．本論文ではこのような非定常な騒音を用いた場合でも，高精度な了解度推定ができる方式を提案し，その汎化性能を評価した．以下に本論文の各章の要約を述べる．

1章　序論

本章では，既存の音声品質について概観し，主に音声の了解度に関する主観評価と客観評価の研究について整理し，本論文の意義について述べる．

2章　バイノーラル音声システムと既存尺度による了解度推定の検討

本章では，筆者がこれまで検討してきた両耳受聴を利用した音声システムをテストモデルとして，4種の騒音を用いた音声了解度の主観評価を行い，既存の16尺度とシグモイド・カーブフィッティングを用いた了解度推定について検討し，以下の課題を導出する．

a. Sustention子音特徴（狭窄音と閉鎖音の聴き分け）は騒音種の影響が顕著

b. 推定条件によって異なる最適な聴覚重みを用いた周波数重み付セグメンタルSNRの推定性能が高い

3章　機械学習を用いた騒音付加音声了解度推定法の概略

本章では2章で明らかとなった課題について，機械学習を用いた解決法を提案する．課題a.に関しては，データベース内の長時間の騒音を3 secのLong Frame（以下，LF）で分割し，個々のLFを1つの騒音とみなして音色を分析する．音色特徴量は，MIR（Music Information Retrieval）で用いられる音響特徴を15次元求め，教師なし学習によるデータ分類アルゴリズムのx–meansクラスタリングを用いることを提案する．

課題b.については，既存の聴覚重みではなく，推定したい主観評価単語に最適な重みを求めるために，帯域別に求めたセグメンタルSNRを特徴量とし，教師つき学習の中から汎化性能が高いとされるサポートベクトル回帰（Support Vector Regression :SVR）を用いることを提案する．

4 章　騒音クラスタリングの検討とその評価

本章では，電子協騒音データベースダイジェスト版から 17 騒音を選択し，3 章で提案した方式で解析する．その結果，騒音クラスタが 3 個作成されることを示す．また，各クラスタから騒音を 32 個選択し，主観評価を行っい，クラスタ間の了解度差は SNR が 0 dB のときの了解度差が 0.4 以上あり，傾向差が顕著であり，分散分析による検定結果でもクラスタ間要因に有意差があり，騒音クラスタリングが有効であることを示す．また，既存の品質尺度 5 種を用いたシグモイド・カーブフィッティングによる推定関数を 5 種作成し，交差検定による RMSE を比較したところ，周波数重み付セグメンタル SNR の推定性能は 2 章と同様に高いことを示す．

5 章　サポートベクトル回帰による了解度推定関数の作成

4 章で検討した 3 つの騒音クラスタごと及びクラスタリングを行わない場合の 4 条件で，SVR を含む L1 正則化回帰手法 5 種（リッジ回帰，Lasso（Least absolute shrinkage and selection operator）を用いた回帰，L1 正則化 RBF(Radial Basis Function) カーネル回帰，線形カーネルによる SVR，RBF カーネルによる SVR）を比較する．教師信号には 4 章の主観評価結果を用い，特徴量にはクリティカルバンドで帯域分割したセグメンタル SNR と 1/3 オクターブバンドで帯域分割したセグメンタル SNR を用いる．交差検定誤差を用いて比較した結果，1/3 オクターブバンドによる帯域分割は全ての回帰手法でクリティカルバントによる帯域分割に劣ることを示す．一方で，交差検定誤差では最適な回帰手法を求めるには至らないことも示す．

6 章　オープンテストによる総合性能評価

4 章で作成した 5 種の既存尺度による推定関数および，5 章で作成したクリティカルバンドセグメンタル SNR を用いた 5 種の推定関数を関数作成時に用いなかった騒音によるオープンテストで比較する．提案法であるクリティカルバンドセグメンタル SNR と RBF カーネルを用いた SVR の組み合わせは，他の推定関数で最も性能が高い fwSNRseg を用いたシグモイド・カーブフィッティングによる推定関数の 0.77 倍の RMSE となり，汎化性能の高い了解度推定関数であることが実証する．

7 章　結論

本章では，本論文を総括し，今後の検討課題について述べる．

Abstract

Today, speech intelligibility tests are important to evaluate mobile and wireless speech systems. In the research and the development of speech communication systems, speech intelligibility has been evaluated as a quality assessment of telephone speech quality. On the other hand, due to the popularity of mobile devices, speech communication has increased in various environments. Therefore, the type of ambient noise to interfere with the speech communication has increased in variety. If the noise environment is different, the prediction of intelligibility is very difficult. Thus, I propose a speech intelligibility estimation method using machine learning technology to solve these problems in this thesis. Following is the content of this thesis.

Chapter 1 is the introduction of this thesis.

In chapter 2, I estimate the intelligibility of binaural speech system using a conventional quality measure. As a result, I found the following issues.

- The phonetic feature sustention is largely influenced by noise.
- Segmental SNR using frequency weighting gives the best estimation performance.

In chapter 3, I propose an intelligibility estimation method including the following.

- Ambient noise clustering using Music Information Retrieval (MIR) features
- Speech intelligibility estimator using the support vector regression (SVR)

In chapter 4, I evaluate the noise clustering method using subjective intelligibility tests. As a result, 3 clusters were generated and significant difference was seen in the cluster factor in the ANOVA test.

In chapter 5, I compare critical bands segmental SNR (cbSNRseg) and 1/3 octave bands SNRseg (obSNRseg) and used the cross–validation RMSE in 5 regression methods including SVR. As a result, the weighted sum of RMSE using cbSNRseg is better than obSNRseg with RMSE reduction factor of about 0.8 compared to all other regression methods.

In chapter 6, I compare the performance of each regression methods in open tests (an unknown noise conditions). As a result, the best regression method is the SVR using the RBF kernel, in which RMSE is reduced by a factor of about 0.77 compared to other regression methods.

In chapter 7 is the summary of this thesis.

目 次

図 目 次

表 目 次

第1章　序論

1.1　研究の背景

我々の生活の中で，音情報が果たしている役割は非常に大きい．これは，ここ 20 年足らずの間に爆発的に普及した携帯電話や，ポータブルオーディオプレイヤー，およびそれらを結びつける通信技術の発展による,「いつでも，どこでも，だれでも」といったユビキタスな音コミュニケーションによって可能となった．歴史を紐解くと，電話の発明は 1876 年にアレクサンダー・グラハム・ベルに，録音再生可能な蓄音機の発明は 1877 年にトーマス・アルバ・エジソンによってなされた．その後，今日までの約 130 年間に多くの技術研究開発によって，これらの音響機器の性能は大きく向上してきた．

これらの音響機器の発展において，機器・システム性能評価指標には必ずと言っていいほど「音質」[1]の項目があり，従来の機器やサービスとの比較が行われている．音を扱うシステムの性能を工学的な見地から評価する方法として，物理指標からシステムの性能を計測する客観評価が行われている．これは，通信や蓄積・再生，拡声器等を用いて，伝えたい音情報と，伝わる音情報をシステムへの入力と出力とみなし，周波数特性，歪率，音声／信号対雑音比（Speech / Signal to Noise Ratio : SNR[2]）等の物理的な指標からシステム性能を評価することである．明確な物理指標を用いることから，理想的な環境では究極的なところまで性能向上が可能と考えられるものの，実際には多様な妨害が入り，不可能である．また，物理刺激と心理的な人間の知覚はなんらかの相関関係にあることが多いものの，完全に線形になることはほとんどなく，物理指標だけで最適なシステム設計は不可能と言ってよい．そこで,「どの程度まで物理指標を改善すれば人間の認知限界未満になるか」という観点での目標値が必要になる．つまり，音の受け手である人間の判断によるシステム評価が必要になる．これを客観評価に対して主観評価という．主観評価の目的は，人間の心理量の計測であり，心理指標による計測と統計処理が多用される．つまり，計測されるものは個々人の心理指標の価値判断値であり，これを多人数の統計量とするため相当な分散がある．このため，主観評価は非常に（人的・金銭的・時間的）コストがかかる．よって主観評価は必要最小限であることが望ましく，実施すべき主観評価の規模が事前にわかることが望ましい．また，

[1]本論文では，人間の発声した音で特に言語情報を含み歌声を除くものを「音声」とし，その品質と電気信号を「音声品質」,「音声信号」と明確に音声とつけて呼ぶ．それ以外の音は「○○音」,「○○音品質」,「○○音信号」と表記する．例外的に「オーディオ信号」,「オーディオ品質」を用いる場合は，フルバンド（上限周波数が 20 kHz 以上を前提とした）の主として音楽・放送用の信号を指す．歌声はその特徴が音声的な手法が馴染むものだけではなく，様々な議論がされている（例えば，大石らによる朗読と歌声の違いに関する分析 [1] がある）．しかし，本論文では歌声の品質自体は扱わないため，オーディオ信号の一種として取り扱うこととする.「音質」と表記する場合は一般用語として，あらゆる音を対象とした,「音の品質」を指す.

[2]主音声と雑音のパワー比．1.2.5 項で詳しく述べる．S を Signal とする場合の N は，電気信号としてのノイズ（雑音）であり，S を Speech とする場合の N は背景騒音であることが多い．本論文は，この二つの立場の横断的な内容であるため,「雑音」は電気信号処理系由来の音,「騒音」は発話者の背景にある主音声以外の音とする．ただし，白色雑音や雑音抑圧といった普通名詞として広く用いられている名詞に関してはこの限りではない.

完成したシステムの性能のみならず，今後のシステムの最適設計に利用可能な様に，心理指標と物理指標の対応付けを行い，主観評価結果を推定できるようにする必要がある．

現在までの音声品質評価は，(背景騒音がほとんどない）環境を想定している電話網の音声品質評価[3]に代表されるように，発話環境と受聴環境の双方がある程度静かな環境を想定しているものが多い．一方で音声強調のための雑音抑圧音声は，会話をする環境が非常に「うるさい」場合の発話を前提としているものの，主音声以外の音を抑圧することを考慮しているため，受聴環境は静かな環境を前提としている．しかし，発話環境，受聴環境にかかわらず，主音声以外の音は何も情報を持たないのであろうか．

著者には，「PHS で通話しているとき，お互いの電話から鈴虫の鳴き声が聴こえる」といった経験があり，そこから話が弾んだことがある．もし，主音声以外の音を全て抑圧してしまえば，この様な「心地よい良い騒音」が埋もれてしまうこともありうる．もちろん，騒音は主音声にとっては妨害であることが基本であり，会話の妨害となる騒音は抑圧した方が良い．では，「心地よい騒音」と「会話の妨害となる騒音」の違いはなんだろうか．

この問題の解は人によって異なる部分も大きいだろうが，著者は会話を前提としている以上，主音声の了解性（聴き取りやすさ）を保つことが絶対条件であると考える．了解度が一定以上あるうえで，発話の音量感や自然さが十分あれば，背景騒音があったとしてもそれほど苦痛にはならない．背景騒音を抑圧する雑音抑圧による音声強調は非常に強力であるものの，方式によってはミュージカルノイズ[4]が発生し，より不自然な音声となる．よって，背景騒音は了解度が十分な値を取る範囲では抑圧する必要は無く，主音声以外の情報を含んだ音声通信が可能になる．このようなシステムの設計のために，騒音の種類による了解度変化を含めた了解度推定が必要である．特に，高性能なモバイルデバイスを用いた情報付加技術である拡張現実（Augmented Reality）[2] の一環として，高臨場感音響技術 [3, 4] を用いた拡張音響現実感（以下，Augumented Audio Reality : AAR と呼ぶ）での利用を考えると，端末で生成した付加音（声）に対する周囲の様々な騒音や残響によるマスキングを考慮しなければならない．よって本論文では，多数の騒音種による了解度の変化を検討する．

了解度は，単語や文章といった音声の言語的手がかりを利用した聴取実験である．固定電話網の品質評価では，発話／受聴の双方がある程度静かな環境を想定することが可能なため，主音声への妨害はほとんど存在しなく，高了解度な通信は容易である．一方で，携帯電話はその可搬性から，騒音の非常に大きい環境での通話が考えられ，発話位置マイクへの背景騒音の混入や，受聴位置での騒音暴露による了解度低下は避けられない．前述した了解度と騒音の関係を明らかにするためには，騒音の種類や音量と了解度の関係をよく検証する必要がある．しかしながら，世の中に存在するすべての騒音について実験を行うのは不可能であり，騒音種が異なっても推定性能が高い（汎化能力[5]が高い）方式を検討する必要がある．

以上のような背景から，本論文では，騒音の影響が非常に大きいと考えられる低 SNR 環境[6]での多種の騒音環境について了解度の主観評価とその推定を検討する．このため，1.2 節では，了解度よりも広い範囲での既存の音質評価研究の概観を述べ，音質評価に関する諸問題を整理する．特

[3]アナログ電話網の時代は，電話網による信号減衰の影響が大きく，音声の明瞭度・了解度による電話網評価が行われたが，現在は他の品質尺度に置き換わっている．詳細は次節で述べる．

[4]非線形信号処理により，周波数領域で特定の周波数に成分が現れたり消えたりする現象によるトーン性のノイズ

[5]推定モデルの作成時に用いたデータだけに対してだけでなく，未知の新たなデータに対しても正しく推定できる能力．

[6]主音声と騒音のパワーが等しくなる SNR 0 dB 以下を主に想定する．

に 2 章以降で用いる主観・客観品質評価指標については詳しく述べる．1.3 節では，本研究の着眼点と目的について述べる．

1.2 音声品質に関する既存の研究事例

本節では音質を,「受聴品質」,「ラウドネス」,「明瞭度」,「了解度」に分け[7]，それぞれについて概説する．また，電話網・放送網の国際的接続性から国際標準が多数作られており，その経緯と概要も述べる．次に，音声の物理的劣化を評価する品質尺度を概説する．最後に，本論文で検討する明瞭度と了解度の予測と推定[8] に関する技術標準とその系譜について述べる．特に本論文で扱う「了解度」に関しては主観評価法とその予測・推定法に関して詳しく取り上げる．

1.2.1 受聴品質

音声の通信として最も初期に普及し，現代でも支配的なシステムは電話網である．電話網の基幹技術は，アナログ通信網，ディジタル通信網，IP（Internet Protocol）通信網と変遷しても，音声情報を「正しく送話し，正しく受聴する」という目的は変わらない．本項では現在用いられている，ディジタル通信網，IP 通信網での電話受聴品質について述べる．特に国際連合の専門機関の一つである国際電気通信連合（International Telecommunication Union，以下 ITU）では，国際電話通信の接続は,「最も品質の悪いところが全体の品質を決めてしまう」という観点から，様々な技術標準を勧告している．本項の前半では電話の受聴品質についての技術勧告を中心に述べる．

次に，オーディオ信号品質ついて述べる．オーディオ品質はその再生対象によって大きく異なる．放送網を前提とすれば，伝送情報量と受聴品質のトレードオフであるし，ポータブルオーディオであれば，保存媒体の容量と受聴品質のトレードオフとなる．この他に，CD やマルチチャネル再生システムを前提とした高品質再生を前提としたシステムの評価も考えなければならない．この様にオーディオの受聴品質は考慮するべき事項が多岐にわたるため，ITU において標準化された方式とその応用例について本項の中盤で述べる．

最後に，受聴品質と関連の深い総合品質モデルについて述べる．

電話網の受聴品質主観評価

電話の受聴品質は，様々な計測法があったものの，現在ではオピニオン評価が用いられる．オピニオン評価を行う手法として，ITU–T Rec. P.800 で定義される「受聴音声の自然さ」を絶対範疇尺度法 (Absolute Category Rating : ACR) で求める [6]．P.800 による主観評価値は MOS（Mean Opinion Score）と呼ばれ，その評価カテゴリは Table 1.1 となる．また，雑音抑圧を用いた音声強調信号に対する主観評価には ITU–T Rec. P.835[7] があり，抑圧後の音声部品質，雑音部品質，総合品質の 3 品質を MOS で求める．また，MOS の評価は実験の枠組みの影響を受けやすいため，ITU–T Rec. P.810 で勧告される基準信号である MNRU（Modulated Noise Reference Unit）[8]

[7]本論文で注目するのは音声品質であるため，これ以外の音楽品質等で良く用いられる「音高感」等の音色解析については割愛する．音色解析については文献 [5] 等に詳しい．

[8]本論文では，主観評価値と 1 対 1 で対応する主観値の推定値を求めることを「推定」，主観値を推定する目安となるインデックス値を求めることを「予測」と呼び区別する．

信号を用い，MNRU 信号の結果から ITU–T Rec. P.830 で勧告される等価 Q 値に変換して他の実験系と比較することが推奨される．

Table 1.1: Mean opinion score(MOS)

MOS	Quality(Japanese)	Impairment(Japanese)	Quality(English)	Impairment(English)
5	非常に良い	わからない	Excellent	Imperceptible
4	良い	わかるが気にならない	Good	Perceptible but not annoying
3	普通	やや気になる	Fair	Slightly annoying
2	悪い	気になる	Poor	Annoying
1	非常に悪い	非常に気になる	Bad	Very annoying

受聴品質の推定

音質評価値の推定アプローチは大きく分けて「物理指標と心理指標を統計的手法で結ぶ推定（以下，統計的推定)」と,「聴覚モデルを用いる推定」に分けられる．また，最近ではこれらの「ハイブリッド型」もみられる．この二つのアプローチは，これまでにディジタル電話網・IP 電話網で伝送された音信号のメディア層による音声コーデック品質の推定のために標準化され，広く用いられている．

MOS の統計的推定

MOS の統計的推定に，伊藤らによる 8 種[9]の客観評価尺度（物理指標）を用いた MOS と音声の明瞭さの主観評価値から求める明瞭度等価減衰量 AEN（Affaiblissement Equivalent pour la Nettete (equivalent ariculation loss) : 仏語）をそれぞれ回帰分析し，自然さと明瞭度の推定がある [9, 10]. これは対数圧伸 PCM（Pulse Code Modulation）や ADPCM（Adaptive Differential PCM）といった波形符号化に対して品質推定性能が高い．同様の手法は雑音抑圧音声の品質評価[10][11, 12] にも用いられている．

PESQ（聴覚モデルによる MOS 推定）

もう一方の聴覚モデルを用いる推定は，伊藤らの検討の後に発展した符号励振線形予測（Code Excited Linear Prediction : CELP）[13] による符号化音声の受聴品質推定のために発展した．ITU では，ITU–T Rec. P.86X シリーズで，電話受聴品質の客観評価法を勧告している．最初に勧告された ITU-T Rec. P.861[14] は，PSQM（Perceptual Speech Quality Measure）と呼ばれ，リファレンス音声との符号化音声の差をバークスペクトルのラウドネス差で表現し，差の大きさから品質を推定している．PSQM はその後，時間方向の整合性を取れるように改良され，主観品質値への変換関数や，広帯域音声[11]にも対応した ITU–T Rec. P.862 の PESQ（Perceptual Evaluation of Speech Quality）シリーズに置き換えられた [15, 16, 17, 18]．PESQ による音質推定のモデル

[9]SNR，セグメンタル SNR（SNRseg），スペクトルひずみ尺度，重み付スペクトルひずみ尺度，LPC ケプストラム距離尺度，スペクトル包絡の尤度比，COSH 距離，重み付尤度比の 8 種．一部は 1.2.5 で述べる．

[10]ハンズフリー電話の開発を目的とすれば受聴品質の一種とみなせる．

[11]7 kHz 帯域

を Fig. 1.1 に示す．まず，入力したリファレンス音声と符号化音声は経路による時間ずれを補正する．次に，リファレンス音声と符号化音声それぞれのバークスペクトルラウドネスを求め，その差を式（1.1）で求め値を出す（ここまでを P.862 で規定）．式中のかっこ内がひずみ量であり，D は減算性ひずみ，A は加算性のひずみである．ひずみ値に乗算する重みが減算性ひずみの方が 0.1 と加算性ひずみの 0.0309 より大きいことから，受聴品質への影響は減算性のひずみの方が大きいことがわかる．この PESQ score は，4.5 から −0.5 であり，推定したい P.800 の MOS の結果である 5 から 1 と対応が取れていない．このためシグモイド型変換関数が P.862.1[16] で規定されている．変換後の値を MOS–LQO（MOS–Listening Quality Objective）といい，P.800 の MOS の推定値となる．PESQ は話者に依存する品質差を正確に予想することができないため，最低でも評価音声は男女 2 名の発話を用いなければならないことが P.862.3 で勧告されている．

$$\text{PESQ score} = 4.5 - (0.1 \times D + 0.0309 \times A) \tag{1.1}$$

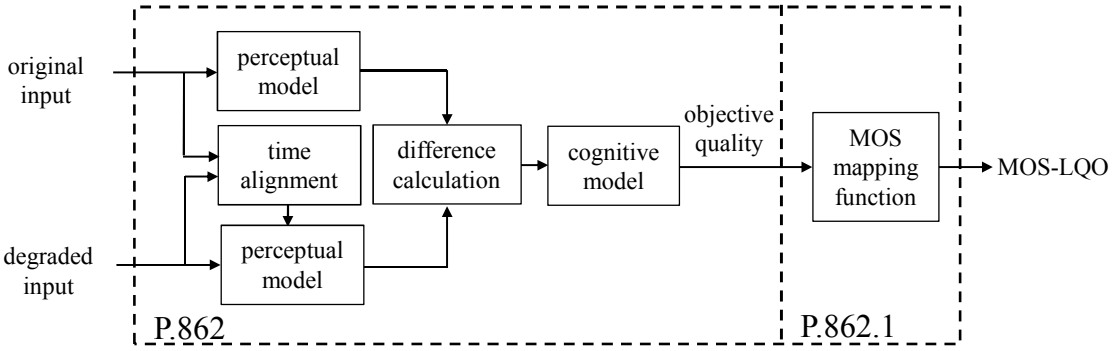


Fig. 1.1: Simplified diagram of the PESQ algorithm

PESQ は受聴品質と相関が高いと考えられる音質評価にも使われ，移動用音源の追尾 [19] や音声認識性能の予測 [20]，音声了解度の推定 [21, 22]，雑音抑圧音声の了解度推定 [23] と言った応用例がある．この様に ITU 勧告に従った使い方ではないものの，電話網以外の客観音質評価としてもデファクトスタンダードとなりつつあるといえる．

POLQA

2011 年には，PESQ の発展・後継規格として，スーパーワイドバンド[12]に対応した ITU–T Rec. P.863 の POLQA（Perceptual Objective Listening Quality Assessment）が勧告され，今後は PESQ に置き換わるものと考えられる [24]．これらの標準化規格により，電話帯域での MOS[6] については，高精度な推定が可能になった．

[12] 14 kHz 帯域

ノンリファレンス型受聴品質評価法

これら受聴品質評価は，劣化前の原音との差を用いるフルリファレンス型の評価方式であるが，評価したい環境において原音が必ずしも手に入るとは限らないことから，原音を用いないノンリファレンス型の評価法も検討されている．電話網向けには，ITU–T Rec. P.563 の 3SQM（Single Sided Speech Quality Measure）がある [25]．3SQM は，伝送系による劣化を含む音声から音声区間検出（Voice Activity Detector : VAD）を用いて音声区間を推定し，雑音性，瞬断，不自然性といった 12 の音響特徴量を求め，回帰式により推定 MOS を出力する．

オーディオ受聴品質の主観評価

前述したように，オーディオ品質の主観評価法は着目点によって異なるため，主観評価法のガイドを ITU–R Rec. BS.1283 で勧告している [26]．多チャンネル音声システムと言ったほとんど劣化の無い高臨場感システムには ITU–R Rec. BS.1116–1[27] を用いる．これは隠れ基準つき三刺激二重盲検法と呼ばれ，基準となる刺激は既知で何度でも再生でき，提示される 2 つの刺激の基準に対する回答を 5 段階での評価語を参照に 0.1 刻みで回答する．回答のスコアは SDG（Subjective Difference Grade）と呼ばれる．

Table 1.2: Subjective Difference Grade(SDG)

Grade	Impairment(Japanese)	Impairment(English)
5	劣化がわからない	Imperceptible
4	劣化がわかるが気にならない	Perceptible, but not annoying
3	劣化が気になるが邪魔にはならない	Slightly annoying
2	劣化が邪魔になる	Annoying
1	劣化が非常に邪魔になる	Very annoying

この他に広く用いられる主観評価法には ITU–R Rec. BS.1284–1 [28] がある．これは基本的には BS.1116–1 と同様の手法だが，単一刺激や一対刺激といった刺激提示について選択肢があり，回答も 5 段階ないし 7 段階の回答で良い．このため主観評価結果の分散が大きくなるため，予備実験の手法も ITU–R Rec. BS.1285 で規定されている [29]．また，映像を伴う場合の評価については ITU–R Rec. BS.1287 で勧告されている [30]. また，中間品質（低ビットレートのオーディオ符号化等）評価用に ITU–R Rec. BS.1534–1 で勧告される MUSHRA 法（MUltiple Stimuli with Hidden Reference and Anchor）[31] がある．これは，基準信号は何度でも再生でき，評価する項目内に必ず含まれる隠れ基準と隠れアンカー[13]が含まれる複数の刺激を 100 点満点で回答する．被験者は少なくとも一つは基準と同等である刺激に 100 点をつけなければならない．以上の様にオーディオ信号の主観評価は，リファレンスに用いる原音自体の録音品質が影響しないように工夫されている．

オーディオ信号の受聴品質評価法

オーディオ用音楽符号化品質の推定法に ITU–R Rec. BS.1387–1 の PEAQ（Perceptual Evaluation of Audio Quality）がある [32]．これは，5 または 11 種類の聴覚モデル出力値と主観評価と

[13]遮断周波数 3.5 kHz のローパスフィルタを書けた信号を含む明らかに劣化している刺激.

のニューラルネットワークによる回帰出力により客観音質値 ODG（Objective Difference Grade）を得る，ODG は前述したオーディオ信号用の主観評価法である ITU–R Rec. BS.1116–1[27] で評価される SDG の推定値である．PEAQ は 3SQM と同様に統計的推定と聴覚モデルを用いる推定のハイブリットな方式である．

PEAQ は本来，MP3（MPEG Audio Layer 3）[33] や AAC（Adaptive Audio Coding）[34, 35] 等のオーディオコーデックで符号化した楽曲評価用であったたため，それ以外の劣化要因に対しては，SDG との相関が悪くなる傾向にあった．この点を改良するため，Huber らによる PEMO–Q（PErception MOdel based Quality estimation）があり [36]，今後の標準化が期待される．

総合品質

受聴品質については，音声／オーディオ信号を問わず標準化され，音響機器の設計開発に広く用いられている．受聴品質は主として，音声／オーディオ信号の符号化によって原音がどの程度自然なまま再生されるかの確認である．音声通信，特に電話網では，双方向通信であることから，会話の双方向性をさらに加味した総合品質の面からの議論がある．総合品質の評価モデルに関して中山，三浦らは Fig. 1.2 に示す要素感覚と総合情緒機能の 2 段階で構成されるモデルを提案している [37, 38, 39, 40]．これは，提示した音刺激 S は人の聴感覚にとらえられ，人間聴覚として普遍的であり，これ以上分割不可能ないくつかの要素感覚（受聴品質の他に音量感や音高感など）としてとらえ，要素感覚の結合としての総合感覚が与えられる．そして，個人や時代環境に応じた欲求を重みとして感情へ変換され，刺激へ対する最終的な反応としての音質を求めるモデルである．このモデルは，分散の大きい総合感覚の直接計測をせずに，安定して評価可能な要素感覚群からの推定となる．

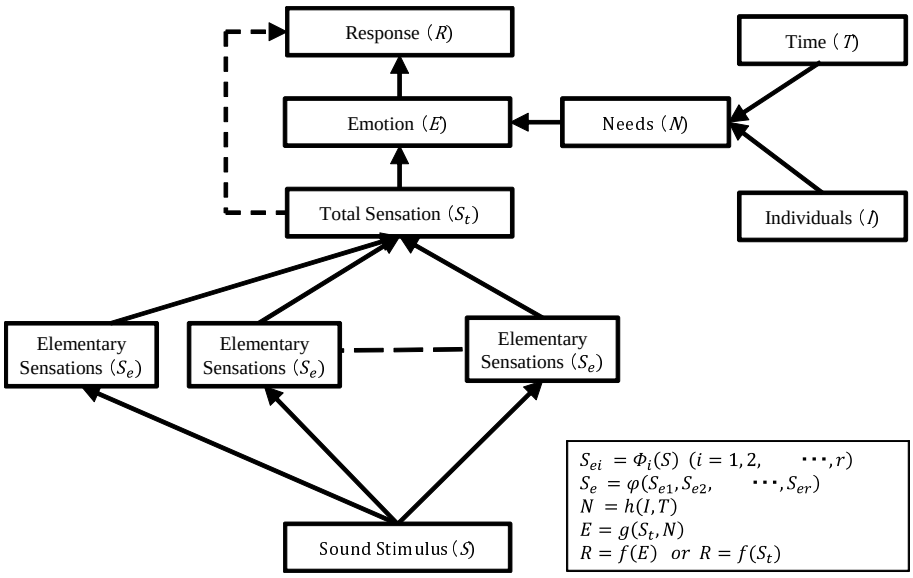


Fig. 1.2: A model of the process of evaluating the quality of reproduced sound

総合品質の主観評価は，被験者側から見れば受聴品質と等価である．放送用途においては，双方向性よりも一方向での配信であるため，受聴品質と総合品質が一致する．電話網においても，双

方向で経路による差が無ければ，受聴機器・受聴環境の違いを除いた劣化は同量であるため，受聴品質と総合品質の主観評価法は区別されない．むしろ，受聴品質の詳細分析のためにFig. 1.2の様な要素分解を用いることが多い．PEAQや3SQMは，このような観点に立って受聴品質のモデルを作成し，客観推定に用いている．本論文では，Fig. 1.2の様な要素分解による音質のモデルを総合音質モデルとし，その主観評価は受聴品質を想定する．

E–model

総合品質モデルを用いた受聴品質の推定は，IP電話網における双方向会話の総合品質推定を求めるE–modelとして，1998年にITU–T Rec. G.107として勧告され，現在も適宜更新されている[41]．E–modelは，音声信号の客観評価ではなく，電話伝送系における，端末要因・環境要因・ネットワーク要因等による20の入力パラメータを用いて，出力であるR値を求める関数を作成する．R値は式（1.2）で求める．ここで，R_oは雑音感，I_sは音量感，I_dはエコー／遅延，I_eは歪／途切れ間，Aはモバイル通信などの利便性調整項である．これは，標準系の基準となるSNRであるR_oから，各要因による品質劣化を心理尺度上で引いた値に利便性調整項を加算して総合品質を求めている．E–modelは，4種の視点からの聴覚モデルによる出力値を用いた品質推定である．

$$R = R_o - I_s - I_d - Ie + A \tag{1.2}$$

1.2.2 ラウドネス

等ラウドネス曲線

総合品質の要素感覚ごとの客観評価については，ラウドネス（音量感）や明瞭度についての評価法がいくつか確立されている．ラウドネス算出の基準となる，周波数ごとのラウドネスをプロットした等ラウドネスレベル曲線は，1930年代のFletcher and Munsonの曲線[42]があり，騒音レベルのA特性等に用いられてきた．こののち，低音域の特性が自遊空間と電話器による受聴とで一致しないことなどから，1950年代にRobinson and Dadsonの曲線[43]が発表され，ISO 226として標準化されている．しかしながら，1985年にRobinson and Dadsonの曲線では，なお1 kHz以下に誤差があることが指摘され，全面改訂され，現在はISO 226（2003）が用いられている[44]．騒音レベルに用いるA特性等もこの結果を基に書き換えられた[45]．これらの曲線は，厳密な条件のもとに行われた主観評価の結果をプロットしたものであり，統計的推定によって心理的な音量感を求めたものである．この曲線を用いて求めるラウドネスレベルは加算性が成り立ち，実用上の利点が大きい．

電話網での利用

電話の音量感については，1960年にCCITT（Comite Consultatif International Telegraphique et Telephonique，ITUの前身）による通話当量（RE : Reference Equivalent）を用いた基準通話系と非測定系の減衰量の差の評価がある．1976年にはCCITTにより，REの後継規格であるラウドネス定格（LR : Loudness Rating）が制定された．LRは基準系，中間基準系（IRS : Inter–mediate Reference System），被測定系により測定する．LRは一定音量の基準系と同一音量になるように

IRS の減衰量を調節し，さらにその基準系と被測定系が同一の音量になるように被測定系の減衰量を調節したときの，IRS と被測定系の減衰量の差によって示される．LR の評価には当初は主観評価が用いられていたものの，1984 年に客観評価装置の規格が完成し，現在では ITU–T Rec. P.76 として勧告されている [46]．LR は比較的簡単なモデルながら，聴覚モデルを用いる推定である．

放送での利用

電話音量以外の音響信号から求めるラウドネスの規格に ISO 532B がある [47]．これは，定常状態のノイズ，または経時変化しない音信号のラウドネス（定常ラウドネス）をチャートによる周波数ごとの値からラウドネスを求めている．放送やオーディオ機器で用いる場合には，定常ラウドネスでは不十分であるため，ITU において，等価騒音レベルの考え方を応用したラウドネス推定アルゴリズムである ITU–R Rec. BS.1770–3[48] および BS.1771[49] が制定され，その放送用途の運用基準は BS.1864[50] となっている.BS.1770–3 および BS.1771 は，主観評価値との相関が高いことが確認されており，一部にゲーティングと言った聴覚モデルを用いる推定の要素があるものの，ほぼ統計的推定である．これは，定常ラウドネスよりも複雑な要因で心理的なラウドネスが定まることによる．

1.2.3 明瞭度

電話網の評価での利用

言語を用いた聴覚検査については歴史が古く，1761 年に Erunaud,1805 年の Itard，1893 年の Urbantschitsch らによる難聴児の訓練の事例があるものの，標準化された手法ではなく実験条件差が非常に大きかったとされる [51]．近代的な明瞭度評価の主観評価については 1929 年からの Fletcher らによる明瞭度試験法（Articulation Testing Method）の検討によって構築された [52].

本邦での明瞭度試験については，1938 年に落合による検討 [53] が先鞭をつけ，戦後の電話網の再構築等の要求により，1957 年に日本音響学会の明瞭度研究会でまとめられた「明瞭度試験法の基準」[54] で標準化された．その後，1966 年に単音節（日本語 195 音を 100 音節で評価），2 連音節，3 連音節の試験用音源テープと試験法が作成され，1982 年に更新された [55].

電話網の明瞭度については世界中で詳細に評価され，1950 年代に CCITT で明瞭度等価減衰量 AEN が勧告された．これは，標準系と測定系の通信線路減衰量ごとに単音明瞭度を測定し，単音明瞭度 80%に対応する標準系と測定系の線路減衰量の差を求める手法である．しかし，特にディジタル電話網の構築による通信品質の向上にから，線路減衰による明瞭度低下が低くなったこと，Fig. 1.2 の総合品質を求める要素感覚としては，明瞭度よりもラウドネスの方が相関が高いことなどを理由として，CCITT による LR の勧告時に AEN に関する内容は参考記載となり，1984 年に削除された．

聴力検査での利用

電話網の品質評価から明瞭度評価は削除されたものの，他分野における品質評価でも明瞭度評価は使われる．Fletcher らの明瞭度試験法 [52, 56] の電話伝送損失が無い系を仮定すると聴力検査

に用いることが可能であるとされ，第二次世界大戦における聴覚障碍者のリハビリに用いる聴力検査法として臨床での利用が検討された．本邦では，難聴研究会（現在の日本聴覚医学会）において「聴力測定法の基準 1956」[57] として研究成果がまとめられ，翌年に「57A 語表，57B 語表」として採用された．

聴力検査での明瞭度検査は，特に補聴器の個人ごとの適合（フィッティング）で広く利用されている．日本聴覚医学会が定めている「補聴器適合検査の指針 2010」[58] では，語音明瞭度曲線または語音明瞭度の測定が必須評価項目であり，雑音を付加したときの語音明瞭度の測定が参考項目として定められている．語音明瞭度試験は臨床で用いることから，なるべく簡便な試験法である必要があるため，100 音節を再編し出現頻度に応じて 50 音節とした S57 語表，とより簡略化して 20 音節とした S67 語表を用いている．補聴であるため，被験者は何らかの聴覚障碍を持ち，障碍を取り除くために補聴器を適合する．つまり，電話網評価の様に単一のシステムに多くの人が合わせるのではなく，特定の被験者に補聴システムを合わせるという観点での主観品質評価であり，統計的な標準値を求める客観評価はあまりなじまない．

その他の利用

上記二つの事例は，人が発話した音声（以下，自然音声）のシステムにおける劣化や聴取者の聴力変動による明瞭度の評価であった．この他に，合成音声や雑音抑圧音声に関する研究開発でも明瞭度試験を利用するようになった．合成音声／雑音抑圧音声は自然音声と大きく異なり，不自然な（機械的な）発話や歪をもつ．自然さに着目した音声の品質劣化を評価する受聴品質では，そもそも原音に音響的・言語的劣化のある合成音声の品質は計測できない．よって，合成音声それ自体の質を求める必要があり，当初は明瞭度評価が用いられた [59]．しかし，初期の合成音声は人間の様な調音結合[14]が十分に行えない問題があったため，調音結合を加味した全ての音節を評価する必要があった．日本語の場合は，母音と子音の数は英語その他外国語よりも圧倒的に少なく語頭・語尾の調音結合を加味しても総音節数は 195 であり，明瞭度試験の規模はそれほど多くならない．一方で，外国語では，子音のクラスタ化，語中や語尾の調音結合でのみ出現する子音なども評価しなければならなく，評価音数が現実的ではない [60]．よって，単語を用いた了解度試験が検討されるようになった．

1.2.4 了解度

英語圏における了解度試験法

音節数の問題から，明瞭度試験から了解度試験に置き換える検討は欧米圏で特に盛んになった．単語を用いた試験は Fletcher らも行っていたが，1944 年に Egan によって，音素バランス（Phonetic nalamce）リスト [61] が提唱され，広く使われている．1958 年に Fairbanks は CVC 型の Rhyme test を考案した [62]．その後，1965 年に House らによって修正され，CVC 型単語 6 種類の中から 1 種類を選択する MRT（Modified Rhyme Test）に改良された [63]．さらに，Voiers によって DRT（Diagnostic Rhyme Test）[64, 65] が考案された．DRT はその後，ANSI（American National Standards Institute）において規格化され，現在でも更新されている [66]．DRT で用いる単語を

[14]連続的に音声を発するときの声道形状の変化で，音響的特徴を変化させる

Table 1.3 に示す．表からわかるように第 1 子音のみ異なる単語対を 6 種類の子音特徴ごとに 16 単語対用意しており，総単語対 96 対で総数 192 単語の単語リストである．対になる音素は Jacobson, Fant，Halle による JFH 音表に従っている [67]．表の行方向は後続母音を統一しており，後続母音ごとの分析にも用いることができる．DRT による了解度は，式（1.3）で求める．$N_{correct}$ は正当数，$N_{incorrect}$ は誤答数，N_{tests} は実験単語総数を示す．この式は二択による正答率バイアス値である正答率 50%への漸近を補正する．

$$\text{DRT intelligibility} = \frac{N_{correct} - N_{incorrect}}{N_{tests}} \times 100[\%] \tag{1.3}$$

Table 1.3: Word pairs used in diagnostic rhyme test(DRT)

Voicing		Nasality		Sustention		Sibilation		Graveness		Compactness	
veal	feel	meat	beat	vee	bee	zee	thee	weed	reed	yield	wield
bean	peen	need	deed	sheet	cheat	cheep	keep	peak	teak	key	tea
gin	chin	mitt	bit	vill	bill	jilt	gilt	bid	did	hit	fit
dint	tint	nip	dip	thick	tick	sing	thing	fin	thin	gill	dill
zoo	sue	moot	boot	foo	pooh	juice	goose	moon	noon	coop	poop
dune	tune	news	dues	shoes	choose	chew	coo	pool	tool	you	rue
vole	foal	moan	bone	those	doze	joe	go	bowl	dole	ghost	boast
goat	coat	note	dote	though	dough	sole	thole	fore	thor	show	so
zed	said	mend	bend	then	den	jest	guest	met	net	keg	peg
dense	tense	neck	deck	fence	pence	chair	care	pent	tent	yen	wren
vast	fast	mad	bad	than	dan	jab	gab	bank	dank	gat	bat
gaff	calf	nab	dab	shad	chad	sank	thank	fad	thad	shag	sag
vault	fault	moss	boss	thong	tong	jaws	gauze	fought	thought	yawl	wall
daunt	taunt	gnaw	daw	shaw	chaw	saw	thaw	bong	dong	caught	thought
jock	chock	mom	bomb	von	bon	jot	got	wad	rod	hop	fop
bond	pond	knock	dock	vox	box	chop	cop	pot	tot	got	dot

これらは単語了解度試験であり，より自然な会話コミュニケーションの評価のために文章リストを用いた文章了解度試験も検討された．特に 1970 年代に老人性難聴評価に用いるための手法がいくつか開発され，M. Oldman による 8 種の短文の意味の伝わり方の 9 段階評価 [68]．D.N. Kalikow による文章リストの検討 [69]，R. Plompga による 13 文のリストによる徴取実験の例がある．これらは，文章を用いた聴力検査であるものの，難聴者の傾向分析向けであり，文章リスト内の難易度や評価の訓練効果についてはあまり言及されなかった．これらとは異なる文章を用いた了解度試験に，Nilssonm による HINT（Hearing In Noise Test）[70] があり，これは自立語 4–6 文節の文章の自立語正答率が 50%となる SNR 値である SRT（Speech Recognition Threshold）を求める．これは文章を用いた評価であるが，自立語正答率を使うこと，SRT を求めることが目的であることから，MRT，DRT と異なり直接了解度を求めてはいない．この他にも，特に人口内耳の開発向けに特定の聴能力を検査するための単語を用いた試験法がいくつかあり，了解度試験に分類される [71].

建築音響分野での了解度

本邦での了解度試験法の確立は，明瞭度試験の簡便さもあり，欧米諸国と比べ遅れていた．建築音響分野では，1953 年に幸田，久我による明瞭度及び了解度の評価が始まった [72]．その後,「明瞭度試験法の基準」[54] が飯田によるの問題提起 [73] があったものの長く使われてきた．了解度に関しては 1984 年の戸井田による野外拡声装置によるエコーとノイズが文章了解度に与える影響の検討 [74] など少数の例しかない．単語リストの構築については，小川による無意味三連音節の検討 [75] や佐藤らによる文章了解度音表 [76] があったものの，試験法としての標準化はみられなかった．

臨床における聴力検査での了解度

続いて，本邦における臨床での聴力検査等に用いる了解度試験について述べる．まず，1989 年に田中らによる補聴器フィッティング用の単語リスト TY–89 の作成がある [77, 78]. これは 2 音節や 3 音節の単語リストや日常生活文，不自然な文などの音源が入っていたものの，単語リスト内の難易度統一（後述する親密度など）を行っていない問題がある．この他に，単音節，成人用単語及び日常会話，幼児用の 2 音節または 3 音節単語と 2 語または 3 語文のリスト，幼児用リストなどが入っているエスコアール社の CI–2004[79] があが，どちらも S57 と S67 を置き換えるところまでは至っていない．さらに，HINT の日本語ローカライズ版として井脇らによる Japanese–HINT がある [80, 81, 82, 83]．これは文章リストの文内の音素バランスは取れており，了解度と英語版 HINT の他，ローカライズした国の評価結果の SRT を直接比較できる利点がある [84] ものの，評価環境（評価室）差の影響がみられることから，これも標準化されていない．以上の様に，補聴器フィッティング・聴力検査分野では，了解度試験が明瞭度試験を置き変えるに至っていないのが現状である．

親密度別単語了解度試験

了解度試験の標準化には音韻論的な音素バランスだけでなく，言語的な難易度の検討が必要との指摘がされている．難易度の統制に用いられる尺度に単語親密度 [85, 86, 87] がある．単語親密度は主観評価による単語のなじみの評定値である．単語親密度は，新明解国語辞典第四版 [88] に収録されている自立語約 80000 語について文字または音声提示による 7 段階の評定値であり，雑音環境下での認知率が異なることが確認されている [89]．このため，了解度試験に用いる単語リストは表内の親密語がある程度統制されていなければ，表内単語の統計解析に正確さを欠くこととなり実用上の問題が出る．

単語親密度に配慮した了解度試験単語リストに，坂本らによる「親密度別単語了解度試験（Familiarity controlled Word lists 2003 : FW03）」[90] がある．これは日本語の語彙特性 [85] で分析された日本語単語のうち，4 モーラの単語を単語親密度を高親密度から低親密度まで 4 分し，親密度グループごとに 1 リストあたり 50 単語の了解度試験用単語リストとして作成された．また，単語認知に用いているモーラの同定 [91]，詳細な SRT の検証 [92]，加齢による聴力低下の影響評価

[93]，臨床向けに 20 単語に削減した FW07[94]，GUI 化へ向けたクローズドセット[15]の検討 [95]，及び心的辞書 [96] の概念を用いた推定 [97] と言った詳細分析がされている．また，様々な音声システムの品質評価に利用されている [12, 98, 99, 100, 101]．さらに，FW03 の単語を用いた了解度試験と併用して，空間伝播するの音声伝送性能を評価する森本らの「聴き取りにくさ」心理指標にも用いられている [102]．FW03 の単語リストのうち高親密度（7.0˜5.5）の例を Table 1.4 に示す.

Table 1.4: Word lists examples included in the familiarity–controlled word lists 2003(FW03)

アマグモ	イマフウ	ウチガワ	オシダシ	オヤモト
ガニマタ	キタカゼ	キュウショク	グウタラ	ケイサツ
ゲンイン	コウフク	ザイガク	サイジツ	ジツブツ
シハライ	シャブシャブ	スタミナ	セツリツ	ソラミミ
タナバタ	ダンタイ	チョウハツ	チンタイ	ツナガリ
デマカセ	ドクヤク	トビバコ	ナツバショ	ニンニク
ネアガリ	ハダイロ	パチンコ	バランス	ヒキダシ
ブランド	フリガナ	ホウタイ	マンルイ	ヤマカジ
ユウワク	ヨクネン	ランパク	リクジョウ	リャクダツ
レンパイ	ワタクシ	カシパン	クスリヤ	ミジンコ

二者択一式日本語音声了解度試験

FW03 はリスト内の難易度の統制は取れているものの，その設計思想は補聴器フィッティングを前提とした了解度試験である．このため，音声システム設計時の設計開発期間における少人数の被験者による繰り返し評価と言った主観評価に慣れることが想定される環境では，慣れによる評価値の向上がみられる [103][16]．このため，十分にトレーニングを積んだスタッフを用意するか，トレーニングの影響が有意ではない評価法が必要になる．前者は電電公社時代の明瞭度試験クルーがこれにあたる．後者の観点から，わずかな練習ですむ了解度試験法として，Voiers による DRT[64, 65] を日本語にローカライズした近藤らによる二者択一式日本語音声了解度試験（以下，JDRT）がある [21, 104]．JDRT は単語親密度を 6.0 以上に統制し，トレーニングの影響が小さいことがわかっている [105]．また，英語版と同じ 96 単語対 192 単語セットを用いた評価から，日本語の音韻特徴を加味した 60 単語対 120 単語で十分なことを確認している [106]. 考慮した日本語の音韻特徴表を Table 1.5 に示す．表中で “+” は特徴が有意な子音，“–” は特徴が無い子音，“0” は特徴がどちらでもないものを示す．表中の特徴から 1 特徴のみ異なる子音のペアを分析する子音特徴とする．JDRT の子音特徴を以下に示す.

Voicing　JFH 分類では vocalic – nonvocalic に相当する，有声音と無声音の分類．有声音は声帯の振動を伴う音，無声音は声帯の振動を伴わない音である．比較的明快な分類である．

[15]了解度試験において，聴取した単語を自由回答させる方式をオープンセット，選択肢の中から回答させる方式をクローズドセットという．クローズドセットの場合，提示する単語数に応じて正答率のバイアス（偶発的正当）を補正しなければならない.

[16]ただし，文献 [103] は話速変換処理の場合の検討である

Nasality　JFH 分類では nasal – oral に相当する，鼻音と口音の分類．鼻音のスペクトルは，口音のスペクトルよりも高いフォルマント密度を示す．つまり鼻音の方はフォルマントがはっきりと現れているが，口音はそれほど鮮明ではない．

Sustention　JFH 分類では continuant – interrupted である．連続性子音 (狭窄音) と中断性子音 (閉鎖音) の分類．狭窄音の音の始まりは緩やかであるが，閉鎖音の音の始まりは急であり，鋭い波頭がある．

Sibilation　JFH 分類では strident – mellow に相当する，粗擦音と円熟音の分類．粗擦音は波形が不規則であり明確なフォルマント領域が認められない．また，規則的な波形と対立する．円熟音はスペクトログラム上で，水平または垂直な縞を形成することがある．

Graveness　JFH 分類では grave – acute であり，抑音と鋭音の分類[17]．もしくは低音調性と高音調性の分類である．抑音はスペクトル上のエネルギーが低周波に集中するが，鋭音は高周波に集中する．

Compactness　JFH 分類では compact – diffuse に相当する集約性と拡散性の分類．集約性はスペクトル上のエネルギーが一つのフォルマント周波数に集中するが，拡散性は分散する．

Average　上記 6 特徴の平均．主として他のコーパスを用いた了解度試験結果との比較，音素特徴に依存しない実験要因の分析の際に用いる．これらの子音特徴を考慮した JDRT の 120 単語セットのリストを Table 1.6 に示す．Table 1.3 と同様に列方向に語頭子音特徴，列方向に後続母音で揃えてあり，日本語の発音特性に合わせて解析ができる．表中の単語は固有名詞を避け，アクセント型が統一された単語で構成されている．一方で，英語版と同様に語頭子音の影響しか見ていないため，語中，語尾の調音結合を加味した分析は原理上できない．

Table 1.5: The Japanese consonent taxonomy

Feature	m	n	z	ǰ	b	d	g	w	r	j	Φ	s	š	č	p	t	k	h	N	t͡s	ç
Voicing	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	+	−	−
Nasality	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
Sustention	−	−	+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	−	−	+
Sibilation	−	−	+	+	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	+	−
Graveness	+	−	−	+	+	−	0	+	−	0	+	−	0	0	+	−	0	0	0	−	0
Compactness	−	−	−	+	−	−	+	−	−	+	−	−	+	+	−	−	+	+	−	−	+
Vowel-like	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−

JDRT も様々な音声アプリケーションの品質評価にも利用されている [21, 107, 108, 109, 110, 111, 112]．JDRT はクローズドセットであり，同一人物での条件を変えた品質評価に対して比較しやすいメリットがある．英語版と同様に了解度は式（1.3）で求める．

[17]第 2 フォルマントが比較的低い音を抑音，高い音を鋭音という．

Table 1.6: Word pairs used in Japanese diagnostic rhyme test(JDRT)

Voicing	Nasality	Sustention	Shibilation	Graveness	Compactness
財–才	万–番	箸–菓子	ジャム–ガム	枠–楽	焼く–沸く
zai–sai	man–ban	hashi–kashi	jam–gam	waku–raku	yaku–waku
抱く-炊く	無い–台	旗–型	着–角	パイ–タイ	貝–パイ
taku–daku	nai–dai	hata–kata	chaku–kaku	pai–tai	kai–pai
議事–記事	ミス–ビス	私利–地理	式–引き	見栄–似え	銀–瓶
giji–kiji	nusy–bisu	shiri–chiri	shiki–hiki	mie–nie	gin–bin
銀–金	見る–ビル	昼–着る	知事–記事	ミス–ニス	キザ–ピザ
gin–kin	miru–biru	hiru–kiru	chiji–kiji	misu–nisu	kiza-piza
髄–粋	無理–鰤	好き–月	中–空	剥く–抜く	黒–プロ
zui–sui	muri–buri	suki–tsuki	chuu–kuu	muku–mili	kuro–puro
寓–食う	無視–武士	砂–綱	純–群	無視–主	タ–ルー
guu–kuu	mushi–bushi	suna–tsuna	jun–gun	mushi–nushi	yuu–ruu
税–生	面–弁	変–剣	シェア–ヘア	面–年	弦–弁
zei–sei	men–ben	hen–ken	shea–hea	men–nen	gen–ben
出刃–手羽	練る–出る	縁–蹴り	シェル–経る	ペン–天	剣–ペン
deba–teba	neru–deru	heri–keri	sheru–heru	pen–ten	ken–pen
象–僧	門–凡	星–蹴り	条–号	毛–脳	語気–簿記
zou–sou	mon–bon	hoshi–keri	jou–gou	mou–nou	goki–boki
誤字–孤児	野良–銅鑼	掘る–凝る	所持–保持	ポロ–トロ	余暇–ろ過
goji–koji	nora–dora	horu–koru	shoji–hoji	poro–toro	yoka–ロカ

1.2.5 音声信号品質

明瞭度と了解度の予測・推定を説明する前に，一般に音声信号の品質尺度として用いられる方式を時間領域，周波数領域，両方を加味した方式に分けて概説する．時間領域からは信号対雑音比 SNR と，短時間のセグメントに区切ったセグメンタル SNR について述べる．音声の周波数領域の品質尺度は，主として音声符号化・音声合成品質の推定に長い間用いられてきた [9, 11]．符号化品質には 1.2.1 項で述べた PESQ に置き換わっているが，雑音抑圧音声品質評価には現在でも使用されている[18]．ここでは周波数領域尺度の代表例として文献 [11] と文献 [114] で解説されている方式から 5 種類の方式について述べる．最後にこれらのハイブリッドとして，周波数重み付セグメンタル SNR について述べる．

時間領域信号品質尺度

SNR SNR は通信工学で広く用いられる音信号を含む信号量と雑音量の比のことで，明瞭度・了解度に限らず通信品質の物理指標である．音声信号処理分野，音声通信分野では信号対雑音比（Signal to Noise Ratio）を SNR[19] と呼び，聴力検査分野，建築音響分野では音声対雑音比を SNR（Speech to Noise Ratio）と呼ぶことが多い．この二つは主信号を主音声とする限りにおいては等価である．本論文では，音声信号の品質を主対象とするため，信号対雑音比を SNR と呼ぶこととする．

[18] 例えば文献 [113] では周波数領域尺度と PESQ を比較している
[19] 電気電子工学分野では S/N，SN 比とも表記する一般的な指標である．

定義式は式（1.4）で示す．$x(n)$ は主信号（音声），$\hat{x}(n)$ は劣化混みの音声信号で，N はサンプル総数を示す．SNR はある程度の長さを持った信号のパワー平均の比であるから，特に極短時間で変動する音声の品質評価には不向きと考えあられ，実際に McDermott らは音声品質の予測には不適当なことを報告している [115]．オーディオ信号においても，聴覚のマスキング効果を積極的に利用した MP3 や AAC では，劣化が SNR に影響しないため，オーディイオ符号化品質推定には使えない．一方で，実験系を構築する際の指標としては音声と雑音の比が明確であり，音質評価指標よりも実験条件を統制する指標として広く利用されている．本論文では，セグメンテーションを行わない長時間の信号パワーを用いた信号対雑音比を SNR と呼ぶ．また，SNR は雑音成分が大きいと値が小さくなり，品質が悪いことを示す．

$$\mathrm{SNR}(x,\hat{x}) = 10\log_{10}\frac{\displaystyle\sum_{n=1}^{N} x^2(n)}{\displaystyle\sum_{n=1}^{N}\{x(n)-\hat{x}(n)\}^2} \tag{1.4}$$

セグメンタル SNR　音声信号はダイナミックレンジの広い信号であり，極わずかな時間で信号パワーが変動する．しかし，SNR は全信号長から求めることから，パワーの大きい部分の寄与率が大きい．この問題の解決のため，音声信号を極短時間のセグメントに分割し，その平均を求めるセグメンタル SNR（以下，SNRseg）がある．定義式は式（1.5）で示す．式中の $x(n)$，$\hat{x}(n)$ は n 番目の分析フレームでの音声と雑音重畳音声 (劣化音) であり，N は分析フレームのサンプル長で，20〜40 msec に設定されることが多い[20]．M は全フレーム数で分析時間長に依存する．SNR と SNRseg は原理的には $\pm\infty$ の値を取りうるため，上限値と下限値を有限にすることが望ましい．Hansen らは雑音抑圧した音声の音質評価に用いるとき，上限を 35 dB，下限を -10 dB とするのが良いとしており，この値を用いることが多い [11]．本論文では，セグメンテーションを行う信号対雑音比を上下限値にかかわらず SNRseg と呼ぶ．SNR と同様に，値が小さい方が劣化が大きい．

$$\mathrm{SNRseg}(x,\hat{x}) = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\displaystyle\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\displaystyle\sum_{n=Nm}^{Nm+N-1}\{x(n)-\hat{x}(n)\}^2} \tag{1.5}$$

周波数領域信号品質尺度

板倉–斎藤距離（最尤スペクトル距離）　周波数領域の音声品質尺度に B.H. Juang による Itakura–Saito 距離尺度（以下 d_{IS} と呼ぶ）[116] がある．これは原音のスペクトルと線形予測によって得られたスペクトルの距離尺度に Itakura–Saito（板倉–斎藤）距離 [117] を用いて音質を評価する．以下の式（1.6）で与えられる．$\vec{a}_\phi$ と $\vec{a}_d$ はそれぞれ原音と劣化音の係数ベクトル，σ_ϕ^2 と σ_d^2 が，それぞれ原音と劣化音の全極フィルタのゲイン，R_ϕ は原音の自己相関行列である．距離指標なので，値が小さい方が性能が良い．

[20] 文献 [9] では 32 msec，文献 [11] では 30 msec，文献 [12] では 25 msec が用いられている．

$$d_{IS}(\vec{a}_d, \vec{a}_\phi) = \left[\frac{\sigma_\phi^2}{\sigma_d^2}\right]\left[\frac{\vec{a}_d R_\phi \vec{a}_d^T}{\vec{a}_\phi R_\phi \vec{a}_\phi^T}\right] + \log\left(\frac{\sigma_d^2}{\sigma_\phi^2}\right) - 1 \tag{1.6}$$

対数尤度比距離　d_{IS} と同様の周波数距離尺度に式（1.7）に示す LLR 距離尺度（Log–Likelihood Ratio measure，以下 d_{LLR} と呼ぶ）がある．これは d_{IS} とほぼ同じだが，分散を利用したゲイン推定が含まれない．このため，d_{IS} と比べ，純粋にスペクトルの形状の比較である．d_{IS} と同様に距離指標なので，値が小さい方が性能が良い．

$$d_{LLR}(\vec{a}_d, \vec{a}_\phi) = \log\left(\frac{\vec{a}_d R_\phi \vec{a}_d^T}{\vec{a}_\phi R_\phi \vec{a}_\phi^T}\right) \tag{1.7}$$

対数断面積比距離　d_{IS}，d_{LLR} と同様に LPC 係数を用いた音声品質指標に LAR 距離尺度（Log Area Ratio distance measure，以下 d_{LAR} と呼ぶ）がある [118]. これは 1979 年に T.P. Barnwell らによって提案された尺度である．LAR は声道を断面積の異なる無損失音響管を接続したものとモデル化し，接続点の断面積比（反射係数）を利用する音声品質尺度である．SNRseg と同様にフレーム分割した結果の平均値により時間変動も取り入れている．d_{LAR} の式を式（1.8）に示す．r_ϕ と $\hat{r}_\phi$ は原音と劣化音声の反射係数，M は分割したセグメント総数，m はフレーム番号を示す．d_{LAR} も距離指標なので値が小さい方が性能が良い．

$$d_{LAR}(r_\phi, \hat{r}_\phi) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left[\log\frac{1+r_\phi(m)}{1-r_\phi(m)} - \log\frac{1+\hat{r}_\phi(m)}{1-\hat{r}_\phi(m)}\right]^2} \tag{1.8}$$

重み付スペクトル傾斜距離　周波数スペクトルの傾斜の差を原音スペクトルと劣化スペクトルの間で比較する距離尺度に，Klatt らによる重み付スペクトル傾斜距離（Weighted Spectral Slope distance measure 以下，d_{WSS} と呼ぶ）がある [119]．式（1.9）に定義式を示す．$S(j,m)$ と $\hat{S}(j,m)$ が原音のスペクトルと劣化音のスペクトル，M はセグメント総数，m はセグメント番号，j は帯域番号で，$W(j,m)$ が重みを示す．d_{WSS} では LPC スペクトルだけでなく，ペリオドグラムによるスペクトルでも比較可能である．また重みも 36 帯域のもの [114, 119]，電話帯域に合わせて 25 帯域にしたもの [11] があり，用途に合わせて使用できる．このため，音声品質だけでなく音楽品質への利用もみられる [120]．d_{WSS} も距離指標なので，値が小さい方が性能が良い．

$$d_{WSS}(S, \hat{S}) = \frac{1}{M}\sum_{m=1}^{M}\frac{\sum_{j=1}^{k} W(j,m)(S(j,m) - \hat{S}(j,m))}{\sum_{j=1}^{k} W(j,m)} \tag{1.9}$$

LPC ケプストラム距離　LPC ベースのスペクトル距離尺度に，LPC ケプストラム係数の差である LPC ケプストラム距離（以下，d_{LPC} と呼ぶ）がある．これはパーセバルの定理から対数距離

尺度に等しく，セグメントごとの LPC ケプストラム係数差の平均値を求める．B.S. Atal による定義式 [121] を式（1.10）に示す．C と $\hat{C}$ は原音と劣化音の LPC ケプストラム係数で，M が最大フレーム数，m がフレーム番号，P が LPC 係数の最大次数，k が LPC 係数の次数となる．d_{Cep} も距離指標なので，値が小さい方が性能が良い．1.2.1 項で述べたように，LPC ケプストラム距離は波形符号化方式を用いた符号化音声の受聴品質の予測性能が高い [9, 10].

$$d_{Cep}(C,\hat{C}) = \frac{1}{M}\sum_{m=1}^{M}\frac{10}{\log 10}\sqrt{2\sum_{k=1}^{P}\left\{C(k,m)-\hat{C}(k,m)\right\}^2} \tag{1.10}$$

時間–周波数領域信号品質尺度

周波数重み付セグメンタル SNR　周波数領域の品質尺度はセグメントの時間平均を取るとはいえ，時間領域品質尺度である SNR，SNRseg と異なり，原信号スペクトルからの僅かな差を評価する方式であるから，騒音や残響による極端な劣化の評価には向かない．一方で，SNR と SNRseg は全帯域の平均であるから，音声聴取に重要な帯域とその他の帯域を均質に扱っている．LPC 係数を用いた尺度はフォルマント周辺を評価することと等価であるため，信号の音声的特徴を評価している．このため周波数領域の音声的特徴に配慮した SNR として，周波数重み付セグメンタル SNR（以下，fwSNRseg）が 1978 年に J.M. Tribolet らによって提案された [122]．定義式を式（1.11）に示す．x と $\hat{x}$ は原音信号と劣化音信号，m がフレーム番号，M がフレーム総数，j が帯域番号，W が帯域ごとの重み，n がサンプル，N がフレーム長を示す．重みと帯域分割については評価対象によって使い分けられており，例えば，雑音抑圧音声について，品質推定に用いる場合の比較 [113] や，了解度推定に用いる場合の比較 [123] がある．fwSNRseg も SNR，SNRseg と同様に値が小さい方が品質が悪い．本論文では，fwSNRseg の重みを明示して，fwSNRseg(重み名) と表記する．

$$\mathrm{fwSNRseg}(x,\hat{x}) = \frac{10}{M}\sum_{m=0}^{M-1}\frac{\displaystyle\sum_{j=1}^{K}W(j,m)\log_{10}\frac{\displaystyle\sum_{n=Nm}^{Nm+N-1}x^2(j,n)}{\displaystyle\sum_{n=Nm}^{Nm+N-1}\left\{x(j,n)-\hat{x}(j,n)\right\}^2}}{\displaystyle\sum_{j=1}^{K}W(j,m)} \tag{1.11}$$

1.2.6　明瞭度・了解度の予測と推定

明瞭度・了解度の予測に使われる標準化された方式として，AI，STI，SII とその経緯について述べる．

AI

Fletcher らの明瞭度と単語を用いた了解度の検討は，1947 年に French らの検討 [56] により明瞭度と了解度の関係が定式化され，1950 年に Fletcher らが改良した [124]．その後 1962 年に Kryter によって明瞭度指数 AI（Articulation Index）が提案された [125]．AI は 1969 年に ANSI S3.5–1969 として標準化されている [126][21]．AI は明瞭度がそのままでは相加性が成り立たないため，物理指標と対応させることを目的として考案された．AI は帯域ごとの明瞭度への貢献具合（明瞭度貢献指数）の総和が音声の明瞭度になるという考えに基づく．AI の定義式を式（1.7）に示す．全バンドにおける明瞭度を AI とした時，20 個に分割した i 番目の音声対雑音比 SNR(i) を求め，その線形結合によって AI を求める．SNR(i) は，信号部／劣化信号部共に，バンドパスフィルタを通して帯域に分割したのちに，式（1.4）と同様に求める．SNR(i) が R dB 以上の場合は，騒音が十分に小さく明瞭度への影響は無いとみなし，その帯域の明瞭度に対する貢献は上限値であるとする．本規格では $R = 30$ となっている．S はシフト項で，SNR が明瞭度に寄与しなくなる最小の SNR との差である．これも AI の規格では $S = 0$ だが，電話網評価以外で AI を利用する際に用いられるので併記した．分割する帯域幅と平均周波数を Table 1.7 に示す.

AI の値は 0.3 以下が「悪い」，0.3 から 0.5 は「普通」，0.5 から 0.7 は「良い」，0.7 より大きい場合は「非常に良い」とされる．このため，後述の様に式中の S と R をいくつに設定するかが規格制定後も議論された.

$$\mathrm{AI} = \frac{1}{20}\sum_{i=1}^{20}\frac{min(\mathrm{SNR}(i), R) - S}{R} \tag{1.12}$$

Table 1.7: Frequency bands of equal contribution to the AI

Number	Frequency limits（Hz）	Mean frequency（Hz）	Number	Frequency limits（Hz）	Mean frequency（Hz）
1	200–330	270	11	1600–1830	1740
2	330–430	380	12	1830–2020	1920
3	430–560	490	13	2020–2240	2130
4	560–700	630	14	2240–2500	2370
5	700–840	770	15	2500–2820	2660
6	840–1000	920	16	2820–3200	3000
7	1000–1150	1070	17	3200–3650	3400
8	1150–1310	1230	18	3650–1250	3950
9	1310–1480	1400	19	4250–5050	4650
10	1480–1660	1570	20	5050–6100	5600

AI の日本語での評価は三浦らによってなされた [127]．日本語では明瞭度貢献度の周波数分布が他言語の分布とやや異なる[22]ことといった，英国英語，米国英語と日本語の傾向差を報告している．また，式（1.12）の R は日本語に対しては 35 dB が良いと報告している [128]．また，佐藤らは帯域騒音（中心周波数が 0.5，1，2，4 kHz の 1/1 オクターブバンドノイズ）と明瞭度評価音声

[21]後述するように，現在では ANSI S3.5–1997 に更新されている.

[22]文献 [127] の図 4.49 参照．他言語は概ね 0.5 から 1.2 kHz に 1 つピークを持つゆるやかな山型なのに対し，日本語は 0.7 kHz と 1.9 kHz に 2 つのピークを持つ.

を別のスピーカから提示し，SNR ごとに AI を用いた明瞭度予測を若年者と高年者に対して行った [129]．その結果，三浦らと同様に $R = 35$，S は騒音レベルと被験者の年代（被験者の聴力）によって最適値を選ぶのが望ましいと報告している．この結果はシフト項 S は,「電話網の評価」や「空気伝送系での評価」と言った評価系による最小可聴レベルを揃える項として有効であることを示している．

以上の様に，AI は伝送系のレスポンスや周波数領域のひずみに対して頑強であるものの，原理的に長時間の音声信号と雑音信号の平均パワーを用いているため，時間変動する残響やエコーなどが及ぼす時間領域での波形ひずみの影響評価には向かない．

STI

建築音響分野では，残響の影響による音の明瞭度・了解度の変化を予測・推定することがホールなどの空間設計に有効である．残響はごく時間変動する加算性の雑音とみなされるため，AI ではその影響を正しく評価できない．1974 年に島原による明瞭度指標 [130] があり，感覚量としての空間伝搬明瞭度を予測した．その後，騒音下による明瞭度試験と予測が行われ，1978 年から 79 年にかけて，植松，曽根らによるランダム騒音の主観評価 [131] とその推定 [132] が行われた．その結果，定常騒音と同様に等価騒音レベルによって音声明瞭度・了解度の変化をうまく表現できることが示された．これに関連し 1979 年には Latham によってホール音響でも SNR に依存して明瞭度が変化することを報告している [133].

これらの成果を踏まえて，1980 年に Steeneken と Houtgast が音場内を伝達する音声波形の包絡線が残響や騒音により変形することに着目し，100%振幅変調した試験音を用いて変調度伝達関数（Modulation Transfer Function : MTF）の変化量を評価する STI（Speech Transmission Index）[134] を提案した．STI のダイアグラムを Fig. 1.3 に示す．STI は音声を模擬したノイズ（以下，模擬ノイズ）を振幅変調し，伝送路を空気伝搬させ多のち，バンドごとに包絡を求める．そして伝送路を経由しない模擬ノイズ自体の包絡と比較し，変調伝達関数（Modulation Transfer Function : MTI）を帯域と変調度ごとに求める．求める帯域幅は，125〜8000 Hz の範囲の 1/1 オクターブバンドで，変調度は 0.63〜12.5 Hz の範囲に 14 あり，この組み合わせは合計 98 個である．これらから見かけの SNR（式（1.7）の SNR(i) に相当）を求め，式（1.7）を $R = 30$，$S = -15$ としてオクターブバンドごとに平均する．最後にオクターブバンドごとに重みをかけて平均し，STI とする．STI はその後 ISO 9921[135] と IEC 60268[136] として標準化され，現在でも更新され続けている．また，STI は 98 個の MTI を求めるため，規格策定当初は計算量の問題があった．そのため様々な簡易計算法が考案され，その中で 9 個の MTF から求める RASTI（RApid STI）が IEC 268–16 として標準化されている [137].

STI の明瞭度予測能力は音韻バランスのとれた日本語の音節明瞭度に対して 5%程度の誤差であるという報告がある [138]．さらに，小涼らによる計算に用いる諸パラメータを日本語に最適化の検討がある [139]．しかし STI では全ての劣化を明瞭度・了解度の低下としているため，残響の中でも初期反射は明瞭度・了解度をむしろ向上させるという結果 [140] と矛盾している．このため 2010 年の翁長による UDP（Useful–to–Derimental ratio by Pulse）といった新物理指標の検討 [99] や，Sander による binoral STI[141] といった改良も検討されている．また，STI を求めるための音声模擬騒音を用いずに，観測信号から MTF をニューラルネットを用いた回帰で予測する手法が検討

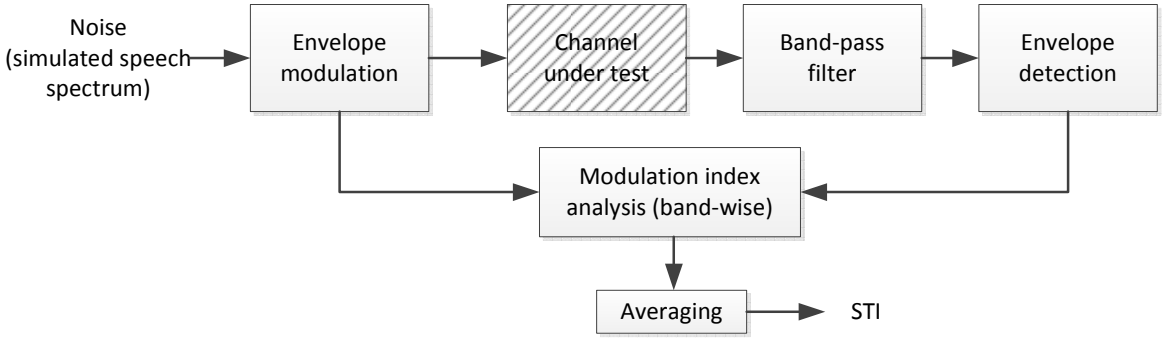


Fig. 1.3: Simplified diagram of the STI measurement

されている [142].

SII

SII（Speech Intelligibility Index）は AI の後継規格で ANSI S3.5–1997 で標準化されている [143]. その名の通り了解度を予測する指標で，STI の計算で用いた手法のいくつかを AI に組み込んでいる．SII の定義式を式（1.13）に示す．$W(i)$ は Band importance function で， Critical band（Table 1.7 とほぼ同等の 20 帯域とそれ以上の帯域で 21 帯域），1/3 オクターブバンド（18 帯域），Equally–contributing critical band（17 帯域），1/1 オクターブバンド（6 帯域）の 4 種帯域分割法ごとの重みが規定されている，1/3 オクターブバンドの時の了解度試験法ごとに標準化されている重みを Fig. 1.4 に，各種明瞭度・了解度試験に合わせた重みを示す[23]．平均的な環境で計測する場合には，average speech を用いることが推奨されている．帯域分割数 i は，各重み系列ごとの帯域の番号を，i_{max} は最大帯域数をそれぞれ示す．$L(i)$ は Speech level distortion factor で，評価したい音声の音量レベル $Sp(i)$ dB と標準発話の音量レベル $nSp(i)$ dB からの差を示し，式（1.14）で示す．本規格で定められている $nSp(i)$ を発声方法ごとに Fig. 1.5 に示す．SNR(i) は音声と雑音の比であることには変わりないが，加齢による最低可聴値の上昇を耳内ノイズと仮定して扱うこととなっている．耳内ノイズの仮定と，音声の発話音量を補正する $L(i)$ が導入されているところが AI，STI と異なる．

$$\mathrm{SII} = \sum_{i=1}^{i_{max}} W(i)L(i)\frac{\mathrm{SNR}(i)+15}{30} \tag{1.13}$$

$$L(i) = min\left[1, 1-\frac{Sp(i)-nSp(i)-10}{160}\right] \tag{1.14}$$

[23] 凡例の試験法については文献 [71] に詳しい.

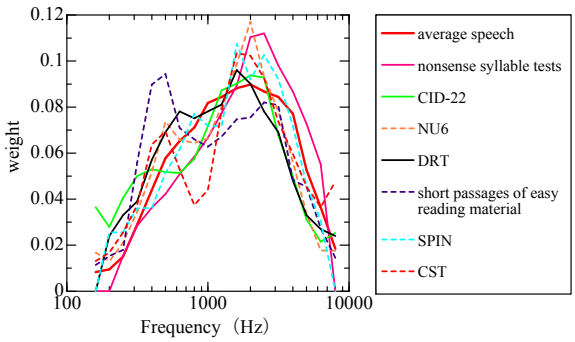


Fig. 1.4: Standardized weights of SII

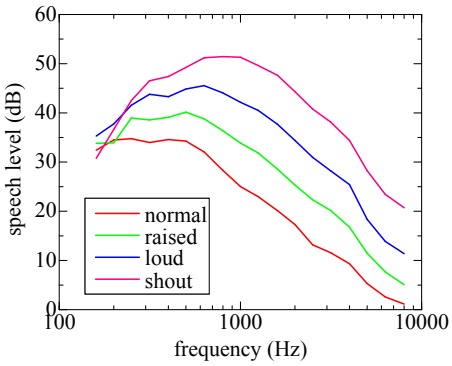


Fig. 1.5: Standardized speech level of SII

1.2.7　了解度推定に関する既存研究の課題

AI，STI と SII は共通して SNR の計算を持つ．1.2.5 項で述べたように，SNR はパワーの大きい部分の寄与率が大きい．つまり，SNRseg の様に音声のダイナミックレンジの広さを考慮するとパワーの小さい部分の考慮，つまり音声の時間変動に追従できる評価法が必要である．この様な観点から，特に聴覚障碍者向けの人工内耳や補聴器の設計指標に用いるために様々な尺度の比較が行われている [113, 123, 144]．比較されている指標は，1.2.5 項で述べた音声品質尺度の他に，SII の聴覚重みを用いた fwSNRseg[123]，劣化音と原音の相互相関を重みにした SNR である CSII（coherence SII）を高域，中域，低域に分けて求め，MARS 法（Multivariate adaptive regression splines）[145] を用いたスプライン平滑化回帰[24]を行った I3[146]，20 のクリティカルバンドごとに正規化共分散距離を重みとした SNR である NCM（Normalized Covariance Metric）[147]，25 のクリティカルバンドごとに重みをつけた SNR を正規化して平均した，AI–ST（Short-Term Articulation Index）[123] などである．文献 [123] の比較で最も性能が良かったのは I3 であった．I3 は SNR をベースとした特徴量から，了解度に対する非線形な回帰を行う手法であり，了解度が SNR だけでは説明できないことを示唆している．さらに，同一の雑音抑圧手法による了解度の国際比較 [148] では，言語ごとの評価単語が異なるなどの要因により，言語観の比較もそのまま行うことが出来ず，言語間の相互変換指標も望まれる．

一方で，STI と SII では重みづけしたクリティカルバンドやオクターブバンドを用いている．このバンドごとの重みは聴覚特性を模擬している．STI，SII 以外の推定で近年広く用いられる PESQ を用いた了解度推定 [21, 113, 149, 150] は，前述したようにバークスペクトル領域でのひずみの比較になるため，人間の聴覚特性を利用した了解度推定とみなせる．この他に，高品位合成音声用に，加藤らによる時間–ラウドネスマーカ表現といった聴知覚特性を用いた了解度推定 [151] といった検討もある．これらは重み付 SNR よりも人間の聴覚を再現しているため，推定性能は高くなるものの，音声を妨害する雑音・騒音が少なかったり，クリーンな音声のみに対応可能といった制限がある．特に「音声を最も妨害するのは音声である」という発想から，スピーチノイズやバブルノイズといった音声を用いた騒音は広く用いられているものの，多様な騒音下での了解度に関する研究例は主観評価，客観評価共に少ないといった課題がある．

[24]非線形関数を部分的に見れば，直線に近似できることに着目した回帰手法．

また，了解度は主観評価によって求めるため，騒音が異なる条件は全て評価しなければ正確な値はわからない．騒音による了解度低下が無いような電話系では 1.2.2 項で述べたように主音声の音量が了解度を決定する．また騒音が発生する系では，AI や SII，STI が SNR を含むように主音声と騒音のパワー比で大局的な傾向はわかる．しかし，聴覚マスキングは SNR を含む指標だけで説明できるエネルギーマスキングだけでなく，より心理的な情報マスキングを考慮する必要がある．特に非定常な騒音ではごく短時間に SNR が大きく変動するため，従来の SNR を中心に据えた指標では不完全である．しかし，あらゆる非定常な騒音環境を評価することは現実的ではないため，必要最低限の騒音数に絞り込む必要がある．田中らは，了解度に影響する騒音の分類に印象評価試験（形容詞対を用いた騒音の音色評価試験）の結果によるクラスタリングを試みている [152]．この他にも周囲環境音認識に関する研究 [153] 等があり，了解度の変化が同傾向な騒音環境を分類することが求められている．

1.3 研究目的と論文構成

1.3.1 着眼点と研究目的

本研究の目的は，多様な騒音環境を模擬し，音声了解度の主観評価を行い，その結果を推定することである．

このためにまず，筆者らがこれまでに検討してきた被験者を中心とした同心円状に騒音がある系を用いて，主観評価結果と 1.2.5 項で述べた音声品質の物理指標及び PESQ を用いた推定結果を比較し，既存の手法による推定の問題点について述べ，見つかった問題点の解決法を提案する．

次に，提案する手法の多種多様な騒音を用いた了解度試験を行い，その主観評価結果を推定する．この時，騒音の種類による了解度への影響は，文献 [152] と同様に複数のクラスタ（了解度変化が同系統な騒音群）に分類されると予想される．例えば，AI，SII でも推定可能と考えられる定常騒音を多く含むクラスタ，時間変動を考慮した推定が必要な極短時間での変動を多く含む非定常騒音のクラスタ，音声的要素の多いクラスタと言ったものが考えられる．本論文では，騒音の種類をいくつかのクラスタに分類し，そのクラスタごとに最適な推定関数を用いることを検討する．これにより，高精度な了解度推定が可能であると考えられる．クラスタリング手法については，機械学習分野において教師なし学習による分類として広く研究されており，異分野での応用例も多い．本論文では騒音の了解度に対する特徴について事前知識なしで分類し，分類結果を用いた推定法の検討に重点を置くこととする．分類に用いる特徴量には騒音の音色情報から求まる特徴と，その環境における発話を用いた音声言語の特徴の二つが考えられる．音色情報を用いることで，評価単語に関わらず騒音の分類が可能になる．本論文では，音色特徴を用いたクラスタリングを行う．言語情報に関しては，推定のための回帰関数をクラスタごとに別途用意することで主観値と対応の取れた精度の高い推定を目指す．

推定関数に用いる回帰関数は，多くの了解度推定で用いられた，線形回帰やロジスティック回帰，シグモイドカーブフィッティングといった線形モデルや一般化線形モデルを用いたパラメトリック回帰による推定関数の作成も検討するが，より複雑な非線形関数への回帰も必要であると考える．パラメトリック回帰は説明変数に用いる物理量[25]と了解度の関係に対して関数の形状を仮定して

[25]多くは SNR を中心としたエネルギーマスキングを説明する物理指標.

適合する．このため，情報マスキングについては関数の形状以外に考慮していない．1.2.7 項で述べた I3 のように，仮に説明変数の物理量がエネルギーマスキングのみであっても，目的変数である了解度に対して，ノンパラメトリックで複雑な非線形関数を用いれば，その非線形な変換は情報マスキングの考慮を模擬すると考えられる．

ノンパラメトリックな回帰の代表例として，機械学習による非線形関数への回帰があり，本論文で検討する．しかし，一般に機械学習を用いた非線形回帰は説明変数に用いる特徴量ベクトルの次元数が多い場合や，サンプル数が少ない場合に過学習による汎化性能の低下が起こる．この様な問題を考慮した回帰手法にサポートベクトル回帰（Support Vector Regression: SVR）がある．SVR はサポートベクトルマシン（Support Vector Machine: SVM）[154] と同様の手法を用いており，正則化，サポートベクトルのマージン最大化，ϵ-不感応関数の利用といった内容を考慮した回帰手法である．機械学習による回帰は，学習データに対して最適な回帰係数の決定であり，特徴量として用いた物理量に対する重みづけである．この重みは，本章のこれまでに述べてきた多くの検討で比較されてきた聴覚重みとは本質的に異なり，了解度推定のためだけの周波数重み[26]をつけることになる．純音または，バンドパス・ノイズを用いた聴覚実験に基づく聴覚フィルタは，言語に依存しない普遍的な人間の聴覚を模擬している．しかし，言語を用いた品質では，単語親密度等の聴覚以外の言語情報による品質も考慮しなければならない．機械学習による回帰は，解きたい問題のメカニズム解析は行えないが，ブラックボックスとして高精度な推定モデルを求めることには有効である．I3 で用いている MARS 法によるスプライン回帰は，特徴量の変化傾向メカニズムは比較的解析しやすいが，SVR よりも汎化性能が低いといった報告（たとえば文献 [155]）もある．本研究は，これまで解析されてきた聴覚メカニズムに基づく了解度予測に対して，たとえブラックボックスであったとしても汎化性能が高い予測ができるのならば，今後の音声アプリケーションの開発に非常に有効であり，よりよい音システムの設計に関する基盤技術となることが期待される．また，機械学習による推定性能の上限が明らかになれば，聴覚のメカニズムに基づいた予測法の発展に対しても目標値になることが期待される．

以上，本研究では，機械学習分野の手法からクラスタリングと SVR を選択し，これらを組み合わせた了解度推定法を提案する．騒音のクラスタリングに用いる特徴量は騒音の音色情報を用い，SVR の特徴量はエネルギーマスキングを説明するセグメンタル SNR を音声聴取のクリティカルバンドごとに用いた特徴量を用いる．そして，検討事例の多い客観音質値を用いたパラメトリックな回帰を用いた了解度推定法に対し，提案法が未知のデータに対する汎化性能の高い手法であることを検証することを本論文の目的とする．

1.3.2 本論文の構成

本論文は本章を含め，全 7 章で構成される．2 章では，被験者を中心とした同心円状に騒音がある系における了解度評価と推定について述べ，騒音下での了解度試験とその推定に関する問題点を整理する．そのために主観評価結果の分析，相関係数の比較，ノイズクローズドな条件での推定比較，ノイズオープンな条件での推定比較を JDRT の子音特徴ごとに行う．3 章では，2 章の結果より騒音下での了解度推定の課題に対する解決法として，機械学習を用いた了解度推定法を提案する．4 章では提案推定法のうち，騒音クラスタリングによる騒音の分類と主観評価による分類

[26]聴覚的実験に基づかないため，本論文で求める重みは聴覚重みと区別して「周波数重み」と呼ぶ．

結果の解析，2 章と同様の手法による推定関数の作成を行う．5 章では，4 章の分類結果を用いた了解度を学習データとして，SVR と他の回帰手法を用いて推定関数の作成とその性能評価について述べる．6 章では，4 章と 5 章で作成した推定関数を用いた了解度推定について，未知データに対する汎化性能を比較する．7 章では全体を総括する．

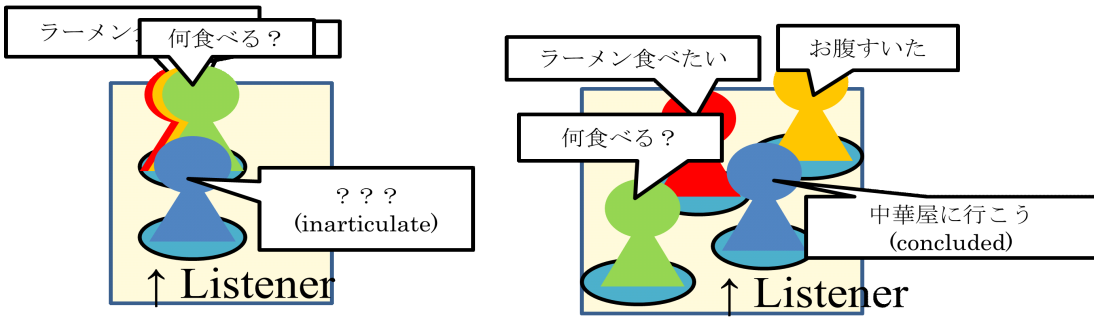
第2章　バイノーラル音声システムと既存尺度による了解度推定の検討

本章では，騒音下での音声了解度試験とその推定のための予備実験として，被験者を中心とした同心円状に妨害騒音を配置した実験系における音声了解度の主観評価結果と，既存尺度による推定について述べる．主観評価には JDRT を，推定には，1.2.5 節で述べた客観音質評価指標と PESQ を説明変数とした非線形最小二乗法によるシグモイド・ロジスティック関数のカーブフィッテイングを用いる．そして，本章の結果より得られた課題について次章以降で解決していく．

2.1　検討する主観評価モデルと実験の設定

2.1.1　実験モデル

本章で検討する主観評価モデルは，Fig. 2.1 に示す筆者らがこれまで検討してきた 3 次元音響会議システムに関する基礎検討 [108, 109, 111] で用いてきたモデルである．本システムは，MIT Media Lab. で公開されている KEMAR–HRIR[156] をドライソース音源[1]に畳み込み，仮想的な音源がその方位にあることを近似する．これにより複数話者発話による輻輳を押さえた音声通信をめざす．このための音声の聞き取りやすさ指標である了解度の主観品質評価を行ってきた [108, 109, 111]．主音声の了解度の主観評価には 1.2.4 項で述べた JDRT を用いる．了解度は JDRT の評価音を聴取し，式（1.3）によって求める．



(a) Conventional system　　(b) Proposed System

Fig. 2.1: Image of proposed system

システムの音像配置イメージを Fig. 2.2 に示す．被験者を中心とした同心円状に仮想音像の騒音が 45°単位で定位される．評価単語を発話する話者音像は被験者の正面（0°）に定位する．騒音は

[1]モノラルの原音信号．本論文では定位感のない信号をドライソースとしている．

話者音像と同位置に定位したときの 2 話者それぞれ 120 評価単語の平均パワーとの SNR_{in}[2]が 0 dB となるように設定した．SNR が正の時は話者より遠い円上にある騒音を，負の時は話者より近い円上にある騒音を近似する．評価信号の生成手順のブロックダイアグラムを Fig. 2.3 に示す．入力した騒音信号に事前に求めてある音声信号とのパワー比を揃える係数 α を乗じ，定位したい方位の HRIR を畳み込む．今回用いた KEMAR–HRIR はダミーヘッドの左右対称性を利用しているため，片側の特性を折り返して用いるため，本論文では右耳を基準とし，左耳には 360°から引いた方位の右耳の HRIR を畳み込む．そして距離を模擬したゲイン 1/a を乗じ[3]，正面に定位した音声と加算する．

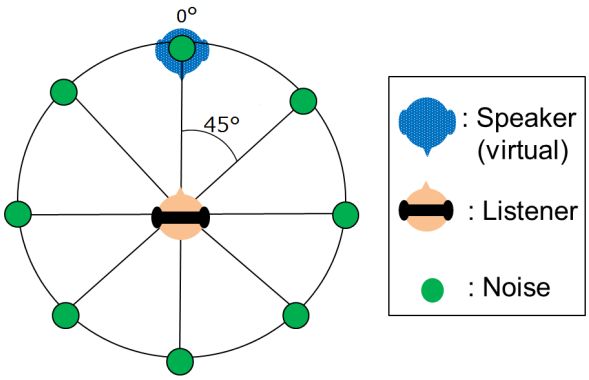


Fig. 2.2: Location of localized sound source

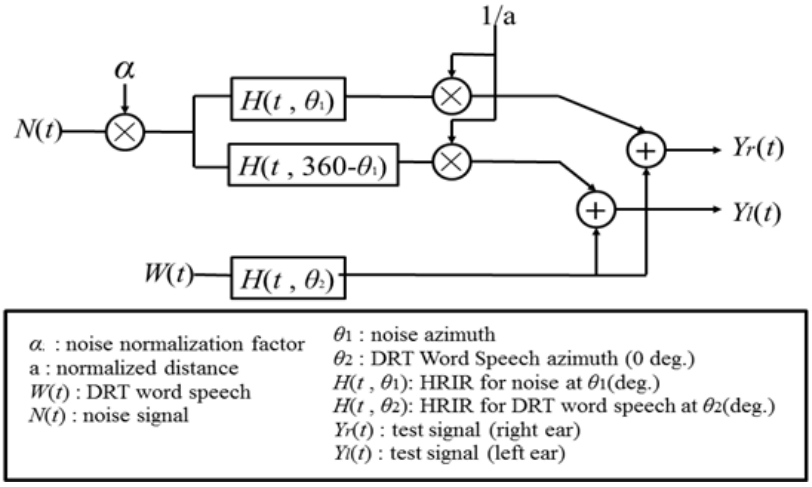


Fig. 2.3: Test signal generation procedure

2.1.2 評価音声と騒音信号

本節では主観評価に用いる評価音声と騒音信号，および実験に用いる SNR_{in} の設定を述べる．

[2]本論文では，実験音声の解析に客観音声品質指標として用いる SNR と区別するために，実験の設定に用いた SNR を SNR_{in} と呼ぶこととする．

[3]被験者から発話者までの距離を a とし，その逆数に相当するゲインを乗算した．

評価音声

評価音声には 1.2.4 項で述べた JDRT の 120 単語コーパスを男女各 1 名分づつ用いる．用いた 2 話者の発音から，鼻音性比較の属性である Nasality 子音特徴比較を行う単語対の man と ban のスペクトログラムを Fig. 2.4 に示す．上は女声で，下が男声，左が man で右が ban である．この単語対は先頭子音以外の音が母音であるため，フォルマントがみられない区間でエネルギーがある区間[4]が第一子音となる．図より，同一話者では単語対間のフォルマント遷移傾向が同じであり，同一単語では子音部のエネルギー分布が話者によらず同傾向であり安定した音声セットであることがわかる．

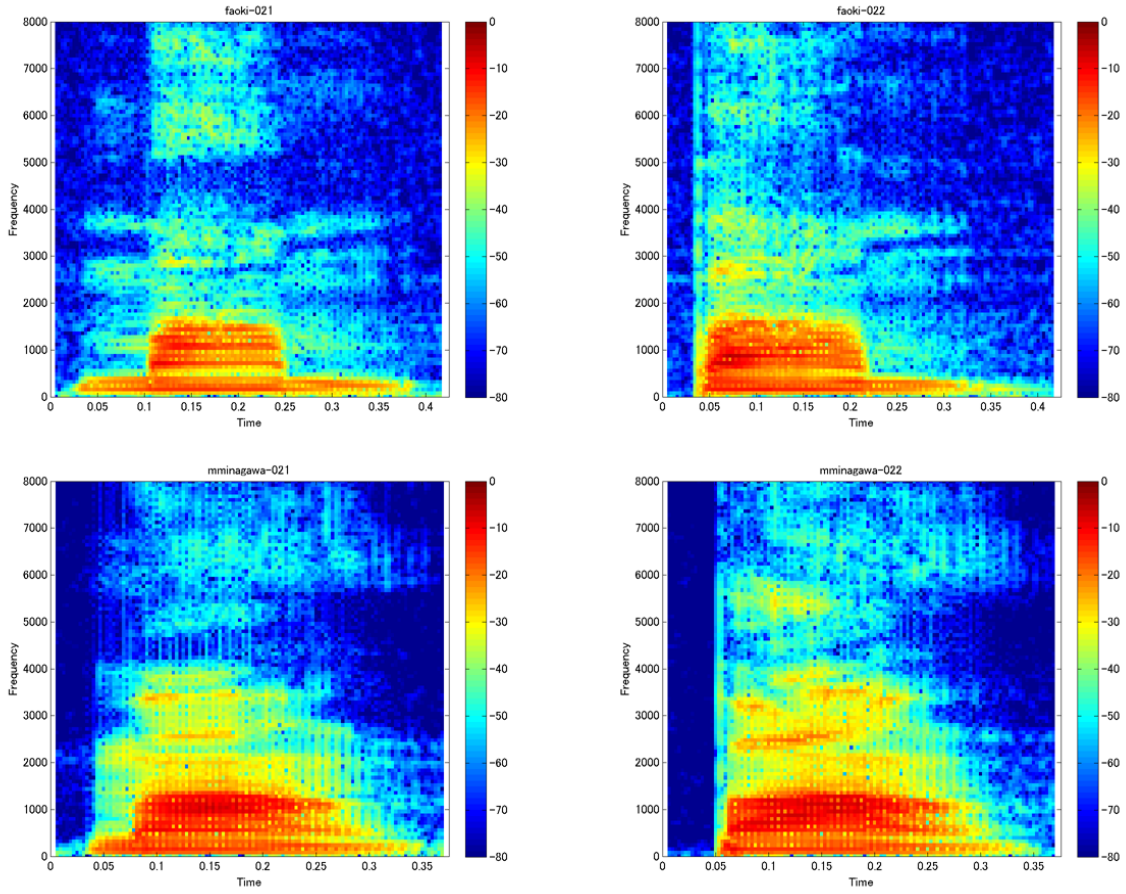


Fig. 2.4: Example of spectrogram man–ban word pair(Nasality)

評価騒音

評価単語音声を妨害する騒音には，TY–89[77] に収録されているマルチトーカノイズ（以下，Babble），白色雑音（以下，White），電子協騒音データベースのダイジェスト版 [157] に収録されている電車騒音（以下，Railway）と幹線道路騒音（以下，Highway）から 1 sec 切り出して使

[4]単語の切り出しについては，文献 [21, 104] の検討時のままであり，男女間で統一されていない．スペクトログラム冒頭の無音区間は子音ではなく発話前の区間である．

用する．電子協騒音データベースはワンポイント・ステレオマイクで録音されているため，定位音源のためのドライソースとするために左右両チャンネルを加算平均してモノラルの音源としてから HRIR を畳み込むこととする．

各騒音を切り出して 2 話者分の評価単語の平均パワーに統制して使用する．統制後の騒音のスペクトログラムを Fig. 2.5 に示す．図より，Babble は 1 kHz 周辺にパワーが集中しており，騒音データベースから選択した 2 種は Babble よりは広く分布している．僅かに Railway の方が，1 kHz 周辺にパワーが集中する傾向にある．White は平坦に分布している．

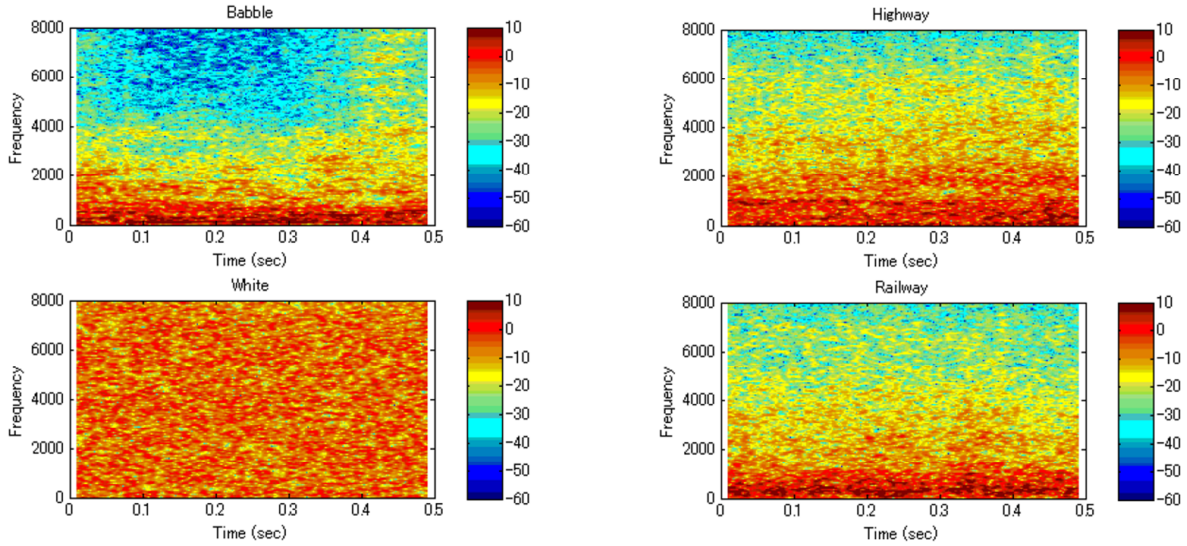


Fig. 2.5: Spectrogram of noise

実験設定

実験の設定を Table 2.1 にまとめる．前節の 4 種の騒音に HRIR を畳み込み，Fig. 2.2 に示した 8 方位に定位した．定位に用いた HRTF スペクトルの例として，DRT 評価音声に畳み込んだ KEMAR–HRTF のスペクトルを Fig. 2.6 に示す．$\mathrm{SNR_{in}}$ は正面 0°の音声と同じ位置にあるときを 0 dB として，6，0，−6，−12 dB として設定する[5]．聴取実験の被験者は各騒音条件がそれぞれ 8 名になるように行った．ただし，騒音条件ごとに被験者は異なる．全ての被験者は 20 代の大学生であり，聴力検査に問題が無いことを確認している．

主観評価は事前に計算機上で評価音に騒音を合成してある音源ファイルを作成しておき，GUI による評価アプリケーションで提示順番をランダムにして被験者に提示する．GUI のスクリーンショットを Fig. 2.7 に示す．また，評価音の再生には，USB 接続のオーディオインターフェース（Roland 社製 UA–25）を介して被験者にヘッドホン（Sennheiser 社製 HD–25II）を用いた．再生時の音声レベルが十分な値になるように，再生音量を統制[6]した．

[5]厳密には，0 dB に合わせて振幅を $\frac{1}{2}$，1，2，4 倍と設定したため，正確には，6.02，0.00，−6.02，−12.04 dB である．

[6]音圧の統制には，受聴位置の正面 1.4 m に設置したラウドスピーカから再生した中心周波数 1 kHz の 1/1 オクターブバンドピンクノイズと統制した．統制した値は受聴位置における騒音レベルが 58 dB となるように設定した．

Table 2.1: Experimental conditions

Label	Noise type	Speaker	SNR_{in}(dB)	Subjects
B	Babble	male 1 female 1	6, 0 −6, −12	8
W	White			
H	Highway			
R	Railway			

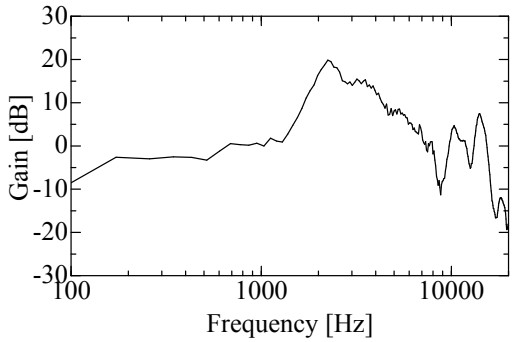


Fig. 2.6: Spectrum of KEMAR–HRTF 0 degrees

Fig. 2.7: Example for selection input dialog box

2.2 主観評価結果

本節では前節で述べた主観評価実験の結果を述べる．まず，実験結果の分散分析結果を示し，評価した要因に有意差がみられたことを示す．次に，了解度と方位角を騒音種ごとに抜き出して騒音の方位角の影響，子音特徴ごとの了解度と SNR の関係を議論し，最後に発話者の性別による違いを議論する．

2.2.1 概要と分散分析結果

主観評価結果の分散分析結果を Table 2.2 に示す．分散分析とは了解度試験の様に複数の被験者による観測値の分散がある実験の変動を誤差変動と各要因およびそれらの交互作用による変動に分解し，要因および交互作用の効果を判定する，統計的仮説検定の一手法である．本実験の実験要因（Source）は，群内要因に話者の性別[7]（G），騒音要因（N），SNR（L），SNR_{in} の変動，騒音の定位方位（A）の 4 要因と，被験者に基づく分散（S），eroor は S を含んだ（）内の要因による誤差（群間要因誤差）である．また，要因二つ以上を × で繋げた項目は交互作用である．SS は

[7]本実験では性別ごとに 1 名づつなので話者差を性別差とみなす.

該当要因の平方和，dfは自由度，MSは平方和を自由度で割った平均平方，Fは各実験要因内誤差と群間要因誤差の比[8]である．pは帰無仮説の確率を表す．pが特定の有意水準よりも小さい時に表下部の凡例に基づいてアスタリスクをつけて示す．

結果より，騒音の定位方位を除く3種の要因では要因単独で有意差がみられた．また，2要因の交互作用では，G×AとN×Dを除く4要因に有意水準5 %以下の有意差がみられた．3要因以上ではG×N×L要因の交互作用に有意差がみられる．本節では主として各要因単独の効果について検証する．

Table 2.2: Results of ANOVA

Source	SS	df	MS	F	p
S:Subject	0.3865438	7	0.0552205		
G:Gender	0.8244257	1	0.8244257	14.678	0.0064^{**}
error[G×S]	0.3931783	7	0.0561683		
N:Noise	0.6191929	3	0.2063976	3.120	0.0478^{*}
error[N×S]	1.3891518	21	0.0661501		
L:SNR_{in}	12.8891291	3	4.2963764	15.546	0.0000^{****}
error[L×S]	5.8037774	21	0.2763704		
A:Azimuth	0.0097085	7	0.0013869	0.013	1.0000
error[A×S]	5.0940508	49	0.1039602		
G×N	0.3517040	3	0.1172347	3.799	0.0254^{*}
error[G×N×S]	0.6480288	21	0.0308585		
G×L	0.2385895	3	0.0795298	6.435	0.0029^{***}
error[G×L×S]	0.2595452	21	0.0123593		
G×A	0.0571498	7	0.0081643	2.107	0.0602^{+}
error[G×A×S]	0.1898460	49	0.0038744		
N×L	0.9568528	9	0.1063170	7.172	0.0000^{****}
error[N×L×S]	0.9339512	63	0.0148246		
N×A	0.1368366	21	0.0065160	0.963	0.5126
error[N×A×S]	0.9950239	147	0.0067689		
L×A	5.6135713	21	0.2673129	4.004	0.0000^{****}
error[L×A×S]	9.8150985	147	0.0667694		
G×N×L	0.3980406	9	0.0442267	4.441	0.0002^{****}
error[G×N×L×S]	0.6274098	63	0.0099589		
G×N×A	0.0579809	21	0.0027610	0.604	0.9107
error[G×N×A×S]	0.6718655	147	0.0045705		
G×L×A	0.0646728	21	0.0030797	0.691	0.8367
error[G×L×A×S]	0.6548666	147	0.0044549		
N×L×A	0.2298512	63	0.0036484	0.649	0.9821
error[N×L×A×S]	2.4792811	441	0.0056220		
G×N×L×A	0.2611546	63	0.0041453	0.865	0.7585
error[G×N×L×A×S]	2.1145251	441	0.0047948		
Total	55.1650044	2047			

$+$ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

[8]群内分散と群間分散は同じ母集団の分散の推定値のため，理想的には1となる．しかし，各群の平均値に差があると群間分散が大きな値となりF値が1よりだいぶ大きくなり，群間要因に差があるかを判別する指標になる．

2.2.2　120 単語平均による分析

方位角の影響

Fig. 2.8 に騒音種ごとに 120 単語平均の了解度と騒音方位角の関係を SNR_{in} ごとに示す．了解度は女性話者と男性話者の平均値である．エラーバーは 95%信頼区間幅を示す[9]．評価値と信頼区間が 100%と 0%を超えることはないため，天井効果とフロア効果は顕著ではない[10]．騒音方位角が了解度に与える影響は，騒音種かかわらず，話者音像と騒音が重なる正面 0°と，定位の前後誤りが発生しやすい後方 −180°が低くなり，45°から 135°，225°から 315°にかけて高くなる傾向がある．方位角の影響は Babble と Railway で特に顕著な落ち込みがみられ，White と Highway では緩やかである．この違いは，Fig. 2.5 のスペクトログラムでわかるように，1 kHz 付近へ騒音のパワー集中が，影響していると考えられる．つまり，パワーが統制された騒音同士でも，その周波数分布によって音声了解度への影響が違うことがわかる．これは 1 章で述べた STI や SII の様な帯域別の了解度貢献度が了解度推定に有効であるだろうことが示唆される．

SNR_{in} ごとに見ると，同一方位角においては SNR_{in} が大きい方の了解度が高く，低い方が小さいというの序列は維持されるものの，同一の SNR_{in} でも方位角によって取る値は大きく異なる．この値は SNR_{in} が小さいほど顕著になり，Railway の −12 dB では了解度 20%から 60%までの広い値を取りうる．これは，SNR_{in} は正面 0°で定義されたものであり，HRIR を畳み込んだことによるノイズパワーの変化，特に左右両耳間での差により，どちらか片耳あたりの相対的な SNR が変化することで主音声が聞き取れるようになったと考えられる．了解度が大局的に SNR に依存することは，1 章で述べたように，AI や STI，SII と言った指標で計測できることが示唆される．

次に，方位角が関係する交互作用のうち，棄却率 5%で有意差がみられた騒音方位角と騒音種の下位検定における単純主効果の結果を Table 2.3 に示す．Noise(0°) は，0°の場合の騒音種 4 種間に有意差がみられるかを検定している．結果より，すべての方位で騒音差が有意差であり，定位した騒音は方位角一つ一つごとに独立した騒音とみなしても，平均値が異なることなり，1 点 1 点の正確な SNR を求め，了解度を推定することが可能であることがわかる．

SNR_{in} と騒音種

方位角の影響は程度の差はあるものの，騒音種によって同傾向であったため，騒音種別に了解度と SNR_{in} を比較する．ここでは，各騒音の同一 SNR_{in} で計測した 8 点の加算平均値を各 SNR_{in} の代表値とする．このため，了解度の分散は被験者要因と方位角要因の 2 種を考えなければいけない．これらは，Table 2.2 の L×A 交互作用である，方位角による分散は前節でその影響を見たため，本節で検証する了解度の分散を被験者による各 SNR_{in} の値の分散とするため，前節で検証した各方位角ごとの信頼区間幅の平均値を各 SNR_{in} 値の被験者間の分散値として議論する．この 95%信頼区間の平均値を Mean 95% Confidence Interval（以下，MCI）と呼ぶ．高橋らは MCI は被験者間の誤差の代表値であり，この値以下になることが音声品質の推定の目標値であるとしている [158]．本論文でも高橋らの検討に倣い，MCI を推定の目標値とする．

[9] 男女話者の同一 SNR_{in} と方位角の結果を被験者ごとに平均してから平均値と信頼区間を求めた．

[10] ただし，了解度が 80%以上と高い範囲に多いため，50%前後の分析結果と比べると，了解度が高い場合は信頼区間が小さくなるため，天井効果はある程度みられていると考えられる．

Table 2.3: Main effect of noise type and SNR_{in}

Source	SS	df	MS	F	p
Noise(0°)	2.7325400	3	0.9108467	9.797	0.0000****
Noise(45°)	1.8591017	3	0.6197006	6.666	0.0003****
Noise(90°)	1.5862527	3	0.5287509	5.687	0.0010****
Noise(135°)	1.8191210	3	0.6063737	6.522	0.0003****
Noise(180°)	1.8468439	3	0.6156146	6.622	0.0003****
Noise(225°)	2.1270687	3	0.7090229	7.626	0.0001****
Noise(270°)	2.7994174	3	0.9331391	10.037	0.0000****
Noise(315°)	3.7323550	3	1.2441183	13.382	0.0000****
error		168	0.0929695		
Azimuth(Babble)	4.0460380	7	0.5780054	7.599	0.0000****
Azimuth(White)	0.1409371	7	0.0201339	0.265	0.9668
Azimuth(Highway)	0.3394633	7	0.0484948	0.638	0.7245
Azimuth(Railway)	1.0968414	7	0.1566916	2.060	0.0496*
error		196	0.0760671		

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

Fig. 2.9 に 120 単語平均での SNR_{in} ごとの了解度と MCI を騒音種ごとに示す．前節でみたように大局的には SNR_{in} によって了解度が定まっているのがわかる．しかし，同一の SNR_{in} 値であっても騒音種の差がみられ，White は他のどの騒音よりも常に低い．Railway の了解度が Babble と Highway に比べて了解度の傾向差がみられる程度に低くなるのは，−12 dB の時で 0 dB と 6 dB ではこの三種はほぼ同じ値になっている．White が Babble と比べて同一の SNR_{in} であっても了解度が低くなることは，文献 [104] と同傾向である．次に MCI を見ると，SNR_{in} の低下と共に増加しており，SNR_{in} が小さい環境ほど被験者間の回答に差があることがわかる．また，White のみ 0 dB で他よりも高い MCI となり，−6 dB と −12 dB では Railway も White と同程度の MCI となる．

文献 [104] の SNR ごとの結果と比べると，本論文の主観評価の結果とほぼ同じノイズレベルで比べた場合に全体的に了解度が改善している．これは方位角による改善を平均しているため，同一 SNR で比較した場合に高い了解度になるという作図上の問題もあるが，同一の了解度となる SNR を比べるとおおむね 5 dB 程度改善している．

SNR_{in} と話者性別

話者性別の違いを検証する．Fig. 2.10 に騒音種ごとに SNR_{in} と了解度の関係を話者性別ごとに示す．プロットは Female と Male に分けてプロットした[11]．エラーバーは 95%信頼区間を示す．JDRT の話者性別差については，文献 [21, 104] では全了解度平均で 10%未満であった．本実験では，White の −12 dB で最大 15%の了解度差がみられる．下位検定における単純主効果の結果を Table 2.4 に示す．結果を見ると G×L 交互作用 (話者と SNR_{in} の交互作用) では，SNR_{in} が −6 dB と −12 dB で有意差がみられるものの，0 dB と 6 dB には有意差がみられない．SNR_{in} の性別差はどちらも有意差だが，これは SNR_{in} の変化によって了解度が有意に変わるという意味であり，前節の結果が性別ごとに分けても成り立っているという意味である．

[11] 男女各 1 名づつで評価したため話者の違いが性別の違いを示す．

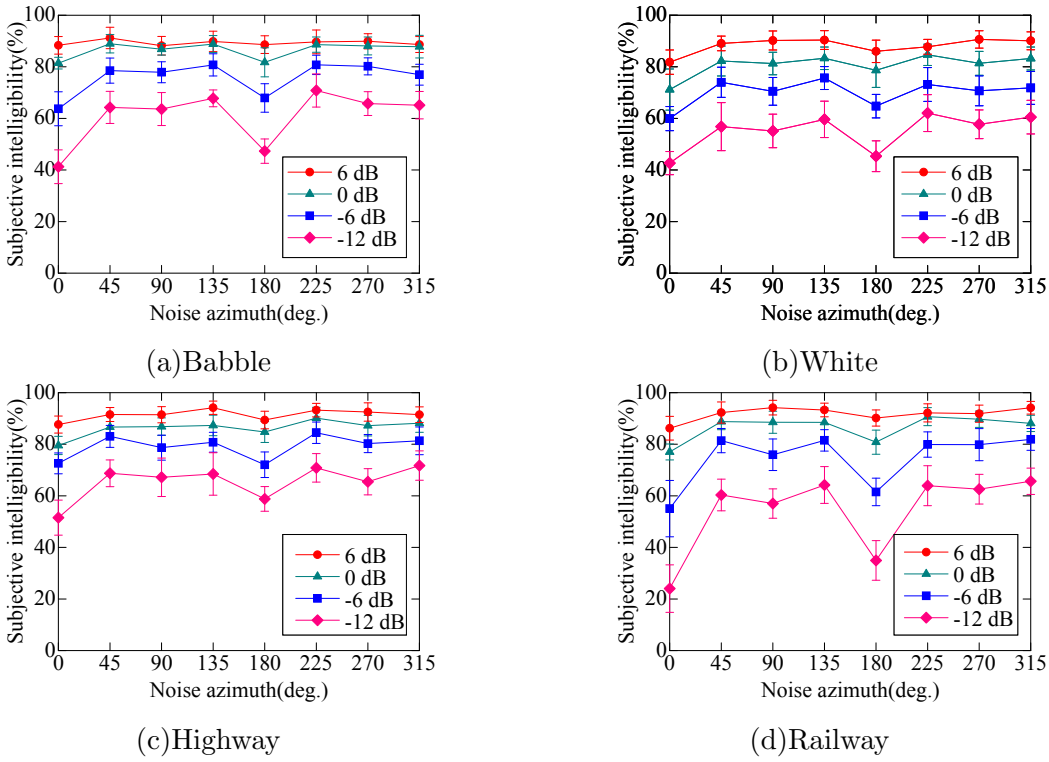
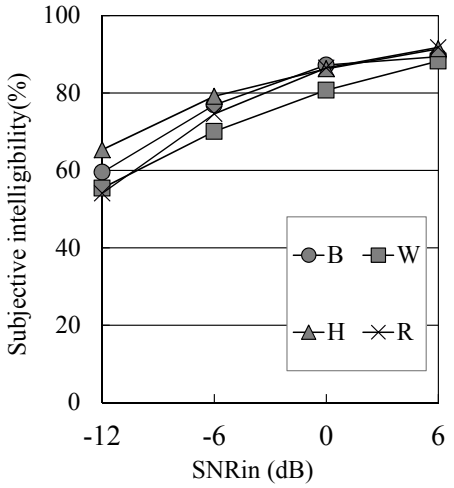


Fig. 2.8: Comparison of intelligibility(120 words average) with various noise localized azimuth

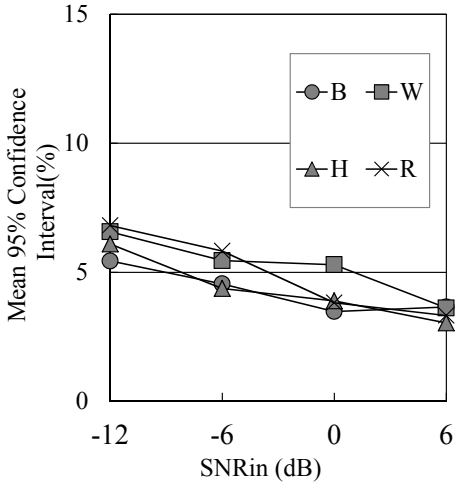
次に，話者性別と騒音種の下位検定における単純主効果の結果を Table 2.5 に示す．結果より，Babble では話者差に有意差はみられないが，White と Railway では有意水準 5%以下の有意差がみられる．これは，Babble と Highway のプロット点のスプラインが近く White と Railway で離れている傾向と一致する．

最後に G×N×L 要因の単純交互作用について，Table 2.6 に示す．本実験はこれまでにみられたように SNR_{in} が低い場合に結果がばらつくので，Gender×Noise (6 dB) だけに有意差がみられるのは直感的には違和感がある．しかし，Fig. 2.9 の MCI の結果より，Babble は 6 dB よりも 0 dB の方が MCI が小さい．つまり，被験者によらず結果が安定している 0 dB の傾向差が他の騒音と異なることが効いている．Gender×SNR (Babble) に有意差がみられるのも同様である．最後の話者と SNR_{in} の交互作用が話者によらず有意なのは，2.2.2 項の結果よりも明らかである．

以上の結果より，SNR が高い範囲（6 dB，0 dB）では話者差が無いものの，SNR が低い範囲（−6 dB，−12 dB）では話者差がみられた．JDRT の基礎検討 [21, 104] では 120 単語平均においては話者差が少なく，子音特徴別の分析で Voicing，Nasality，Graveness の 3 種において話者差がみられるという結果だった．本論文では実験系が仮想音響系であること，騒音種が増えていること，同一の話者音源であるが，性別ごとに 1 話者づつであることといった違いがあり，単純比較できないものの，騒音方位角ごとの騒音種が独立であるという結果より全 32 騒音（4 騒音 ×8 方位）であり，SNR_{in} も片耳あたりに換算すれば 32 パターン（SNR4 値 ×8 方位）あることから，基礎検討の結果に換算して，特に 0 dB 以下の範囲を細かく評価し，子音特徴ごとの傾向差が全体平均にも表れてきたのではないかと考えられる．次節では子音特徴ごとの SNR_{in} と了解度，MCI，男女差を見ることで，この仮説が正しいか検証する．

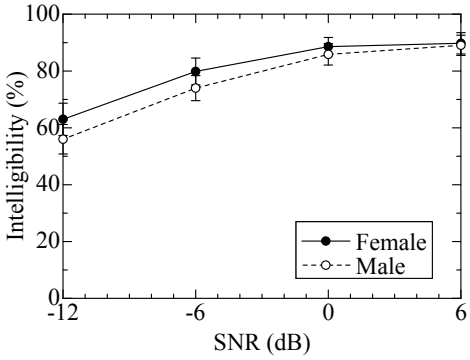


(a)Intelligibility

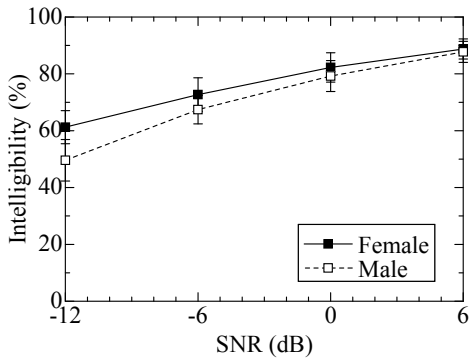


(b)MCI

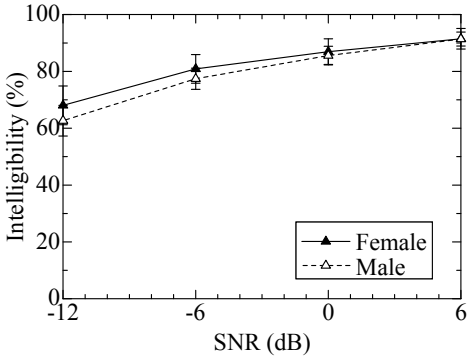
Fig. 2.9: Comparison of intelligibility and MCI with various noise type (120 words average)



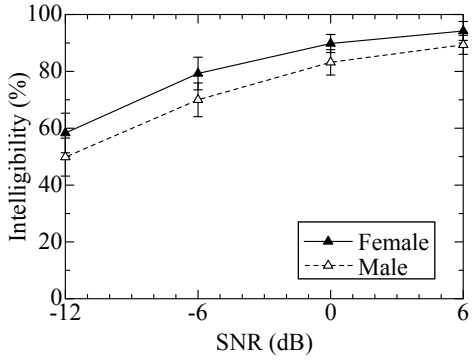
(a)Babble



(b)White



(c)Highway



(d)Railway

Fig. 2.10: Comparison of intelligibility(120 words average) with gender type

Table 2.4: Main effect of gender and SNR_{in}

Source	SS	df	MS	F	p
Gender(6 dB)	0.0334675	1	0.0334675	1.436	0.2409
Gender(0 dB)	0.0604579	1	0.0604579	2.593	0.1185
Gender(−6 dB)	0.4023736	1	0.4023736	17.261	0.0003^{****}
Gender(−12 dB)	0.5667163	1	0.5667163	24.311	0.0000^{****}
error		28	0.0233116		
$\mathrm{SNR_{in}}$(Female)	5.4237975	3	1.8079325	12.523	0.0000^{****}
$\mathrm{SNR_{in}}$(Male)	7.7039212	3	2.5679737	17.788	0.0000^{****}
error		42	0.1443648		

$+$ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

Table 2.5: Main effect of gender and noise type

Source	SS	df	MS	F	p
Gender(Babble)	0.0026637	1	0.0026637	0.072	0.7909
Gender(White)	0.2887990	1	0.2887990	7.766	0.0095^{**}
Gender(Highway)	0.1271632	1	0.1271632	3.420	0.0750^{+}
Gender(Railway)	0.7575038	1	0.7575038	20.371	0.0001^{****}
error		28	0.0371860		
Noise(Male)	0.4077380	3	0.1359127	2.802	0.0514^{+}
Noise(Male)	0.5631589	3	0.1877196	3.870	0.0157^{*}
error		42	0.0485043		

$+$ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

Table 2.6: Main effect of gender, noise type and SNR_{in}

Source	SS	df	MS	F	p
Gender × Noise (6 dB)	0.5224652	3	0.1741551	11.470	0.0000^{****}
Gender × Noise (0 dB)	0.0840108	3	0.0280036	1.844	0.1454
Gender × Noise (−6 dB)	0.0852132	3	0.0284044	1.871	0.1408
Gender × Noise (−12 dB)	0.0580554	3	0.0193518	1.275	0.2884
error		84	0.0151838		
Gender × SNR (Babble)	0.4789813	3	0.1596604	15.121	0.0000^{****}
Gender × SNR (White)	0.0323316	3	0.0107772	1.021	0.3877
Gender × SNR (Highway)	0.0657153	3	0.0219051	2.075	0.1097
Gender × SNR (Railway)	0.0596019	3	0.0198673	1.882	0.1389
error		84	0.0105590		
Noise × SNR (Female)	0.7931337	9	0.0881260	7.112	0.0000^{****}
Noise × SNR (Male)	0.5617598	9	0.0624178	5.037	0.0000^{****}
error		126	0.0123918		

$+$ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

2.2.3 子音特徴別分析

次に，JDRT の 6 子音特徴ごとに分析を行う．Fig. 2.11 から Fig. 2.18，Fig. 2.19 から Fig. 2.22 に各特徴の了解度と MCI を SNR_{in} ごとに騒音種別と話者性別に分けて示す．JDRT の基礎検討 [21, 104] や，英語版の結果 [64, 65] からも明らかなように，SNR_{in} と子音特徴の関係は 6 種類それぞれ異なる．またこれらの先行研究で明らかなようにスピーチノイズと白色雑音の間で傾向差が顕著な子音特徴がみられるかも検証する．最後にダイオティック系の評価の文献 [21] のとの違いをまとめることでバイノーラル系特有の課題を検討する．

Voicing

Voicing は有声音と無声音の聞き分けであり，単語対の聞き分けがしやすい子音特徴である．有声子音はエネルギーがあり，無声子音はエネルギーが小さい．つまり有声子音の方が低 SNR 騒音下でも聞き分けやすいと予測される．Fig. 2.11 の了解度は，SNR_{in} と騒音種によらず高い．MCI は，−12 dB では MCI が 10%前後とやや大きくなるものの，他の子音特徴と比ると MCI の増加は少ない．これは SNR_{in} の変化に対して平均了解度が高いため，被験者間の分散も小さくなり MCI は小さい値となる．

次に Fig. 2.11 の平均了解度，MCI 共に男女差が顕著で，女性の方が了解度が固く MCI が小さい．了解度の違いに関しては，文献 [21] における JDRT の基礎分析とも同傾向である．

Nasality

Nasality は鼻音と口音の聞き分けであり，Voicing と同様に明確な子音特徴である．鼻音は子音区間がわずかに長く，口音区間は短い．騒音種間の了解度差は Fig. 2.13 にみられるように，Babble が他よりも小さく，White が高い．この二種の間には約 20%の差がある．Highway と Railway はこの二つの間であるが，Railway は −12 dB で Babble と同程度に急峻に低下している．−6 dB までは Highway と同傾向であり騒音種による傾向差がある．文献 [21] でも Nasality は 0 dB 未満の低 SNR 環境で騒音種差がみられており，スピーチノイズ同様に 1 kHz にパワーが集中する Railway 騒音の特徴がみらえたものと考えられる．

男女差は Voicing より顕著であり，Fig. 2.14 の結果より −12 dB では了解度 20%の差がある．また，MCI も男女差が顕著であり，6 dB を除いて 5%程度の差がある．文献 [21] と比べても男女差が明確になっている．

Sustention

Sustention は子音スペクトルの継続性に関する聞き分けで，比較的明瞭である．また，JDRT の基礎検討 [104, 21] や，英語版の結果 [64, 65] においても，スピーチノイズと白色雑音の差が明確にみられた子音特徴である．騒音種間の了解度差は Fig. 2.15 にみられるように，White が明確に低い値を取り，Railway がそれに続く．Highway と Babble は同傾向であり，−12 dB で了解度 40%と騒音種間差は明確である．MCI も White が常に高く，それ以外の 3 種は騒音間差はないものの，Voicing，Nasality と比べて 5%程度高く，騒音の影響を受けやすい子音特徴であることがわかる．

男女差はFig. 2.16よりSNR_{in}が低下するほど差が徐々に開き，−12 dBでは10%程度の差がみられるが，Voicing,Nasalityほど顕著な差ではなく，MCIは男女差が全くみられない．これは文献[21]と比べると，SNR_{in}が低下するごとに男女差が開いていくことに違いがある．

Sibilation

Sibilationは波形の不規則性に関する分類であるが，あまり明確ではない．Fig. 2.17の了解度変化はVoicing同様，SNR_{in}同様にほとんど変化せず，Whiteが僅かに低下する程度である．MCI変化もWhiteは顕著に増加するものの，他はほとんど変化していない．また，Fig. 2.18より，男女差もみられない．これも文献[21]と同傾向である．

Graveness

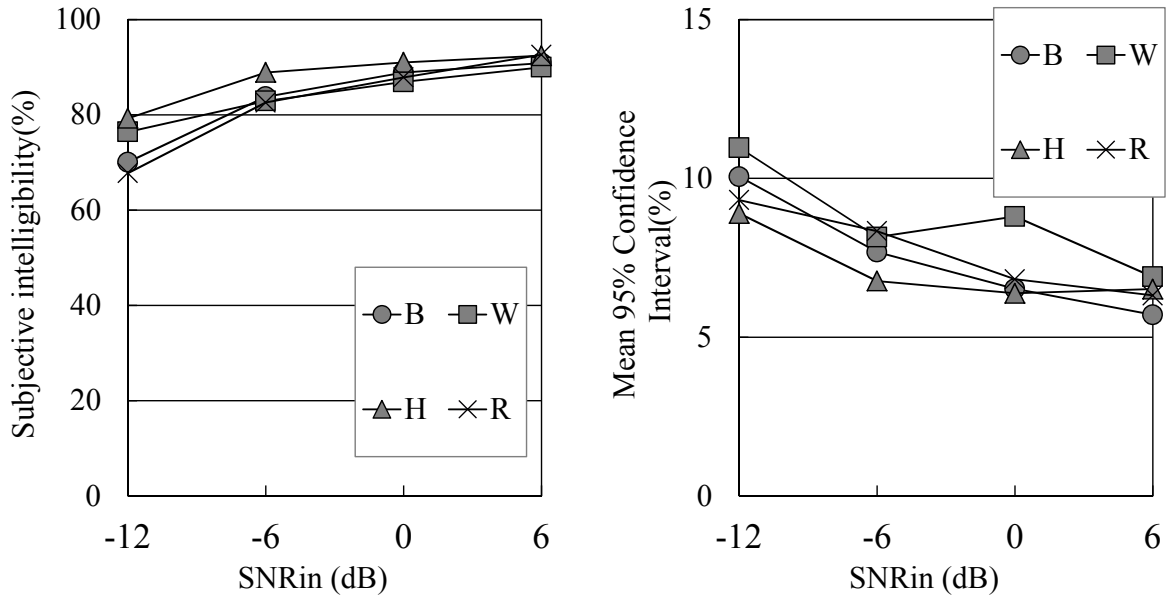
Gravenessは抑音と鋭音の聴き分けである．Fig. 2.19の了解度変化はSustentionと同様に，SNR_{in}の減少に合わせて了解度が減少する．騒音種の違いは，Whiteがその他3種よりも0 dB以下で常に10%以上低い．MCI変化は騒音種の違いは少ないものの，0 dBから−6 dBにかけてHighwayの変化が急峻である．文献[21]の結果と比べると，騒音種による違いがほとんどみられなくなった．

次に，Fig. 2.20より男女差は了解度，MCI共に男女差は無い．文献[21]ではわずかに男女差がみられたが，本論文では異なる結果が得られている．文献[21]の結果との違いを考察する．本実験では主音声にもHRIRを畳み込んでいるため，声質が変換されている．Fig. 2.6のHRTFスペクトルから明らかなように，2 kHz周辺のゲインが最も大きいため，Gravenessの特徴である第2フォルマント周波数周辺の聴取が改善されたものと考えられる．このため，文献[21]での音声部に何も処理をしていない実験と比較し，騒音に対する頑健さが改善していると考えられる．

Compactness

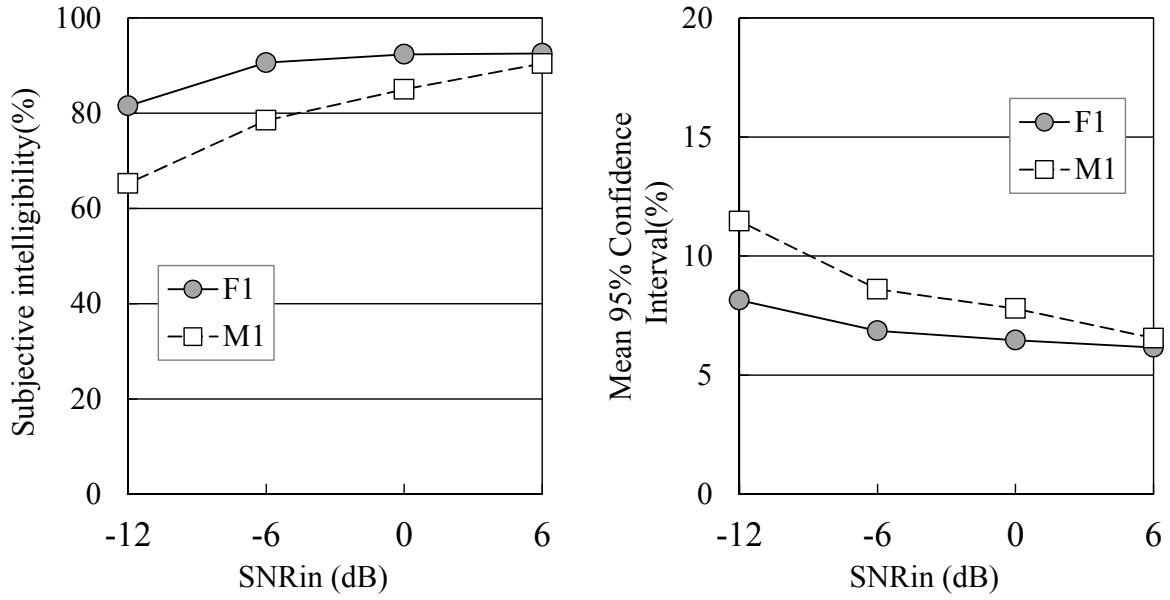
Compactnessはスペクトルのエネルギーの特定周波数への集中がある音とそうでない音の分類である．Fig. 2.21より，了解度の騒音種間の差は少ないものの，Highway騒音は他の騒音よりも了解度低下が緩やかになり，−12 dBで他よりも10%高い．MCIはHighwayとRailwayの2種が−12 dBと−6 dBの変化が大きい．男女差を見てもFig. 2.22のMCIもHighway，Railwayは，Babble，Whiteに比べ変化が急峻になっており，環境騒音の影響が出やすい子音特徴であるといえる．

男女差については平均了解度，MCI共に大きくはないが，他の特徴と異なり，男性話者の方が了解度が高い．文献[21]では男女差が全くみられなく傾向差がある．この傾向差については話者数を増やして同様の実験を行い，詳細を検証する必要がある．



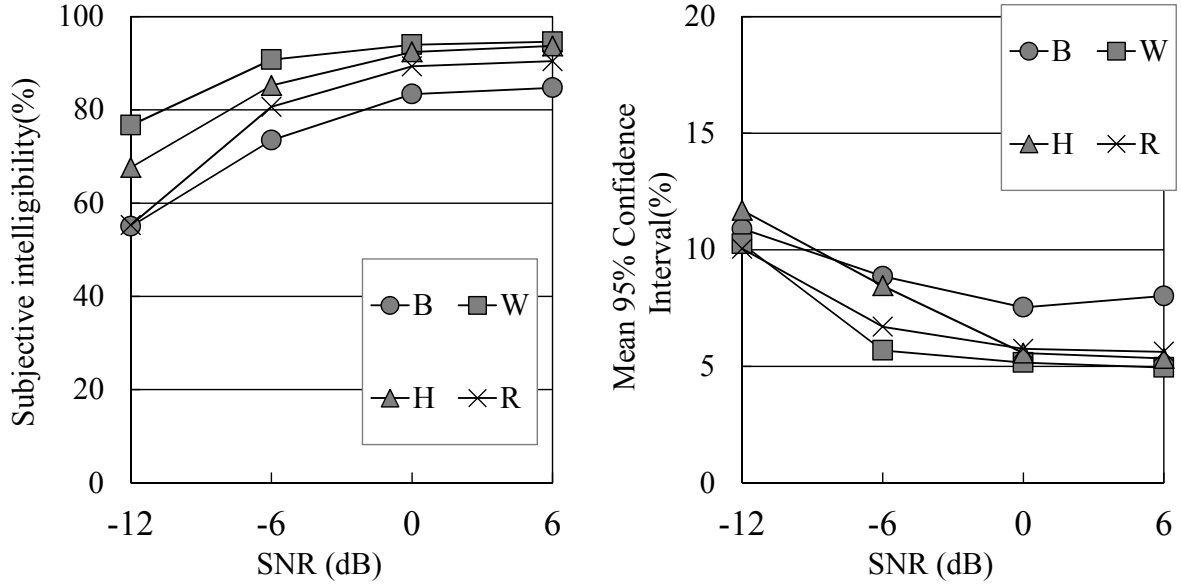
(a)Intelligibility (b)MCI

Fig. 2.11: Comparison of intelligibility and MCI with various noise type (Voicing)



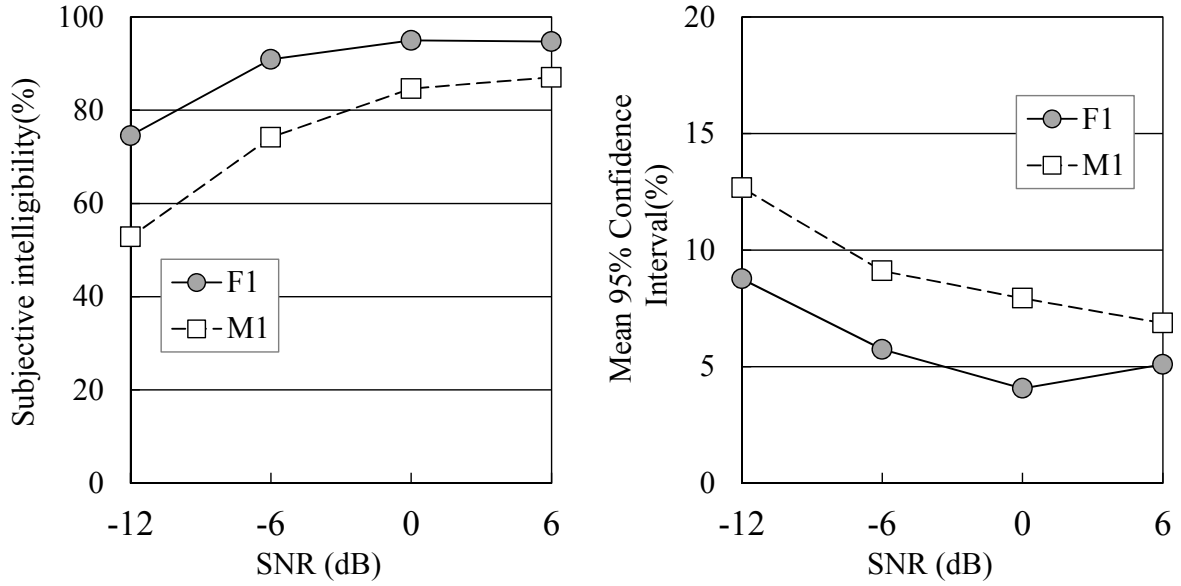
(a)Intelligibility (b)MCI

Fig. 2.12: Comparison of intelligibility and MCI with gender (Voicing)



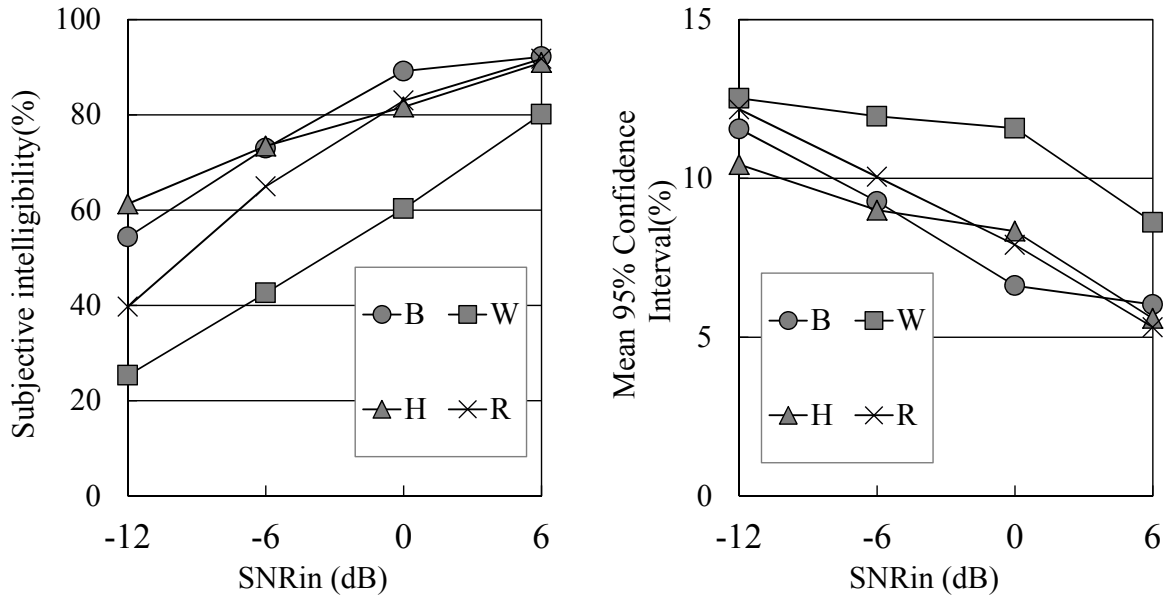
(a)Intelligibility (b)MCI

Fig. 2.13: Comparison of intelligibility and MCI with various noise type (Nasality)



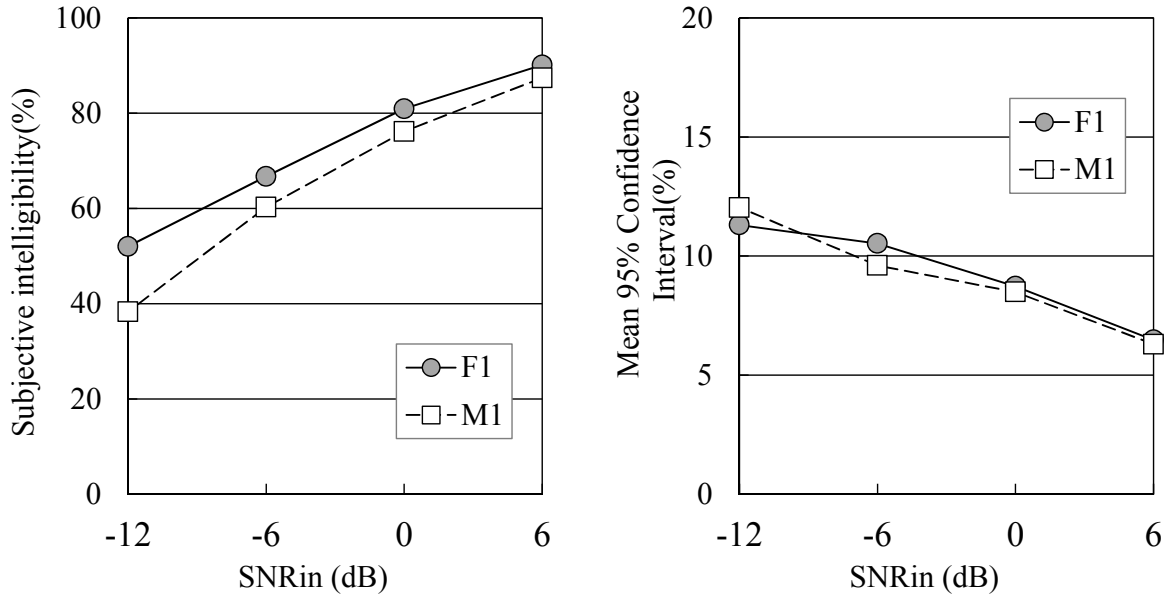
(a)Intelligibility (b)MCI

Fig. 2.14: Comparison of intelligibility and MCI with gender (Nasality)



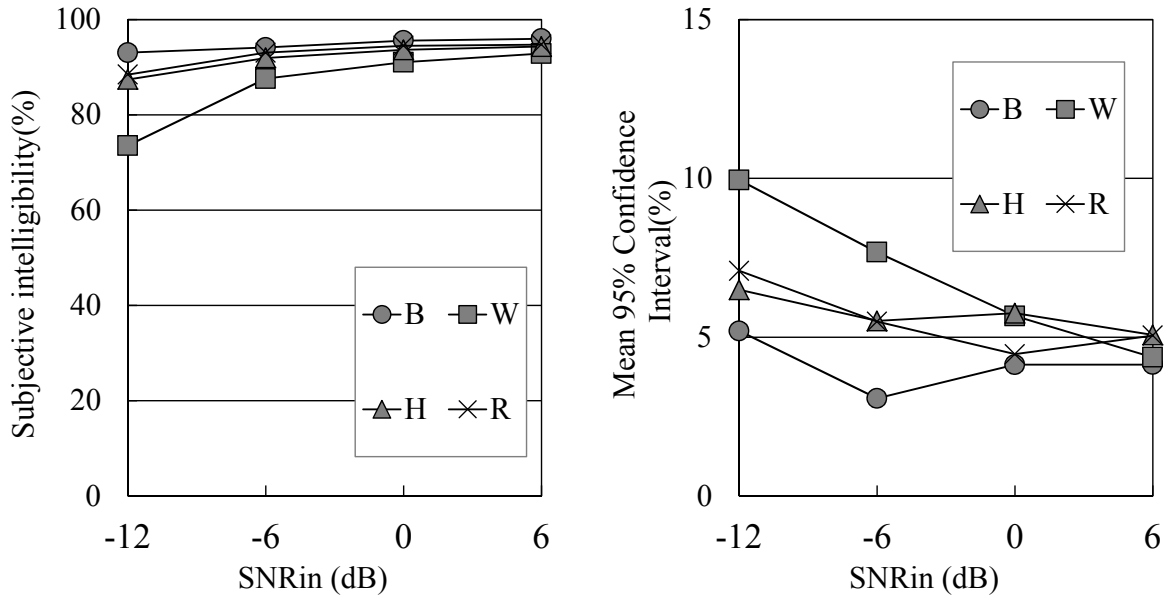
(a)Intelligibility (b)MCI

Fig. 2.15: Comparison of intelligibility and MCI with various noise type (Sustention)



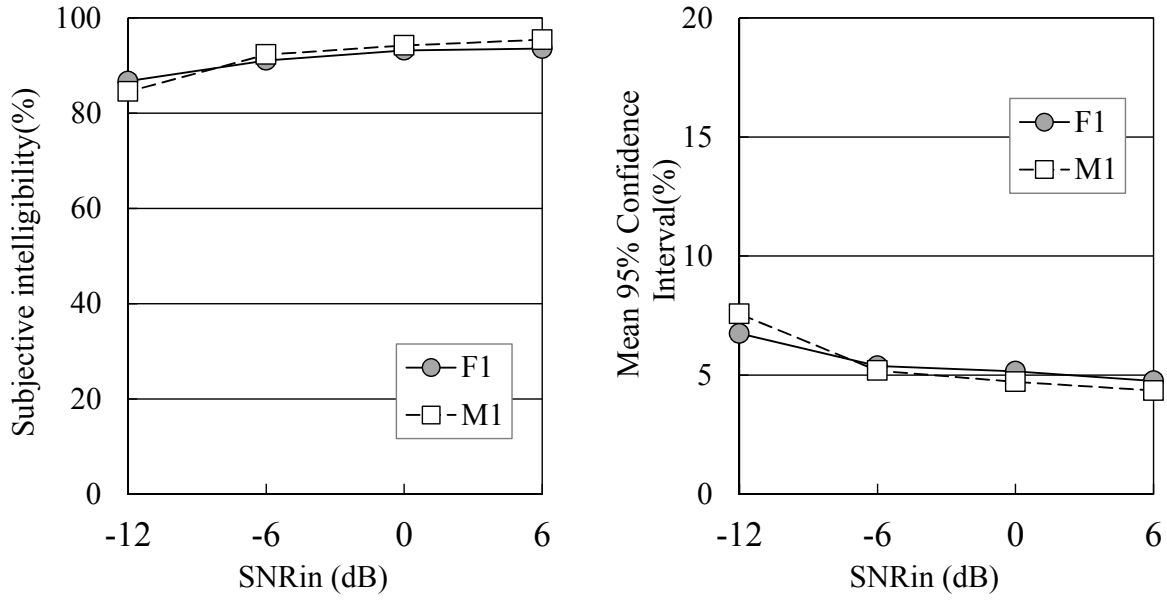
(a)Intelligibility (b)MCI

Fig. 2.16: Comparison of intelligibility and MCI with gender (Sustention)



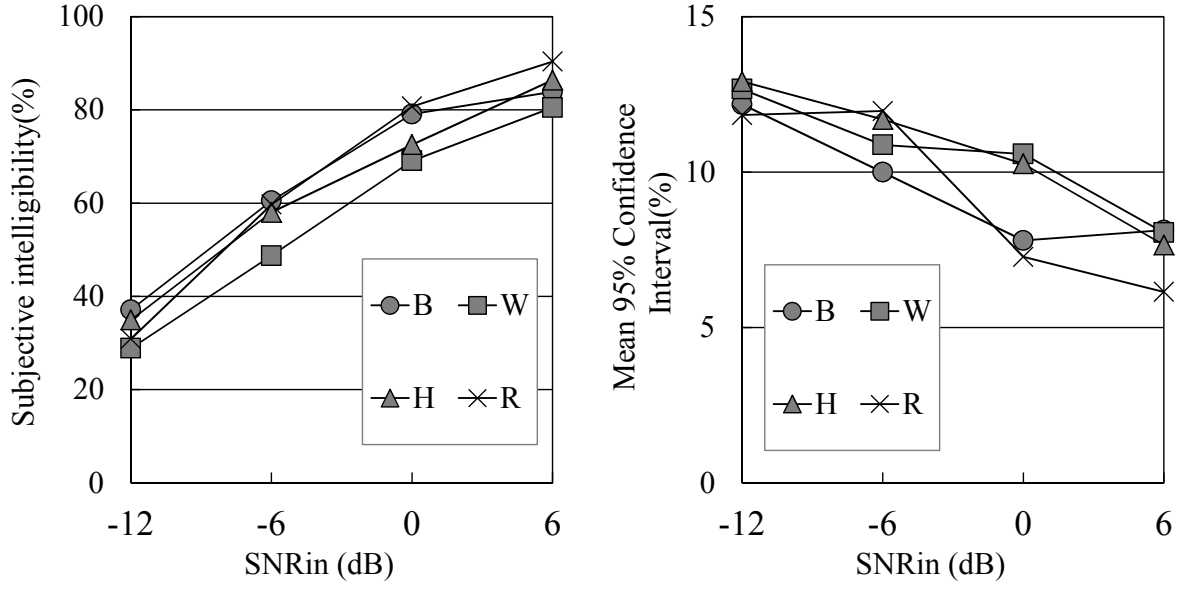
(a)Intelligibility (b)MCI

Fig. 2.17: Comparison of intelligibility and MCI with various noise type (Sibilation)



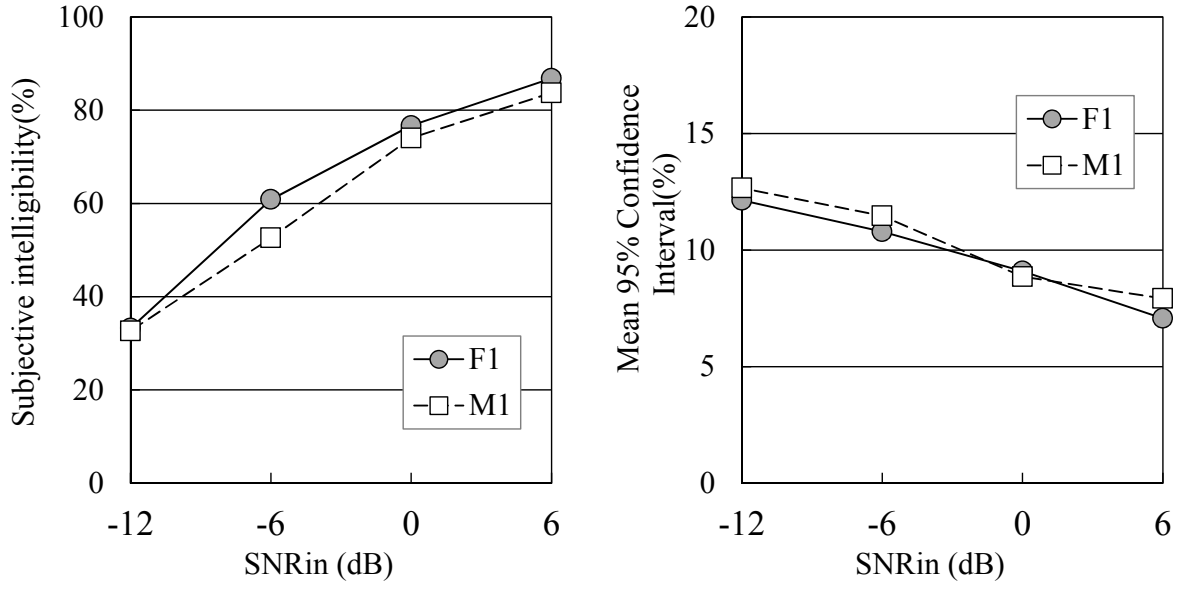
(a)Intelligibility (b)MCI

Fig. 2.18: Comparison of intelligibility and MCI with gender (Sibilation)



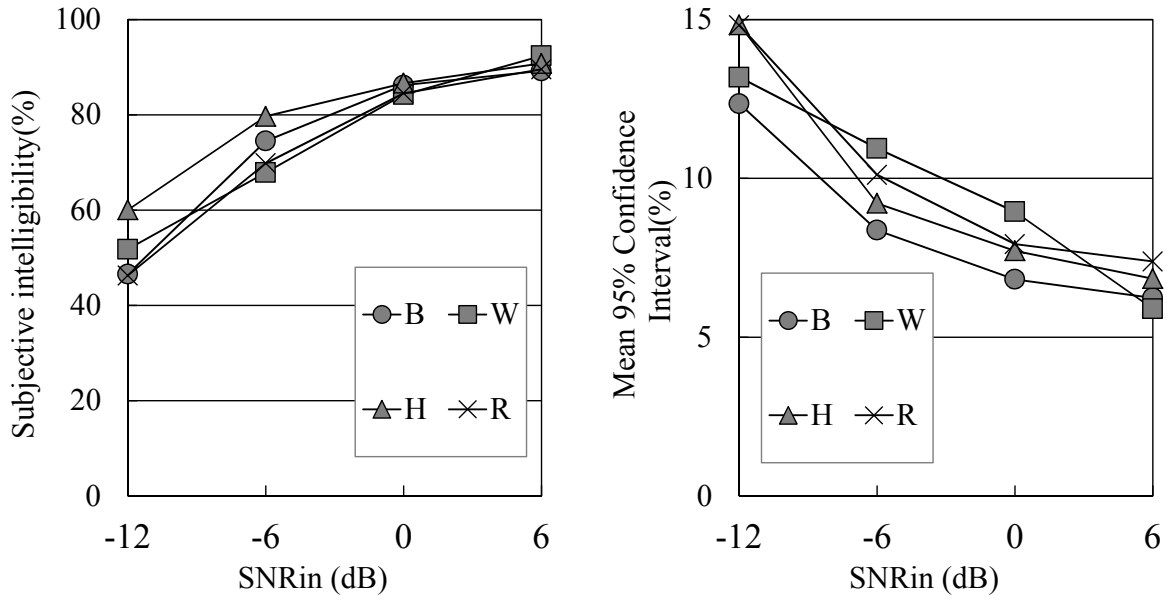
(a) Intelligibility (b) MCI

Fig. 2.19: Comparison of intelligibility and MCI with various noise type (Graveness)



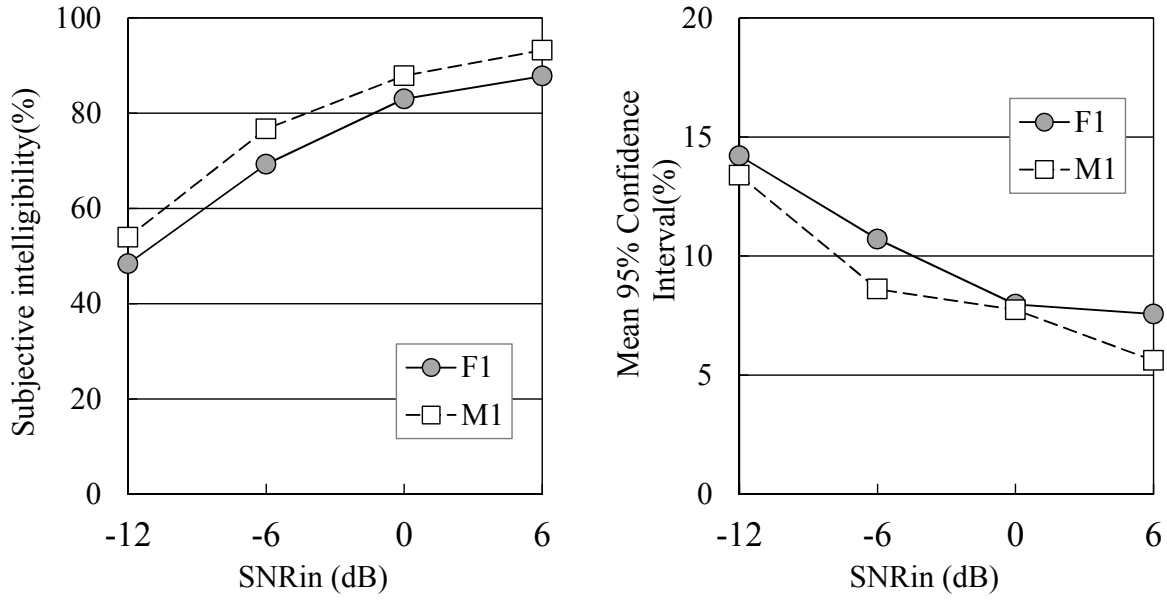
(a) Intelligibility (b) MCI

Fig. 2.20: Comparison of intelligibility and MCI with gender (Graveness)



(a) Intelligibility (b) MCI

Fig. 2.21: Comparison of intelligibility and MCI with various noise type (Compactness)



(a) Intelligibility (b) MCI

Fig. 2.22: Comparison of intelligibility and MCI with gender (Compactness)

2.2.4 先行研究との比較

Table 2.7 に文献 [21] と本論文の主観評価結果との騒音種と話者性別の結果の比較を行う．表中の “+” は各図より明らかな差がみられたもの，“–” は各図より明らかな差がみられないもの，それ以外は，顕著な特徴を持つ要素を記載した．low SNR は 0 dB 未満の SNR での傾向を指す．繰り返しになるが，文献 [21] は主音声に HRIR を畳み込まないダイオティック系での実験であり，騒音種，実験 SNR，話者数[12]が異なる．

結果より，本実験は雑音種に関して白色雑音が得意な傾向になりやすい．また，話者性別差の影響は文献 [21] と比較すると出易い傾向にあるが，話者性別の差が，$\mathrm{SNR_{in}}$ によって序列が入れ替わるのは，Sibilation の −12 dB のみであり，男女の主観評価値の平均値を了解度とみなしても実用上の問題は少ないと考えられる．

次節以降，了解度の推定を行っていくが，子音特徴による傾向差が主観評価によって既に明らかになっているため，AI，STI，SII の様に全明瞭度平均，全了解度平均を推定するだけでなく，子音特徴ごとの推定も検討していくこととする．また，騒音種については傾向差を見るために 4 種類ごとに検討し，話者差については男女平均値を用いることで代表値とする．騒音方位角と $\mathrm{SNR_{in}}$ については同一の騒音種内の個別の騒音とみなし，1 騒音種に 32 音（$\mathrm{SNR_{in}}$4 種 × 方位角 8 種）あることとする．

Table 2.7: Comparison with previous studies

	previous work		this paper	
phonetic feature	noise type	speaker gender	noise type	speaker gender
120 words	+	−	+	+
voicing	−	+	−	+
nasality	low SNR	low SNR	+	+
sustention	white	−	white	−12 dB
sibilation	low SNR	−	white	−
graveness	+	+	white	−
compactness	low SNR	−	ambient noise DB	+

+ : remarkable difference, −: not remarkable, other: remarkable factor

2.3 了解度と客観音声品質評価法との相関分析

本節では，前節の結果を推定するために，1.2.5 項で述べた客観音質評価尺度と PESQ の中から，了解度推定に適した品質評価尺度を検討する．まずは尺度ごとの音質値と了解度の相関分析を行い，了解度と相関の高い尺度を選択し，推定実験に用いる尺度を決定する．

2.3.1 使用する音声品質尺度

今回実験で用いる客観音質評価尺度は以下の 16 指標である．SNR から AIseg までが評価値と音質値に正の相関があるものであり，d_{Cep} から SNRloss(S) までが，負の相関を持つものである．

[12]本実験で用いた話者 2 名を含んでいる．

各指標で音質値を求める際，PESQ を除くと音声区間検出（Voice Activity Detection: VAD）を含んでいない．STI の様な空間伝搬を考慮すると，VAD を用いた正確な音源位置特定は不可欠である．しかし，本章で検討している音声システムは，提示音声と外部騒音を同時に録音でき，VAD による音声位置検出が決定的な性能差とはなりにくい．このため，音声区間は既知であるとして，2.3.2 項で説明する各指標のベターイヤースコアを用いることとする．

また，音声品質評価尺度は元来，音声帯域（遮断周波数 3.4 kHz ないし，7 kHz）での評価のために設計されている．そのため，主観評価は KEMAR–HRIR に合わせて 44.1 kHz のサンプリングレートにアップサンプルして音源を作成したが，音声品質については各尺度に合わせてダウンサンプリングを行った．特に明記していない場合はサンプリングレート 16 kHz とした．

SNR　本実験では，主音声と騒音のレベル統制を $\mathrm{SNR}_{\mathrm{in}}$ として統制した．$\mathrm{SNR}_{\mathrm{in}}$ は JDRT の 120 単語の 2 話者の 240 単語平均パワーとノイズのパワーが等しくなる値を 0 dB とし，そこからゲインを調整し −12，−6，0，6 dB を設定している．音質指標として用いる SNR は 240 単語それぞれの SNR を求め平均する．このため，単語ごとに式（1.4）を計算し，平均とすることから，SNRseg と同様，特定のパワーが強い単語に全体が影響されることが無くなる．このため，単語ごとの SNR 平均を比較する指標の一つとした．

SNRA　SNR は周波数ごとの重みづけの無い指標である．そこで聴覚重みとして騒音計などで利用される A 特性重みを用いた SNR を求めた．A 特性は，1.2.2 項で述べた等ラウドネス曲線の 40 phon の逆特性であり，人間の音量感に近い値が得られると考えられる．

SNRseg　セグメンタル SNR は，式（1.5）で実装した．分割フレーム長は 20 msec とした．実装は文献 [11] のソースコードに基づき，最大値が 35 dB，最小値が −10 dB とした．

fwSNRseg(A)　SNRA 同様に A 特性を用いて式（1.11）の fwSNRseg を使用した．分割フレーム長は SNRseg と同様に 20 msec とした．セグメントに分割した音声に A 重みを乗じて，文献 [11] のソースコードに基づき，最大値が 35 dB，最小値が −10 dB として SNRseg を求めた．

fwSNRseg(C)　fwSNRseg(C) は，文献 [113, 123] で検討された，SII の重みを改良し，25 のクリティカルバンドごとに求めた重みのうち，子音向けのものを利用した fwSNRseg を用いた．子音重みと文章重みを Fig. 2.23 に示す．重みと分割帯域以外は SNRseg と同様に分割フレーム長は 20 msec とした．また，文献 [113, 123] と同様に各帯域の SNR 値は −15 dB から 15 dB に制限した．

fwSNRseg(S)　fwSNRseg(S) は Fig. 2.23 の文章重みを用いた fwSNRseg で，重みと分割帯域以外は SNRseg と同様に分割フレーム長は 20 msec とした．文章重みは子音重みよりも低い帯域を重視しており，母音の聞き取りに配慮した重みである．文献 [113, 123] と同様に各帯域の SNR 値は −15 dB から 15 dB に制限した．

AI　AI は式（1.12）で実装した．クリティカルバンドは Table 1.7 を用いた．定義式中の R は ANSI の当初の基準通り 30 dB とし，S は 0 とした．本論文の実験系では，SNR が 0 dB 以下にな

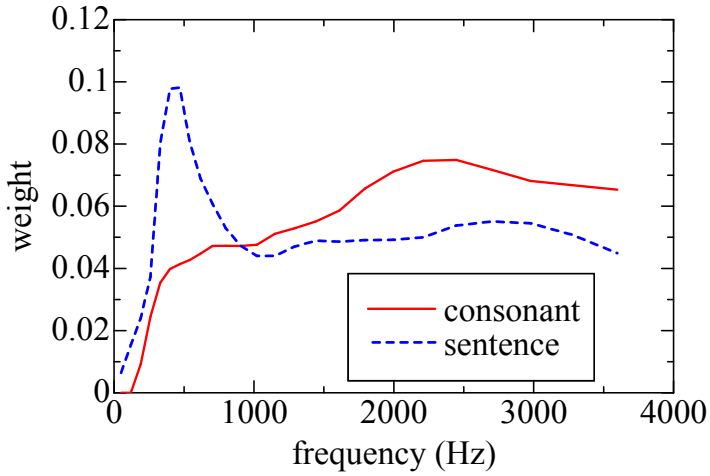


Fig. 2.23: Band–importance functions

ることが多いため，出力値は 0 以下となる．一方で，他の指標との比較に後述する正規化処理を行うこと，推定に用いる場合は了解度への変換関数を回帰によって求めることを考慮すると，AI の出力値を 0 から 1 のインデックスとして利用せず，そのまま用いても支障がないため，本論文では求まった AI 値をそのまま用いることとした．

AIseg　AIseg は式（1.12）中の SNR 計算部を SNRseg に置き換えた指標であり，今回の評価に合わせて筆者が作成した．SNRseg 同様に，フレーム長は 20 msec とし，AI と同じ Table 1.7 のクリティカルバンドを用いた．出力値に対する扱いも AI と同様である．

PESQ　PESQ は広帯域音声用の P.862.2 を用い，P.862.1 の変換式を用いて MOS–LQO を求めた．PESQ は元々音声受聴品質の推定方式だが，DRT の主観品質の推定 [149] や雑音抑圧音声の推定 [100] においてある程度有効なことがわかっている．

$\mathbf{d_{Cep}}$　d_{Cep} は式（1.10）の LPC ケプストラム距離を用いた．LPC 係数の次数は 16 次とし，分割フレーム長は 20 msec とし加算平均値を用いた．

$\mathbf{d_{LAR}}$　d_{LAR} は式（1.8）の LAR 距離を用いた．ただし，反射係数を LPC を用いたスペクトルから求めた．この際の LPC 係数の次数は 16 次とした．音声の分割フレーム長は 20 msec とし加算平均値を用いた．

$\mathbf{d_{LLR}}$　d_{LLR} は式（1.7）を用いた．実装は [11] のソースコードを用いたため，サンプリング周波数を 8 kHz にダウンサンプリングしてから用いた．分割フレーム長は 30 msec である．

$\mathbf{d_{IS}}$　d_{IS} は式（1.6）を用いた．実装は [11] のソースコードを用いたため，サンプリング周波数を 8 kHz にダウンサンプリングしてから用いた．分割フレーム長は 30 msec である．

dWSS　d_{WSS} は式（1.9）を用いた．実装は [11] のソースコードを用いたため，サンプリング周波数を 8 kHz にダウンサンプリングしてから用いた．また，傾斜を求めたスペクトルは LPC 係数 10 次で求めた．分割フレーム長は 30 msec である．

SNRloss(C)　SNRloss は，J. Ma らが提案している雑音抑圧音声の品質評価指標で，原音と，劣化信号を信号処理して求めた雑音抑圧音声の SNR から了解度を予測する指標である．SNRloss は距離尺度と同様に音質とスコアに負の相関があり，0 から 1 の値を取る．また，計算時の重みに Fig. 2.23 の子音または文章重みを用いる．SNRloss(C) は子音重みを用いた場合である．これ以外の設定は J. Ma らが公開しているソースコードの設定に従った．

SNRloss(S)　SNRloss(S) は，SNRloss(C) と同様に，Fig. 2.23 の文章重みを用いた音声品質指標である．

2.3.2　評価信号と正規化品質

評価信号の設定

客観評価信号には主観評価音源である騒音が重畳された劣化信号[13]と，JDRT コーパスの原信号を用いる．後述する各品質評価値は JDRT の 120 単語ごとに求め，120 単語全平均の推定には前スコアの平均値，各子音特徴ごと推定では特徴ごとの 20 単語の値を用いる．

前節の結果では，発話者の性別によって主観評価値に傾向差がみられる子音特徴があった．しかし，Fig. 2.16 の Sustention での −12 dB を除けば[14]，$\mathrm{SNR_{in}}$ の値によって男女差が入れ替わることが無いことを加味して，両話者の平均了解度と音質値の加算平均値を用いることとする．

ベターイヤースコア

主観評価音源は両耳受聴に基づくシステムを対象としたため，左右の両耳に提示する騒音の音圧は，定位した方位ごとに異なる．一方で評価する了解度は一次元の尺度であるため，音質値は左右別に求まる．文献 [22] において Beerends らは，PESQ を用いて，両耳受聴系の音声了解度を推定するために，スコアの良い方の耳側のスコアを採用するベターイヤースコアによる推定を行っている．本論文でも両耳の音声品質値の代表値としてベターイヤースコアを用いる．音声品質尺度によって，評価値と音質値に正の相関があるもの（SNR，PESQ 等）と負の相関があるもの（例，d_{IS} 等の距離指標）があるため，各指標に合わせて音質の良いスコアを選択する．

正規化了解度と正規化品質値

16 品質指標のうち，JDRT の各子音特徴了解度ごとに最適な音声品質尺度を検討する．このため，音声品質尺度間を正当に比較できるように音質値の正規化を行う．各音声品質尺度によって求まった音質値の正規化には以下の式（2.1）と式（2.2）を用いる．式（2.1）は評価値と音質に正の相関がある指標に用い，式（2.2）は負の相関がある場合に用いる．この変換により，評価値の

[13]原音信号との比較を行うため，品質評価の観点から，騒音が付加された劣化信号とみなす．
[14]Sustention は全体的に了解度が高く，−12 dB の男女差も了解度 2%程度である．

中で最も音質が悪い値が 0 となり，最も音質が良い値が 1 となる．また，主観評価実験で求めた了解度も式（2.1）で正規化する．JDRT は前節の結果から明らかなように，同一の $\mathrm{SNR_{in}}$ でも，子音特徴ごとに了解度値の分布が異なり，Voicing や Sibilation の様に高い値に飽和する特徴もあれば，Sustention，Comactness の様に広く分布する特徴もある．よって，子音特徴ごとの変化の最小値と最大値を 0 から 1 に正規化する．この変換により，尺度ごとの単位が無次元化され，散布図による指標間の比較が視覚的にも行いやすい[15]．

指標間の比較は式（2.3）のピアソンの積率相関と式（2.4）ケンドールの順位相関 [159] で行う[16]．式（2.3）の R はピアソンの積率相関係数，$I(n)$ は正規化了解度 $Q(n)$ はサンプル番号 n における正規化音質値で，$\overline{I}$，$\overline{Q}$ はそれぞれの加算平均値である．同じ平面において，了解度と客観音質値が対応が良ければ（十分に線形な関係であれば），原点を通る対角線上にプロットされ，相関係数は 1 となる．つまり相関係数の大小で客観音質指標の比較が可能となる

一方，ピアソンの積率相関計数は正規分布を仮定したパラメトリックな分析であるため，了解度の様な 0 から 1 の範囲のみに分布する分布では，天井／フロア効果といった最大値と最小値近傍では分布の正規性が仮定できず，ピアソン相関との対応が悪いとされる．2.2 節の主観評価結果より天井／フロア効果は顕著に発生していないと考えられるが，全体的に評価値が 50%より高く，80%〜90%程度のサンプルが多いため，ノンパラメトリックな相関分析手法であるケンドールの順位相関計数での比較も必要である．式（2.4）の τ がケンドールの順位相関計数，P は 2 変数間の大小関係が一致する組数，n は総サンプル数である．順位相関も正規化了解度と正規化音質値を用いて求める[17]．そして，両相関係数における最良尺度および，特徴的な傾向を持った尺度を選択し，了解度推定を行うこととする[18]．本論文では，以後，ピアソンの積率相関係数を R，ケンドールの順位相関係数を τ と呼ぶこととする．

$$\text{normalized objective score} = \frac{\text{raw score} - \text{min score}}{\text{max score} - \text{min score}} \tag{2.1}$$

$$\text{normalized objective score} = \frac{\text{raw score} - \text{max score}}{\text{min score} - \text{max score}} \tag{2.2}$$

$$R = \frac{\sum (I(n) - \overline{I})(Q(n) - \overline{Q})}{\sqrt{\sum (I(n) - \overline{I})^2} \sqrt{\sum (Q(n) - \overline{Q})^2}} \tag{2.3}$$

$$\tau = \frac{2P}{\frac{1}{2}(n-1)} - 1 = \frac{4P}{n(n-1)} - 1 \tag{2.4}$$

[15] この他に計測値から平均値を引き，標準偏差で割る標準化も一般に利用される．

[16] ケンドールの順位相関はピアソンの積率相関と導出方法が異なり，導出法の異なる相関係数の値は比較できない．しかしながら，係数の順位の比較は行える．よって「ピアソンの積率相関は全尺度の中でも高いが，ケンドールの順位相関が全尺度の中で低い」といった比較を行う．

[17] どちらの相関係数も正規化処理を行わない場合と絶対値は変わらない．

[18] 文献 [150] における Liu らの検討においても，両相関係数は若干異なる傾向を示しており，一般に広く使われる積率相関だけでなく順位相関の比較が必要である．

2.3.3　了解度と音声品質の相関

主観評価によって求めた了解度と音声品質値との相関を議論する．前項で取り上げた音声品質尺度 16 種を比較し，騒音種と JDRT の子音特徴ごとに最適な音声品質尺度を比較する．比較した結果，相関が高かった尺度を用いて次節で了解度の推定を行う．本稿で着目する騒音種の影響による了解度変化を推定するために，品質尺度と了解度の相関関係を以下の 2 パターンに分けて考える．

- 全 4 騒音を混合した相関
- 騒音種ごとに分けた相関

全騒音混合条件で相関が高ければ，騒音種に関わらず了解度を推定することが可能な尺度と考えられる．これは最も望ましい尺度であるが，1.2.7 項で述べたように,「音声を最も妨害するのは音声」という経験則から環境騒音等と同一のパワーを持つ音声の方が明瞭度・了解度の低下は大きい．このため次善の基準として，騒音が既知であれば出せうる最大性能として騒音種ごとに分けた時の相関も議論する．

2.3.4　全騒音混合条件のピアソンの積率相関とケンドールの順位相関

まずは全騒音混合条件での 120 単語平均について詳細に述べ，次に子音特徴ごとに相関係数の一覧表を基に特徴のある尺度について述べる．

120 単語平均

全騒音混合時の 120 単語平均の積率相関係数 R を 16 指標ごとに Table 2.8 に示す．また，16 指標ごとの正規化了解度と正規化音質値の散布図を Fig. 2.24(a)〜(p) に示す．図中では騒音種ごとにプロットを変えてあるが，相関係数は全ての騒音種を混合して求めてあり，正規化に用いる最大値，最小値は全評価値から求めた値を用いた．

Table 2.8 の結果より，表上段の SNR を用いた尺度（SNR，SNRA，SNRseg，fwSNRseg(A)，fwSNRseg(C)，fwSNRseg(S)，AI，AIseg，以下混合して SNR 系尺度と呼ぶ）はどれも 0.8 以上の高い相関を持つ．この中で最も積率相関が高いのは SNRA で，それにわずかに劣る SNRseg と続く．以下，個別の傾向を分析する．SNR，SNRA，SNRseg の三種はほぼ同傾向で，了解度，音質値がどちらも 0.5 以上では相関が高いものの，0.5 未満の開きが大きい．fwSNRSeg(A) も同様に音質が 0.5 未満は開きがあり，0.5 以上では了解度の方が音質要理やや値が大きく，プロットはやや上に凸である．これら 4 種は White と Babble の傾向差が大きく，Highway と Railway はその間のスコアを取る傾向にある．一方で同一騒音のプロットはよく序列化されており，これら 4 種は騒音種が限定される環境では了解度推定に有利なものの，騒音が未知の環境には不利であると予想される．fwSNRseg(C) と fwSNRseg(C) は客観音質が 0.5 未満の開きは小さいが，fwSNRseg(C) はやや上に凸のプロット傾向があり，fwSNRseg(S) は直線的である．AI と AIseg も客観値 0.5 未満での了解度との開きがみられるが，プロットの凸傾向はみられない．上下関わらず，凸傾向がみられるものは音質尺度由来の要因であれば本節で比較したい最適尺度選択の問題となり，了解度の主観評価値が要因であればそもそも線形性が仮定できないと考えられるため，R による評価

では不十分である．つまり，了解度と音質の間に線形性があるのか，非線形性があるのかの検討が必要である．このため，他の子音特徴も含めこの非線形性がみられるか了解度と音質の関係を検証する必要がある．

PESQ は 0.789 とある程度の相関がある．これは音質値 0.5 未満の騒音種間差が大きいため，相関が悪くなったことが原因である．d_{Cep}，d_{LAR}，d_{LLR}，d_{IS} のスペクトル距離 4 種はどれも全体の R が 0.5 未満であるが，騒音種ごとに見た場合，序列は維持されているため，騒音種が特定の環境での了解度推定に利用できると考えらえる．これらの尺度は元来，音声符号化や合成音声といった，劣化の少ない音声の品質評価尺度であるため，本実験で加算した音声よりもパワーのある騒音環境の音声品質を十分に表現できないと考えられる．d_{WSS} についてはスペクトル距離の中では相関が高いが，SNR 系と比べると相関が悪いのは，Babble に関して他の騒音と明確に異なる傾向があるためである．SNRloss2 種は SII の重みを利用しているものの，相関は悪い．これは，本来の用途と異なり，雑音抑圧処理を行っていない音声を評価したためと考えられる．

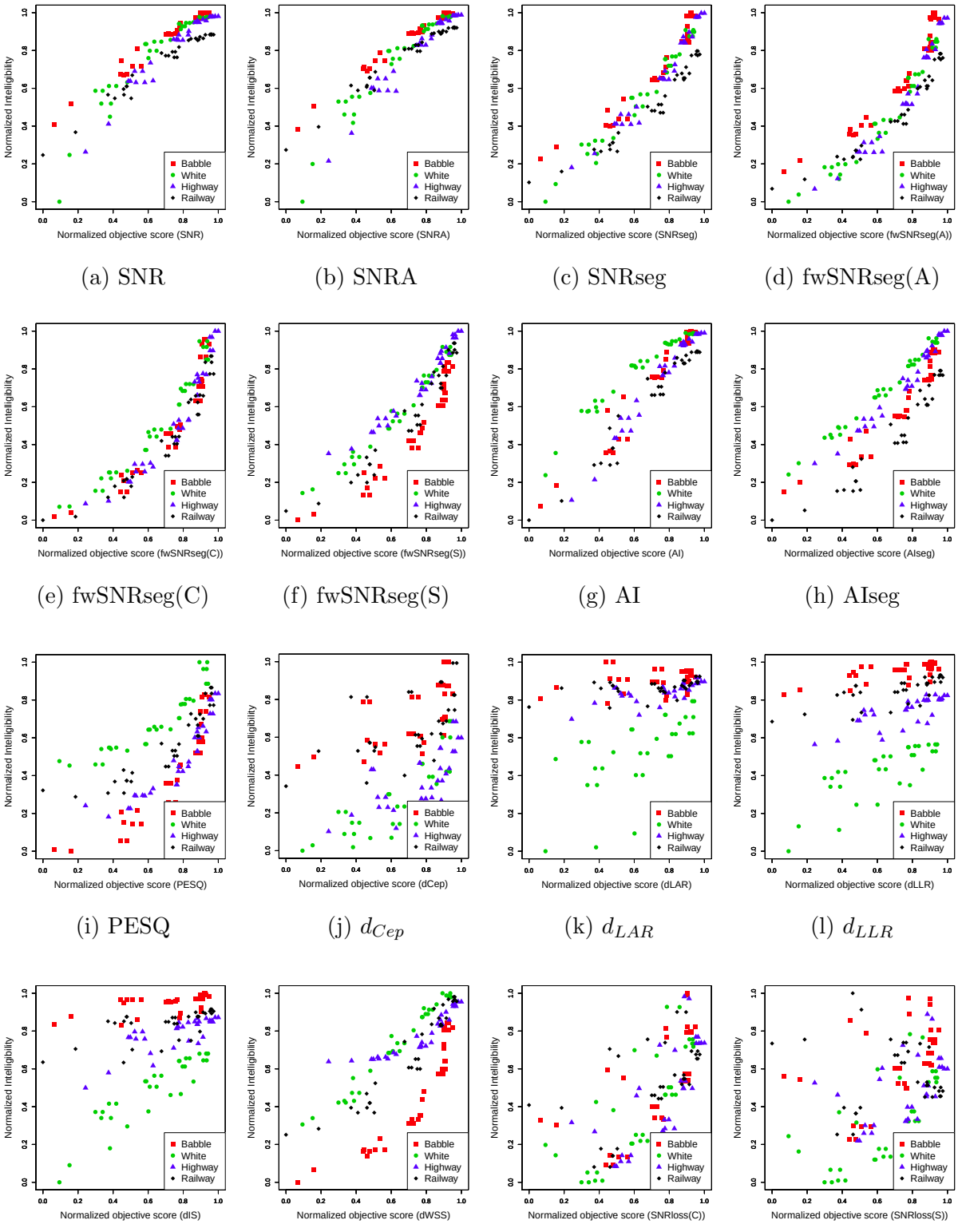
最後に Table 2.9 にケンドールの順位相関係数 τ の結果を示す．R と異なり fwSNRseg(C) が 0.834 と最もよく，fwSNRseg(A) が次いで 0.805 で，SNRseg が 0.8 である．τ が 0.8 を超えるのはこの 3 種のみであり，他の SNR 系尺度よりも明らかに高い．PESQ，d_{WSS} は 0.6 程度と SNR 系に次ぐ順位であるのは R と変わらない．R と最も異なるのは，セグメンテーションを伴う SNR 系尺度の方がそれ以外の SNR 系尺度よりも相関が高いという結果になったことである．これは，セグメンテーションを伴わない尺度にみられたプロットの下に凸の傾向が客観音質評価法由来のものであり，了解度と音質値の関係は線形ではなく，積率相関による比較が困難であることを示唆している．また SNRseg，fwSNRseg(A)，fwSNRseg(C) で特に顕著なやや上に凸の傾向は了解度の天井効果によるものであり，了解度と音質値の非線形性が示唆される．

Table 2.8: Pearson correlation(R) between normalized intelligibility(120 words average) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.932	0.942	0.934	0.922	0.922	0.913	0.914	0.866
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.713	0.494	0.378	0.404	0.493	0.773	0.697	0.416

Table 2.9: Kendall rank correlation(τ) between normalized intelligibility(120 words average) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.746	0.782	0.800	0.805	0.834	0.759	0.726	0.699
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.577	0.358	0.268	0.257	0.299	0.609	0.564	0.300



(a) SNR (b) SNRA (c) SNRseg (d) fwSNRseg(A)

(e) fwSNRseg(C) (f) fwSNRseg(S) (g) AI (h) AIseg

(i) PESQ (j) d_{Cep} (k) d_{LAR} (l) d_{LLR}

(m) d_{IS} (n) d_{WSS} (o) SNRloss(C) (p) SNRloss(S)

Fig. 2.24: Comparison between normalized intelligibility(120 words average) score and normalized objective speech quality score

Voicing

Table 2.10 に Voicing の R 一覧を示す．Voicing の了解度と音質の相関については 120 単語平均とほぼ同傾向で，SNR 系の指標の相関が高いが，全体的に R の値は小さくなっている．これは，Fig. 2.11 にみられるように，正規化前の了解度の範囲が狭いためである．Voicing の相関分析の代表例として，Fig. 2.25 に相関の高い SNR と AI，特異な傾向がみられた例として fwSNRseg(C) と PESQ の結果を示す[19]．SNR は音質値が 0.2 以下に主観値との対応が悪い点があるが，AI はほぼ対角線上にプロットされている．AI は式（1.12）の定義より，内部計算に SNR を用いているため，この違いは Table 1.7 の帯域分割による貢献である．fwSNRseg(C) は 120 単語平均と同様に上に凸の傾向にあり，了解度と音質値に非線形性があることが示唆される．また，PESQ の結果では，騒音種による傾向差が顕著であり，騒音ごとの順序尺度として PESQ は有効であるが，混合条件には向かないことを示唆している．Table 2.11 に Voicing の τ の結果を示す．120 単語平均同様，SNR，SNRA と AI は R では上位の尺度となるものの，τ ではセグメンテーションを伴う指標の方が相関が高い．この結果からも了解度と音質の非線形性が示唆され，セグメンテーションを伴う指標の了解度推定性能の高さが予想される．

Table 2.10: Pearson correlation(R) between normalized intelligibility(Voicing) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.878	0.870	0.822	0.797	0.806	0.808	0.860	0.817
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.564	0.203	0.288	0.239	0.453	0.654	0.564	0.472

Table 2.11: Kendall rank correlation(τ) between normalized intelligibility(Voicing) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.675	0.677	0.684	0.684	0.692	0.654	0.620	0.639
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.422	0.172	0.329	0.219	0.355	0.509	0.466	0.399

Nasality

Table 2.12 に Nasality の R を示す．Nasality の了解度と音質の相関は 120 単語平均，Voicing よりもさらに低い．ここでも SNR 系の指標の相関が高いが，重み付 SNR4 種のうち fwSNRseg(S) を除く 3 種は重みを伴わない他の 3 種よりもやや相関が低い傾向にある．一方で，PESQ と d_{WSS} の相関は SNR 系の fwSNRseg と AI 以外よりも良い．この傾向は，Fig. 2.4 に示すスペクトログラムより鼻音と口音のききわけである Nasality の音韻特徴が 500 Hz 未満にみられる第 1 フォルマントが子音区間から始まる鼻音とそうでない口音の聞き分けであることに由来する．このため，母音の聞き取りに重点を置いた fwSNRseg(S) や AI の相関が比較的高くなっている．また，この他のスペクトル距離一部の指標では負の相関となっており，Nasality の推定には向かない．Nasality の了解度と相関の高かった例として，Fig. 2.26 に fwSNRseg(S)，PESQ，d_{WSS}，中程度の相関がみられた例として，SNRseg の結果を示す．PESQ 以外はどれも騒音種ごとの傾向差が目立ち，同

[19]他の子音特徴も含め，すべての尺度の結果は A に掲載した．

一音質値でも騒音種によって了解度は大きく異なっている．また，fwSNRseg(S) は 120 単語平均，Voicing における fwSNRseg(C) と同様に上に凸のカーブとなっており，了解度と音質値に非線形性があることが示唆される．Table 2.13 の τ の結果からも，AI，AIseg，fwSNRseg(S)，PESQ，d_{WSS} の相関が高いことからこれらの尺度による Nasality 推定性能が高くなるものと考えられる．

Table 2.12: Pearson correlation(R) between normalized intelligibility(Nasality) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.745	0.697	0.741	0.674	0.780	0.856	0.907	0.862
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.828	0.015	−0.169	−0.247	−0.264	0.803	0.340	−0.394

Table 2.13: Kendall rank correlation(τ) between normalized intelligibility(Nasality) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.510	0.466	0.542	0.480	0.573	0.631	0.652	0.664
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.634	0.031	0.178	0.300	0.343	0.620	0.258	0.274

Sustention

Table 2.14 に Sustention の R 一覧を示す．Sustention も SNR 系の相関は高いが，AI と AIseg は 0.683 と 0.656 とやや相関が悪い．一方で，d_{IS} の相関が 0.824 であり，SNR 系と同等の相関がある．Fig. 2.27 に fwSNRseg(A)，fwSNRseg(C)，d_{IS}，d_{WSS} の結果を示す．図より明らかなように条件 White も含めて対角線上にプロットされる d_{IS} に対し，fwSNRseg(C) と d_{WSS} は 4 種の騒音のうち White は異傾向となる．fwSNRseg(A) も White と他の騒音の傾向差はみられるものの，fwSNRseg(C) ほど顕著ではない．d_{IS} の性能が高いのは，式（1.6）で LPC スペクトルの係数ベクトルだけでなく，ゲインも見ているため，スペクトル包絡の形状と音量の時間変動を評価しているためと考えられる．Table 2.15 に Sustention の τ の結果を示す．最も相関の高い尺度は 0.709 の fwSNRseg(A) であることは変わらない．d_{LLR} と d_{IS} の相関が比較的高いのも R と同傾向である．全体として，尺度の序列は R との差はない．

Sibilation

Table 2.16 に Sibilation の R を示す．SNR 系の中でもセグメンテーションを行わない SNR，SNRA と d_{IS} は 0.8 前後と比較的良いものの，他の尺度は 0.7 未満である．Fig. 2.28 に SNRA，fwSNRseg(A)，d_{LLR}，d_{IS} の結果を示す．SNRA と fwSNRseg(A) との比較は同一の A 重みを用いた SNR のセグメンテーションの有りと無しが比較できる．結果より，SNRA の方が線形に変化し，fwSNRseg(A) は上に凸のプロットとなる．これは，Fig. 2.17 の主観評価より Sibilation はそもそも了解度と $\mathrm{SNR_{in}}$ の関係が非線形であり，R で評価するのは適さないといえる．スペクトル距離で比較的相関の高い d_{LLR} と d_{IS} は White の分散が大きいものの，その他の分散はやや小さくなる傾向にある．Table 2.17 に Sibilation の τ の結果を示す．R では SNR，SNRA と SNRseg 等との間に大きな差があったが，τ ではほとんど同程度の相関であるといえる．また，スペクトル

Table 2.14: Pearson correlation(R) between normalized intelligibility(Sustention) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.807	0.838	0.853	0.866	0.769	0.746	0.683	0.656
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.463	0.728	0.644	0.764	0.824	0.619	0.587	0.515

Table 2.15: Kendall rank correlation(τ) between normalized intelligibility(Sustention) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.638	0.681	0.672	0.709	0.623	0.539	0.542	0.500
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.398	0.558	0.402	0.594	0.639	0.443	0.468	0.377

距離尺度も SNRseg 等と同程度の値であり，ある程度の相関はあることがわかる．また R でも相関が極端に低かった PESQ と d_{WSS} の相関はやはり低い．

Table 2.16: Pearson correlation(R) between normalized intelligibility(Sibilation) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.799	0.814	0.653	0.650	0.528	0.466	0.548	0.493
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.109	0.490	0.626	0.676	0.780	0.281	0.391	0.377

Table 2.17: Kendall rank correlation(τ) between normalized intelligibility(Sibilation) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.486	0.488	0.460	0.494	0.391	0.276	0.364	0.320
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.099	0.409	0.397	0.485	0.489	0.155	0.337	0.317

Graveness

Table 2.18 に Graveness の R を示す．結果より，SNR 系は相関が高く，次いで PESQ と d_{WSS} が続く傾向は Voicing，Nasality と同傾向である．Fig. 2.29 に SNR，SNRseg，PESQ，d_{IS} の結果を示す．SNR は相関係数が 0.9 以上あるものの，やや下に凸のプロットで，SNRseg の方が対角線に近い．この二つの様に SNR 系は騒音種間差はほとんどみられないが，PESQ は騒音種によってプロット傾向に差がみられる．d_{IS} は PESQ よりもこの傾向が顕著である．これらの傾向も Voicing，Nasality と近い．Voicing と Nasality の主観評価結果である Fig. 2.11，Fig. 2.13 と比べ，Graveness の主観評価結果である Fig. 2.19 は $\mathrm{SNR_{in}}$ に対して了解度変化が大きい．Graveness で比較してるのは単語対間の第 2 フォルマント周波数であるため，スペクトル距離の客観音質指標では，騒音のスペクトルの違いが評価値に反映されやすく，騒音傾向差が出たものと考えられる．Table 2.19 に Graveness の τ を示す．R と比べて SNR，AI の順位がやや低くなるのはこれまで通

りだが，SNRA は最も相関の高い fwSNRseg(A) に次ぐ．つまり Graveness には A 特性重みによる SNR による推定性能が高くなると予想される．

Table 2.18: Pearson correlation(R) between normalized intelligibility(Graveness) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.902	0.923	0.936	0.948	0.941	0.913	0.905	0.827
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.730	0.625	0.376	0.363	0.380	0.765	0.671	−0.011

Table 2.19: Kendall rank correlation(τ) between normalized intelligibility(Graveness) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.732	0.787	0.754	0.802	0.777	0.738	0.706	0.614
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.565	0.493	0.306	0.236	0.207	0.573	0.480	0.019

Compactness

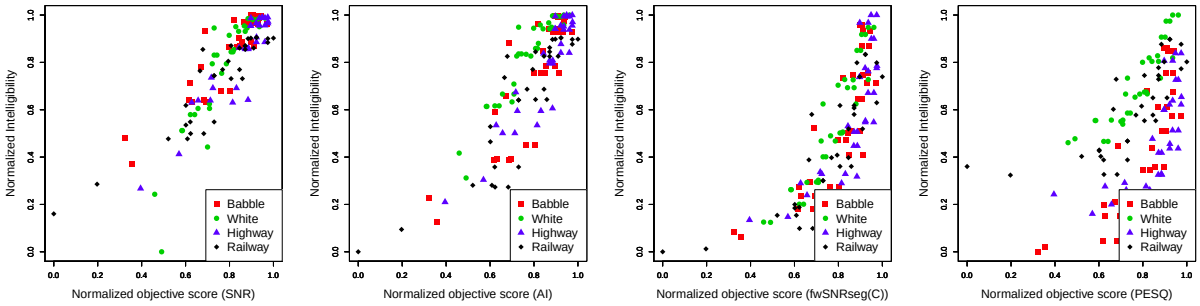
Table 2.20 に Compactness の相関係数一覧を示す．Compactness の結果は Graveness とほとんど同傾向である．Compactness で評価する音韻特徴も Graveness と近い特定周波数への集中度であるから，品質評価値も同傾向になったものと考えられる．Fig. 2.30 に Graveness の結果でも示した SNR，SNRseg，PESQ，d_{IS} の結果を示す．これらは同尺度の Graveness の結果と比べてもほとんど同じ傾向にあることがわかる．Table 2.21 の τ の結果も，120 単語平均，Voicing と同様に SNR と SNRA は R に比べ上位とならない．このため，了解度と音質値の間は線形ではなく非線形な関係にあるといえる．

Table 2.20: Pearson correlation(R) between normalized intelligibility(Compactness) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.906	0.899	0.905	0.874	0.913	0.933	0.938	0.907
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.789	0.344	0.131	0.196	0.277	0.818	0.623	0.151

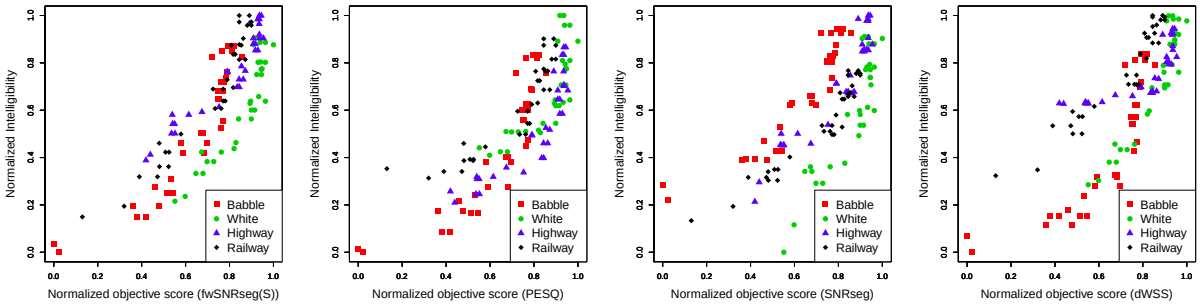
Table 2.21: Kendall rank correlation(τ) between normalized intelligibility(Compactness) score and normalized objective speech quality score

SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
0.737	0.746	0.763	0.744	0.824	0.771	0.750	0.761
PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
0.657	0.259	0.012	0.112	0.092	0.673	0.447	0.100



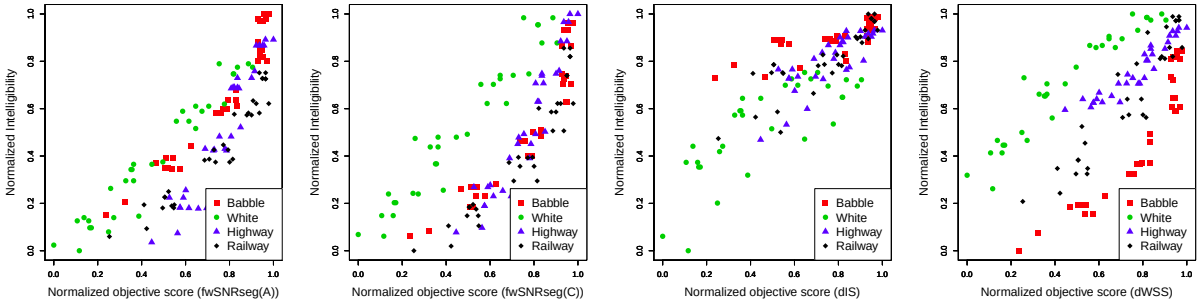
(a) SNR (b) AI (c) fwSNRseg(C) (d) PESQ

Fig. 2.25: Comparison between normalized intelligibility(Voicing) score and normalized objective speech quality score



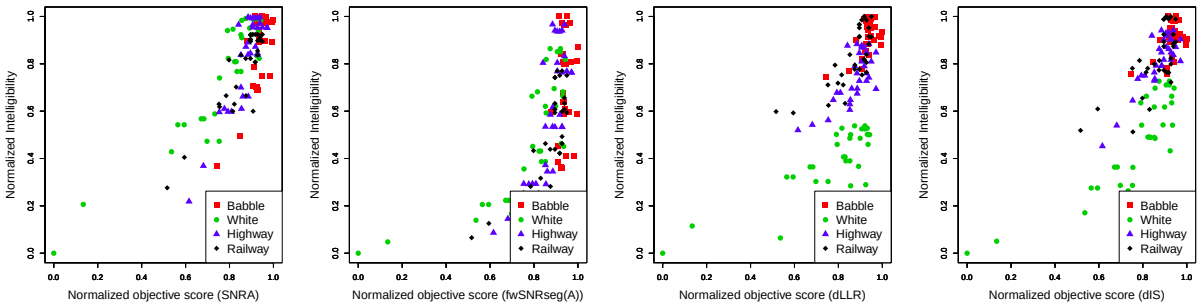
(a) fwSNRseg(S) (b) PESQ (c) SNRseg (d) d_{WSS}

Fig. 2.26: Comparison between normalized intelligibility(Nasality) score and normalized objective speech quality score



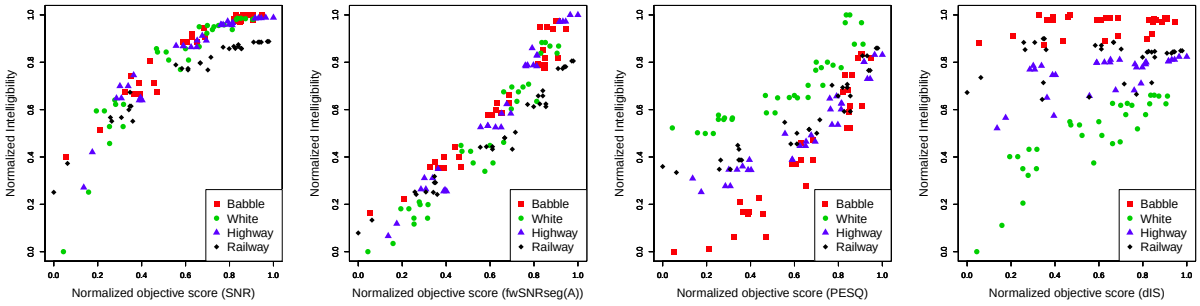
(a) fwSNRseg(A) (b) fwSNRseg(C) (c) d_{IS} (d) d_{WSS}

Fig. 2.27: Comparison between normalized intelligibility(Sustention) score and normalized objective speech quality score



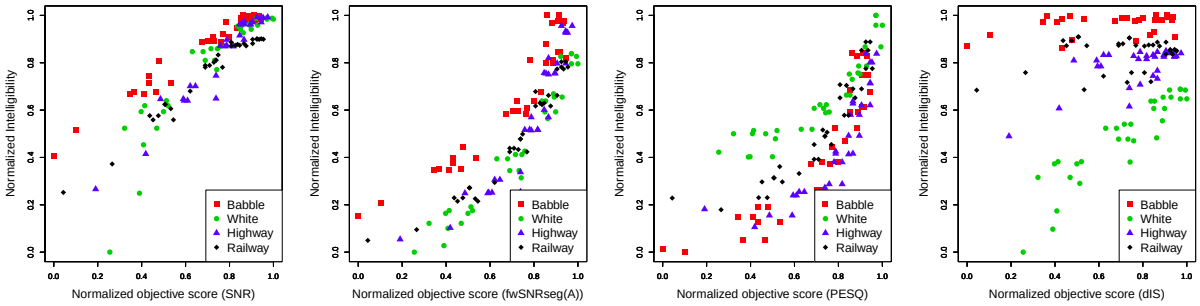
(a) SNRA (b) fwSNRseg(A) (c) d_{LLR} (d) d_{IS}

Fig. 2.28: Comparison between normalized intelligibility(Sibilation) score and normalized objective speech quality score



(a) SNR (b) fwSNRseg(A) (c) PESQ (d) d_{IS}

Fig. 2.29: Comparison between normalized intelligibility(Graveness) score and normalized objective speech quality score



(a) SNR (b) fwSNRseg(A) (c) PESQ (d) d_{IS}

Fig. 2.30: Comparison between normalized intelligibility(Compactness) score and normalized objective speech quality score

全騒音混合条件のまとめ

Table 2.22 に全騒音を混合した条件の子音特徴ごとに R と τ が最良の結果であった尺度と特異な傾向が認められた尺度をまとめる．R，τ 共に最良結果は全て SNR 系尺度だが，R はセグメンテーションを伴わない SNR，SNRA，AI が選択されることが多いが，τ はすべてセグメンテーションを伴う SNR 系尺度であった．各子音特徴ごとの分析でも Sustention と Graveness を除き，R は高いが，τ がやや低くなる傾向にある尺度がみられたため，心理尺度である了解度と物理尺度である客観音声品質値は，線形ではなく非線形な関係であるといえ，ピアソンの積率相関係数 R による分析は向かず，ケンドールの順位相関係数 τ で評価するのが良いと考えられる．Sustention と Graveness については，R，τ 共に最良尺度が fwSNRseg(A) であった．これは，Fig. 2.27 と Fig. 2.29 でみられたように，騒音ごとの傾向差が他の尺度と比べ出にくかったことが原因である．

以上の結果から，何らかの聴覚重みを用いた周波数重み付セグメンタル SNR を用いた了解度推定の性能が高いことが期待される．

SNR 系以外の尺度で子音特徴ごとに，特徴的な結果がみられた尺度をまとめる．120 単語平均，Voicing，Nasality，Graveness，Compactness は PESQ または d_{WSS} が SNR 系尺度に次ぐ性能であった．d_{WSS} は Liu らの検討 [150] で相関が最も高いことが多かったのは比較対象が全て本節で検討したような騒音や残響が固定される系での順位相関であり，本実験も同様の結果である．Sustention，Sibilation の 3 種は JDRT で聞き分けている音韻特徴が特定周波数への集中であり，d_{Cep}，d_{LLR}，d_{IS} といったスペクトル距離尺度も有効であった．Voicing 以下の 6 音韻特徴では，いくつかの尺度についてのみ了解度と音質値のプロットを示したが，相関係数があまり高くない（τ が 0.7 未満）尺度は騒音種内の序列はみられるものの，騒音種間の傾向差を表現できない尺度であることが多かった．次節では騒音ごとに順位相関係数 τ を比較する．

Table 2.22: Objective measures with highest correlation in all noise mixed condition

phonetic feature	Pearson correlation(R)	Kendall correlation(τ)	notable results
120 words	SNRA	fwSNRseg(C)	PESQ, d_{WSS}
Voicing	SNR	fwSNRseg(C)	PESQ, d_{WSS}
Nasality	AI	AIseg	PESQ, d_{WSS}
Sustention	fwSNRseg(A)	fwSNRseg(A)	d_{Cep}, d_{LLR}, d_{IS}
Sibilation	SNRA	fwSNRseg(A)	d_{LLR}, d_{IS}
Graveness	fwSNRseg(A)	fwSNRseg(A)	PESQ, d_{WSS}
Compactness	AI	fwSNRseg(C)	PESQ, d_{WSS}

2.3.5 騒音ごとのケンドールの順位相関

本節では騒音種ごとに正規化了解度と正規化音質値のケンドールの順位相関係数 τ を比較する．各子音特徴ごとの相関係数を騒音条件（B:Babble，W:White，H:Highway，R:Railway）ごとにまとめる．

120 単語平均

Table 2.23 に 120 単語平均での各騒音ごとの τ を示す．SNR 系尺度と PESQ，d_{WSS} は全ての騒音で 0.8 以上の相関があり，相関が高い．PESQ と d_{WSS} は Table 2.9 の全騒音混合条件ではそれぞれ 0.577，0.609 であったことから，これらは同一騒音における品質評価においては了解度と音質値の関係が保持される．d_{Cep}，d_{LAR}，d_{LLR}，d_{IS} は Babble，Railway 条件では相関が低いくなり，White，Highway の相関は高くなる傾向がみられた．これは Fig. 2.5 で見たように，騒音ごとのスペクトル特徴が原因と考えられる．

Table 2.23: Kendall rank correlation(τ) between normalized intelligibility(120 words average) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.845	0.853	0.845	0.853	0.861	0.853	0.853	0.853
W	0.878	0.878	0.878	0.878	0.910	0.902	0.869	0.918
H	0.878	0.869	0.878	0.878	0.902	0.902	0.910	0.902
R	0.865	0.882	0.865	0.882	0.874	0.849	0.849	0.849

	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.833	0.494	0.114	0.490	0.543	0.869	0.551	0.412
W	0.886	0.653	0.486	0.665	0.747	0.902	0.665	0.567
H	0.890	0.535	0.535	0.645	0.612	0.902	0.649	0.461
R	0.825	0.498	0.363	0.592	0.539	0.825	0.482	0.151

Voicing

Table 2.24 に Voicing 子音特徴での各騒音ごとの τ を示す．全体的に 120 単語平均と同傾向ではあるが，Babble と Railway の τ の低下は SNR 系尺度でも顕著にみられる．

Table 2.24: Kendall rank correlation(τ) between normalized intelligibility(Voicing) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.616	0.641	0.616	0.641	0.657	0.633	0.633	0.633
W	0.830	0.830	0.830	0.830	0.847	0.830	0.838	0.847
H	0.738	0.730	0.738	0.738	0.705	0.705	0.713	0.705
R	0.749	0.766	0.749	0.766	0.733	0.720	0.737	0.720

	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.616	0.420	0.351	0.490	0.539	0.649	0.453	0.396
W	0.822	0.569	0.380	0.524	0.765	0.847	0.560	0.454
H	0.684	0.426	0.590	0.537	0.631	0.672	0.553	0.430
R	0.687	0.382	0.460	0.464	0.592	0.720	0.469	0.411

Nasality

Table 2.25 に Nasality 子音特徴での各騒音ごとの τ を示す．全体的に 120 単語平均と同傾向ではあるが，Babble と Railway の τ の低下は SNR 系尺度でも顕著にみられる．

Sustention

Table 2.26 に Sustention 子音特徴での各騒音ごとの τ を示す．全騒音混合時と比べると，SNR 系尺度は 0.500〜0.709 から，0.767〜0.871 と大幅に改善している．また PESQ と d_{WSS} はそれぞれ 0.398，0.443 から，0.75 以上とこちらも大幅に改善している．以上の結果より，Sustention は騒音種の影響差が大きく，全騒音混合条件では了解度推定が行いにくい子音特徴であるといえる．

Table 2.25: Kendall rank correlation(τ) between normalized intelligibility(Nasality) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.808	0.800	0.808	0.800	0.800	0.808	0.808	0.808
W	0.621	0.621	0.621	0.621	0.682	0.691	0.682	0.699
H	0.765	0.724	0.757	0.757	0.769	0.794	0.777	0.794
R	0.749	0.766	0.749	0.766	0.733	0.720	0.737	0.720
	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.788	0.461	0.127	0.049	0.135	0.808	0.584	0.298
W	0.526	0.399	0.526	0.477	0.682	0.658	0.448	0.637
H	0.777	0.421	0.371	0.220	0.084	0.777	0.593	0.031
R	0.687	0.382	0.460	0.464	0.592	0.720	0.469	0.411

Table 2.26: Kendall rank correlation(τ) between normalized intelligibility(Sustention) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.767	0.776	0.767	0.776	0.776	0.767	0.767	0.767
W	0.847	0.847	0.847	0.847	0.847	0.834	0.871	0.847
H	0.832	0.815	0.832	0.823	0.864	0.860	0.864	0.860
R	0.794	0.810	0.777	0.794	0.786	0.777	0.777	0.777
	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.763	0.604	0.024	0.637	0.612	0.784	0.518	0.412
W	0.830	0.773	0.487	0.442	0.597	0.834	0.659	0.659
H	0.860	0.664	0.521	0.705	0.680	0.860	0.668	0.603
R	0.761	0.654	0.207	0.736	0.724	0.786	0.523	0.133

Sibilation

Table 2.27 に Sibilation 子音特徴での各騒音ごとの τ を示す．Sibilation は全騒音混合条件ではどれも τ が 0.5 未満だったが，SNR 系と d_{IS}，d_{WSS} では White で 0.6 以上とわずかに改善している．しかしながら，Babble 条件の τ は最大でも 0.3 程度であり，やはり了解度と音質値の相関は悪い．これは，Fig. 2.17 で見たように，そもそも $\mathrm{SNR_{in}}$ で了解度がほとんど変化しないためである．

Table 2.27: Kendall rank correlation(τ) between normalized intelligibility(Sibilation) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.284	0.276	0.284	0.276	0.292	0.300	0.300	0.300
W	0.648	0.648	0.648	0.648	0.639	0.631	0.615	0.648
H	0.496	0.517	0.496	0.480	0.501	0.492	0.484	0.492
R	0.534	0.542	0.534	0.542	0.546	0.529	0.529	0.529
	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.280	0.206	0.010	0.255	0.186	0.292	0.292	0.219
W	0.594	0.471	0.385	0.492	0.607	0.615	0.475	0.402
H	0.484	0.365	0.463	0.455	0.488	0.496	0.382	0.311
R	0.496	0.307	0.332	0.529	0.416	0.508	0.340	0.088

Graveness

Table 2.28 に Graveness 子音特徴での各騒音ごとの τ を示す．SNR 系は全騒音混合条件では AIseg を除き 0.7 以上の相関がみられたが，騒音ごとには 0.8 以上となりより相関が高くなった．また，PESQ と d_{WSS} は全騒音混合条件では 0.565，0.573 であったが，どちらも 0.8 程度に改善している．これまでの子音特徴と同様に PESQ と d_{WSS} は騒音種が固定される条件での相関が高い尺度であることがわかる．White に限定すれば，d_{LLR}，d_{IS} も τ が 0.7 程度と高くなっているが，Babble では 0.3 未満であるため，騒音種の影響を顕著に受けている．

Table 2.28: Kendall rank correlation(τ) between normalized intelligibility(Graveness) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.787	0.811	0.787	0.811	0.811	0.787	0.770	0.787
W	0.823	0.823	0.823	0.823	0.852	0.856	0.823	0.864
H	0.850	0.866	0.850	0.858	0.866	0.866	0.891	0.858
R	0.850	0.875	0.867	0.875	0.855	0.850	0.842	0.842

	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.778	0.570	0.378	0.243	0.047	0.803	0.411	0.055
W	0.832	0.652	0.529	0.733	0.701	0.832	0.635	0.493
H	0.846	0.646	0.478	0.592	0.570	0.842	0.498	0.298
R	0.826	0.510	0.453	0.199	0.064	0.834	0.363	0.523

Compactness

Table 2.29 に Compactness 子音特徴での各騒音ごとの τ を示す．全騒音混合時と同様に Compactness は Graveness と同様の傾向を示し，SNR 系は τ が概ね 0.8 以上となり，PESQ と d_{WSS} はそれぞれ 0.657，0.673 から 0.8 程度に改善した．d_{LLR}，d_{IS} の White は良いものの，特に他の騒音で τ が低めになる傾向も共通している．

Table 2.29: Kendall rank correlation(τ) between normalized intelligibility(Compactness) score and normalized objective speech quality score by noise type

	SNR	SNRA	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg
B	0.800	0.800	0.800	0.800	0.792	0.792	0.800	0.792
W	0.843	0.843	0.843	0.843	0.888	0.884	0.852	0.893
H	0.851	0.834	0.851	0.872	0.884	0.880	0.884	0.880
R	0.866	0.849	0.866	0.849	0.862	0.862	0.870	0.862

	PESQ	d_{Cep}	d_{LAR}	d_{LLR}	d_{IS}	d_{WSS}	SNRloss(C)	SNRloss(S)
B	0.771	0.486	0.041	0.429	0.253	0.800	0.539	0.392
W	0.852	0.667	0.401	0.663	0.762	0.884	0.585	0.328
H	0.855	0.549	0.072	0.462	0.387	0.880	0.557	0.188
R	0.849	0.390	0.258	0.423	0.094	0.849	0.361	0.213

騒音別相関係数のまとめ

Table 2.30 に子音特徴ごとの各騒音条件で最も τ が高い尺度を示す．同値の尺度は全て掲載した．ほとんどの子音特徴で騒音によらず SNR 系尺度が選択されたが，一部の組み合わせでは d_{WSS} が最も τ が高い．また SNR 系尺度では周波数重みを伴う尺度（SNRA，fwSNRseg(A)，fwSNRseg(C)，

fwSNRseg(S)，AI，AIseg）がほとんどであり，一部 SNR と SNRseg が選択されているが，SNR が単独で最良尺度になることはなかった．d_{WSS} も含め，何らかの聴覚重み付けをした音質評価尺度が了解度との相関が高くなる傾向にある．また，Table 2.22 の全騒音混合条件ではセグメンテーションを伴った尺度の τ が高かったことから，了解度の高精度な推定には聴覚重みを用いたセグメンタル SNR を用いる必要があるといえる．実際に，SII の重みを用いた fwSNRseg(C) は全騒音混合条件も含め，τ の高い尺度であった．d_{WSS} はスペクトル距離変化を評価する尺度ではあるが，時間ごとのパワースペクトルの変化尺度であるため，fwSNRseg に近い発想の尺度であり，τ が高くなったと考えられる．この他のスペクトル尺度では d_{LLR} と d_{IS} が騒音 White を用いた時に τ が高い傾向にあったが，Babble 条件では τ は低くなった．これは元来，音声スペクトルの変化を見る尺度であり，Babble 条件では妨害もスピーチノイズであるから，τ が低下したと考えられる．PESQ は文献 [21, 22, 23] では了解度推定に用いられているものの，最良の尺度ではない．しかし，PESQ は規格内に VAD による音声区間推定を含むため，本実験の様に音声区間が既知である場合以外も加味すると有効な尺度であることには変わりない．次節では本結果を用いて音質評価尺度を選択肢，了解度の推定を行う．

Table 2.30: Objective measures with highest correlation in each noise condition

phonetic feature	Babble	White	Highway	Railway
120 words	fwSNRseg(C)	AIseg	AI	SNRA, fwSNRseg(A)
Voicing	fwSNRseg(C)	fwSNRseg(C), AIseg, d_{WSS}	SNRA, fwSNRseg(A)	SNRA, fwSNRseg(A)
Nasality	SNR,SNRseg AI,AIseg	AIseg	fwSNRseg(S), AIseg	SNRA, fwSNRseg(A)
Sustention	d_{WSS}	AI	SNR, SNRA, SNRseg fwSNRseg(A), AIseg	SNRA
Sibilation	fwSNRseg(S), AI AIseg	AIseg	AI	fwSNRseg(C)
Graveness	SNRA, fwSNRseg(A) fwSNRseg(C)	fwSNRseg(S)	fwSNRseg(S)	fwSNRseg(A)
Compactness	SNR, SNRA, SNRseg fwSNRseg(A), AI, d_{WSS}	AIseg	fwSNRseg(C), AI	AI

2.4 非線形回帰による了解度推定

2.3.3 節で了解度と尺度の相関係数を比較した．本節では順位相関 τ の高い尺度を選択し，了解度の推定実験を行う．

2.4.1 回帰手法と了解度推定

回帰関数

推定には，了解度を目的変数に，客観音質値を説明変数とした回帰分析を行う．全騒音条件の相関係数比較より，了解度と客観音質値には線形ではなく非線形の関係があることがわかってお

り，回帰には最小二乗法によるシグモイドカーブフィッティングを行う．回帰する関数は式（2.5）のシグモイド・ロジスティック関数を用いる．

シグモイド・ロジスティック関数は本来，二値のみが観測される特性関数を連続関数に置き換えるために利用される．このシグモイド・ロジスティック関数にパラメトリックな回帰としてカーブフィッティングするためには，了解度の分布がシグモイドカーブで近似できる変化を持つことが前提になる．そこで了解度の変化傾向について考える．そもそも了解度は単語が「全て聴こえる」を上限とし,「全く聴こえない」を下限とする分布を持つ連続量の心理量である．上限値と下限値が明確に決まるのは，個々の単語まで分解して考えれば,「聴こえる」と「聴こえない」の二値になるからである．このため，多くの単語聴取結果の平均値である了解度は二値にはならないが，同一条件の単語の聴取結果を平均する行為自体が，　一般線形化モデルのロジスティック回帰のためのスプラインを作成していることと等価である．また，本来であれば，単語の平均ではなく，全ての単語の聴取結果を混合して一般線形化モデルとしてロジスティック回帰を行うことが妥当である．しかし，単語ごとの言語特性の差による心理的なマスキング効果の差を考慮せずに「聴こえる」と「聴こえない」の二値問題とすると単語セット内の言語特性の分散によって回帰結果が大きく変わる可能性がある．了解度試験は言語情報による聴取効果[20]はある程度統制されているものの，完全に等しいわけではない．このため特定単語ごとのロジスティック回帰は行えるが，全ての単語を用いたロジスティック回帰は妥当ではない．よってカーブフィッティングによって単語セット平均値への回帰を行うこととする．

式（2.5）は Fig. 2.31 に示す様に最大値 a で中心座標 $(-b/c,\ a/2)$ に変曲点を持ち，変曲点で点対称な関数である．本論文では，式（2.5）を各条件ごとに当てはめ，最小二乗法により b と c を求める．主観評価結果ではどの子音特徴においても最大値が 1 になることはなかったが，応用を考えると，騒音の影響が全くないとみなされる時（了解度が全く低下しない条件）も必要であると判断し，$a = 1$ とした．最小二乗法の評価値は，後述するテストデータの主観評価による了解度と推定値の最小二乗誤差（RMSE:Root Mean Squared Error）とし，RMSE が最小になる b と c を用いる．

$$y = \frac{a}{1 + \exp\left(b + cx\right)} \tag{2.5}$$

推定性能評価指標

推定性能の評価指標には，式（2.6）に示す，主観評価による了解度と推定関数によって求まる推定値の RMSE を用いる．RMSE が小さければ，了解度と推定値が近いことを示す．また，RMSE と 2.2 節で述べた MCI を比較し，RMSE<MCI が成り立てば推定性能が十分であるとみなすことができる [158]．これは推定値が，統計的代表値（平均値）の信頼区間内にあることを示し，主観評価の分散の範囲内で推定できているという解釈である．

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(Sub.Intell. - Est.Intell.\right)^2}{N}} \tag{2.6}$$

[20] 単語親密度や音素バランスといった聴覚以外の単語統制.

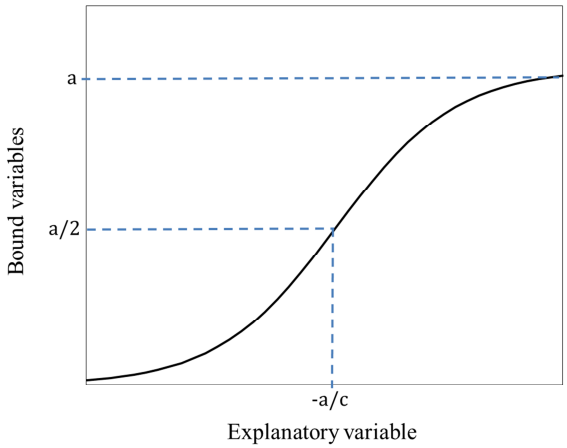


Fig. 2.31: Example for logistic function

学習データとテストデータ

客観音質値は，正規化前の実測値を用いる．また推定関数の作成には，騒音種ごとに計測点 32 点を 2 分割し，半分の 16 点を関数作成の学習データとし，残り半分を推定実験用のテストデータとした．ノイズクローズドテストの全騒音混合条件は 4 騒音の学習データ 64 点全てを用いて推定関数を作成し，4 騒音のテストデータ 64 点全てを推定した．ノイズオープン条件では，トレーニングデータに用いた騒音以外はすでに分割してあるテストデータを用いて較する．

推定に用いる客観音質指標

推定に用いる客観音質指標には，騒音種によらず順位相関 τ の高い尺度として SNR 系から，SNRseg，fwSNRseg(A)，fwSNRseg(C)，fwSNRseg(S)，AI，AIseg を選択し，スペクトル距離に基づく尺度からは PESQ と d_{WSS} を選択する．これらの尺度から，子音特徴ごとに最も推定性能の高い尺度を選択する．

2.4.2　ノイズクローズドテスト

JDRT の子音特徴ごとに騒音が既知の場合（以下，ノイズクローズドテスト）の推定結果を整理する．ここで，騒音が既知というのは推定関数作成に用いた学習データの騒音条件と，推定に用いるテストデータの騒音条件が同一であることを示す．ノイズクローズドテストで推定性能の低い尺度は騒音が未知の場合（ノイズオープンテスト）の推定結果も低くなることが予想されるため，まずはノイズクローズドテストで音質評価尺度を比較する．次に，子音特徴ごとにノイズクローズドテストで推定性能が高かった尺度を用いて騒音がノイズオープンテストで比較する．

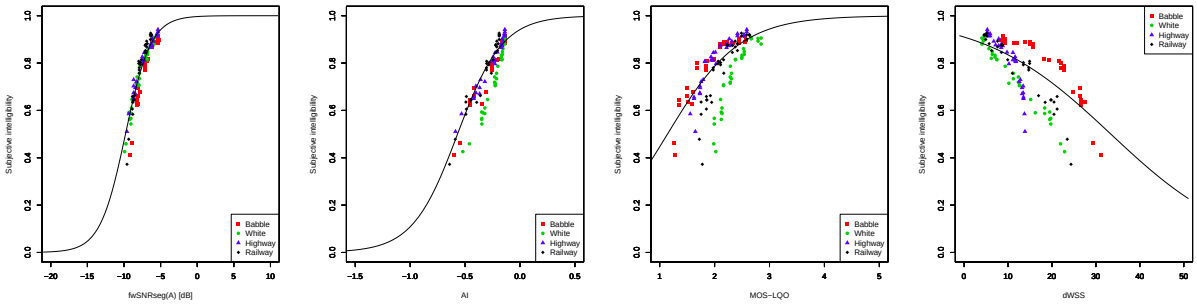
120 単語平均

120 単語平均の RMSE 一覧と MCI を Table 2.31 に示す．表中の B〜R は騒音ごとの推定結果で，Mean は騒音ごとの結果の加算平均値，All は全騒音混合条件での RMSE である．Mean と ALL の MCI は同じ値とした．結果より，PESQ と d_{WSS} を除けば全ての尺度が騒音種ごと及び Mean で RMSE<MCI を満たす．PESQ は White と Highway で満たし，Babble と Railway で僅かに満たさない．d_{WSS} は White のみ満たすが，他は満たさない．All では，AI と AIseg が満たさないが，他の SNRseg と各 fwSNRseg は RMSE<MCI を満たす．つまり，SNRseg または各 fwSNRseg であれば，騒音の種類の影響は小さいということがわかる．これらの尺度には，2.4.3 項で学習条件とテスト条件の騒音種を入れ替えた影響について考察する．

推定関数の例として，Fig. 2.32 に最も推定性能が高い fwSNRseg(A)，中程度の性能（Mean で条件を満たし，All で満たさない）であった AI，推定性能は低いが，他の文献でも使用される PESQ，d_{WSS} の例を示す[21]．fwSNRseg(A) 以外は全騒音で学習した推定関数に対し，騒音種ごとに傾向が異なることが明確であり，特に PESQ と d_{WSS} では騒音種ごとの推定関数が必要なことがわかる．以上の結果より，120 単語平均の推定には SNRseg または fwSNRseg(A) が良いことがわかる．

Table 2.31: Comparison of RMSE by objective quality measures along with the MCI per noise (120 words average)

	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.029	0.023	0.031	0.030	0.034	0.036	0.044	0.064	0.043
W	0.028	0.028	0.025	0.016	0.041	0.019	0.043	0.035	0.052
H	0.013	0.014	0.020	0.013	0.031	0.010	0.029	0.054	0.043
R	0.019	0.017	0.034	0.027	0.034	0.033	0.051	0.061	0.049
Mean	0.022	0.020	0.028	0.022	0.035	0.025	0.042	0.054	0.047
All	0.032	0.035	0.033	0.046	0.048	0.064	0.080	0.106	0.047



(a) fwSNRseg(A) (b) AI (c) PESQ (d) d_{WSS}

Fig. 2.32: Objective quality score and estimate function(120 words average)

Voicing

Voicing の RMSE 一覧と MCI を Table 2.32 に示す．MCI は 120 単語平均よりも大きな値となるため，d_{WSS} の Railway を除き騒音別の場合はどの尺度も RMSE<MCI を満たす．騒音種別にみる

[21]図中の各騒音のプロットでは学習データとテストデータを区別していない．以下，子音特徴別の結果も同じ．

と，White 条件の RMSE がどの尺度でも小さい．これは Table 2.24 の τ が White 条件は他の騒音よりも高い傾向と合致する．しかし，τ が最も低いのは Babble 条件であり，Babble の RMSE が大きい傾向にはあるが，必ずしも τ の序列とは一致していない．また，全騒音混合条件でも d_{WSS} 以外は RMSE<MCI を満たす．また，120 単語平均と異なり，SNRseg，fwSNRseg(A)，fwSNRseg(C)，fwSNRseg(S)，AI は Mean と All の差が小さい．これは Fig. 2.11 より，Voicing については騒音間の傾向差がほとんどないことが関係している．Fig. 2.33 に SNRseg，fwSNRseg(C)，AI，PESQ の客観音質値と All 条件の推定関数を示す．SNRseg と fwSNRseg(C) では騒音種ごとの傾向差が小さいが，AI と PESQ では傾向差がみられる．以上の検討より，Voicing 推定には SNRseg または fwSNRseg(A) が良い．

Table 2.32: Comparison of RMSE by objective quality measures along with the MCI per noise (Voicing)

	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.048	0.046	0.047	0.048	0.040	0.049	0.053	0.071	0.075
W	0.022	0.022	0.019	0.019	0.033	0.020	0.026	0.028	0.087
H	0.038	0.038	0.039	0.038	0.029	0.037	0.043	0.061	0.071
R	0.043	0.041	0.045	0.046	0.051	0.049	0.054	0.091	0.077
Mean	0.038	0.037	0.038	0.038	0.038	0.039	0.044	0.063	0.078
All	0.035	0.036	0.037	0.039	0.042	0.045	0.064	0.082	0.078

Nasality

Nasality の RMSE 一覧と MCI を Table 2.33 に示す．Nasality も Voicing 同様に騒音別に見た場合は d_{WSS} 以外で RMSE<MCI を満たす．騒音種ごとの RMSE は Voicing と同様に White 条件で低く，Babble 条件で高くなっている．しかし，Table 2.25 の τ は Babble が高く，White が低い傾向である．また，Mean は SNRseg，fwSNRseg(A)，fwSNRseg(C) は 0.037〜0.042 の低い値になるが，All では 0.065〜0.081 と 2 倍程度 RMSE が増加している．一方で，fwSNRseg(S) は Mean が 0.042，All が 0.044 とほぼ同じ値である．fwSNRSeg(C) は Table 2.13 で最も τ が高く 0.631 であったため全騒音条件でも RMSE が小さいと考えられるが，d_{WSS} も 0.620 と相関は高かったものの，RMSE は他に比較して大きい．Nasality の結果からは τ だけで推定性能を予測することは困難であることがわかる．d_{WSS} は τ にかかわらず RMSE が大きくなる．スペクトル距離系の尺度は騒音種（スペクトル劣化の影響）が固定された条件で τ が低くなっているのが要因である．Fig. 2.34 に fwSNRseg(A)，fwSNRseg(C)，fwSNRseg(S)，PESQ の客観音質値と All 条件の推定関数を示す．fwSNRseg(S) を除いて騒音種の違いがみられることから，Nasality 推定には fwSNRseg(S) が良いことがわかる．

Sustention

Sustention の RMSE 一覧と MCI を Table 2.34 に示す．結果より d_{WSS} も含め，騒音別条件では RMSE<MCI が成り立っている．騒音ごとの傾向は，White 条件の RMSE が他の騒音よりも大きく，Voicing，Nasality と異なる傾向にある．Voicing と Nasaloty では RMSE が大きかった d_{WSS} が RMSE<MCI を満たしている．しかし，全騒音を混合した条件 ALL では SNRseg と fwSNRseg(A) だけが RMSE<MCI を満たし，fwSNRseg(C) がほぼ同値であるものの，他の尺度は 0.1 以上の

Table 2.33: Comparison of RMSE by objective quality measures along with the MCI per noise (Nasality)

	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.058	0.057	0.062	0.056	0.056	0.062	0.063	0.106	0.088
W	0.037	0.027	0.021	0.015	0.042	0.022	0.035	0.015	0.065
H	0.038	0.038	0.048	0.038	0.044	0.032	0.046	0.054	0.078
R	0.024	0.025	0.036	0.028	0.028	0.031	0.048	0.043	0.070
Mean	0.039	0.037	0.042	0.034	0.042	0.037	0.048	0.054	0.075
All	0.071	0.081	0.065	0.057	0.044	0.053	0.057	0.095	0.075

RMSE であり，十分な推定が行えていないと考えられる．これは騒音の傾向差が大きく，全騒音混合条件と騒音別の τ の値が異なったことからも妥当である．このため，Sustention は騒音種が限定される条件で無ければ推定できない子音特徴である．Fig. 2.35 に SNRseg，fwSNRseg(A)，fwSNRseg(C)，fwSNRseg(S) の結果を示す．どの尺度においても White が別傾向としてプロットされているのがわかる．All 条件の RMSE の小ささとプロットの傾向から Sustention 推定には，SNRseg か fwSNRseg(A) が良い．

Table 2.34: Comparison of RMSE by objective quality measures along with the MCI per noise (Sustention)

	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.024	0.020	0.040	0.037	0.048	0.037	0.053	0.048	0.084
W	0.063	0.058	0.067	0.060	0.088	0.062	0.098	0.073	0.112
H	0.037	0.033	0.028	0.029	0.064	0.047	0.032	0.038	0.083
R	0.042	0.038	0.043	0.049	0.073	0.056	0.079	0.075	0.089
Mean	0.041	0.037	0.045	0.044	0.068	0.050	0.066	0.059	0.092
All	0.080	0.080	0.094	0.112	0.111	0.120	0.142	0.168	0.092

Sibilation

Sibilation の RMSE 一覧と MCI を Table 2.35 に示す．騒音別条件では，PESQ の White，Railway と d_{WSS} を除き RMSE<MCI が成り立っている．Fig. 2.17 より，本実験系では Sibilation はほとんど了解度変化が無く，Table 2.27 においても，Babble は τ が低かったが，SNR 系尺度では RMSE は 0.025 程度と十分に小さかった．また，SNRseg と fwSNRseg(A) では Mean 条件と All 条件の差が小さく，fwSNRseg(C) と fwSNRseg(S) では Mean と All で RMSE に 0.015 程度の差がつくことから，Sibilation 推定に適した尺度は SNRseg または fwSNRSeg(A) である．Fig. 2.36 に示す All 条件の推定関数と音質値も，SNRseg または fwSNRSeg(A) は緩やかなカーブを描くが，fwSNRseg(C) と fwSNRseg(S) はほぼ直線上であり，推定関数としてほとんど機能を果たしていない．

Graveness

Graveness の RMSE 一覧と MCI を Table 2.36 に示す．騒音別条件では，d_{WSS} の Railway を除き RMSE<MCI が成り立っている．SNR 系尺度では Babble と White 条件の RMSE がやや高い傾向にある．All 条件では SNRseg 系 4 種が RMSE<MCI を満たす．特に fwSNRseg(A) と fwSNRseg(C) は Mean 条件と All 条件の RMSE 差が 0.01 程度と小さい．Fig. 2.37 に SNRseg，fwSNRseg(C)，

Table 2.35: Comparison of RMSE by objective quality measures along with the MCI per noise (Sibilation)

	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.024	0.024	0.024	0.024	0.024	0.024	0.025	0.039	0.041
W	0.060	0.066	0.067	0.067	0.054	0.061	0.078	0.120	0.069
H	0.040	0.039	0.031	0.041	0.039	0.035	0.032	0.044	0.057
R	0.044	0.043	0.044	0.044	0.045	0.044	0.046	0.063	0.055
Mean	0.042	0.043	0.041	0.044	0.040	0.041	0.045	0.067	0.056
All	0.046	0.044	0.055	0.060	0.054	0.056	0.062	0.103	0.056

AI，PESQ の All 条件の推定関数と音質値のプロットを示す．AI と PESQ は騒音傾向差がみられ，SNRseg と fwSNRseg(C) は騒音差がほとんどみられない．

Table 2.36: Comparison of RMSE by objective quality measures along with the MCI per noise (Graveness)

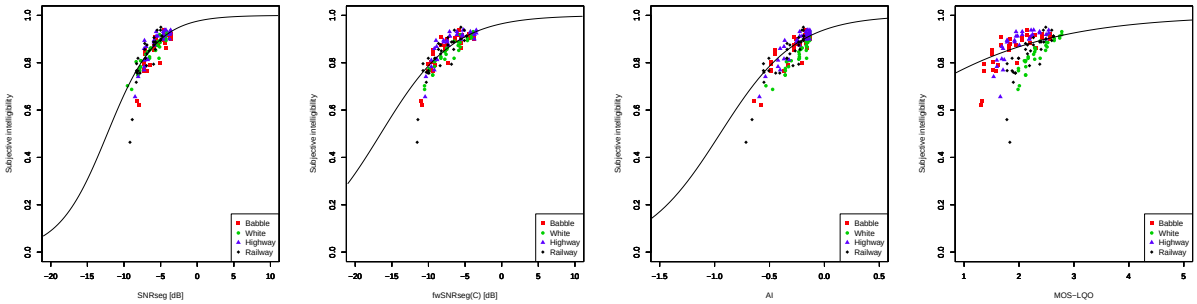
	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.047	0.037	0.051	0.049	0.066	0.061	0.070	0.081	0.095
W	0.047	0.061	0.043	0.035	0.050	0.034	0.082	0.071	0.105
H	0.038	0.048	0.038	0.032	0.061	0.041	0.063	0.091	0.106
R	0.036	0.033	0.043	0.039	0.062	0.055	0.085	0.098	0.093
Mean	0.042	0.045	0.044	0.039	0.060	0.048	0.075	0.085	0.100
All	0.064	0.057	0.055	0.077	0.085	0.118	0.139	0.167	0.100

Compactness

Compactness の RMSE 一覧と MCI を Table 2.37 に示す．騒音別条件では，全ての尺度で RMSE<MCI が成り立っている．しかし，AI,AIseg，PESQ，d_{WSS} の 4 種は Mean と All の RMSE 差が 2 倍程度あり，SNRseg，fwSNRseg に比べ推定性能が悪い．Mean では fwSNRseg(S) が 0.039 と最も RMSE が小さいが，All では 0.077 まで増加している．fwSNRseg(C) は Mean では 0.044 と fwSNRseg(S) は僅かに大きいものの，All では 0.055 と最も RMSE が小さい．Fig. 2.38 に fwSNRseg(C)，fwSNRseg(S)，AI，PESQ の All 条件の推定関数と音質値のプロットを示す．

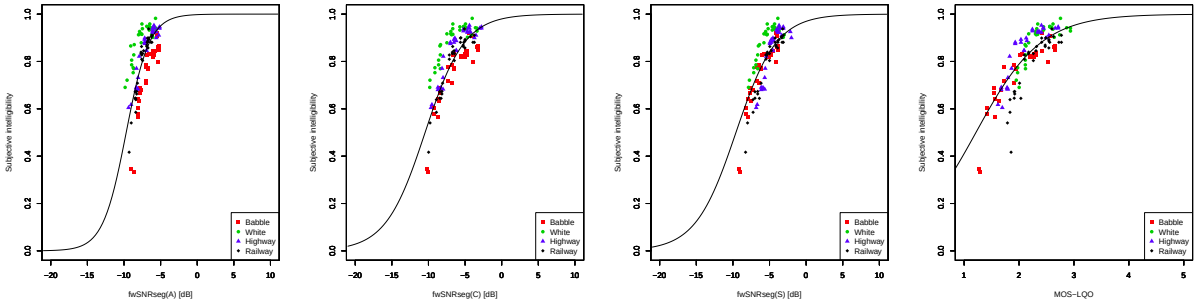
Table 2.37: Comparison of RMSE by objective quality measures along with the MCI per noise (Compactness)

	SNRseg	fwSNRseg(A)	fwSNRseg(C)	fwSNRseg(S)	AI	AIseg	PESQ	d_{WSS}	MCI
B	0.047	0.037	0.051	0.049	0.066	0.061	0.070	0.081	0.084
W	0.047	0.061	0.043	0.035	0.050	0.034	0.082	0.071	0.097
H	0.038	0.048	0.038	0.032	0.061	0.041	0.063	0.091	0.097
R	0.036	0.033	0.043	0.039	0.062	0.055	0.085	0.098	0.101
Mean	0.042	0.045	0.044	0.039	0.060	0.048	0.075	0.085	0.095
All	0.064	0.057	0.055	0.077	0.085	0.118	0.139	0.167	0.095



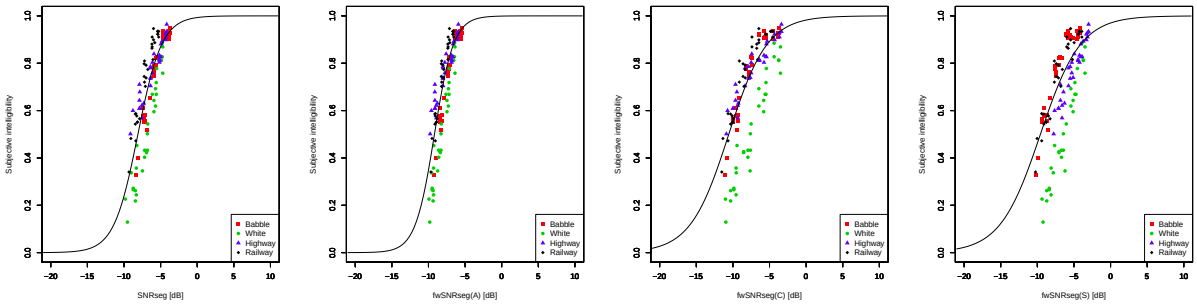
(a) SNRseg (b) fwSNRseg(C) (c) AI (d) PESQ

Fig. 2.33: Objective quality score and estimate function(Voicing)



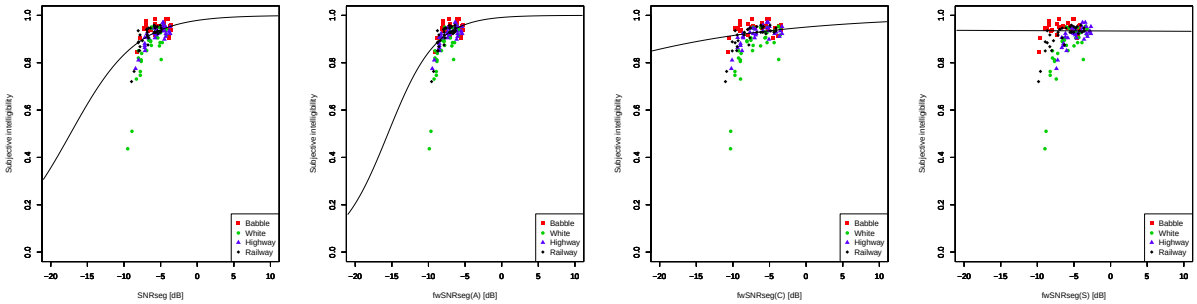
(a) fwSNRseg(A) (b) fwSNRseg(C) (c) fwSNRseg(S) (d) PESQ

Fig. 2.34: Objective quality score and estimate function(Voicing)



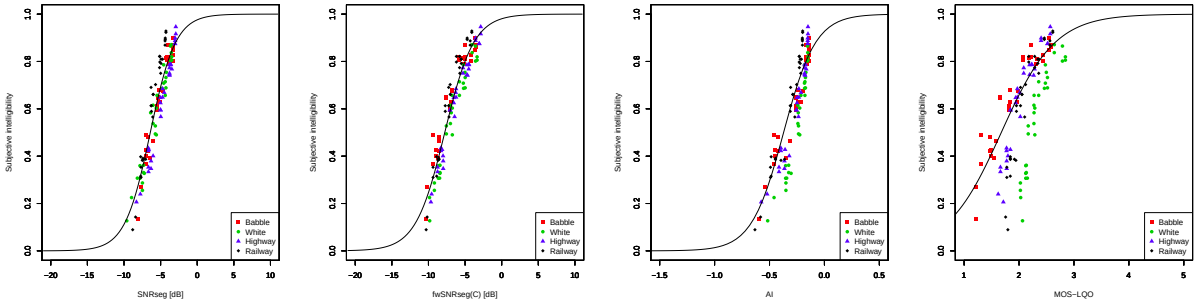
(a) SNRseg (b) fwSNRseg(A) (c) fwSNRseg(C) (d) fwSNRseg(S)

Fig. 2.35: Objective quality score and estimate function(Sustention)



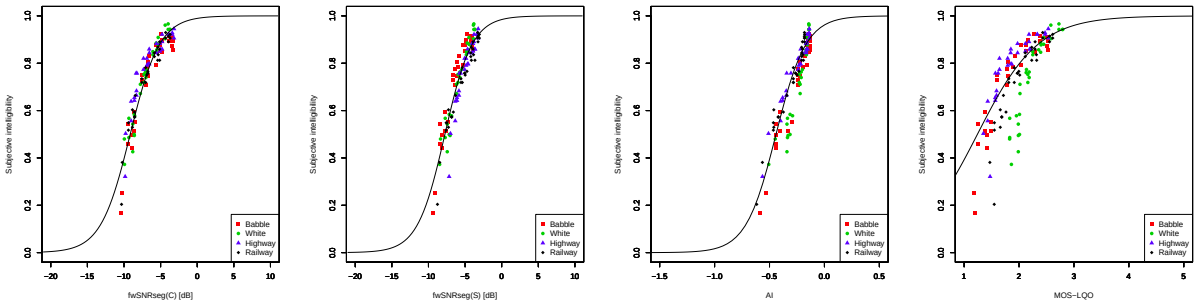
(a) SNRseg (b) fwSNRseg(A) (c) fwSNRseg(C) (d) fwSNRseg(S)

Fig. 2.36: Objective quality score and estimate function(Sibilation)



(a) SNRseg (b) fwSNRseg(C) (c) AI (d) PESQ

Fig. 2.37: Objective quality score and estimate function(Graveness)



(a) fwSNRseg(C) (b) fwSNRseg(S) (c) AI (d) PESQ

Fig. 2.38: Objective quality score and estimate function(Compactness)

ノイズクローズドテストのまとめ

ノイズクローズドテストの子音特徴ごとにRMSEが最小だった客観音質尺度をTable 2.38に示す．騒音種ごと（per noise）と全騒音混合条件（pooled noise）を分けて提示する．per noise条件は4騒音条件の平均値とした．結果より，PESQとd_{WSS}以外のSNR系尺度が選択されている．PESQとd_{WSS}は順位相関係数τは子音特徴によってはSNR系尺度とほぼ同等であることもあり，相関は決して悪くない．特にNasalityのWhiteといった，d_{WSS}が最もτが高かい条件ではRMSEが0.015と十分な推定性能を持っていることもあるが，全体としては推定性能が悪い．PESQもd_{WSS}程ではなかったものの，最良尺度になることはなかった．AIはNasalityでは騒音種別では0.042と中程度であったものの，All条件との差が0.002であり，Nasalityについては有効な尺度である．Sibilationでは，騒音種別とAllとの差が開くため，個別騒音ごとに最適な尺度である．AI以外の最良尺度はSNRsegに乗算する重みの最適化問題である[22]．以上のの結果から，現在，明瞭度／了解度の予測に使われるSII（AI），STIはその内部処理にSNRを持つことから，より推定性能を高めるためには，セグメンタルSNRを用いる方が良いことが本実験結果より示唆される．

また，順位相関係数τによる推定性能予測は，SNRsegの重み選択では参考になる指標と言えるが，d_{WSS}といった異なる物理指標との推定性能比較には向かなかった．これは各客観音質評価尺度ごとの音質値自体の変動とあてはめた関数の形状による．あてはめた関数と各プロットとの距離が広ければRMSEは大きくなる．つまり，客観音質の値域が広いd_{WSS}は回帰による了解度推定には向かない．

RMSE<MCIの基準については，騒音種別の場合はほとんどの尺度が満たした．RMSEとMCIの比較は，120単語平均の様に一人あたりの結果が十分に安定する場合は有効ではあるが，子音特徴別に見た場合は20単語の平均となるため，2.2.3節での結果のようにMCIが大きな値になる傾向にある．また，本節の比較はノイズクローズドテストであり，全体的にRMSEが小さくなる．よって，ノイズクローズドテストではRMSE<MCIの基準だけでは不十分であり，ノイズオープンテストの性能をきちんと評価する必要がある．

次節では，per noise条件とpooled noise条件の最良尺度を用いてノイズオープンテストを行う．

Table 2.38: Best intelligibility estimating measure for each phonetic features (noise closed test)

	Best measure	
	Per noise training	Pooled noise training
120 words	fwSNRseg(A)	SNRseg
Voicing	fwSNRseg(A)	SNRseg
Nasality	fwSNRseg(S)	AI
Sustention	fwSNRseg(C), AIseg	fwSNRseg(A)
Sibilation	AI	fwSNRseg(A)
Graveness	fwSNRseg(S)	fwSNRseg(C)
Compactness	fwSNRseg(S)	fwSNRseg(C)

[22]重み無，A，C，S，AIの5種

2.4.3 ノイズオープンテスト

ノイズオープンテストでは，特定の騒音条件で作成した推定関数を用いて別の騒音条件を推定する．ノイズオープンテストの性能が高ければ，全ての騒音に対し，1 つの推定関数で予測できることとなり，最も望ましい了解度推定関数であることがわかる．

120 単語平均

Table 2.39 に fwSNRseg(A) の，Table 2.40 に SNRseg を用いたノイズオープンテストの結果を示す．表の行方向が学習データを示し，列方向がテストデータを示す．学習データとテストデータが一致するマスはノイズクローズドテストの結果を再掲した．表下部の mean は各表のうち，オープンテストの結果 12 サンプルの平均値である．また，MCI は各騒音の MCI の平均値とした．オープンテストの条件で RMSE<MCI を満たすことができれば本手法の推定性能が十分であるといえる．また，平均値を求めるのに使った 12 サンプルの標準偏差を SD として併記する．同一の子音特徴において標準偏差が小さい方が，テストセットごとの極端な RMSE 変動が無いとみなせる．以上の表の見方は他の子音特徴でも共通する．

結果より，どちらも RMSE<MCI を満たさないが，fwSNRseg(A) の方が平均 RMSE が 0.01 ほど大きいが，RMSE の標準偏差は fwSNRseg(A) の方がわずかに小さい．そこで，両尺度の推定結果のうちノイズオープンテストである 12 点を対応のある t 検定で有意差を見たところ，$t = 1.5626$, $p = 0.1464$ であり平均値に有意差は無かった．ただしこれは fwSNRseg(A) と SNRseg に差が無いということであり，本実験で比較した尺度で了解度推定が行えないことを意味しない．また，両尺度とも，推定関数とテスト条件によって RMSE が極端に大きくなることがみられ，推定関数の騒音依存性がある．特に fwSNRseg(A) では Babble 条件で作成した関数で他の騒音を推定したときの RMSE が 0.064〜0.074，SNRseg では Railway 条件で作成した関数で他の騒音を推定したときの RMSE が 0.053〜0.098 と RMSE が大きくなる．以上の結果より，本章では，他の実験系に用いる場合では必ずしも SNRseg が最良の結果にならないと言う前提で，120 単語平均の最適尺度は SNRseg とする[23]．この結果は，何らかの最適な聴覚重みづけを行うことでより高精度な推定が行えることを示唆する．また，騒音種と最適な推定関数との組み合わせを実現するために，了解度推定関数を複数用意し，騒音環境ごとに最適な関数を選択する方式の検討が必要である．

Voicing

Table 2.41 に fwSNRseg(A) の結果を，Table 2.42 に SNRseg の結果を示す．表の読み方は 120 単語平均と同様である．結果より，どちらの尺度でも RMSE<MCI を満たす．しかしこれは，2.4.2 項で述べたように，子音特徴ごとの MCI は 120 単語平均と比べ大きな値になることも原因である．平均値は 0.003 程度の差をつけて fwSNRseg(A) の方が良く，標準偏差も 0.0033 程度小さい．2 尺度の平均値を対応のある t 検定で検定すると，$t = 2.1063$，$p = 0.05895$ であり，棄却率 5%よりわずかに大きい．このため，この 2 尺度の差は有意な傾向[24]にあるが，有意差では無い．騒音種別

[23]本論文では各子音特徴ごとに最適尺度を選択する．このため，統計的に有意差が無い組み合わせでは，どちらの尺度でも有効であるということから他の基準も用い，必ずどちらかを選択する．

[24]$p > 0.05$ 有意差ではないが，他の子音特徴も含め，p が 0.05〜0.06 になることが多く，このような表記とした．これはクローズドテストの per noise と pooled noise の最上位同士の比較を行ったため，あまり差が大きい比較ではないことが原因である．本項では 2 指標のどちらかを選ぶとした場合の目安としての検定であるため，有意傾向という表現を用い，選択するうえでの参考指標とした．

にみると，どちらの尺度も Babble 条件はノイズクローズテストよりも RMSE が小さい騒音があり，Highway 条件では他の条件の推定性能が悪いなど，120 単語との違いがみられる．本論文では Voicing 推定時には僅かに SNRseg の方が推定性能が高いと結論付ける．

Nasality

Table 2.43 に fwSNRseg(A) での，Table 2.44 に SNRseg での結果を示す．どちらも RMSE<MCI を満たしていない．また平均 RMSE は 120 単語平均，Voicing よりも大きく，fwSNRseg(A) では 0.1 を超える．SNRseg は平均 RMSE も 0.0805 と小さく，標準偏差も fwSNRseg(A) の 0.0487 と比べ半分の 0.0201 と小さい．この原因は fwSNRseg(A) の Highway での White を推定時の RMSE が 0.217 と非常に大きくなったことが原因である．しかし，この 2 方式の平均差の対応のある t 検定の結果は $t = -2.177$，$p = 0.05213$ であり，Voicing 同様に有意な傾向にあるものの有意差ではない．このため，Nasality については検討課題はあるものの SNRseg の方が良いとする．Nasality も推定関数と騒音の組み合わせで RMSE が増大する傾向がみられるため，騒音ごとに最適な推定関数を選択数必要がある．

Sustention

Table 2.45 に fwSNRseg(C)，Table 2.46 に AIseg，Table 2.47 に fwSNRseg(A) の結果を示す．3 尺度とも共通して全騒音平均で RMSE<MCI を満たさない．fwSNRseg(A) 以外は標準偏差が 0.1 以上あり，分散が大きい．特に，White 条件で推定関数を作成し，他の騒音条件を推定する組み合わせと，White 条件以外で推定関数を作成し，White 条件を推定した場合に RMSE が大きい．fwSNRseg(A) は Babble 条件と White 条件との組み合わせで RMSE が大きくならないため，平均 RMSE と標準偏差が小さくなった．3 尺度の差を尺度を要因とした 1 要因の分散分析で検定したところ，$F = 5.789$，$p = 0.0096$ となり尺度間に有意差がみられる．このため下位検定における多重比較で組み合わせごとの結果を比較する．fwSNRseg(A) と AIseg は $t = 3.385$，$p = 0.0026652$ であり，棄却率 1%以下で有意差がみられるた．fwSNRseg(C) と AIseg には $t = 1.992$，$p = 0.0588992$ で Voicing，Nasality 同様に棄却率 5%よりわずかに大きい．fwSNRseg(A) と fwSNRseg(C) は $t = 1.393$，$p = 0.1776172$ であり，有意差はみられない．以上の結果より，Sustention は fwSNRseg(A) と fwSNRseg(C) を用いるのがく．AIseg と比べた際に有意差のある fwSNRseg(A) を本論文の最適尺度とする．騒音種と推定関数の選択は Voicing，Nasality よりも必要であり，さらに他の聴覚重みを使用するなど推定性能の向上のために必要な課題が多い．

Sibilation

Table 2.48 に AI の結果を，Table 2.49 に fwSNRseg(A) の結果を示す．両尺度とも全騒音平均で RMSE<MCI を満たさないが，標準偏差は 0.057 前後でありほぼ同等である．どちらの尺度も White 条件を推定したオープンテストの RMSE が 0.17 以上と非常に大きい．一方で他の騒音条件では共通して RMSE が大きくなることはみられない．2 尺度間の対応のある t 検定の結果は，$t = 1.9638$，$p = 0.07532$ であり，有意差は無い．以上の結果より Sibilation の推定では 2 尺度の差はみられなく，また MCI と比較して RMSE も十分小さくなっていない．Table 2.27 から，相関係数 τ は 0.5 程度，Fig. 2.28 より，音質の変化に対し，了解度はほとんど変動していないため，そもそも Sibilation は推定しづらい子音特徴であるといえる．このため，既存の尺度はどの尺度を

用いても Sibilation の推定性能が高くならない．本論文では，全騒音混合条件で AI よりも順位相関係数 τ が高かった fwSNRseg(A) を選択することとする．

Graveness

Table 2.50 に fwSNRseg(S)，Table 2.51 に fwSNRseg(C) の結果を示す．fwSNRseg(S) は全騒音平均で RMSE<MCI を満たさないが，fwSNRseg(C) は満たしている．標準偏差も fwSNRseg(C) は fwSNRseg(S) の半分である．対応のある t 検定の結果は $t = -2.184$，$p = 0.0515$ であり，ほぼ 5%であるが，わずかに大きい．最大の傾向差は，fwSNRseg(S) では Highway 条件で関数を作成した際の Babble 条件の推定とその逆で RMSE が 0.2 以上と非常に大きくなるが，fwSNRseg(C) ではみられず，騒音種の影響が少ない．以上の結果より Graveness 推定には fwSNRseg(C) が適している．

Compactness

Table 2.52 に fwSNRseg(A)，Table 2.53 に fwSNRseg(C) の結果を示す．どちらの尺度も RMSE<MCI を満たす．また，標準偏差はどちらも 0.02 と同程度で，対応のある t 検定でも $t = -1.9004$, $p = 0.08389$ と有意差がみられなかったことから，Compactness については，fwSNRseg(S) または fwSNRseg(C) のどちらでも十分な推定が可能である．平均 RMSE と標準偏差が小さい fwSNRseg(S) を本論文では最良尺度として選択する．

Table 2.39: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(A) score with noise open test (120 words average)

		Test			
		B	W	H	R
Train	B	0.032	0.064	0.074	0.071
	W	0.075	0.032	0.023	0.054
	H	0.088	0.046	0.021	0.064
	R	0.088	0.045	0.043	0.028

mean: 0.0612 MCI: 0.047 SD: 0.0197

Table 2.40: RMSE between subjective intelligibility and estimated intelligibility using SNRseg score with noise open test (120 words average)

		Test			
		B	W	H	R
Train	B	0.041	0.044	0.029	0.067
	W	0.056	0.033	0.017	0.056
	H	0.060	0.036	0.018	0.048
	R	0.098	0.068	0.053	0.027

mean: 0.0528 MCI: 0.047 SD: 0.0209

Table 2.41: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(A) score with noise open test (Voicing)

		Test			
		B	W	H	R
Train	B	0.068	0.029	0.041	0.095
	W	0.066	0.028	0.036	0.086
	H	0.096	0.066	0.060	0.128
	R	0.065	0.033	0.043	0.076

mean: 0.0645 MCI: 0.078 SD: 0.0317

Table 2.42: RMSE between subjective intelligibility and estimated intelligibility using SNRseg score with noise open test (Voicing)

		Test			
		B	W	H	R
Train	B	0.071	0.031	0.043	0.091
	W	0.061	0.029	0.034	0.074
	H	0.092	0.059	0.058	0.118
	R	0.068	0.032	0.034	0.076

mean: 0.0615 MCI: 0.078 SD: 0.0284

Table 2.43: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(A) score with noise open test (Nasality)

		Test			
		B	W	H	R
Train	B	0.087	0.099	0.067	0.075
	W	0.102	0.019	0.109	0.121
	H	0.181	0.217	0.044	0.099
	R	0.065	0.129	0.048	0.036

mean: 0.1093 MCI: 0.075 SD: 0.0487

Table 2.44: RMSE between subjective intelligibility and estimated intelligibility using SNRseg score with noise open test (Nasality)

		Test			
		B	W	H	R
Train	B	0.081	0.098	0.074	0.069
	W	0.112	0.077	0.088	0.081
	H	0.095	0.088	0.091	0.078
	R	0.091	0.054	0.038	0.038

mean: 0.0805 MCI: 0.075 SD: 0.0201

Table 2.45: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(C) score with noise open test (Sustention)

		Test			
		B	W	H	R
Train	B	0.048	0.237	0.065	0.077
	W	0.288	0.079	0.297	0.335
	H	0.082	0.249	0.038	0.080
	R	0.055	0.278	0.063	0.054

mean: 0.1756 MCI: 0.092 SD: 0.1126

Table 2.46: RMSE between subjective intelligibility and estimated intelligibility using AIseg score with noise open test (Sustention)

		Test			
		B	W	H	R
Train	B	0.052	0.329	0.088	0.122
	W	0.349	0.081	0.306	0.459
	H	0.068	0.258	0.058	0.165
	R	0.138	0.408	0.141	0.067

mean: 0.2359 MCI: 0.092 SD: 0.1322

Table 2.47: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(A) score with noise open test (Sustention)

		Test			
		B	W	H	R
Train	B	0.027	0.091	0.158	0.125
	W	0.077	0.066	0.213	0.200
	H	0.112	0.220	0.041	0.070
	R	0.122	0.203	0.088	0.044

mean: 0.1401 MCI: 0.092 SD: 0.0564

Table 2.48: RMSE between subjective intelligibility and estimated intelligibility using AIseg score with noise open test (Sibilation)

		Test			
		B	W	H	R
Train	B	0.035	0.197	0.069	0.087
	W	0.137	0.085	0.090	0.105
	H	0.036	0.180	0.055	0.073
	R	0.030	0.180	0.052	0.069

mean: 0.1032 MCI: 0.056 SD: 0.0578

Table 2.49: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(A) score with noise open test (Sibilation)

		Test			
		B	W	H	R
Train	B	0.035	0.197	0.069	0.088
	W	0.052	0.117	0.030	0.042
	H	0.037	0.180	0.056	0.073
	R	0.031	0.173	0.050	0.067

mean: 0.0852 MCI: 0.056 SD: 0.0620

Table 2.50: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(S) score with noise open test (Graveness)

		Test			
		B	W	H	R
Train	B	0.058	0.159	0.224	0.143
	W	0.150	0.043	0.101	0.052
	H	0.239	0.086	0.042	0.093
	R	0.137	0.046	0.122	0.049

mean: 0.1294 MCI: 0.100 SD: 0.0601

Table 2.51: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(C) score with noise open test (Graveness)

		Test			
		B	W	H	R
Train	B	0.064	0.093	0.075	0.073
	W	0.126	0.053	0.047	0.128
	H	0.093	0.059	0.042	0.100
	R	0.082	0.090	0.083	0.050

mean: 0.0875 MCI: 0.100 SD: 0.0238

Table 2.52: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(S) score with noise open test (Compactness)

		Test			
		B	W	H	R
Train	B	0.062	0.060	0.109	0.091
	W	0.111	0.079	0.072	0.084
	H	0.128	0.084	0.049	0.064
	R	0.101	0.057	0.063	0.045

mean: 0.0741 MCI: 0.095 SD: 0.0195

Table 2.53: RMSE between subjective intelligibility and estimated intelligibility using fwSNRseg(C) score with noise open test (Compactness)

		Test			
		B	W	H	R
Train	B	0.067	0.058	0.099	0.065
	W	0.062	0.043	0.080	0.057
	H	0.115	0.068	0.068	0.084
	R	0.073	0.046	0.083	0.058

mean: 0.0854 MCI: 0.095 SD: 0.0230

ノイズオープンテストのまとめ

Table 2.54 に各子音特徴ごとに選択した最良の尺度を示す．t–test は比較した 2 尺度（Sustention は 3 尺度のため分散分析の結果）の有意差検定結果を，MCI rule は RMSE<MCI の基準を満たしたかどうか，noise trend はノイズ差の傾向が大きくみられたかどうかを示す．有意差検定において，有意確率 p が $0.05< p <0.06$ の時は，有意な傾向にあるとした．結果より，MCI 基準を満たした Voicing，Nasality，Graveness，Compactness は表に掲載した SNRseg，fwSNRseg(C)，fwSNRseg(S) を用いた場合に十分な精度で推定できる．この時，有意差検定の結果が有意でない場合は比較したもう一方の尺度であってもかまわないことを意味する．MCI 基準を満たさなかった 120 単語平均，Sustention，Sibilation はどれもノイズ種による傾向差が大きかった．このため，各騒音条件ごとに最適な推定関数を選択する必要がある．Sustention は，fwSNRseg(A) が他の尺度と比べても有意に推定性能が高かったが，それでも MCI 基準を満たしていない．このため，Sustention は騒音の種類の影響を非常に受けやすく，騒音条件による推定関数の選択のためのテストセットとして用いることができる．Sibilation に関しては他の子音特徴と異なり主観評価値がほとんど変動しないため，主観評価に用いる $\mathrm{SNR_{in}}$ 等も含めた検討が必要である．

120 単語平均は 2.4.2 項でも述べたように，子音特徴別より値が小さく，他の子音特徴と同程度の平均 RMSE でも MCI 基準を満たさなかった．120 単語平均は子音特徴ごとの 6 倍の単語を用いた分散であり，大数の法則に則り平均値に収束したものと考えられる．このため，MCI が他の子音特徴と比べて小さな値になると考えられる．つまり，120 単語を基準とすれば他の子音特徴の MCI は十分収束していなく，他の子音特徴を基準とすれば 120 単語平均の MCI は小さすぎる．JDRT の基礎検討 [21, 104] においても，子音特徴や被験者ごとの回答の分散については十分議論されていなく，MCI 基準は他の子音特徴と分けて考える必要がある．

騒音条件ごとに最適な推定関数を選択るために，騒音種ごとのスペクトル形状等の物理的特徴を事前に求め，同傾向な騒音種ごとに最適な推定関数を用いる騒音のクラスタリングによる推定関数分岐が考えられる．また，クローズドテストの結果も踏まえると，JDRT の子音特徴別の了解度予測には，SNRseg の重みを子音特徴と騒音の組み合わせによって最適なものを選択する必要がある．しかし，本論文で比較した聴覚重みは代表的なものを選択したが，全てを網羅していない．また，JDRT に最適な重みが既存の尺度として存在しているとも限らない．そのため，ある程度の数の主観評価結果をプールし，その傾向から最適な重みを求めることが必要であり，機械学習による重みづけを次章以降で検討する．

Table 2.54: Best estimate measure for each phonetic feature and noise dependency by noise open test

Phonetic feature	Best measure	t–test	MCI rule	Noise dependency
120 words	SNRseg	not significant	MCI>RMSE	high dependency
Voicing	SNRseg	significant dependency	MCI<RMSE	low dependency
Nasality	SNRseg	significant dependency	MCI<RMSE	high dependency
Sustention	fwSNRseg(A)	significant	MCI>RMSE	high dependency
Sibilation	fwSNRseg(A)	not significant	MCI>RMSE	high dependency
Graveness	fwSNRseg(C)	significant dependency	MCI<RMSE	low dependency
Compactness	fwSNRseg(S)	not significant	MCI<RMSE	low dependency

significant trend: $0.05< p <0.06$

2.5 まとめ

以上，本章ではバイノーラル音声システムの了解度主観評価と既存尺度を用いたパラメトリック回帰による了解度推定を行い，JDRT の子音特徴ごとに最適な既存尺度を選択した．要点を以下に示す．

- 主観評価の結果は，先行研究である文献 [21] と同様に JDRT の子音特徴ごとに大きく異なり，騒音種の影響も子音特徴ごとに異なる（Table 2.7）．このため，子音特徴ごと，騒音種ごとに了解度推定を行う必要がある．
- 騒音の方位角と $\mathrm{SNR_{in}}$ の組み合わせは，同一騒音種内の個別の騒音とみなすことができ，ベターイヤースコアの客観音質値の違いとして扱える．
- 既存の 16 種の客観音質評価指標の比較には，主観評価による了解度と，客観音質値の順位相関 τ を用いた比較である程度絞り込める．また，騒音 4 種を混合した場合と，騒音種ごとの順位相関は尺度によって傾向が異なる．SNRseg や fwSNRseg，AI 等は騒音混合条件でも相関が高く，PESQ や d_{WSS}，d_{IS} といったスペクトル距離に基づく尺度は騒音混合条件で相関が低い．
- d_{WSS} はスペクトル距離尺度の中で，多くの子音特徴で主観評価値との順位相関は高い．しかし，客観音質の値域が広いものはシグモイド関数のカーブフィッティングによる推定では RMSE が増加する．
- 子音特徴ごとに最適な聴覚重みは異なるが，音声品質尺度に fwSNRseg で重み無（SNRseg）を含む何らかの聴覚重み用いた尺度を選択し，了解度を推定すると，RMSE が小さく推定性能が高くなる．JDRT の子音特徴ごとの最適な指標は Table 2.54 に示した．
- 推定性能が低い Sustention に関して，推定関数作成に用いる騒音条件と推定したい騒音条件のスペクトルがある程度近い必要がある．このため，騒音ごとに最適な推定関数を選択する騒音クラスタリングによる推定関数分岐が必要である．

次章では，上記の課題を解決する了解度推定の手法について提案する．

第3章　機械学習を用いた騒音付加音声了解度推定法の概略

本章では，2 章で検討した既往尺度の問題点を解決するため，機械学習を用いて騒音下音声了解度を推定する方法を提案する．そして提案法で利用する要素技術について概説する．

3.1　提案する了解度推定法

3.1.1　解決すべき課題

2 章の結果より，特に Sustention については以下の課題があり，機械学習やデータマイニングで使われる手法を用いながら，本章以降で解決に向けて検討する．

- (a) 騒音種による最適な推定関数の選択
- (b) 推定に最適な聴覚重みの作成

騒音クラスタリングの提案

(a) は，騒音信号の了解度に対する特性[1]を基に，最適な推定関数を選択する必要がある．2 章では，同一騒音であれば順位相関が高かった．つまり，$\mathrm{SNR_{in}}$ と騒音方位角は音質と了解度の序列にほとんど違いが無いため，より傾向差が大きい騒音種の傾向が近いものの自動分類が必要である．

音信号の分類の代表例に音声認識がある．音声信号の言語情報を解析する音声認識には，メル周波数ケプストラム係数（MFCC）とそのデルタパラメータ，および言語情報を用いた分類・認識を行う [114]．騒音を分類するためには，教師データ[2]を必要とする判別分析と，教師データを必要としないクラスタリングがある．2 章では 4 種の騒音を比較したが，判別分析・クラスタリング問わず，判別器作成のためのデータとしては少なすぎる．理想的には遍く多数の騒音を分析する必要があるが，これも現実的ではない．そこで，電子協騒音データベース [157] の全騒音を分析することで，ある程度の騒音数を確保することとする．

また，本論文で検討している了解度は単語を用いて評価するため，騒音への単語挿入タイミングを考慮する必要がある．なぜなら，了解度変化は AI や SII，STI といった指標の内部処理に用いている SNR で説明されるエネルギーマスキング[3]で大局的には説明されるが，単語単位の聴取

[1] 2 章を例にすると，Babble ならば音声の様な騒音，White なら平坦なスペクトル包絡，Highway や Railway は自然音に近いといった騒音信号の周波数特性，時間特性によるマスキング効果の違い．

[2] 入力データに対して理想的と考えられる出力のこと．本論文では，主観評価で求める了解度を推定するため，教師データは了解度となる．

[3] マスカーとなる信号の周波数領域，時間領域のエネルギー特性による音声マスキング．

能力（単語了解度）は音声の調音結合や親密度，文中の単語間のつながりといった効果による心理的なマスキング効果である情報マスキングも考慮しなければならないためである．特に非定常な長時間の自然騒音[4]の場合は，音声情報の非定常さと騒音のエネルギー変動の非定常さの組み合わせとなるため，大局的なエネルギーマスキングでは説明できない了解度変動が起きていると考えられる．このため，同一の騒音として収録された音源であっても全ての区間で了解度変動が均質になるとは限らない．そこで，データベース上では同一騒音として扱われる各種騒音を了解度試験に用いるのに妥当な時間長に分割し，1 つ 1 つを独立した騒音として扱い（以下，Long Frame : LF と呼ぶ），その了解度への影響を評価する．

次に，教師データが事前に必要な判別分析では，事前に了解度試験を行う必要がある．了解度試験は評価単語セット単位での騒音の影響を均一化するために，上述のように分割した LF 一つにつき JDRT のフルセットであれば 120 単語，1 子音特徴でも 20 単語の評価が必要であり，膨大な数の評価単語が必要である．このため，事前に教師データが必要な判別分析ではなく，教師データを用いない分類であるクラスタリングの利用を検討する．この時用いる特徴量としては，了解度試験を行う前の騒音 LF だけで求まる特徴であることが望ましい．そこで，騒音信号の音色特徴を用いたクラスタリングを検討する．騒音の音色を用いるのは，音声了解度は言語情報に基づく品質だが，騒音信号自体は言語情報を含まない[5]ため非言語情報として扱うことができるためである．

音色情報を用いて音楽の主観的な分類を行う技術に，音楽情報検索 MIR（Music Information Retrieval）がある．MIR で用いられる統計量は，言語情報を含むとは限らない音楽信号の主観印象を客観推定するのに用いられる特徴である．このため，騒音のような言語情報を含まない音の解析にも有効であると考えられる．以上の点を考慮した騒音クラスタリングについて検討し，作成したクラスタリングモデルの精度の確認に JDRT の Sustention 単語セットを用いた主観評価を行う．

SVR を用いた了解度推定の提案

(b) はクラスタリング後の了解度推定関数について 2 章で用いた各種重みの fwSNRseg を用いたシグモイドカーブフィッティング（パラメトリック回帰）による推定関数よりも推定性能を向上させることを目指す．他の子音特徴も含め，各種重みの fwSNRseg を用いた回帰の性能は高く，帯域ごとの SNRseg を既存の聴覚重みにとらわれず結合することを考える．新しい推定関数は 3.1.1 項の騒音クラスタごと及びクラスタ分けしない条件で作成し，両者を比較することで，騒音クラスタリングの了解度推定に与える影響を検証する．

fwSNRseg を用いた回帰は，聴覚重みを用いたエネルギーマスキングの考慮であり，人間の言語認知特性を利用した情報マスキングについては十分考慮していない．騒音クラスタリングも情報マスキングを考慮するための了解度への傾向差分類の検討であるが，騒音の音色の影響に限定されるためまだ不十分である．そこで回帰関数をシグモイドカーブフィッティングの様なパラメトリック回帰から，ノンパラメトリックな機械学習による非線形関数への回帰にすることを検討する．しかし，一般にこの様な非線形回帰は説明変数に用いる特徴量ベクトルの次元数が多い場合や，サンプル数が少ない場合に過学習による汎化性能の低下が起こる．この様な問題を考慮した回帰手

[4]騒音 DB を含む自然環境で起こりうる騒音．

[5]Babble 等のスピーチノイズは何らかの言語情報を持つことがありうるが，その言語情報が了解度へ影響を与えるかどうかまでの解析は本論文では扱わない．本論文のスピーチノイズは，音声の特徴をもった非言語騒音として扱う．

法にサポートベクトル回帰（Support Vector Regression: SVR）がある．SVR はサポートベクトルマシン（Support Vector Machine: SVM）[154] と同様の手法を用いており，正則化，サポートベクトルとのマージン最大化，ϵ-不感応関数の利用といった内容を考慮した回帰手法である．

SVR の基本概念を Fig. 3.1 に示す．入力特徴量ベクトル x に隠れ層で教師データから作成した回帰係数（特徴量次元ごとの重み）x_1〜x_n を乗じ，出力層で出力値とする．これは，Fig. 1.2 の総合音質評価モデルと同様の構造を持っている．つまり，SVR の出力値を了解度などの主観的な音質とすれば，人間の主観評価を模擬しているとみなせる．了解度は言語を用いた音の品質評価であるため，言語を用いた音声信号の特徴による SVR は，人間の知覚を模擬した了解度推定とみなせる．しかし，機械学習的手法であるため，特徴量ごとの回帰係数から相対的な特徴の重みを分析することは可能であるが，聴覚的に意味のある重みを得ることはできない問題がある．本論文では，1.3 節でも述べたように聴覚的なメカニズムの解析よりも了解度推定精度を目指すことを目的とするため，機械学習の手法から SVR を選択し了解度推定に適合することを検討する．用いる特徴量には帯域ごとのセグメンタル SNR とする．これは 2 章で検討した各種重みの fwSNRseg と同様に帯域ごとの SNRseg に対する最適な重みづけを検討することに近似される[6]．本論文では，正確な聴覚重みを求めることではなく，了解度の推定性能を向上に寄与する周波数特性を求めることを検討するため，機械学習によって求まる重みが現実の聴覚特性との整合性が無くても良いと考える．以上の観点に基づいた騒音クラスタごとに作成した SVR による了解度推定関数を用いた了解度推定法について提案し，その要素技術を概説する．

3.1.2　提案する了解度推定法の手順

Fig. 3.2 に提案する了解度推定法の全体の流れを示す．まず，騒音信号から 15 次元の音響特徴量を解析し，入力騒音のクラスタ番号を求める．次に，騒音信号と音声信号を加算し，セグメンタル SNR クラスタごとに SVR による了解度推定関数（以下，SVR 推定関数）から推定了解度を求める．クラスタリングモデルとクラスタごとの SVR 推定関数はそれぞれ事前に求めておく．本論文では 4 章でクラスタリングモデルの作成，5 章で SVR 推定関数の作成を行い，6 章で未知データに対する推定性能や 2 章で検討した尺度との比較を行う．

Fig. 3.3 に SVR 部の詳細を示す．まず，騒音信号と DRT 音声を主観評価設定で加算した評価信号から，SVR の特徴量として Fig. 2.23 に示した 25 帯域で帯域分割したセグメンタル SNR（以下，cbSNRseg）を求める．次に，クラスタ番号を基に事前に求めてある推定関数を選択し，特徴量から推定了解度を得る．推定関数は，クラスタごとにそれぞれ 1 つの計 3 個を作成しておく．推定関数の学習に用いる教師データには了解度の主観評価値を用いる．

提案法は騒音クラスタリング部と SVR による推定関数を用いるため，クラスタリングによる誤差と推定関数の推定誤差の両方を含む．本論文ではまず騒音クラスタリングを行い（4 章前半），その主観評価値を求め（4 章後半），そこから SVR 推定関数を作成し（5 章），提案法全体の評価（6 章）を行う．これは騒音データベースを用いた了解度の主観評価が膨大にならないようにすることを主目的としている．よって，騒音クラスタリングの結果を主観評価で確認したのちは，騒音クラスタリング自体の誤差は無いものとみなし，SVR 推定関数による推定誤差を提案法の推定誤

[6]教師信号に了解度を用いるため，fwSNRseg と同じ重み付のエネルギー比ではなく，了解度が直接求まる．帯域ごとの回帰係数は聴覚重みとカーブフィッティングしたシグモイド・ロジスティック関数の積と等しい．

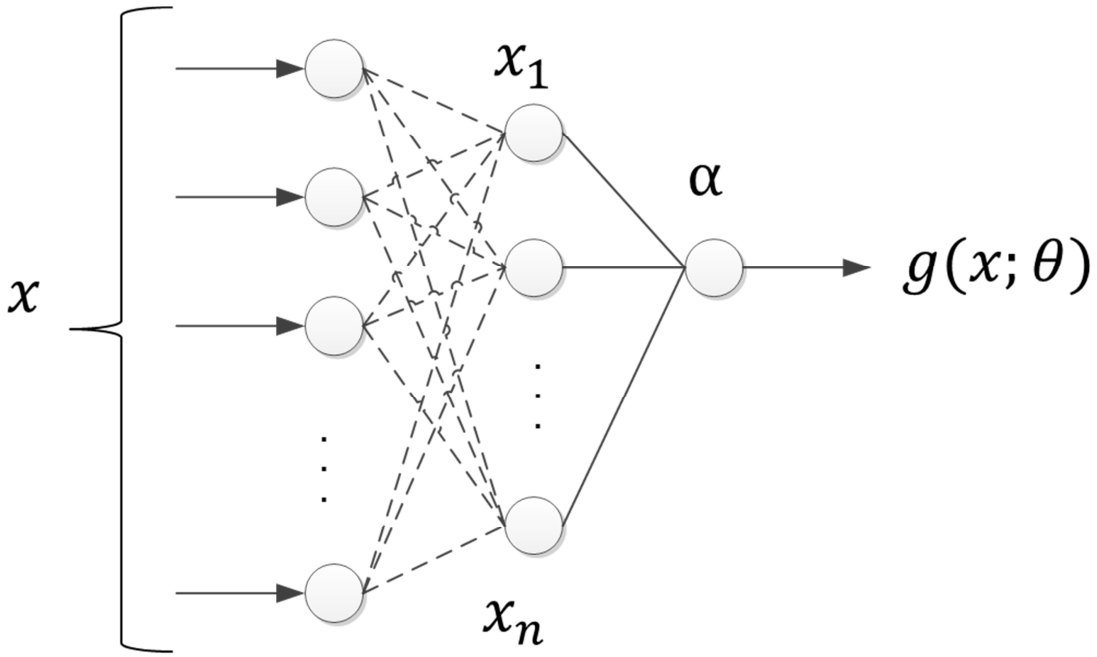


Fig. 3.1: Basic concepts of SVR/SVM

差とみなす．SVR はそれ自体が非常に汎化性能の高い回帰手法であるから，騒音クラスタリングの性能がある程度以上であれば，提案法全体の推定誤差は小さくなると考えられる．よって，騒音クラスタごとの推定関数の RMSE の重み[7]付き平均と，騒音クラスタリングを考慮しない推定関数による RMSE を比較する．そして，騒音クラスタリングを用いた RMSE の方が用いない場合よりも小さくなることを目標とする．

[7]本論文では，クラスタごとのサンプル数と総サンプルから求めた重みを用いることとした．

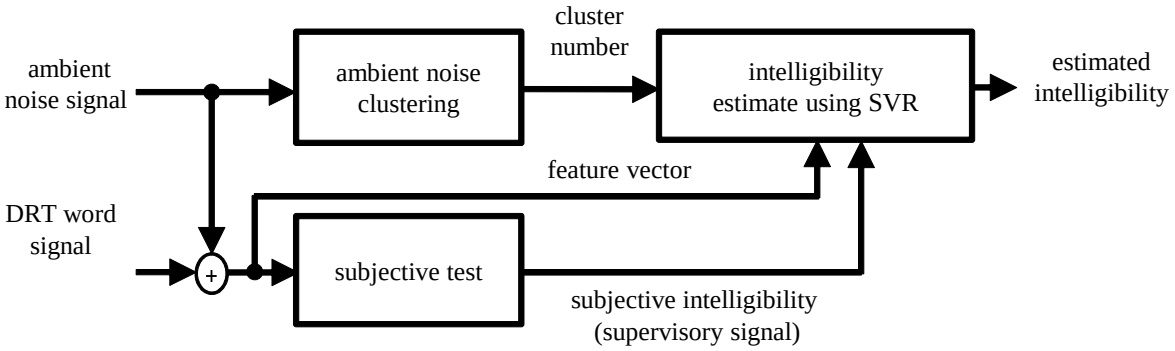


Fig. 3.2: Overview of the proposed intelligibility estimation system

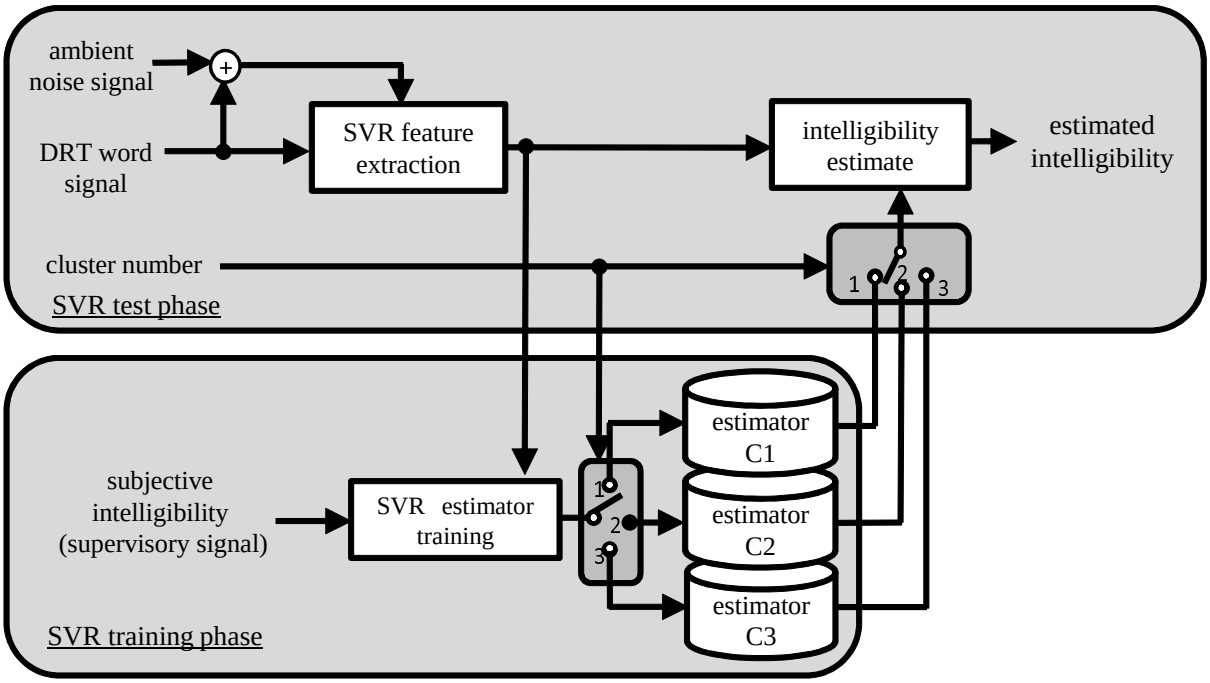


Fig. 3.3: Intelligibility estimation flow

3.2 クラスタ分析

3.2.1 クラスタ分析の種類

クラスタ分析はデータ集合をデータ間の類似度を基に，いくつかのクラスタ（データのグループ）に分けるデータ解析手法の一つである．クラスタを作成する際に，正解となる教師データを与えずに自動的に分類し，データの解析を行う．クラスタ分析には，階層的クラスタリングと非階層クラスタリング（分割最適化クラスタリング）に分けられ，非階層クラスタリングはさらに，データが属するクラスタを必ず 1 つ定めるハードクラスタリングと複数に属することを許容するソフトクラスタリングに分けられる．クラスタリングアルゴリズムとその分類を Table 3.1 に示す．

階層的クラスタリングは，個々のデータが一つのクラスタであるところから出発し，類似度の近いデータを統合して徐々にクラスタを減らし，目的のクラスタ数まで統合していくアルゴリズムである．結果の分析には Fig. 3.4 に例示するデンドログラムを用いる．比較的古典的なアルゴリズムで，単連結法，完全連結法，群平均法，ウォード法，重心法，メディアン法といったアルゴリズムがある．

非階層クラスタリングは，データ分割の程度を評価する評価関数を用いて，評価関数に対する最適解（最適分割）による分割を行う手法である．もっとも代表的なアルゴリズムに k–means[160] がある．k-means は各データの属するクラスタを 1 つだけ求めるハードクラスタリングである．ハードクラスタリングにはこのほかに，k–mean の分割数を自動で推定する x–means[161]，グラフ分割問題として解く，スペクトルクラスタリング [162, 163] がある．複数のクラスタに所属することを許容するソフトクラスタリングには，Fuzzy c–means[164]，pLSI(probabilistic Latent Semantic Indexing)[165]，NMF(Non–negative Matrix Factorization)[166] といったアルゴリズムがある．

Table 3.1: Examples of clustering algorithm

Hierarchical clustering	single linkage method, complete linkage method, group average method, Ward's method, centroid method, median method
Partitional optimization clustering	k–means, x–means, Fuzzy c–means, spectral clustering, mixture distribution model, pLSI, NMF
Hard clustering	hierarchical clustering, k-means, x-means, spectral clustering
Soft clustering	Fuzzy c-means, mixture distribution model, pLSI, NMF

3.2.2 k–means

非階層クラスタリングの代表例として k–means について述べる．k–means は以下の手順で実行される．また参考例として Fig. 3.5 に k–means の基本的動作を図解する．

1. K 個のクラスタの代表値，$c_i(i = 1 \cdots K)$ をランダムに作成する (3.5 1st step，図中では c_i を×で表記)．

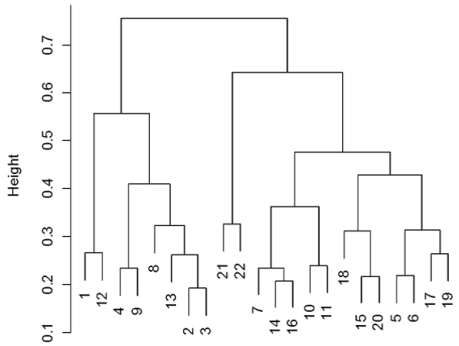


Fig. 3.4: Examples of dendrogram

2. 各データ x_n に対し，全てのクラスタ代表値 c_i までの距離を求め，最も近い c_i に対するクラスタを x_n が属するクラスタとする (3.6 1st step のデータ色参照).

3. 各クラスタの重心（セントロイド）を求め，c_i を重心座標に更新し，2 を繰り返す (3.5 2nd step).

4. 2,3 を繰り返し，クラスタ内のデータに変更が無くなれば終了する (3.5 3rd step〜Last step).

以上の動作によって，k–means はクラスタのセントロイドをクラスタの代表値として更新し続けることで，データに割り当てるクラスタを最適化する．この時の評価関数は以下の式（3.1）で求まる．ここで $\|\|$ はユークリッドノルムを示す．式（3.1）は単調非増加（単調減少）となるため，必ず解が見つかものの，局所解と大域最適解との区別はつかない．Fig. 3.5 と同一のデータに別の初期代表値を与えた例を Fig. 3.6 に示す．クラスタ数は同じながら，Last step でセントロイドの座標が異なるため，各データが所属するクラスタが異なる．

$$Err(C_i) = \sum_{i=1}^{K} \sum_{x \in C_i} \| x - C_i \|^2 \tag{3.1}$$

全てのクラスタ共通の標準偏差 σ と単位行列 I であらわされる共分散行列 $\sigma^2 I$ を考えると，各クラスタごとのセントロイド C_i の多変量正規分布 $f(x; C_i, \sigma^2 I)$ を用いて式（3.2）の混合分布を得る．

$$f(x) = \sum_{i}^{K} \alpha_i f(x; C_i, \sigma^2 I) \tag{3.2}$$

この混合分布のデータ集合 X に対する最尤推定を EM アルゴリズムで行うと，データのクラスタへの割り当てを $[0, 1]$ の範囲にある実数は混合分布と EM アルゴリズムでは許されるが，k-means では，一つのデータはいずれか一つのクラスタの要素にしかならない．クラスタが全て超球状になる．また，そのクラスタの半径はほぼ等しくなることが暗黙的に仮定されている．よってこの仮定に合わないクラスタは抽出されない．

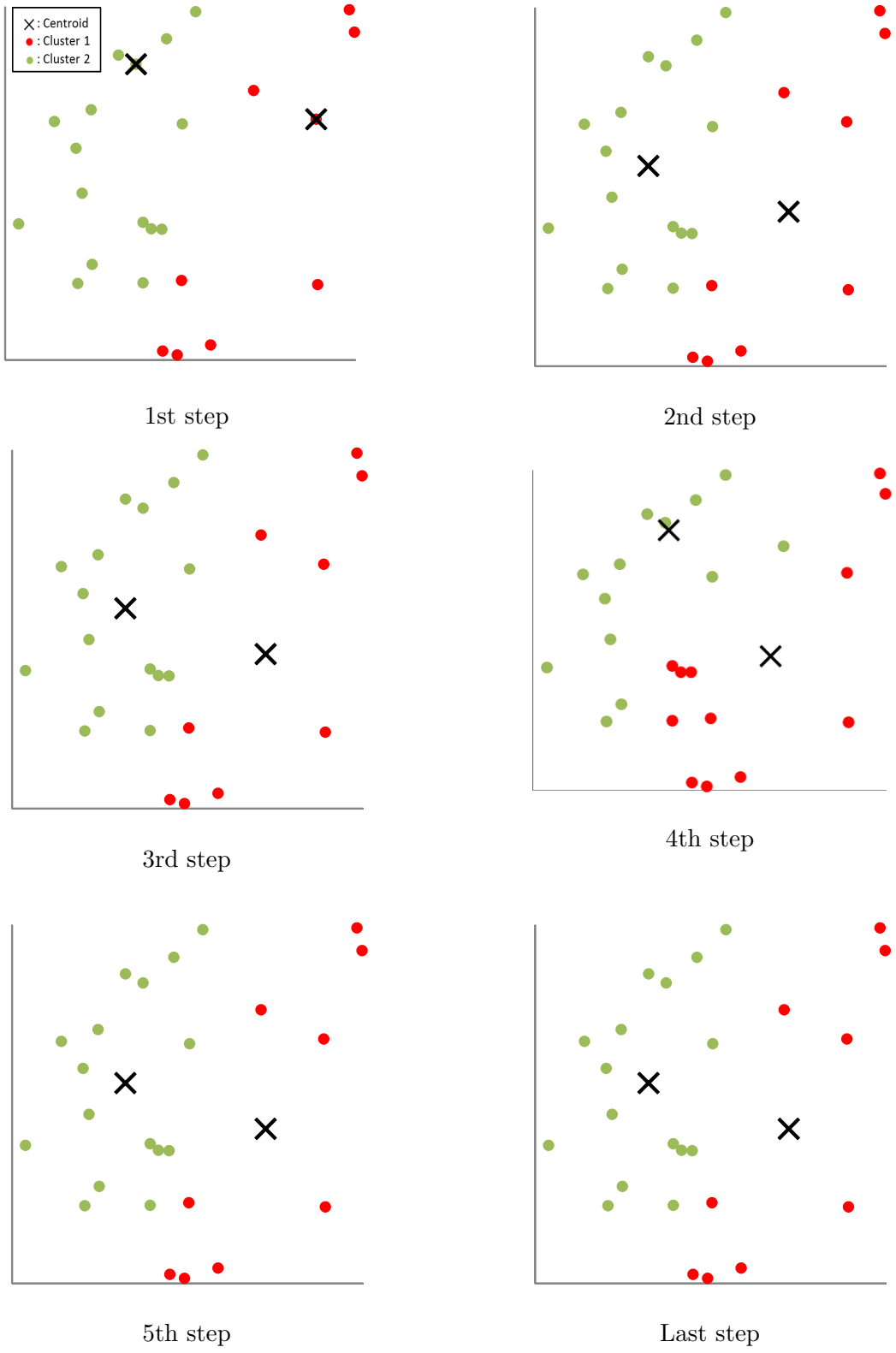
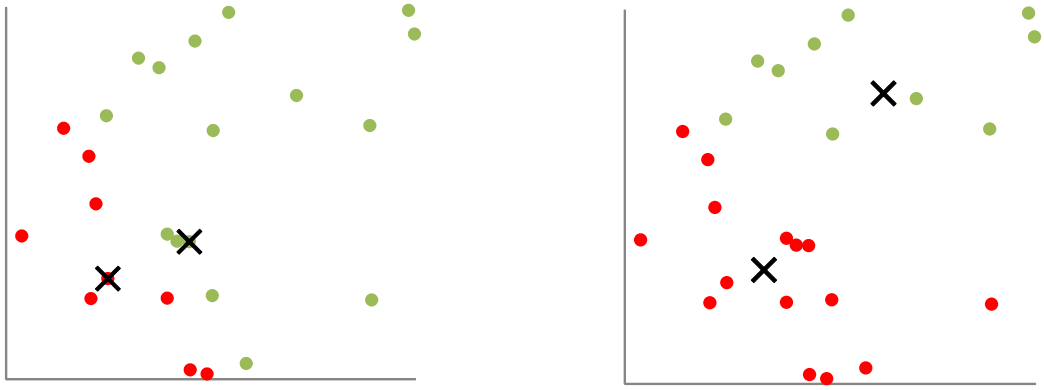


Fig. 3.5: Examples of k–means (1)



1st step　　Last step

Fig. 3.6: Examples of k–means (2)

3.2.3 x–means

x–means は，事前にクラスタ数を与えて分類する k–means によるクラスタリングを改良したもので，全データを小さな数のクラスタ数に設定した k–means によって分類し，各クラスタをさらに分割できるかどうかベイズ情報量基準（Bayesian information criterion: BIC）によって判断する．このため，いくつのクラスタに分割されるか事前にわからないため x–means と呼ばれる．x–means の例を Fig. 3.7 に示す．図では，k–means によって 2 クラスタに分けたのちに，緑色のクラスタをさらに 2 分割し，その片側をさらに分割して合計 4 つのクラスタを作成している．x–means は k–means を基に作られた方式であるため，k–means と同様に，初期値に解が依存することが知られており，特に BIC を基にした細分時に影響する．

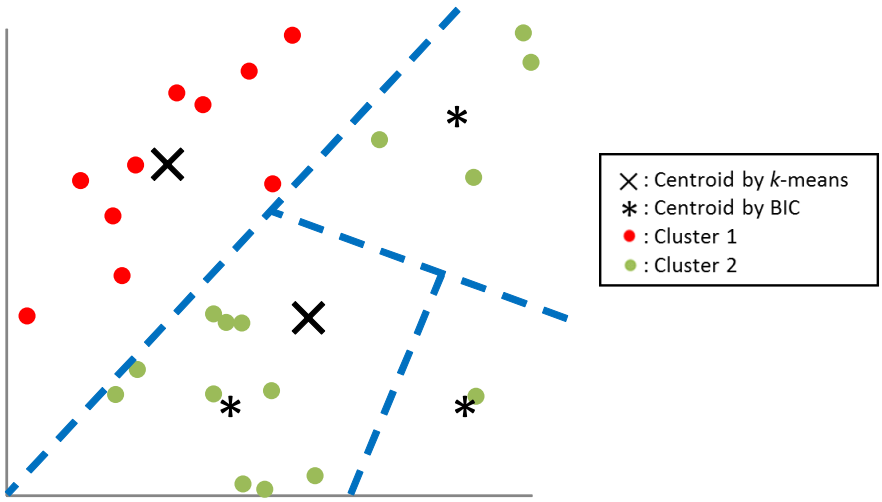


Fig. 3.7: Examples of x–means

3.3 サポートベクトル回帰

SVR は，SVM[154] を用いた回帰であり，高い汎化性能を持つ．本論文では，SVM と SVR を簡単に実装できるように開発された LIBSVM[167, 168] を，統計言語 GNU R[169, 170] 上で動作するようにした e1071 パッケージ [171] を用いた．

SVR に用いる特徴量ベクトルを $D : \{(x_i, y_i), i = 1, \cdots, N\}$ として回帰関数を特徴空間への写像 $\Phi(x)$ を用いて以下の様に表す．ここで，$\langle , \rangle$ は内積を，w は l 次元の係数ベクトル，b はバイアス項を示す．

$$f(x) = \langle w, \Phi(x) \rangle + b \tag{3.3}$$

この時の目的関数 Q は，トレーニング誤差を表すスラック変数 ξ^{ρ},$\xi^{*\rho}$ を用いて，以下のように定式化される．

$$\begin{aligned} &\text{minimize} \quad Q(w, b, \xi, \xi^*) = \\ &\frac{1}{2}||w||^2 + \frac{C}{\rho} \sum_{i=-1}^{M} (\xi^{\rho}{}_i + \xi_i^{*\rho}) \end{aligned} \tag{3.4}$$

$$\text{subject to} \begin{cases} y_i - \langle w, \Phi(x_i) \rangle - b \leq \epsilon + \xi_i \\ \langle w, \Phi(x_i) \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{3.5}$$

ここで ρ が 1 のときを L1 SVR と呼び，LibSVM では L1 SVR を扱う [168]．ϵ は $f(x)$ の広がりを示し，ϵ チューブと呼ばれる．ϵ チューブとスラック変数のイメージを Fig. 3.8 に示す．図より，$f(x) \pm \epsilon$ の範囲外での教師データの許容範囲がスラック変数である．C はマージンと教師データの近似誤差を制御するコストパラメータであり，大きな値を取ると過学習となり凡化性能が低下する．ϵ と C は回帰全体に影響するハイパーパラメータとして，最適な値を探索する．

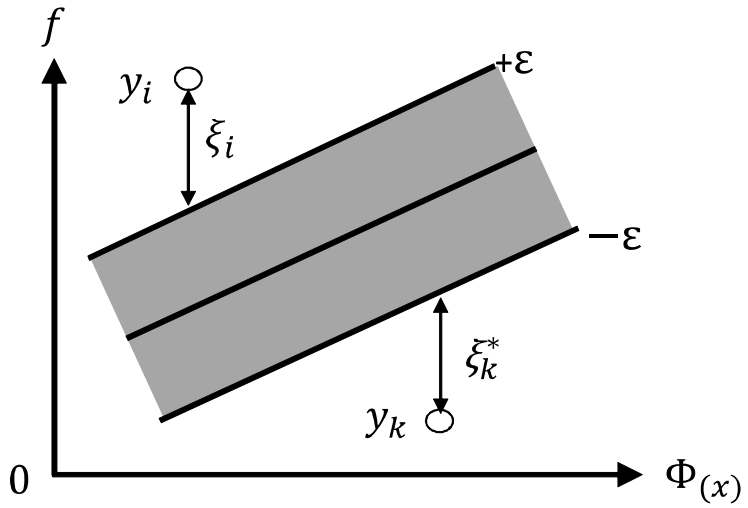


Fig. 3.8: ϵ tube and slack variable

次に，$\rho = 1$ のとき，式 (3.4)，(3.5) の内積をカーネル関数 $K(x, y) = \langle \Phi(x), \Phi(y) \rangle$ に置き換え，ラグランジュ未定係数法を用いて 2 次計画問題として解くと下 2 式が得られる．2 次計画法として解くことが可能であるため，大域最適解が求まる．α_i, α_i^* はラグランジュ乗数である．

$$\begin{aligned} &\text{maximize} \quad Q(\alpha, \alpha^*) = \\ &-\frac{1}{2} \sum_{i,j=-1}^{M} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ &- \epsilon \sum_{i=-1}^{M} (\alpha_i + \alpha_i^*) + \sum_{i=-1}^{M} y_i (\alpha_i - \alpha_i^*) \end{aligned} \tag{3.6}$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{M} (\alpha_i + \alpha_i^*) \\ 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C \end{cases} \tag{3.7}$$

また，式 (3.4) は最終的に式 (3.8) になる．

$$f(x) = \sum_{i=1}^{M} (\alpha_i + \alpha_i^*) K(x_i, x_j) + b \tag{3.8}$$

本論文では，使用するカーネル関数 K は，式 (3.9) の線形カーネルと，式 (3.10) の RBF（Radial Basis Function）カーネルを比較する．RBF カーネルの γ はハイパーパラメータとして，次章の設定で最適な値を探索する．一般に，SVM と SVR は，特徴量の数が多いときに RBF カーネルのような高次元へのマッピングが有効とは限らなく，線形カーネルで十分な場合があるとされる [167]．このため，了解度推定に線形カーネルで十分な精度が出るのならば，ハイパーパラメータが多い RBF カーネルを用いる必要はないため，本論文で比較する．

$$K(x, x') = x^T x' \tag{3.9}$$

$$K(x, x') = exp(-\gamma ||x - x'||^2) \tag{3.10}$$

3.4 交差検定

2 章では，集めた客観音質値と主観評価値のデータを 2 分割し，片方を学習データ，もう片方をテストデータとし，テストデータの推定性能でクローズドテストを行った．これは未知のデータの推定性能で尺度ごとの推定関数の性能を比較するためである．SVR を含む機械学習による回帰でも，学習に用いたデータへの過度の適応（過学習）を防ぐために同様の手法を用いる．一般に学習に用いるデータは推定したいデータよりも圧倒的に少なく，分割の仕方による性能差が考えられる．特に学習データはある程度の数が無いとそもそも十分な回帰が行えない．少数のデータしかない場合は，学習に多くのデータを割いた場合のテストデータが不足する．このためデータセットを複数のブロックに分割し，1 ブロックをテストデータに，残りを学習データとした回帰を分割した数で行い，その平均を取る交差検定法 [172] が用いられる．交差検定法のイメージを Fig. 3.9 に示す．分割するブロックあたりのデータは同数程度になるように設定する．交差検定は回帰だけでなく識別にも用いられ，機械学習研究における基本的なモデル作成手法である．

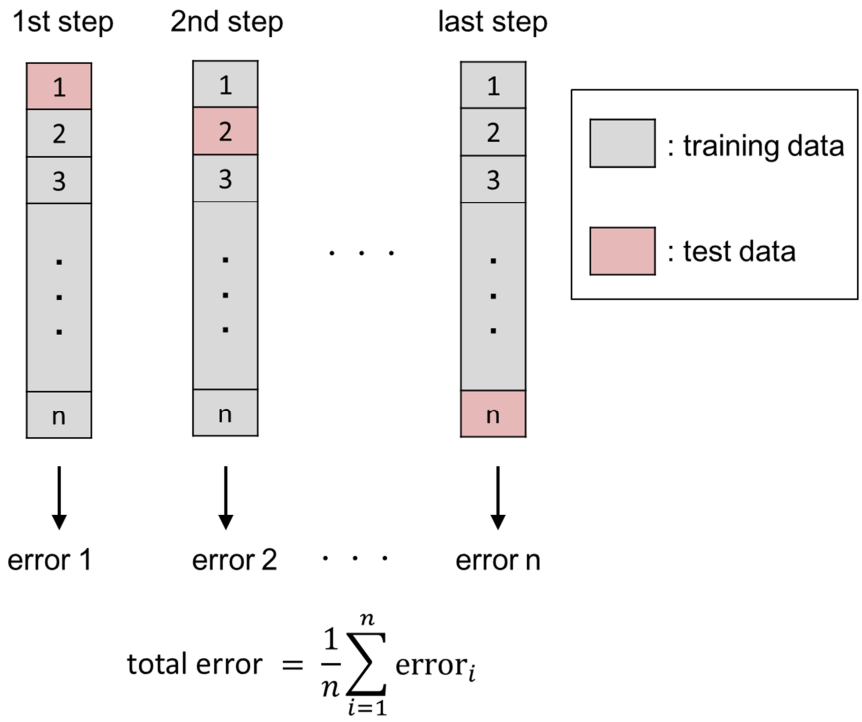


Fig. 3.9: Example of cross–validation

3.5 まとめ

2 章で課題となった，騒音種の影響を受けやすい JDRT の Sustention 子音特徴の推定性能向上について検討し，以下の内容について新たな手法を提案した．

1. 騒音種による最適な推定関数の選択

2. 推定に最適な聴覚重みの作成

1 については，騒音信号の特徴量を用いて分類する騒音クラスタリングを提案した．騒音クラスタリングの詳細と主観評価における性能分析は 4 章で行う．1 については，最適な回帰係数を SVR を用いて求める推定法について提案した．SVR を用いた推定関数の作成とその評価については 5 章で述べる．最後に提案推定システムの総合評価については 6 章で述べる．

第4章　騒音クラスタリングの検討とその評価

本章では，本論文で用いる騒音について述べたのち，3 章で提案した騒音クラスタリングを実装し，分類結果と主観評価結果を比較する．また，2 章と同様の手法で推定関数の作成を行う．

4.1　検討する騒音データベース

4.1.1　解析する騒音種

本論文では，4 章と 5 章のモデル作成では電子協騒音データベース [157] のダイジェスト版を，6 章のオープンテストではフルセット版を用いる．データベースに収録されている騒音のうち，検聴表を基にステレオ録音されている音源を選択した．ダイジェスト版ではフルセット版の中から代表的な箇所を抜き出したセット，その中から Table 4.1 に示す 18 種を使用した．フルセット版は Table 4.2 に示す．項目数がダイジェスト版より少ない 13 種になっているのは，フルセット版で 1 項目の騒音であった騒音のうち代表的なところをで分割したためで，ダイジェスト版で使用される騒音はすべて含まれる．

Table 4.1: JEIDA Noise database (digest set)

name				
exhibition booth 1	exhibition booth 2	telephone booth	factory 1	factory 2
sorting facility	highway 1	highway 2	crossing	crowd
bullet train	train	computer room	air conditioner 1	air conditioner 2
air duct	elevator hall 1	elevator hall 2		

Table 4.2: JEIDA Noise database (full set)

name				
exhibition booth 1	exhibition booth 2	telephone booth	factory	sorting facility
highway & crossing	crowd	bullet train	train	computer room
air conditioner	air duct	elevator hall		

4.1.2　騒音間のパワー統制

収録時にマイクバイアスされている騒音は，平均パワーを引いてバイアスをキャンセルをした．そののち，騒音種（ダイジェスト版ではCDトラック毎，フルセットでは同一騒音種のCD2枚ごと）の平均音圧と音声のパワーを統制する．Fig. 4.1とFig. 4.2にダイジェストとフルセット版の騒音統制フローを示す．図中で用いたdB(A)とdB(C)はそれぞれA特性，C特性のパワー比をである．データベースに収録されている基準ノイズ[1]を用いて18，13種の音源のC特性パワーをそろえた．そして，次章で述べる主観評価で使用する音圧に設定するために，日本語DRTの評価単語120単語の2話者分，240単語の平均音声パワーとA特性パワー[2]が等しくなるようにゲイン調整を行った．ゲイン調整後の騒音をSNRが0 dBと定義した．よって，同一の騒音種でもLFごとのレベルが異なる．

ダイジェスト版では，列車走行音2種（新幹線，在来線）に関しては走行位置と非走行位置のパワー差が顕著であったため，走行音を切り出して使用した．フルセット版では，後述するLFごとの騒音のA特性パワーの小さい下位10%は破棄し，上位90%を用いた．基準が異なるのは，ダイジェスト版では新幹線と列車を除いて，騒音内のパワーはほぼ統制されていたが，フルセットではパワーの分散が大きかったためであり，セットごとにパワー統制を変えた．

[1]中心周波数1.2 kHzの1/1オクターブバンドノイズ．

[2]騒音間のパワー統制には，電子協騒音データベースの仕様に合わせて，大音量向けのC特性を用いた．一方，評価単語は通常発話を録音したものである．本論文では主観評価で用いる音声に合わせ，C特性ではなく，一般用途向けのA特性で統制した．

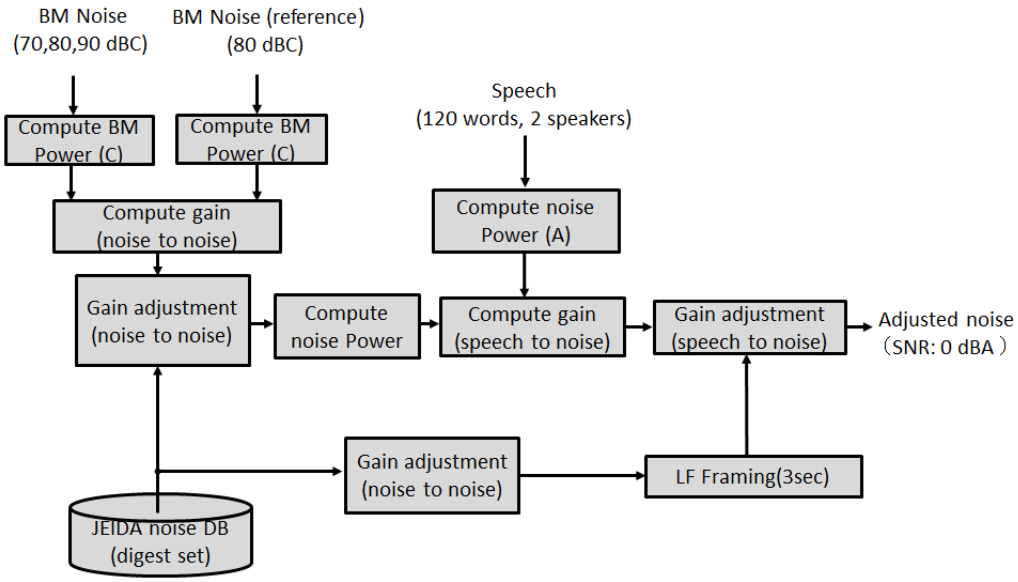


Fig. 4.1: Signal power adjusting flow(digest set)

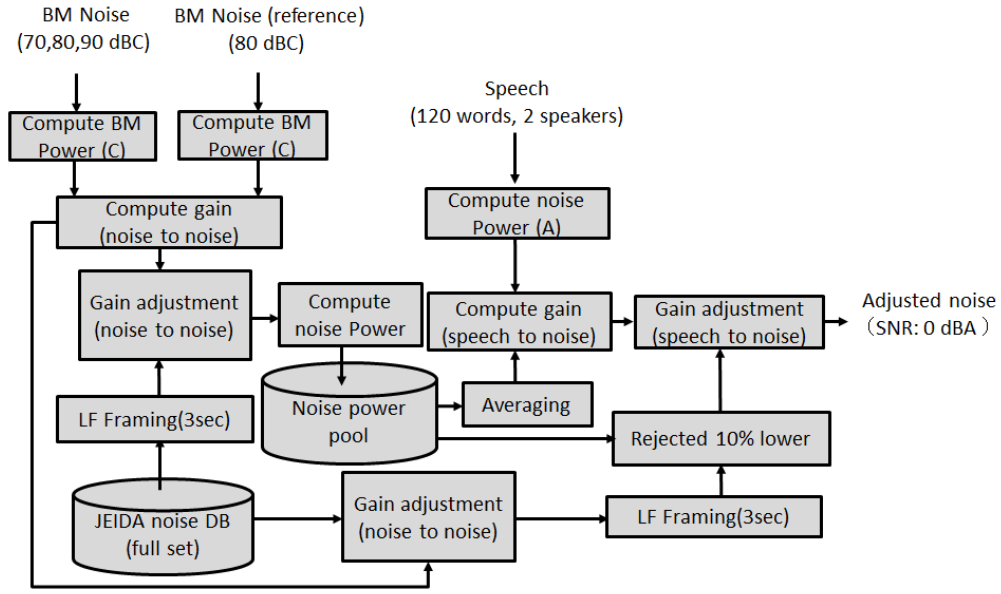


Fig. 4.2: Signal power adjusting flow(full set)

4.2 LF 分割による騒音クラスタリング

前節で分類した騒音について，JDRT の Sustention 単語セットを用いて了解度試験を行う．

4.2.1 LF 分割

本論文では音声了解度試験結果が同傾向になる騒音の分類を目的とするため，了解度試験で用いる目的音の時間よりも極端に長い騒音は必要なく，長時間収録した騒音音源のうち，了解度が悪くなる条件を評価できればよい．電子協騒音データベースダイジェスト版は，CD トラック 1 つにつき，騒音が 1 種録音されている．CD のトラックごとに了解度試験に合わせて 3 sec で区切って，その一つ一つを独立した騒音音源とみなして解析を行う．3 sec に分割した騒音をロングフレーム（以下，LF）と定義する．また，特徴量を求める際に使用する短時間フーリエ変換で用いるフレーム長さを 0.1 sec とし，本論文では LF と区別してショートフレーム（以下，SF）と呼ぶ．各騒音音源の冒頭部と終端部のそれぞれ 2 sec は無音部があるため，分析から除外した．騒音音源の LF と SF のイメージを Fig. 4.3 に示す．

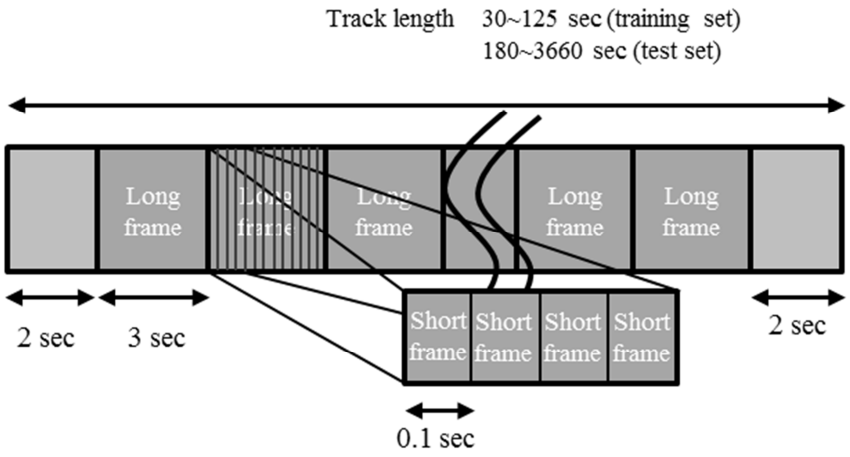


Fig. 4.3: Image of the frame segmentation

4.2.2 音色特徴量

本研究で用いる MIR 特徴量は MIRtoolbox[173] に収録されているものを利用した．MIRtoolbox は，MIR 分野で広く使われる特徴量を MATLAB 上で扱えるようにしたライブラリであり，本論文では収録されている中から 8 種類の特徴量を選択した．選択した特徴名と説明を Table .4.3 に示す．音源に対して 1 つの値を求める Tempo と Attack time，Loudness 以外の 6 特徴は，50% のオーバーラップがある SF で求めた特徴量の平均値と標準偏差を取り，合計で 15 次元の特徴量ベクトルを LF ごとに求めた．SF で求める特徴量は，Zero–cross を求めた際に零交差が存在しないなどの値が求まらない SF については平均と標準偏差を求める際に除外する．特徴量は式 (4.1) 次元ごとに全 18 騒音の最大値と最小値で正規化を行った．

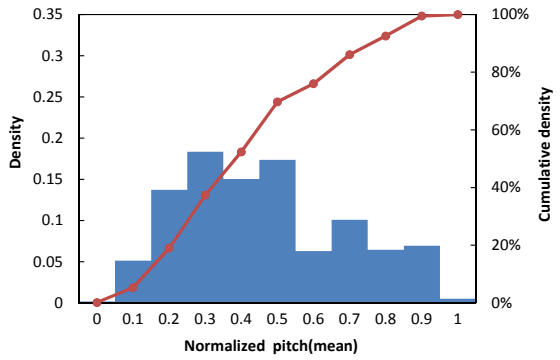
$$\text{normalized MIR feature score} = \frac{\text{MIR feature score} - \text{min score}}{\text{max score} - \text{min score}} \tag{4.1}$$

選択した特徴量を概説する．Pitch は音源の基本周波数 (F_0) であり，自己相関法で SF ごとに求めている．Zero–cross は SF ごとに時間波形で零交差（極性変化）の数を求める基本周波数導出法の一つである．Brightness は SF ごとのパワースペクトルの 1.5 kHz 以上のエネルギー割合である．Flatness，Spectral centroid，Spread spectrum はそれぞれ SF ごとのパワースペクトルの平坦さ，重心，標準偏差で，パワースペクトルの形状を求めている．Tempo，Attack time は波形の時間特徴で，それぞれ時間波形の自己相関ピークから求めたテンポと，時間波形のピークとノッチの間隔を求めている．Loudness だけは MIRtoolbox に含まれる特徴量ではなく，ITU–R BS.1770–2 で求めたラウドネスレベル（Loudness, K–weighting, Full Scale:LKFS）を用いた．MIRtoolbox では，音の大きさに対応した特徴量として解析信号の RMS パワーが含まれている．しかし，音の大きさに関してラウドネスレベルでは，ゲーティング処理により信号中に存在するものの，人間には聞こえていない部分のパワーを除いている．このため単純な RMS パワーよりも人間の主観とよく対応しており，より最適な特徴量と考えられるラウドネスレベルを採用した．

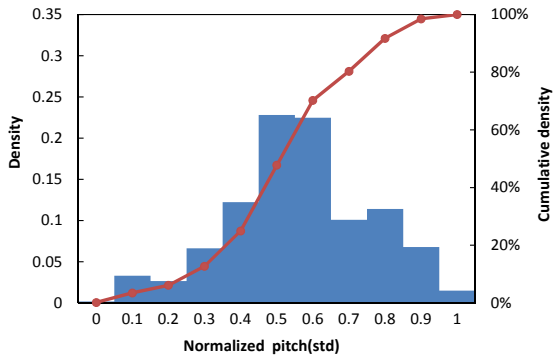
Fig. 4.4(a)〜(i) に各特徴ごとに正規化特徴量の頻度ヒストグラムと累積度数分布を示す．フリードマン検定を用いてヒストグラムを基に 15 の特徴間の有意差検定を行うと，$\chi^2(14) = 23.6819, p = 0.05004$ であり，ヒストグラムに要約した結果でも $p \simeq 0.05$ であること，605 騒音の実測値を用いた場合には $\chi^2(14) = 5564.573, p = 2.2e^{-16}$ であった．実測値を用いた検定でも特徴間に有意差がみられない組み合わせも一部には見られたが，独立した特徴量であるとして騒音クラスタリングに用いることとする．

Table 4.3: MIR features

feature name	content
Pitch	fundamental frequency (F_0)
Zero–cross	time waveform sign–change rate
Brightness	high frequency energy (over 1.5 kHz)
Flatness	flatness of power spectrum
Spectral centroid	centroid of power spectrum
Spread spectrum	standard deviation of power spectrum
Tempo	tempo by autocorrelation
Attack time	time difference between notch and peak of the spectral envelope
Loudness	ITU–R BS.1770–2

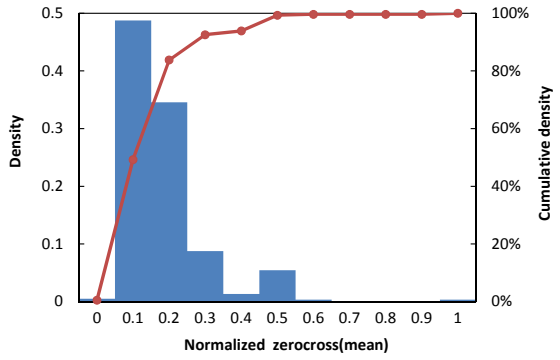


Mean

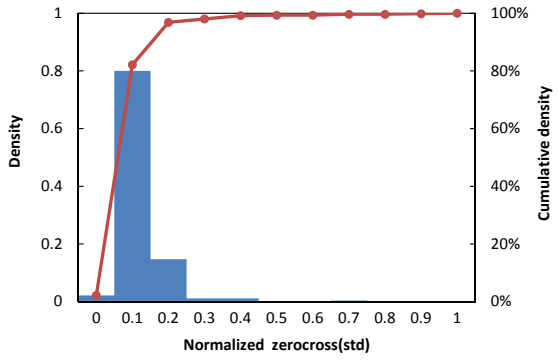


Standard deviation

(a) Pitch

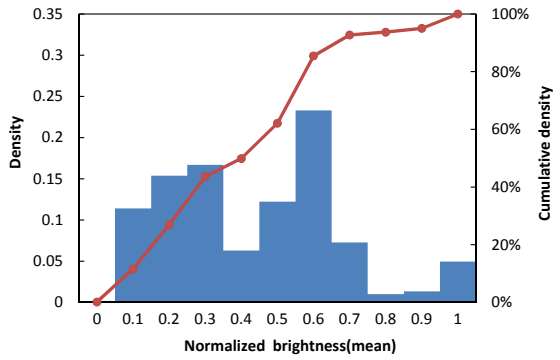


Mean

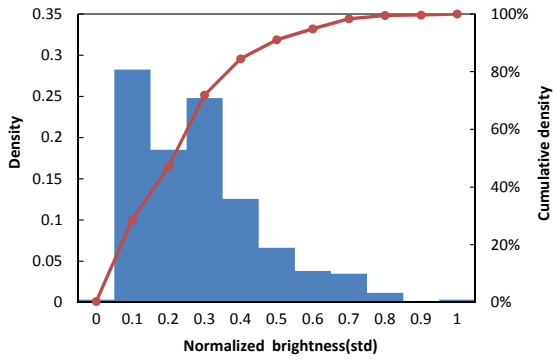


Standard deviation

(b) Zero–cross



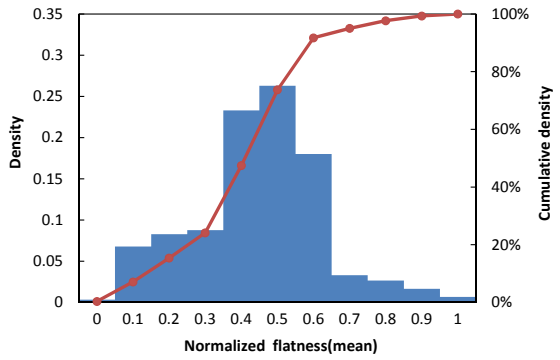
Mean



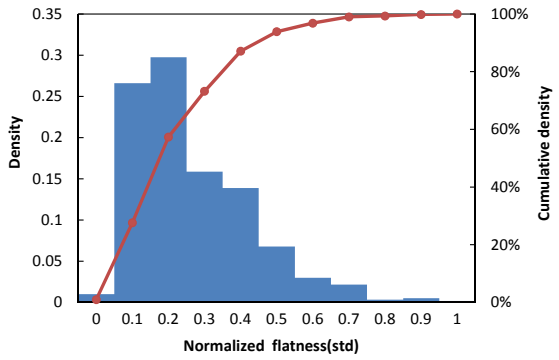
Standard deviation

(c) Brightness

Fig. 4.4: Histogram of MIR features

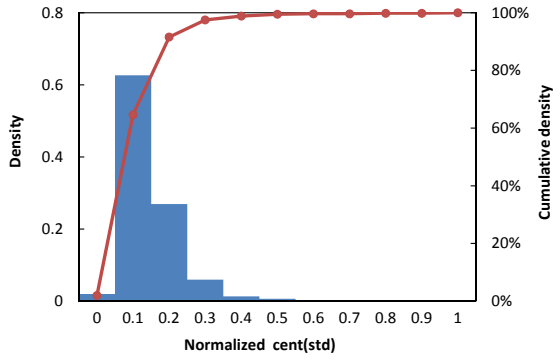


Mean

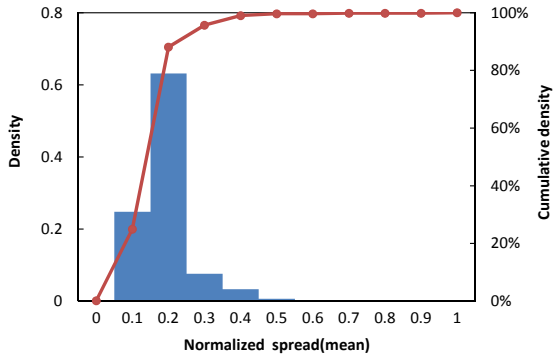


Standard deviation

(d) Flatness

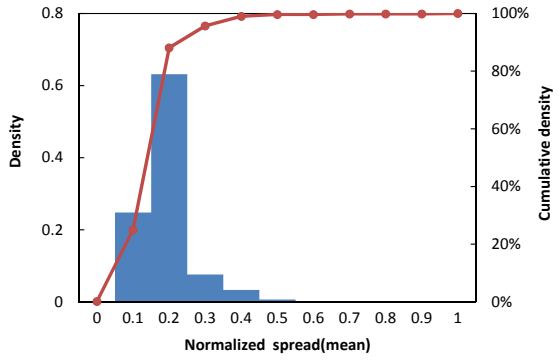


Mean

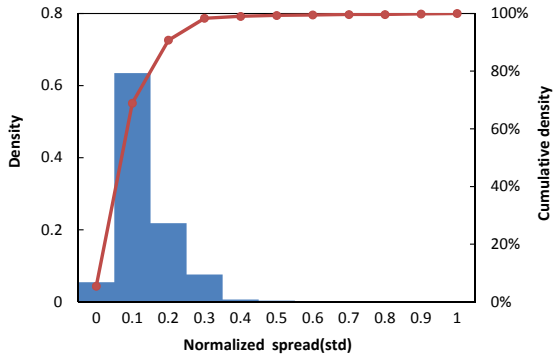


Standard deviation

(e) Spectral centroid



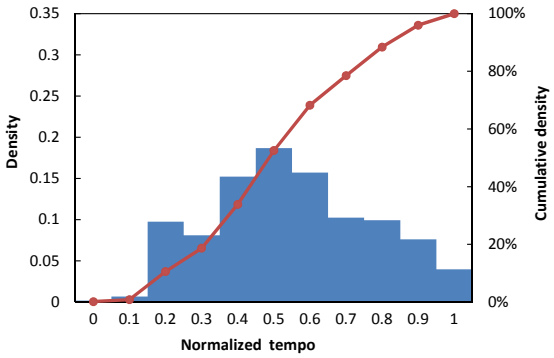
Mean



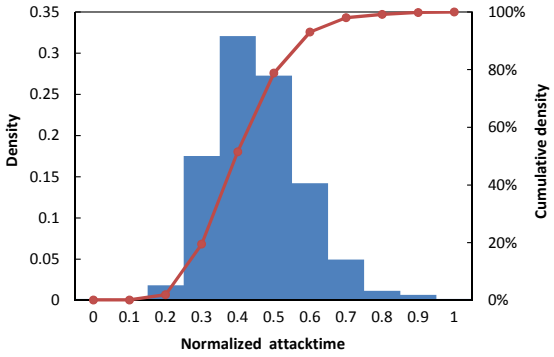
Standard deviation

(f) Spread spectrum

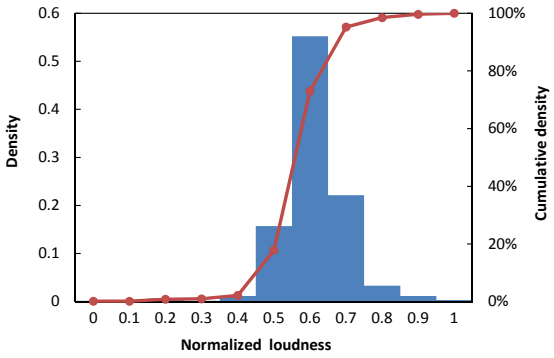
Fig. 4.4: Histogram of MIR features(cont'd)



(g) Tempo



(h) Attack time



(i)Loudness

Fig. 4.4: Histogram of MIR features(cont'd)

4.2.3 クラスタリングアルゴリズム

一般にクラスタリングの結果を解析するには，分類する評価値（本論文では了解度）が事前に求まっており，それに対しどこまで正しく分類されたかを評価する．しかし，了解度を求めるためには，評価環境ごとに数十〜数百単語（本論文では 20 単語）の評価が必要であり，事前に全ての環境の了解度を求めておくことは非常に困難である．このため，適切であると考えられる特徴量で LF のクラスタリングを行い，結果の妥当性を主観評価で確認することとする．また，事前にクラスタ数がいくつになるかわからないため，クラスタリングアルゴリズムの選択を考える必要がある．この場合,「クラスタ数を変化させながら，最適なクラスタ数を見出す」手法と,「分類するクラスタ数を与えずにクラスタリングが可能なアルゴリズムを選択する」手法の二つが考えられる．本論文では，明確な基準でクラスタ数を定めることが可能な後者の手法によるクラスタリングアルゴリズムの x–means 法を採用した．

x–means クラスタリングはデータマイニグツール Weka [174] に含まれているアルゴリズムを用いる．セントロイドと特徴ベクトルとの距離計算にはユークリッド距離を用いた．一般に，x–means アルゴリズムは初期乱数によって生成されるクラスタ数が変わることから，本論文では乱数列を切り替え，最も選択されたクラスタ数のうち，BIC が最大になる結果を採用した．その他のパラメータは初期設定を利用した．

4.2.4 分類結果

騒音クラスタリングの結果を Table 4.4 に示す．x–means 法は初期設定で与える乱数によって結果が変化する．初期値を複数与え，傾向を解析したところ，全体の 70%がクラスタ数 3 となり，残りは 2 か 4 だった．本論文では最頻値であるクラスタ数 3 を採用し，用いる乱数セットは BIC が最も大きくなる設定とした．各クラスタを以後 C1〜C3 と呼ぶ．騒音名と，各クラスタに所属した LF 数を示す．空欄は該当クラスタに所属した LF が無いことを示す．前述のとおり，列車走行音 2 種は走行音を切り出したため，LF 数は他より少ない．Total は各クラスタの LF 総数である．

C1 は，全 LF の半数が分類されるクラスタであり，騒音レベルの時間変動がみられる騒音が多く分類されているが，時間変動が少ない computer room の LF もすべて分類されている．computer room は定常ではあるが，パワースペクトルの 4 kHz 付近に常に極所ピークを持つため，Flatness や Spectral centroid，Spread spectrum が他の定常騒音と異なる結果になった．また，factory 2 の様な間欠騒音も分類されている．全体として，“非定常” または，“うるさい” 騒音が含まれるクラスタと言える．C2 は，定常騒音である air conditioner 2，air duct の LF がすべて分類され，比較的定常である telephone booth，crowd の最頻分類クラスタである．bullet train に関しては，新幹線通過中の LF がすべて所属している．全体的に，“定常” な騒音が含まれるクラスタと言える．C3 は x-means 法で C2 から分割されたクラスタであるため，特徴量の傾向は C2 に近く全体の数も少ない．exhibition booth 1，exhibition booth 2 の音声や雑踏音が非常に複雑に混ざった騒音と air conditioner 1 の殆どの LF が分類される．

次節以降の主観評価で用いる LF のスペクトログラムは付録 B の Fig. B.1 に示す．

Table 4.4: Number of LF by clustering (digest set)

name	C1	C2	C3	LF num
exhibition booth 1		1	36	37
exhibition booth 2		8	29	37
telephone booth	4	33		37
factory 1	32	4		36
factory 2	36			36
sorting facility	35	1		36
highway 1	36			36
highway 2	34	2		36
crossing	22	8	6	36
crowd	1	35		36
bullet train	5	1		6
train	2	17	1	20
computer room	37			37
air conditioner 1		2	35	37
air conditioner 2		36		36
air duct		37		37
elevator hall 1	34	4		38
elevator hall 2	31			31
Total	309	189	107	605

Table 4.5: LFs used for the subjective test

	C1	C2	C3	Total
Num.	13	14	5	32
Percentage	0.406	0.438	0.156	1.000

4.3 主観評価結果との比較

前節で分類した騒音について，JDRT の Sustention 単語セットを用いて了解度試験を行う．

4.3.1 主観評価設定

主観評価には，騒音の影響が大きかった Sustention の 20 単語を女性話者 1 名分用いて主観評価を行う．Sustention 評価単語対を Table 4.6 に再掲する．評価に用いる LF は，Table 4.4 の結果より，全 18 騒音の分類されたクラスタから一つづつ（Table 4.4 の空欄以外）LF を選択し，32 個の LF で主観評価を行う．クラスタごとの選択した LF 数と総数に対する割合は Table 4.5 に示す．主観評価に用いた同一の騒音中の LF でも，同一クラスタに分類される LF は複数あるため，各ク

ラスタのセントロイドまでの距離が最も近い LF をそのクラスタの代表 LF とした[3].

Table 4.6: Sustention word list(cont.)

ハシ（hashi）–カシ（kashi）	ハタ（hata）–カタ（kata）
シリ（shiri）–チリ（chiri）	ヒル（hiru）–キル（kiru）
スキ（suki）–ツキ（tsuki）	スナ（suna）–ツナ（tsuna）
ヘン（hen）–ケン（ken）	ヘリ（heri）–ケリ（keri）
ホシ（hoshi）–コシ（koshi）	ホル（horu）–コル（koru）

音声と騒音の A 重み付パワー比は 4.1.2 項で述べた方法を用いて統制したのちに，A 特性を用いた SNR(A)[4]で −20，−10，0，10，20 dB になるように計算機上で加算した．この実験条件は 2 章の実験系と大きく異なるが，実験する騒音数を増やすこと，2.2.2 項の結果より，騒音の方位角よりも音声との $\mathrm{SNR_{in}}$ が重要であったことから，方位角を削除した．また，各騒音の SNR(A) は Fig. 2.15 の結果を基に，主観評価結果に天井効果とフロア効果が確実に入るように考慮した．これにより，次章で検討する SVR において学習データ内の了解度変化に非線形性が含まれることが期待される．

3 sec の LF に対し評価音声が短いことから，音声の埋め込み位置は LF の冒頭 0.1 sec を除く区間で，SNR(A) が小さくなるタイミングで音声と合成した．埋め込み位置探索は，評価音声のうち最も長い単語の半分の時間（最長モーラを想定）とした．

本実験における総評価単語数は，評価 LF 数が 32，SNR(A) が 5 種，評価単語が 20 単語の 3200 単語であり，これに騒音を加算していない原音 20 単語をリファレンスとして加え，被験者一人当たり 3220 単語の評価を行った．また，急峻な音圧変化による被験者への負担を減らすため，各 LF の冒頭 100 msec に対し緩やかに音圧が上昇していく時間窓を掛けた．実験音声はコンピュータから Roland 社製 USB オーディオインターフェース UA–25EX を介し，Sennheiser 社製ヘッドホン HD–25II で提示した．全被験者に対する提示音圧は一定とし，騒音を加算していない原音が十分聴こえる音圧で実験を行った．なお，被験者は 20 代男性 10 名である．

4.3.2 主観評価結果

主観評価の結果について，クラスタ間に着目した場合と騒音種間に着目した場合とに分けて述べる．

クラスタ間の結果

Fig. 4.5 にクラスタ別に SNR(A) ごとに平均した了解度と MCI を示す．平均了解度についているエラーバーは該当 SNRSNR(A) での最大値と最小値である．また，騒音を付加していない原音の了解度は 0.975 となり，提示音圧の不足による了解度低下は見られなかった．−20 dB では了

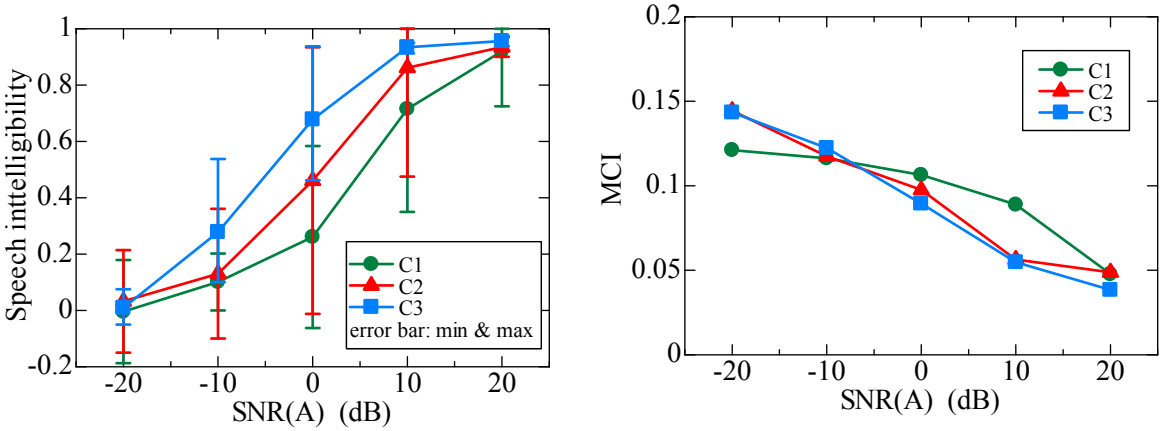
[3]3.1.1 項でも述べたように，理想的には全ての LF の主観評価値を求めることが望ましいが，1 人当たりの評価単語数が 6 万を超えてしまい現実的ではない．

[4]本章以降は実験設定の SNR を A 特性を用いて計算し，設定音圧も異なるため，2 章の $\mathrm{SNR_{in}}$ とではなく SNR(A) とする．

解度がマイナスになっているサンプルが多い．なお，これは DRT の調整正答率を求める際のチャンスレベル補正による[5]．

結果より，SNR(A) が 20 dB と −20 dB では天井効果とフロア効果により，クラスタ間の差が無くなっているが，SNR(A) が −10〜10 dB の範囲ではクラスタごとの平均値が明確になり，0 dB のときにクラスタごとに了解度が 0.2 づつ異なる．天井効果とフロア効果の影響が無い −10〜10 dB の範囲で，SNR(A) とクラスタの違いを被験者内要因とした 2 要因の分散分析を行った．分散分析の結果を Table 4.7 に示す．SNR(A) とクラスタ共に p <0.001 で有意差がみられた．特にクラスタ間の差が有意であるため，下位検定の結果を比較する．Table 4.8 に多重比較の結果を示す．結果より，SNR(A) と 3 クラスタのすべての組み合わせで p <0.001 の有意差がみられた．各クラスタの代表騒音の主観評価で各クラスタ間の差が有意であり，クラスタごとの了解度推定関数を用いた了解度推定が有効であると考えられる．

MCI の SNR（A) ごとの傾向は，C2 と C3 はほぼ重なり，C1 は僅かに異なるが有意差はない（$F(2, 14) = 0.339, p = 0.7220$）．C1 は SNR(A) が −10 dB，−20 dB のときの了解度が小さく，フロア効果で MCI が増加しなく，10 dB が C2，C3 よりも大きいのは C2，C3 と異なり天井効果が得られなかったためと考えられる．クラスタごとの了解度推定の目標値となる全 SNR(A) 混合での MCI 一覧を Table 4.9 に示す．all はクラス分けを考慮しない全サンプルでの MCI を示す．この値を了解度推定の目標値とする．



(a) Intelligibility　　(b) MCI

Fig. 4.5: Comparison of intelligibility and MCI with cluster

[5]2 章や文献 [21] での結果と異なり，本論文では SNR(A) を −20 dB まで設定したこと，評価話者が 1 名であることが影響している．また，単語対による騒音の影響差も見られたものの，詳細な分析を行うには，サンプル数が十分ではなかった．一方，クラスタごとの了解度差，推定性能差は十分に見られたことから，本論文では了解度がマイナスの場合も含め，そのまま解析する．

Table 4.7: Results of ANOVA by noise cluster

Source	SS	df	MS	F	p
S:Subject	0.3190383	9	0.0354487		
L:SNR(A)	3.2717975	2	1.6358987	134.396	0.0000****
error[L×S]	0.2191006	18	0.0121723		
C:Cluster	0.5576314	2	0.2788157	25.124	0.0000****
error[C×S]	0.1997586	18	0.0110977		
L×C	3.2920924	4	0.8230231	66.979	0.0000****
error[L×C×S]	0.4423632	36	0.0122879		
Total	8.3017819	89			

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

Table 4.8: Main effect of SNR(A) and noise cluster

Source	SS	df	MS	F	p
SNR(A)(C1)	1.7773867	2	0.8886933	72.550	0.0000****
SNR(A)(C2)	2.2328876	2	1.1164438	91.143	0.0000****
SNR(A)(C3)	2.5536156	2	1.2768078	104.235	0.0000****
error		54	0.0122493		
Cluster(−10 dB)	3.3909058	2	1.6954529	142.581	0.0000****
Cluster(0 dB)	0.2232088	2	0.1116044	9.386	0.0003****
Cluster(10 dB)	0.2356091	2	0.1178046	9.907	0.0002****
error		54	0.0118911		

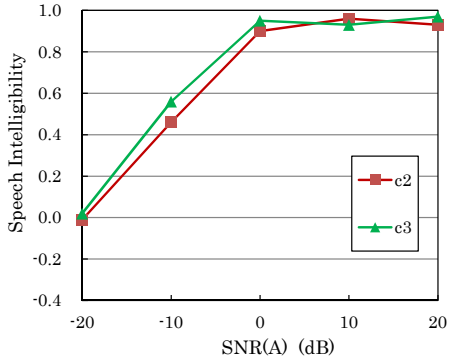
+ p <0.1, * p <0.05, ** p <0.01, *** p <0.005, **** p <0.001

Table 4.9: Average MCI by clustering

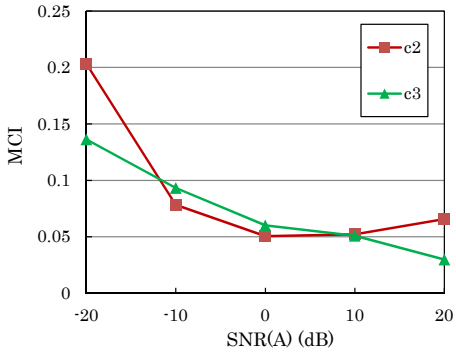
C1	C2	C3	all
0.0961	0.0929	0.0897	0.0937

騒音種とクラスタの関係

次に，同一騒音を分割した LF で，異なるクラスタに分類された場合に了解度へ与える影響が異なるのかどうかを確認する．個々の LF の主観評価結果について Fig. B.1 に示す．分析した全 18 騒音のうち，4 騒音についてはクラスタ間に有意差が見られなかった Fig. 4.6 に騒音種ごとに了解度平均値と MCI を示す．exhibition booth 1 の様にクラスタ間に差が無い例と，exhibition booth 2 の様にクラスタ間の差が顕著な例がある．exhibition booth 1，crowd，air conditioner 1，elevator hall 1，の 4 種はクラスタ間に有意差が見られなかった．この他に同一騒音種で 2，3 クラスタに分かれたものは，全部または一部の SNR(A) で有意差がみられた．MCI は Fig. 2.15 の結果と同様に，SNR(A) と負の相関を持つものと，factory 2 の様に，0 dB または ±10 dB で最大となるものに分かれる．−20 dB 以外で MCI 最大になる騒音は了解度もフロア効果が見られ，−10 dB とほぼ同値になる．全ての LF で，天井効果またはフロア効果のどちらか，あるいは両方が見られており，非線形関数での回帰を検討する必要がある．

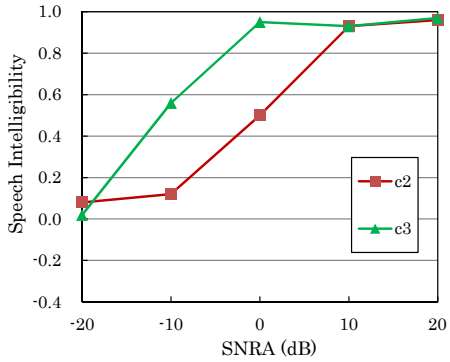


Intelligibility

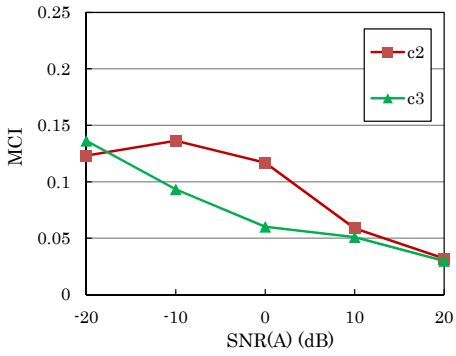


MCI

(a) exhibition booth 1

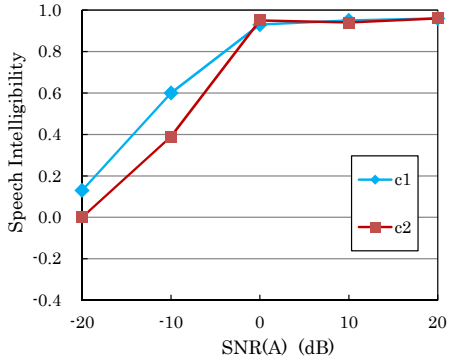


Intelligibility

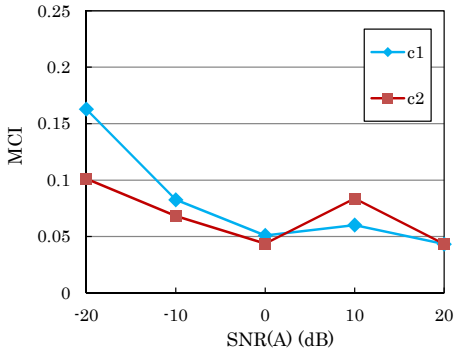


MCI

(b) exhibition booth 2



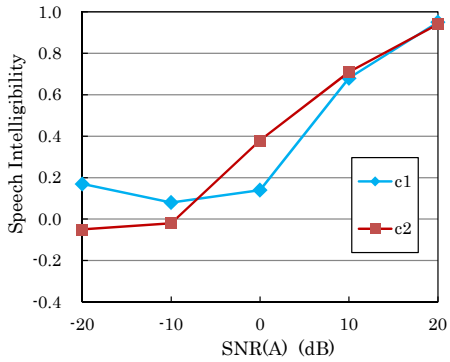
Intelligibility



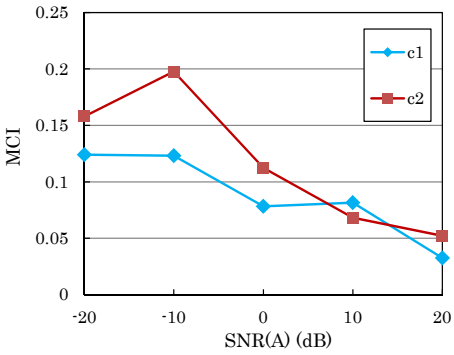
MCI

(c) telephone booth

Fig. 4.6: Comparison of intelligibility and MCI with various noise type.

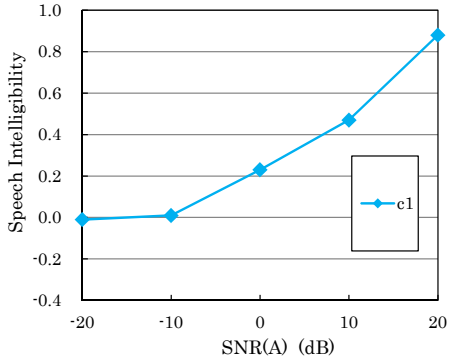


Intelligibility

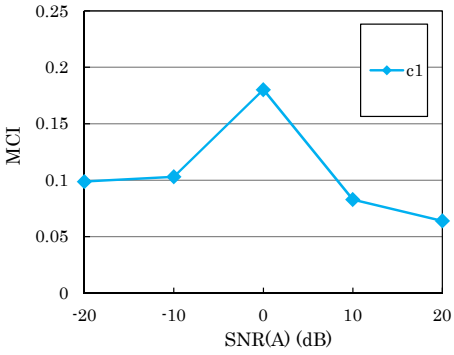


MCI

(d) factory 1

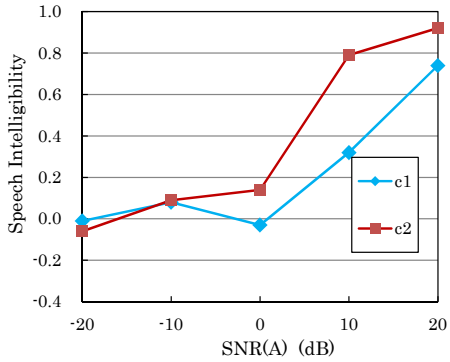


Intelligibility

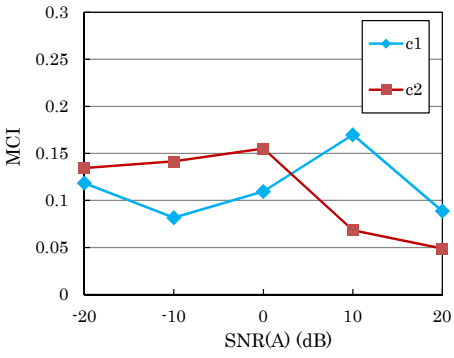


MCI

(e) factory 2



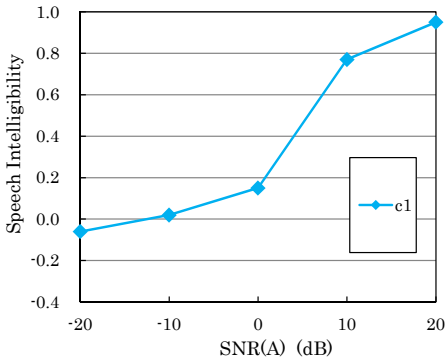
Intelligibility



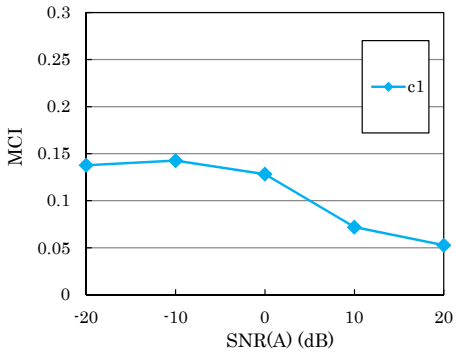
MCI

(f) sorting facility

Fig. 4.6 Comparison of intelligibility and MCI with various noise type (cont'd)

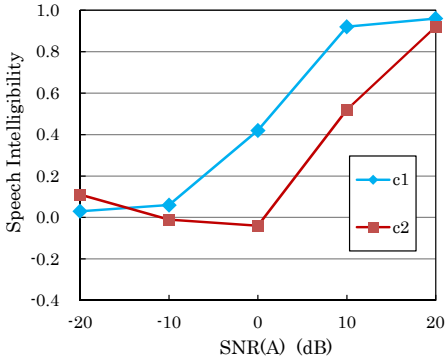


Intelligibility

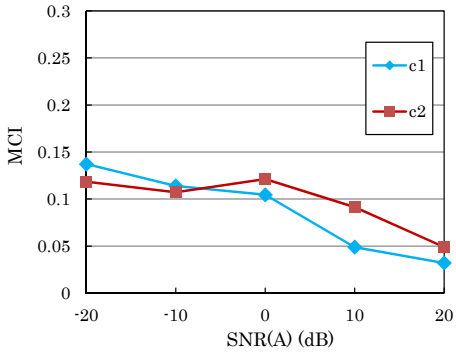


MCI

(g) highway 1

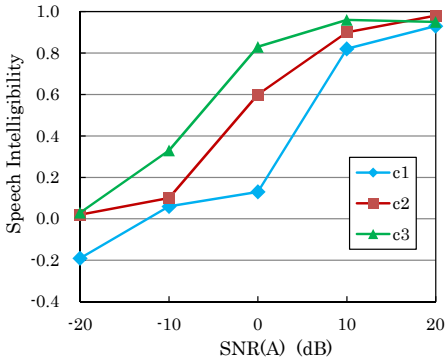


Intelligibility

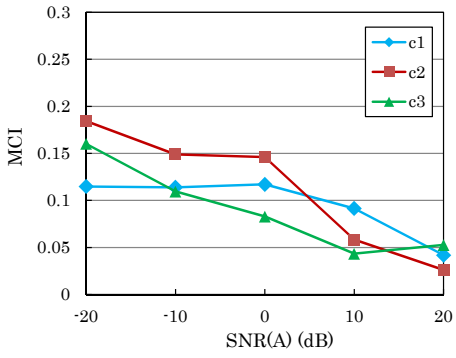


MCI

(h) highway 2



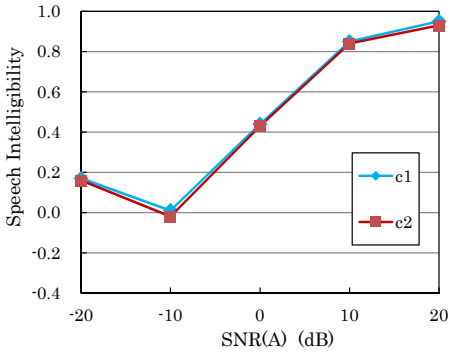
Intelligibility



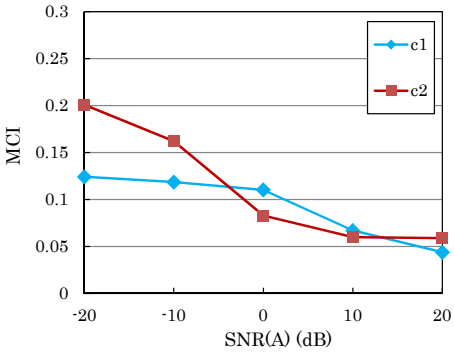
MCI

(i) crossing

Fig. 4.6 Comparison of intelligibility and MCI with various noise type (cont'd)

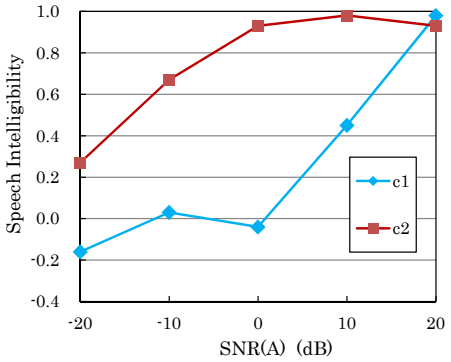


Intelligibility

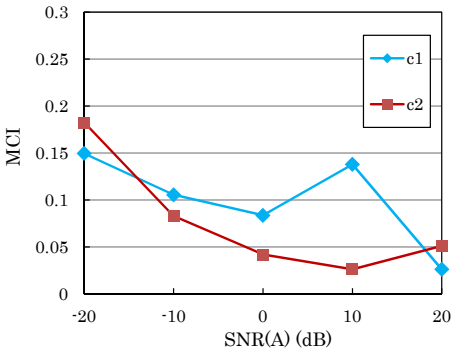


MCI

crowd

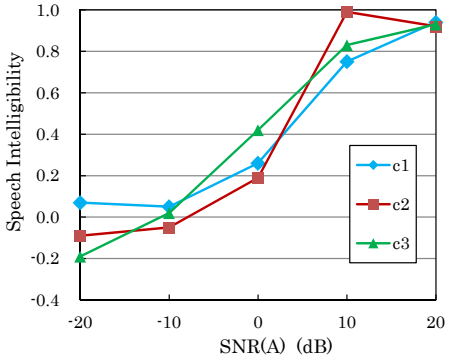


Intelligibility

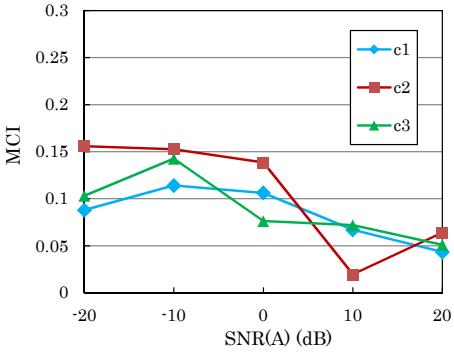


MCI

bullet train



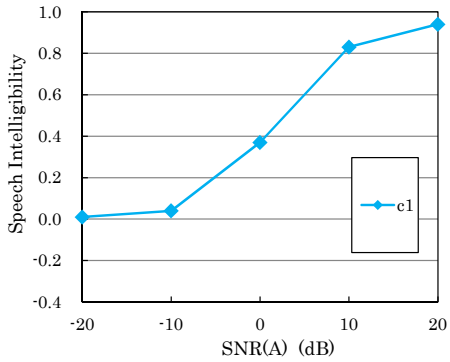
Intelligibility



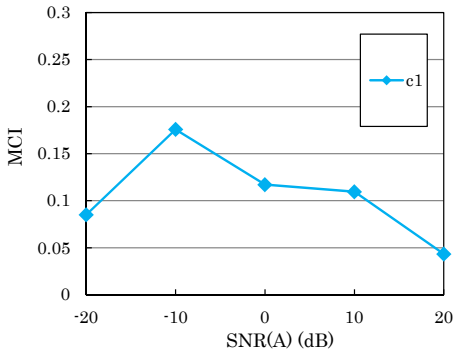
MCI

train

Fig. 4.6 Comparison of intelligibility and MCI with various noise type (cont'd)

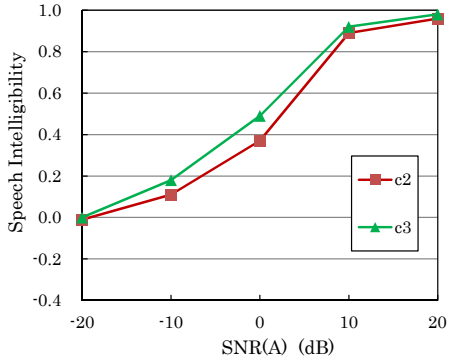


Intelligibility

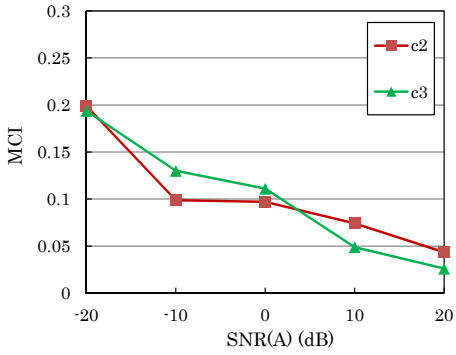


MCI

(j) computer room

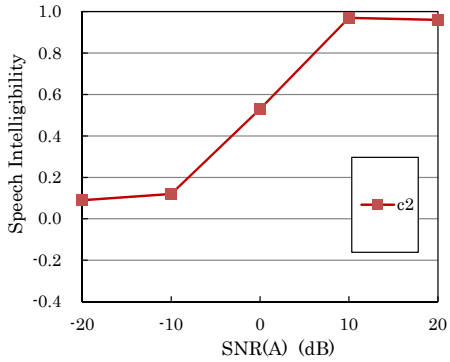


Intelligibility

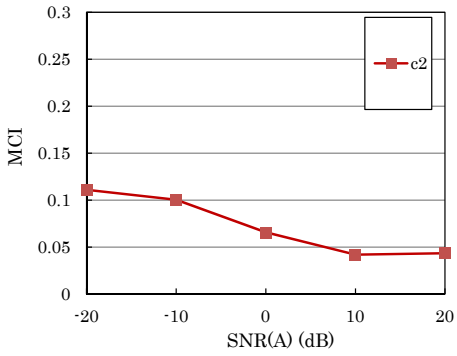


MCI

(k) air conditioner 1



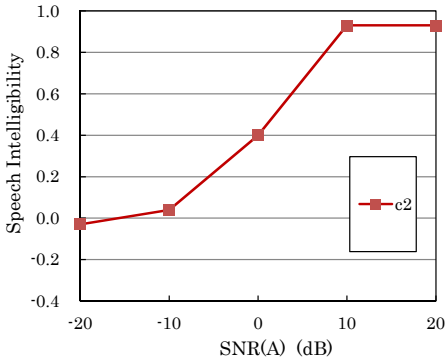
Intelligibility



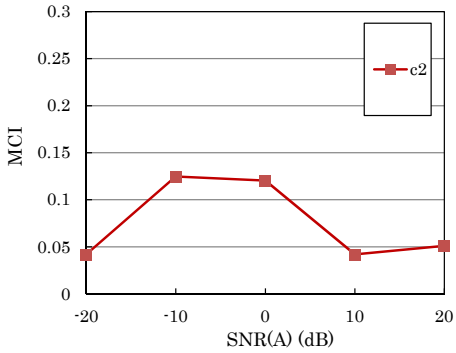
MCI

(l) air conditioner 2

Fig. 4.6 Comparison of intelligibility andMCI with various noise type (cont'd)

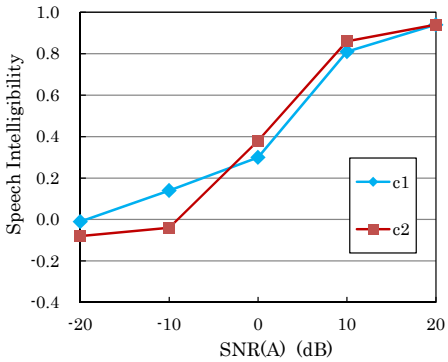


Intelligibility

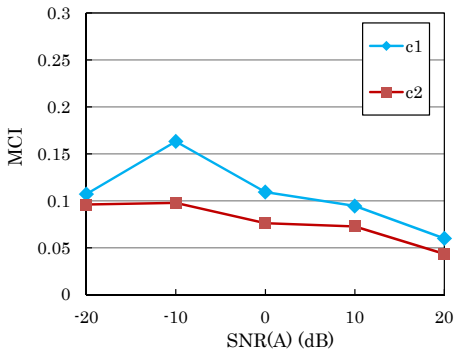


MCI

(m) air duct

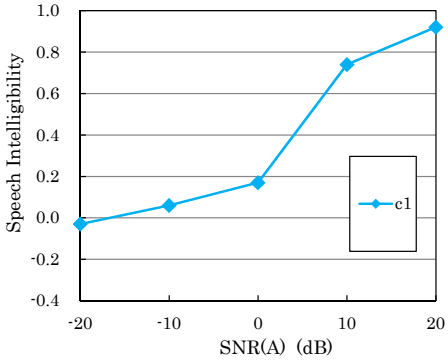


Intelligibility

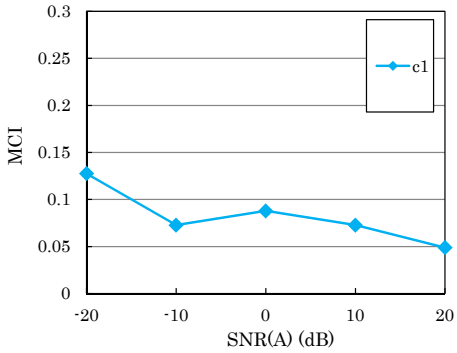


MCI

(n) elevator hall 1



Intelligibility



MCI

(o) elevator hall 2

Fig. 4.6 Comparison of intelligibility andMCI with various noise type (cont'd)

4.4 パラメトリック回帰による推定

騒音クラスタごとのSVRによる推定関数（以下，提案法）と2.3.1節で述べた既存のfwSNRsegを5種類[6]による順位相関係数の比較と，シグモイドカーブフィッティングによるパラメトリック回帰による推定性能比較を行う．

4.4.1 客観音質値と順位相関係数

まず，主観了解度と客観音質値との散布図を比較し，その順位相関係数τを比較する．

客観音質値

主観評価値と客観音質の散布図を，Fig. 4.7に示す．図中に全騒音混合条件での推定関数を併記する．SNRsegは主観了解度0.4以上の分散が大きく，AIsegは推定関数近傍で分散が大きくみえる．他の3尺度はこれらよりも分散が少なくみえる．fwSNRseg(A)は客観音質値が-10 dBに固まっている．これは2.3.1項で述べた客観音質評価法の設定をそのまま用いたため，下限SNRが-10 dBと主観評価系に対して大きかったため，値が固定されたと考えられる．

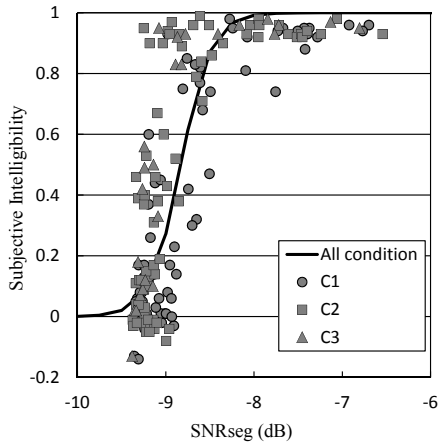
順位相関係数

次に，2.3.5項でも検討した順位相関係数τを比較する．τは式（2.4）Table 4.10に尺度ごとのτを示す．SNRsegを除けば，ほぼ同程度の値である．2章ではτが高くてもRMSEが小さいとは限らなかったが，τの低いものはRMSEが大きかったため，SNRsegの推定性能は低くなることが予測される．

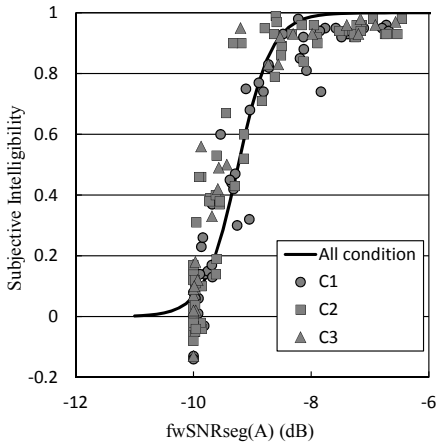
Table 4.10: Kendall rank correlation(τ) between intelligibility (sustention) score and objective speech quality score by noise cluster

	C1	C2	C3	pooled noise
SNRseg	0.597	0.526	0.766	0.554
fwSNRseg(A)	0.764	0.720	0.831	0.748
fwSNRseg(C)	0.770	0.740	0.867	0.769
fwSNRseg(S)	0.772	0.729	0.853	0.763
AIseg	0.755	0.686	0.813	0.724

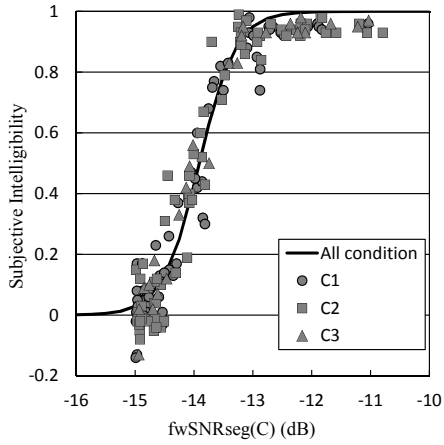
[6]重み無のSNRseg，fwSNRseg(A)，fwSNRseg(C)，fwSNRseg(S)，AIseg.



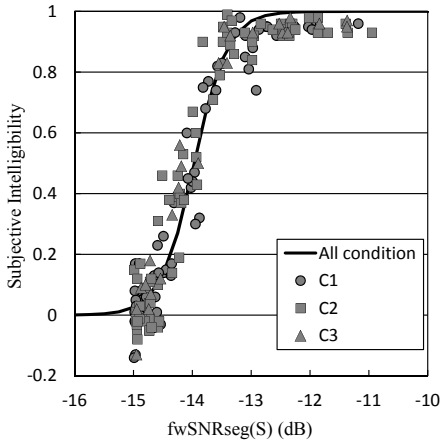
(a) SNRseg



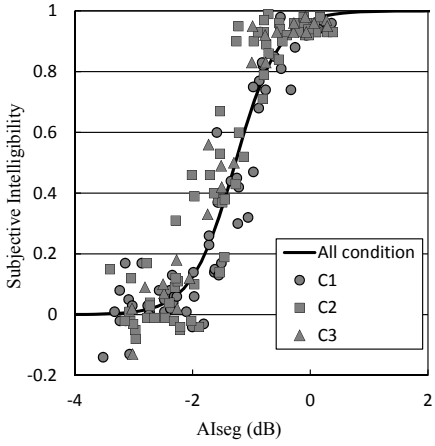
(b) fwSNRseg(A)



(c) fwSNRseg(C)



(d) fwSNRseg(S)



(e) AIseg

Fig. 4.7: Objective quality score and estimate function(All cluster multi condition)

4.4.2　既存尺度を用いた騒音クラスタ別推定関数作成

既存尺度による音質評価値の求め方は，cbSNRseg，obSNRseg と同様に各 LF 中の音声成分のパワーがほぼ 0 となるフレームを除いた平均音質値とする．推定関数の作成は，2.4.1 項で検討した推定関数を騒音クラスタ別に作成する．作成する関数は式（2.5）を用いたシグモイドカーブフィッティングで，トレーニングに用いる教師データに 4.3 節で検討した主観評価値を用いる．また，次章で検討する SVR と同様に 10 回の交差検定で回帰係数を作成する[7]．

4.4.3　推定結果の比較

5.3.4 節と同様に交差検定の RMSE を比較する．Table 4.11 比較する 5 種の客観評価尺度を用いた推定の交差検定による RMSE を示す．Weighted sum は 3 クラスタの RMSE の重み付平均で，重みは Table 4.5 の Percentage を用いた．Multi は全クラスタを混合して同様に推定関数を作成した場合の RMSE である．また，表最下段に主観評価結果の MCI を記載する．まず，順位相関係数 τ の低かった SNRseg は，やはり RMSE が 0.1 以上と非常に悪く推定に適していない．つまり，了解度の推定には何らかの聴覚重み付けまたは機械学習による回帰係数の決定が必要であることがわかる．fwSNRseg(C) と fwSNRseg(S)，AIseg の三指標は概ね RMSE が小さい．fwSNRseg(A) を除けば，Weighted sum の方が Multi よりも RMSE が小さく，騒音クラスタリングの効果が見られる．MCI と比較すると，SNRseg と fwSNRseg(A) の C2 を除けば RMSE<MCI となり，推定関数の性能を十分満たしている．これら 5 種の客観音質を用いたシグモイドカーブフィッティングと SVR その他の回帰手法との比較は 6 章のオープンテストの結果で比較する．

Table 4.11: RMSE of 10–fold cross–validation by sigmoid fitting estimation

Measure	C1	C2	C3	Weighted sum	Multi
SNRseg	0.123	0.188	0.161	0.157	0.176
fwSNRseg(A)	0.069	0.121	0.088	0.095	0.088
fwSNRseg(C)	0.062	0.078	0.065	0.070	0.072
fwSNRseg(S)	0.063	0.081	0.072	0.072	0.075
AIseg	0.066	0.097	0.063	0.079	0.090
MCI	0.0961	0.0929	0.0897	0.0937	0.0937

4.5　まとめ

了解度変化傾向の近い騒音種に分類するため，以下の 3 点を考慮した騒音クラスタリングを検討し，主観評価と既存尺度による推定関数の作成を行った．

- 長時間騒音信号の LF 分割

[7]一般に，線形回帰と一般線形化回帰では，関数の形状が固定されるため，交差検定を考慮しても汎化性能はほとんど変わらない場合が多い．本論文では，SVR 等の回帰手法との比較のため交差検定の回数を揃えることとした．実際に交差検定を用いない回帰の RMSE と交差検定による RMSE に差はほとんどみられなかった．

- MIR 特徴量による騒音の音色解析
- x-means クラスタリングによるクラスタ数の自動分割

その結果，以下のことが明らかになった．

- 騒音クラスタは 3 つ作成される．
- 主観評価の結果，クラスタ内の平均値間には有意な差がある．
- 既存の客観音質を用いたシグモイドカーブフィッティングによる推定では騒音クラスタリングの効果はあまり大きくない．
- 聴覚重み付けの無い SNRseg では交差検定の RMSE が大きく，何らかの聴覚重みが不可欠である．

クラスタ内の了解度の平均値間に有意な差があり，クラスタごとの了解度差が明確に見られたため，同一の SNR であっても了解度が低い環境と，了解度が高い環境を分類することは可能である．このため，騒音クラスタごとに了解度推定関数による推定は，単一の了解度推定関数による推定よりも精度が良くなる場合がみられた．次章では本章の結果を用いて，騒音クラスタごとに SVR 等を用いた了解度推定関数を作成する．

第5章　サポートベクトル回帰による了解度推定関数の作成

SVR は多変量回帰の中でも汎化性能が高いとされる手法である．本章では SVR 及び L1 正則化を用いた他の回帰手法を用いたノンパラメトリック回帰による推定関数を作成し，交差検定で比較する．

5.1　SVR 特徴量

SVR の特徴量には評価信号を帯域分割し，帯域ごとのセグメンタル SNR を用い，最適な回帰係数を SVR の学習結果より求める．

5.1.1　帯域分割法

聴覚理論に基づいた音声信号の帯域分割法はいくつも検討されている．本論文では SII の技術標準に組み込まれている 2 種の帯域分割方式を比較する．帯域分割法以外はすべて同様に求めた特徴量に対し，交差検定の RMSE が最小になるハイパーパラメータ[1]を選択した推定関数の回帰係数を採用する．帯域分割は，令分割法ごとの中心周波数とバンド幅で設計したバンドパスフィルタを計算機上で作成し，信号に畳み込むことで帯域制限を行った評価信号とする..

(a) 1/3 オクターブバンド

(b) Fig. 2.23 のクリティカルバンド

(a) は騒音計をはじめ広く用いられている．今回は SII で用いる場合と同様に中心周波数は 0.16 kHz〜8 kHz の 18 帯域を用いる．以下，本論文では，この分割法による特徴量を obSNRseg と呼ぶこととする．(b) は fwSNRseg(C)，fwSNRseg(S) と同様の帯域分割であり，学習によって求めた回帰係数と，SII の聴覚重みとの比較に最適である．本論文ではこの分割法による特徴量を cbSNRseg と呼ぶ．本論文では以後，これら二つの帯域分割法式の違いを別個の特徴量として比較する．

特徴量はどちらも帯域ごとに式 (1.5) のセグメンタル SNR を用いる．ここで $x(n)$ と $\hat{x}(n)$ は，n 番目の分析フレームでの原音声と雑音重畳音声（劣化音）で，M は音声区間のフレーム総数を，N は分析長で，騒音クラスタリングで用いた SF 長である 100 msec に設定した．また，特徴量を求める際に音声成分のパワーがほぼ 0 となるフレームを除いて平均を求めた．

[1]事前確率を決めるパラメータや確率モデル全体に影響を与えるパラメータ．ハイパーパラメータが定まると確率モデルの分布が定まる．

5.1.2 特徴量正規化

特徴量の SNRseg は，帯域ごとに SNRseg のベターイヤースコア（2.3.2 参照）を求め，0 から 1 に正規化し，SVR による回帰係数の決定を行う．本論文で用いる LIBSVM[167, 168] は特徴量を標準化して用いるため，観測された値に対しては正規化の処理は本来必要ない．しかし，2 章の結果から，客観音質値と値域が了解度の変化範囲が重要になることは明らかである．4.3.2 項の学習セットに用いる主観評価結果より，主観評価結果は了解度が十分に低い 0 近傍から 1 まであり，主観評価値の分散は十分であるため，客観音質値の値域を変化させ最適な値を設定した推定性能の向上を検討する．正規化には式（5.1）を用いる．min value と max value は帯域ごとの下限 SNR 値と上限 SNR 値である．上限 SNR 値は騒音が無い状態の SNR に該当する．2 章で用いた fwSNRseg(C) と fwSNRseg(S) は，−15 dB から 15 dB に制限した．精度の高い推定には正規化上限以上の了解度が 1，最大値以下の了解度が 0 であることが望ましい．本提案法では，最大値を 0〜30 dB の 5 dB ごと，最小値を −10〜−50 dB の 5 dB ごとの計 63 組設定し比較する[2]．

$$\text{norm. SNRseg} = \frac{\text{SNRseg} - \text{max value}}{\text{min value} - \text{max value}} \tag{5.1}$$

5.2 SVR と他の L1 正則化を用いた回帰との比較

SVR は機械学習を用いたノンパラメトリック回帰では一般に汎化性能が高い有効な手法である．本節では SVR の汎化性能が特に了解度推定問題において有効である理由を考察するために，L1 正則化を用いる各種回帰法との比較を行う．

5.2.1 L1 正則化を用いた他の回帰手法

SVR の汎化性能が高い理由に以下の三点が挙げられる．

(a) 正則化を用いることで（実質的に）変数選択を行う

(b) ϵ-不感応関数を用いるロバストな回帰

(c) 回帰関数とサポートベクトルとのマージンを最大化する明確な基準

本論文では，SVR が了解度予測に最適な回帰手法であることを確認するため，(a) について他の正則化を用いた回帰手法との比較を行う．比較対象としてリッジ回帰（以下，Ridge）[175] と Lasso（Least absolute shrinkage and selection operator）を用いた回帰（以下，Lasso）[176] を GNU R[169, 170] の glmnet パッケージ [177] で実装した[3]．正則化パラメータ λ は SVR 同様に 10–fold の交差検定で決定する．また，SVR はカーネル法を用いた回帰の一形態であるとみなすことができる．そこで，L1 正則化と RBF カーネルを用いたカーネル回帰（以下，Kernel）[178] も比較対象に加える．Kernel は ϵ-不感応関数とサポートベクトルの選択を含まず正則化パラメー

[2] バンドごとに最適値が異なると考えられるが，組み合わせ数が膨大になるため，本論文ではすべてのバンドで同じ値を用いることとした．

[3] glmnet はリッジ回帰と Lasso を用いた回帰の一般化であるが，本論文では両者の中間になるパラメータは比較しなかった．

タを用いた正則化のみを実装したため，(b) と (c) による回帰の頑健さを考慮した回帰精度の比較ができる[4]．Kernel も正則化パラメータ λ と RBF カーネルの γ を交差検定で最適なパラメータを設定する．

5.3 推定関数の作成

SVR を用いて了解度推定関数を特徴量ごとに作成し比較する．その際のカーネル関数に線形カーネルと RBF カーネルを用いる場合を比較する．学習に用いるハイパーパラメータの組み合わせは交差検定を用いて最適値を探索する．推定関数は C1〜C3 の各クラスタごとと，比較対象に 3 クラスタの全データを用いた multi 条件の 4 推定関数を作成する．multi 条件による RMSE と C1〜C3 の推定関数によるクラスタごとの分布重み付平均を比較し，騒音クラスタリングを用いた推定性能を検証する．クラスタごとの重み付平均に用いる重みは，Table 4.5 のクラスタごとの所属割合を用いる．

5.3.1 推定条件

本節では，2.2 節で求めたクラスタごとの主観評価結果を推定する．各クラスタの LF ごとに特徴量を求める．C1〜C3 の各クラスタで主観評価を行った LF 数はそれぞれ，13，14，5 個あり，主観評価実験で設定した SNR が 5 種なので，各クラスタごとに主観評価値と対応させて cbSNRseg をそれぞれ 65，70，25 組求める．特徴量とハイパーパラメータの設定には，式 (5.2) で定義する，教師信号である主観評価で求めた了解度（$Sub.Intell.$）と推定関数によって求めた推定了解度（$Est.Intell.$）の RMSE が最少になるものを選択する．式中の N はサンプル数で，各クラスタごとに求めた特徴量の組数となる．

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(Sub.Intell. - Est.Intell.\right)^2}{N}} \tag{5.2}$$

5.3.2 探索する SVR のハイパーパラメータ

探索する SVR のハイパーパラメータは，ϵ と C，RBF カーネルを用いる場合は γ も求める．探索範囲は ϵ が 10^{-4}，10^{-3}，10^{-2} の 3 値[5]，C は $2^{-4} \leq C \leq 2^{5}$，γ は $10^{-7} \leq \gamma \leq 1$ の範囲をそれぞれ 512 等分し，全ての組み合わせで 10–fold の交差検定を行い，交差検定の RMSE が最も小さい組み合わせを正規化の上下限の各組み合わせの代表値とし，正規化上下限の全組み合わせの中から最も RMSE が小さくなるハイパーパラメータと正規化上下限の組み合わせを採用する．

[4](b) の性能評価だけを考慮すると，MM 推定を用いたロバスト回帰ななどとの比較が考えられる．しかし，提案特徴量である cbSNRseg が 25 次元であり，騒音クラスタ C3 は騒音 LF が 5 個で実験用 SNR が 5 種類の 25 サンプルしかない．このため，C3 の評価正則化を伴う回帰手法でなければ一般化逆行列を用いる必要がある．本論文では正則化を含むカーネル回帰と比較することで (b) と (c) を総合した性能比較を行うこととした．

[5]より広い範囲を詳細に探索したが，ほぼ変化がないためこの 3 値とした．

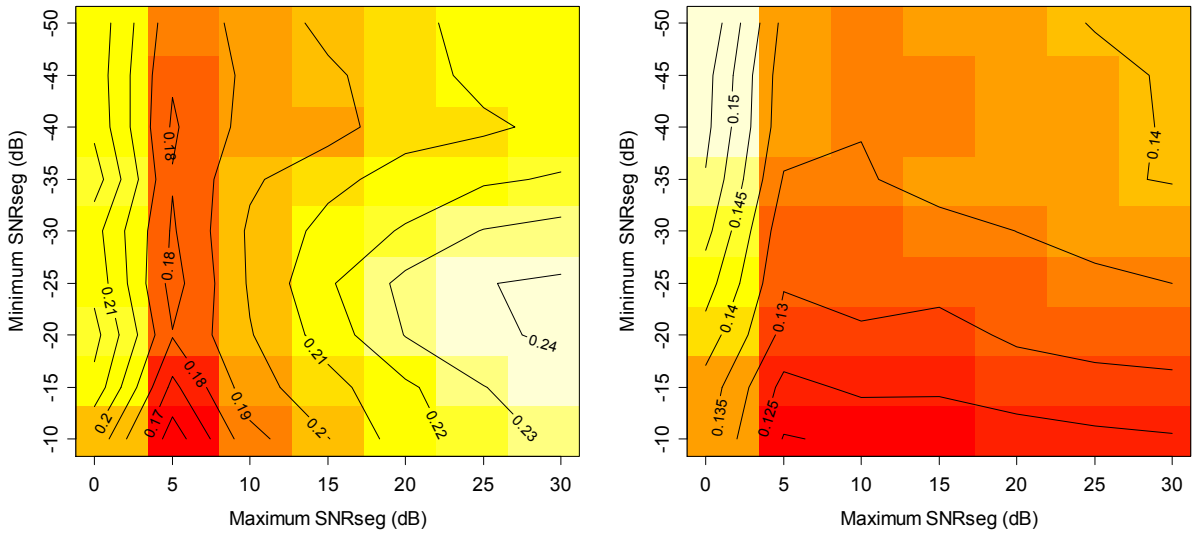
5.3.3 特徴量の正規化範囲の検討

Fig. 5.1 と Fig. 5.2 に SVR の特徴量の違いによる正規化に用いた最大値，最小値と RMSE の関係を示す．図中で色が濃い方が RMSE が小さいことを示す．また，RMSE より求めた等高線を参考に示す．obSNRseg は概ね上限値が 5 dB，下限値が −10 dB に近づくほど RMSE が低下している．C1 と C3 の線形カーネルは傾向差がみられ，それぞれ下限値が −25 dB，−40 dB の RMSE も RMSE が減少傾向にある．cbSNRseg ではどちらのカーネルでも上限値が 0 dB，下限値が −10 dB で RMSE が最小になる．各条件ごとの最適な上限値と下限値を Table 5.1 に示す．表には比較対象の L1 正則化を用いた回帰手法の結果も示す．L1 正則化を持ちた回帰手法は手法ごとの最適値は異なるが概ね最小値は −10 dB，最大値は 5 dB になる傾向にある．

以上の結果より，SVR の場合は上限 SNR と下限 SNR の幅が狭い方が交差検定の RMSE が小さいことがわかる．これは，Fig. 4.5 より，SNR（A) が −20 dB の様な極端に悪い環境では了解度は 0 に，20 dB の様な極端に良い環境では了解度が 1 にそれぞれ飽和することから，了解度が 0 から 1 に変動する範囲に制限された SNR を用いることで，RMSE が減少していると考えられる．以後本論文では，全ての回帰手法で Table 5.1 に示した最適値を用いることとする．

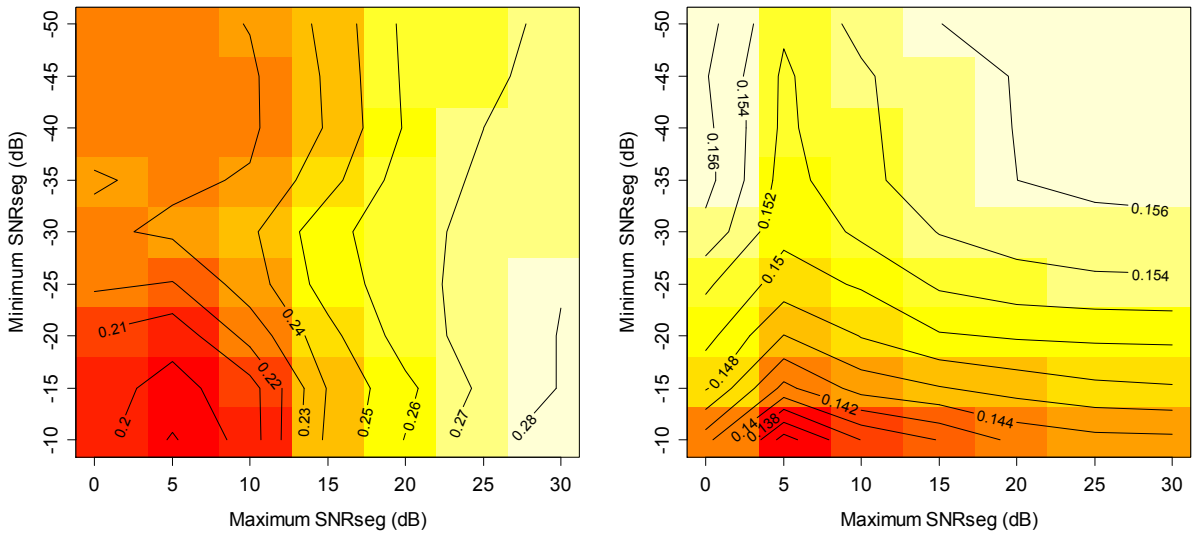
Table 5.1: Combination of the maximum / minimum SNRseg value

regression method	feature vector	C1		C2		C3		multi	
		max	min	max	min	max	min	max	min
SVR(linear)	obSNRseg	5	−10	5	−10	5	−40	5	-10
	cbSNRseg	0	−10	0	−10	0	−10	0	-10
SVR(RBF)	obSNRseg	5	−10	5	−10	5	−10	5	-10
	cbSNRseg	0	−10	0	−10	0	−10	0	−10
Lasso	obSNRseg	5	−10	25	−10	10	−40	5	-10
	cbSNRseg	5	−15	5	−10	0	−15	0	-10
Ridge	obSNRseg	5	−10	5	−10	0	−10	5	-10
	cbSNRseg	5	−10	5	−10	5	−10	0	−10
Kernel	obSNRseg	10	−20	0	−10	5	−10	0	-10
	cbSNRseg	0	−25	20	−45	0	−10	5	−30



Liner RBF

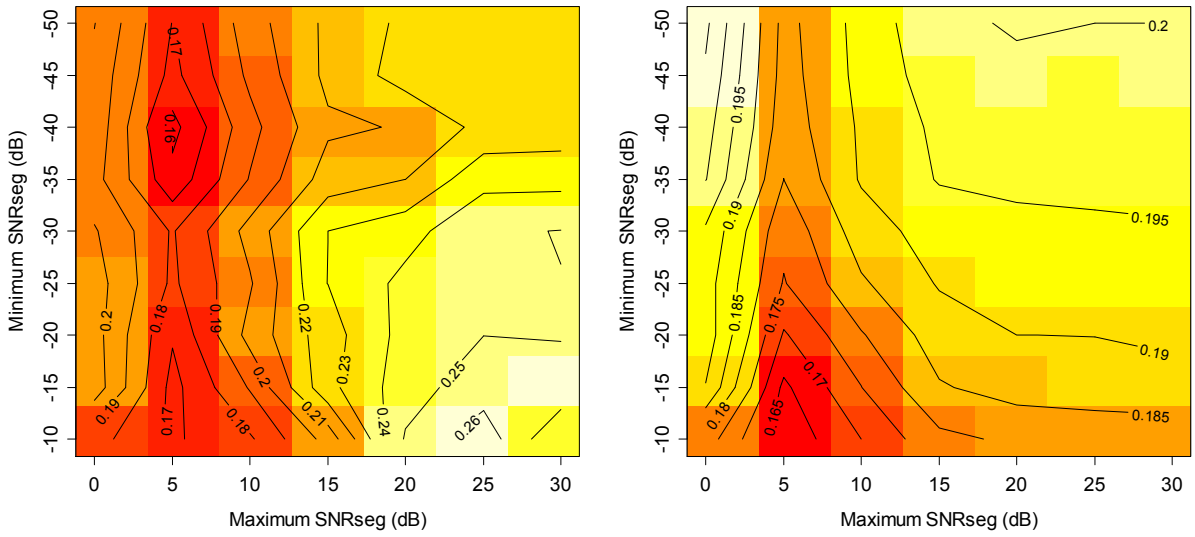
(a) obSNRseg C1



Liner RBF

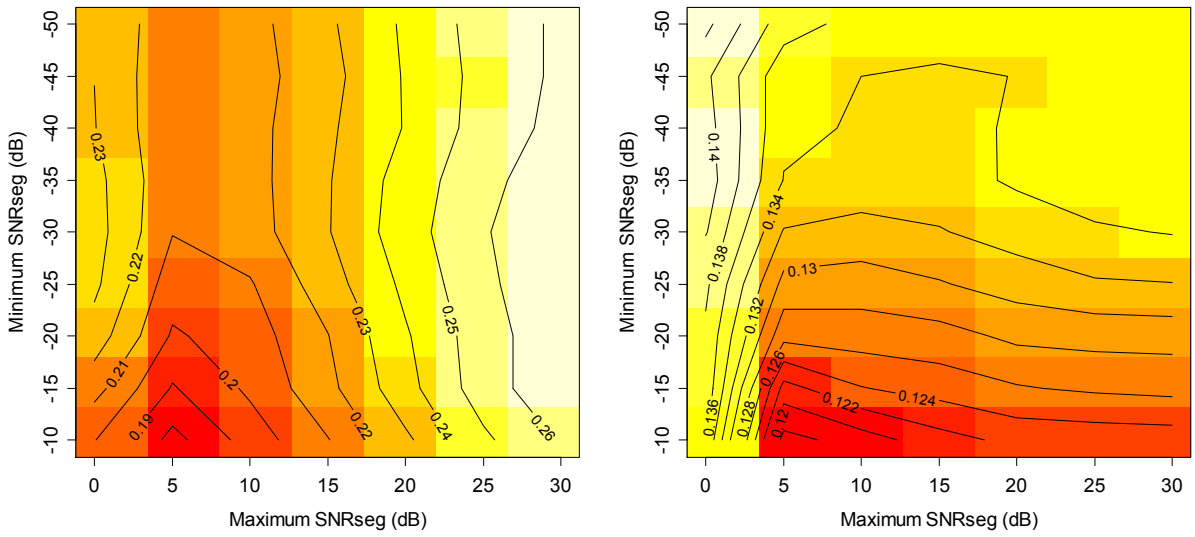
(b) obSNRseg C2

Fig. 5.1: Relationship between RMSE and maximum / minimum obSNRseg



Liner RBF

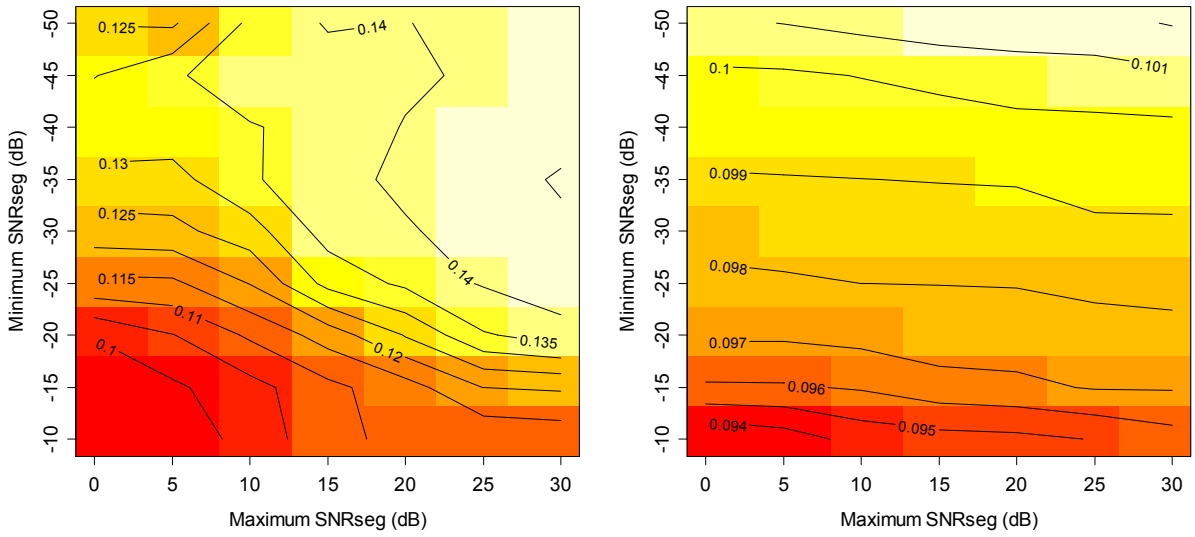
(c) obSNRseg C3



Liner RBF

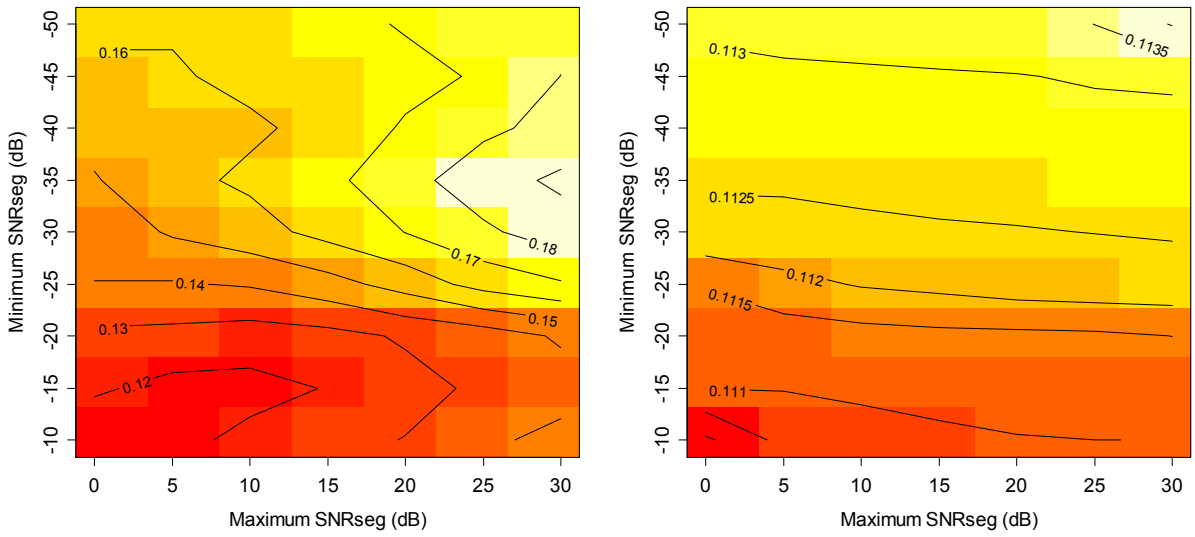
(d) obSNRseg all

Fig. 5.1 Relationship between RMSE and maximum / minimum obSNRseg(cont'd)



Liner RBF

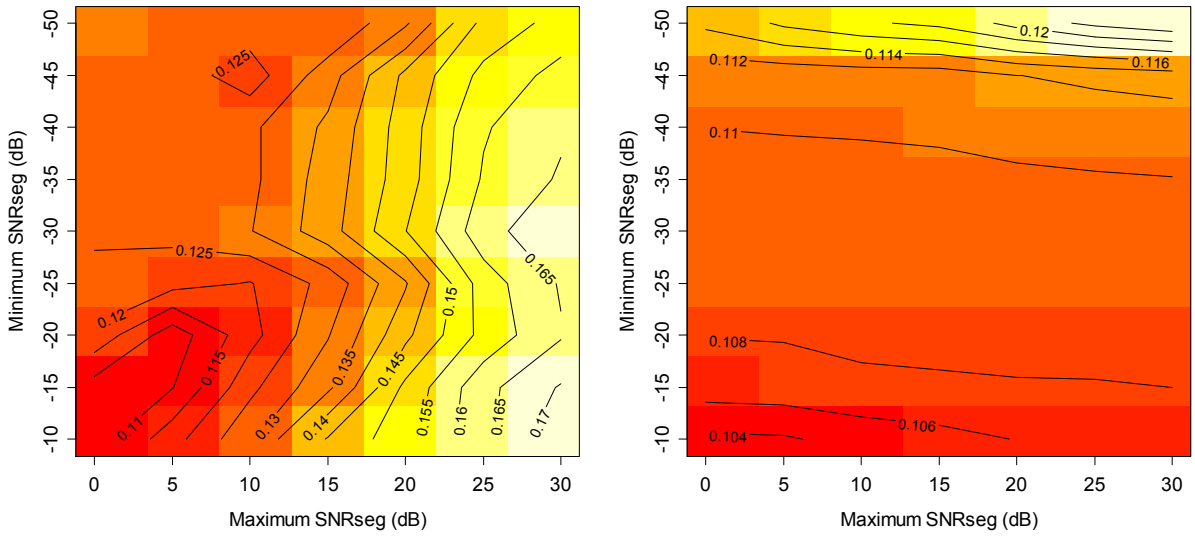
(a) cbSNRseg C1



Liner RBF

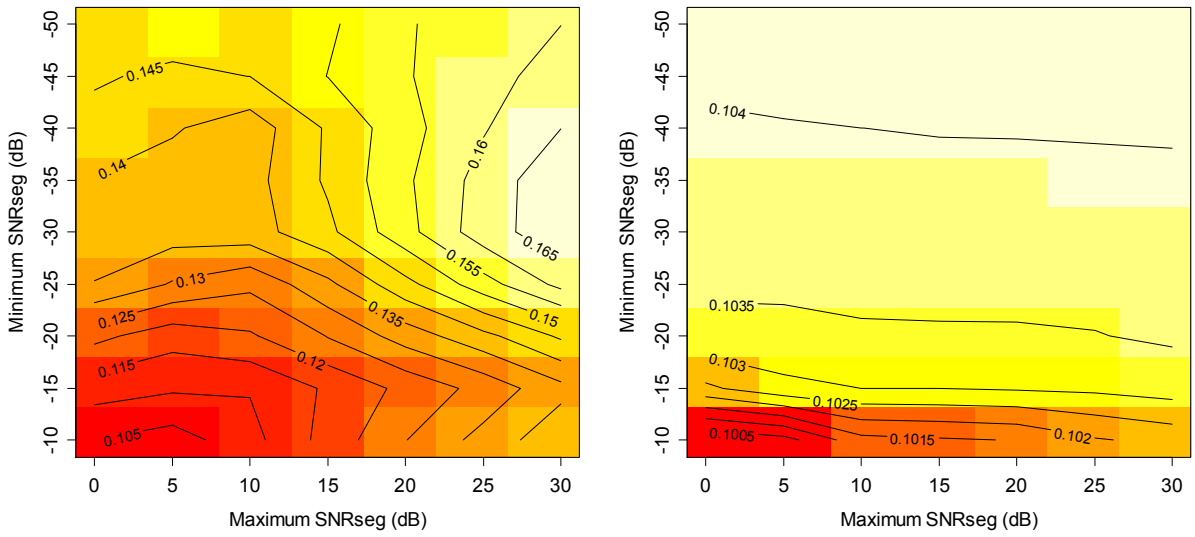
(b) cbSNRseg C2

Fig. 5.2: Relationship between RMSE and maximum / minimum cbSNRseg



Liner RBF

(c) cbSNRseg C3



Liner RBF

(d) cbSNRseg all

Fig. 5.2Relationship between RMSE and maximum / minimum cbSNRseg(cont'd)

5.3.4 推定結果・特徴量の比較

回帰手法ごとの最適なハイパーパラメータを用いた 10–fold 交差検定の RMSE と相関係数を比較する.

5.3.5 比較結果

Table 5.2 と Table 5.3 に Lasso，Ridge，Kernel の 3 種の回帰と SVR の線形カーネル（以下，SVR(Linear)）と RBF カーネル（以下，SVR(RBF)）を回帰手法ごとに特徴量が cbSNRseg と obSNRseg の場合に分け，クラスタごとの推定結果を示す．表の C1〜C3 は各クラスタの RMSE であり，Weighted sum は Table 4.5 のクラスタ別の主観評価サンプルの割合を乗じた RMSE の和，Multi はクラスタリングを用いず，全てのサンプルで学習したときの交差検定の RMSE である．二つの表に示した結果を特徴量，クラスタリングの効果，回帰手法のそれぞれ着目した場合に分けて考察する．SVR のハイパーパラメータと特徴量の正規化に用いた値の組み合わせは，付録 C に示す.

Table 5.2: RMSE of cross-validation using cbSNRseg

Regression method	C1	C2	C3	Weighted sum	Multi
Ridge	0.097	0.116	0.079	0.102	0.108
Lasso	0.103	0.125	0.100	0.112	0.112
Kernel	0.086	0.100	0.062	0.088	0.093
SVR(linear)	0.097	0.112	0.108	0.105	0.103
SVR(RBF)	0.095	0.111	0.105	0.104	0.100
MCI	0.0961	0.0929	0.0897	0.0937	0.0937

Table 5.3: RMSE of cross-validation using obSNRseg

Regression method	C1	C2	C3	Weighted sum	Multi
Ridge	0.143	0.239	0.117	0.181	0.177
Lasso	0.173	0.231	0.149	0.195	0.205
Kernel	0.117	0.155	0.064	0.126	0.147
SVR(linear)	0.155	0.189	0.158	0.170	0.177
SVR(RBF)	0.118	0.132	0.158	0.130	0.117
MCI	0.0961	0.0929	0.0897	0.0937	0.0937

特徴量の比較

cbSNRseg と obSNRseg を比較すると，全ての条件で obSNRseg の方が RMSE が大きい．Ridge と Lasso は cbSNRseg を用いることで obSNRseg に対し，C2 の RMSE がそれぞれ 0.485 倍，0.542 倍と大きく改善している．この他の回帰手法でも Weighted sum の改善率は 0.8 倍程度になることから，特徴量は cbSNRseg が良いといえる.

クラスタリングの効果の比較

騒音クラスタリングを用いたことによる推定関数の RMSE の重み付平均である Weighted sum が Multi の RMSE より小さければ，騒音クラスタリングが了解度推定に効果があったといえる．結果より，Table 5.2 の cbSNRseg では Weighted sum が Multi の差はほとんどみられないが，Table 5.3 の obSNRseg では Lasso と Kernel で Weighted sum の方が小さく，他は Multi の方が大きい．このため騒音クラスタリングの効果は本章の結果からは読み取れない．騒音クラスタリングの効果の解析については次章のオープンテストで行う．

回帰手法の比較

回帰手法ごとの Weighted sum はどちらの特徴量を用いても Kernel が最もよい．SVR(RBF) はこれに次ぐ値だが，5.2.1 項で述べたように SVR は ϵ-不感応関数を用いて過学習による影響を押さえている．一方，Kernel は正則化以外に過学習の影響を考慮しておらず，特に RMSE が 0.088 と小さい Kernel と cbSNRseg の組み合わせは，過学習を起こしている可能性が憂慮される．同様の理由で，SVR(linear) と Ridge，Lasso の比較も obSNRseg では SVR(linear) が良いものの，cbSNRseg ではほぼ同じ値であり回帰手法の優劣を比較できない．よって回帰手法の比較もオープンテストで行う．

MCI との比較

4.4.3 項で検討した既存尺度を用いたシグモイドカーブによる推定関数では，概ね RMSE<MCI を満たしていた．しかし，Table 5.2 の RMSE が小さい cbSNRsegn の結果においても RMSE<MCI を満たす組み合わせは少ない．これは交差検定が用いたサンプルだけの誤差ではなく，未知のサンプルに対する誤差も考慮しているため，MCI よりも大きくなることが原因である．特に過学習を起こしやすいノンパラメトリック回帰では交差検定による RMSE の増加はパラメトリックな回帰よりも大きくなる傾向にある．よって，ノンパラメトリック回帰では，交差検定による RMSE と MCI の比較は妥当ではないと考える．

5.4 まとめ

騒音クラスタごとに SVR を用いた了解度推定関数を作成した．SVR の特徴量には帯域ごとの SNRseg を用い，1/3 オクターブバンドでの分割（obSNRseg）とクリティカルバンドでの分割（cbSNRseg）を比較した．また SVR に用いるハイパーパラメータと SNRseg の上下限値を比較した．その結果，以下の内容が明らかになった．

- obSNRseg では上限 SNR が 5 dB，下限 SNR が −10 dB（線形カーネルの C3 のみ −40 dB）が良く，cbSNRseg では上限 SNR が 0 dB，下限 SNR が −10 dB が良い．しかし，L1 正則化を用いた回帰では，回帰手法とクラスタのごとに最適値が異なり，あまり明確な傾向はみられない．

- 交差検定の RMSE を用いて cbSNRseg と obSNRseg を比較すると，全ての回帰手法で cbSNRseg の方が RMSE が小さくなるため，用いる特徴量は cbSNRseg が良い．
- 騒音クラスタリングの効果は回帰手法ごとに Weighted sum と Multi を比較したが，cbSNRseg ではほとんど違いが無い．obSNRseg は回帰手法によって傾向が異なる．
- SVR 以外の回帰手法も含め，了解度推定に最適な回帰手法は交差検定の RMSE からは定めることができない．

以上の結果より，提案特徴量である cbSNRseg は有効であることはわかったが，機械学習を用いることの是非は判断できなかった．次章では推定関数作成に用いなかったデータによるオープンテストを行い，回帰手法間の総合的な汎化性能の比較と騒音クラスタリングの効果を検証する．

第6章　オープンテストによる総合性能評価

提案方式のオープンテストとして，モデル作成に用いなかった騒音下での了解度を推定し，主観評価結果と比較する．オープンテストであるため，提案法の騒音クラスタリングと SVR の学習に用いた条件に対し，未知の情報が増えるような評価対象が必要である．また，性能評価のために，クラスタリング精度の分析に用いる分散分析や推定値との RMSE や相関係数の算出のために，オープン条件の主観評価値も必要となるためこれらも含めて本節で詳細に述べる．

6.1　評価 LF のランダムサンプリング

5 章の推定関数の作成の教師データに用いた 4 章の主観評価値は，クラスタリング精度の検証が主目的であり，特に同一騒音種のクラスタ違い（Table 4.4 の行方向の比較）を考慮して主観評価用のサンプリングをしたため，データベースの騒音分布の再現を考慮していない．また，他の了解度評価による検討の結果との比較のために，主観評価に用いる LF と評価単語の SNR を複数設定して主観評価を行った．このため，本論文で検討している了解度推定では，同一の騒音 LF でも SNR が異なる場合は別のサンプルとして扱った．これは，SNR を複数設定することで，受聴者と騒音源との距離の違いなどによる主観的な音量差を考慮したものであり，騒音と受聴者の位置といった条件を考慮しないならば，疑似的に騒音数を増加させることに近似できると考えたためである．つまり，SNR の違う同一騒音を複数用意して主観評価を行うことは，機械学習で用いられるブートストラップと同様の発想とみなし，32 個の騒音 LF から 160 個の主観評価値のあるサンプルを騒音クラスタに分割して 5 章の推定関数の教師データとした．

本章で検討する了解度推定に用いるオープンテスト用データとしては，多数の騒音種に対する汎化性能評価が重要であり，SNR を複数設定することによるみかけの騒音 LF 数増加よりも，実際の騒音 LF 数を増加させることが望ましい．そこで，電子協騒音データベース [157] のフルセット版から作成した騒音 LF を用いることでトレーニングデータに無い未知の騒音を再現する．この時の主観評価設定に用いる SNR については，トレーニングに用いた主観評価では天井効果と床効果が確実に発生するように極端に高い SNR と低い SNR を含めた複数の SNR 値を設定する必要があったのに対し，テストデータに行う主観評価では現実的にシステム評価に用いたい環境の模擬だけで良い．このため，トレーニングデータの結果から了解度分布の広い SNR 値を 1 つ選択して主観評価値を求め，推定関数の性能評価に用いるテストデータとすることにした．設定する SNR は，Fig. 4.5 の結果より，一つの SNR 値でも騒音 LF の違いにより了解度が広く分布している −10 dB と 0 dB の中間にあたる −5 dB を評価に用いる設定とする．

推定関数をトレーニングしたデータは，フルセットの中から一部を抜き出したダイジェスト版から分割し作成した 605 個の LF である．フルセット版から主観評価に用いる LF を抜き出すのにダイジェスト版と同様に，4.1.2 項で述べた方法で音量を統制し，4.2 節で述べた方法でステレオ録

音された音源を 3 sec の LF に分割すると 28328 個の LF に分割された．28328 個の騒音 LF から，ランダムサンプリングで騒音 LF を抽出し，全体の騒音分布を再現できるような主観評価セットを作成する．サンプル数 n は式（6.1）を用いて決定する．ここで母集団総数 $N = 27928$，標本誤差 $E = 0.05$，統計信頼度 $Z = 0.95$，母集団予想比率 $P = 0.95$ とした場合，$n = 380$ となる．本論文ではこの値よりやや多い 400 サンプルを乱数を用いて抽出した．

$$n = \frac{N}{\left(\frac{E}{Z}\right)^2 \left\{\frac{N-1}{P(1-P)+1}\right\}} \tag{6.1}$$

6.2 騒音クラスタリング結果

騒音クラスタリングに用いる特徴量は 4.2.2 項で検討した 15 特徴量を同様に LF ごとに計算し使用する．ただし，式（4.1）の正規化に用いる最大値と最小値は，ダイジェスト版によるモデル作成時の値を用いる．このためフルセットの版の正規化後の値は必ずしも 0 から 1 の範囲に収まらないが，そのまま解析を行うこととした．

Table 6.1 に 4 章で検討した騒音クラスタリングモデルでのフルセット版騒音 LF の分類結果を示す．Full set が全 LF，Test set が 6.1 節で検討したランダムサンプリングによる 400 LF の結果を示す．Digest set は Table 4.4 の結果を再掲した．Training set は Digest set から主観評価用に抜き出したサンプルの結果を示す．num. が LF 数で，percentage は LF 総数に対する割合を示す．Test set は C2 がわずかに Full set より割合が 0.033 多くなる傾向にあるが，概ね一致している．また，Full set，Test set 共に Digest set に近い割合であるが，Training set は C2 の割合が多く，C1 が少ない．これは主観評価サンプルの抽出法の問題である．

以上の結果より，提案した騒音クラスタリングによる LF 分類はフルセット版の騒音に対しても，LF の数では同程度に有効であると言える．次節では Test set 主観評価により，分類傾向を解析する．

Table 6.1: Number of LF by clustering

	Full set		Test set		Digest set		Training set	
Cluster	Num.	Percentage	Num.	Percentage	Num.	Percentage	Num.	Percentage
C1	14242	0.503	198	0.495	309	0.510	13	0.406
C2	8983	0.317	140	0.350	189	0.312	14	0.438
C3	5103	0.180	62	0.155	107	0.177	5	0.156
Total	28328	1.000	400	1.000	605	1.000	32	1.000

6.3 主観評価

6.3.1 実験条件

主観評価には，4.3 節と同様に Table 4.6 の JDRT の Sustention 単語対を同一の女性発話者で 1 名分用いる．ランダムサンプリングによって抽出した LF は，4.1.2 節で述べた様に，騒音種全体の平均ノイズパワーと音声のパワーが等しくなるように統制した．この値を SNR(A) で 0 dB として主観評価を行った．

この他の評価設定は 4.3 節に準じて設定した．評価に用いる 3 sec の LF に対し評価音声が短いことから，音声の埋め込み位置は LF の冒頭 0.1 sec を除く区間で，SNR(A) が小さくなるタイミングで音声と合成した．埋め込み位置探索は，評価音声のうち最も長い単語の半分の時間（最長モーラを想定）とした．

総評価単語数は，評価 LF 数が 400，SNR が 1 種，評価単語が 20 単語の 8000 単語であり，これに騒音を加算していない原音 20 単語を 5 回分で 100 単語をリファレンスとして加え，被験者一人当たり 8100 単語の評価を行った．また，急峻な音圧変化による被験者への負担を減らすため，各 LF の冒頭 100 msec に対し緩やかに音圧が上昇していく時間窓を掛けた．実験音声はコンピュータから Roland 社製 USB オーディオインターフェース UA–25EX を介し，Sennheiser 社製ヘッドホン HD–25II で提示した．全被験者に対する提示音圧は一定とし，騒音を加算していない原音が十分聴こえる音圧とし，4.3 節と同じ値にして実験を行った．なお，被験者は 20 代男性 8 名で，うち 7 名が 4.3 節の主観評価の被験者と重複する．

6.3.2 実験結果

Fig. 6.1 に主観評価結果を LF ごとに求めた SNR(A) と共に示す[1]．トレーニングに用いたデータを同様に処理したもの[2]を Fig. 6.2 に示す．Fig. 6.2 は，SNR(A) の変化に対し了解度が 1 と 0 に飽和する天井効果と床効果が顕著に見られるのに対し，Fig. 6.1 は，SNR(A) が −30 dB から 0 dB の間で了解度は 0 から 1 まで広く分布する．これは 6.1 節で検討した主観評価の基準に用いる SNR 値の違いが実験意図通りであったことを示す．どちらの図も騒音クラスタ間の境界が重なる LF は多いが，大局的には SNR(A) の変化に対し同傾向である．また，クラスタごとに SNR(A) による了解度の変化に傾向差がみられ，Fig. 4.5 の結果と同様，同一の SNR(A) の了解度は概ね $C3 > C2 > C1$ となる．このテストデータの主観評価値を推定する．

Table 6.2 に Test set と Training set の MCI をクラスタ別に示す．被験者数の少ない Test set の方が MCI が小さくなるのは天井付近のサンプルが多く，床効果がほとんど発生していないため，平均値が高くなり，MCI も小さくなったと考えられる．

テストセットのクラスタごとの平均値差をシェッフェの一対比較法で多重比較したところ，C1 と C2 の間には $p < 0.001$ で有意差がみられたが．C1 と C3 の間は $p = 0.116$，C2 と C3 の間は $p = 0.070$ であり有意差はみられない．この結果は C3 のサンプル数が C1 と C2 に比べ少ないことが原因と考えられる．

[1]本論文では JDRT 全 120 単語の 2 話者分の音声パワーに対して 0 dB を求めているため，1 話者の sustention だけの SNR は 0 dB が中心にならない．

[2]Fig. 4.5 の SNR は長時間平均値であり，0 dB の設定も Sustention の 20 単語 SNR(A) は特徴量に用いた cbSNRseg と同様に音声区間に対して求めたため，両者の SNR 値は一致せず順序のみしか比較できない．

Table 6.2: Average MCI by clustering (full set)

	C1	C2	C3	all
Test set	0.0835	0.0683	0.0695	0.0781
Train set	0.0961	0.0929	0.0897	0.0937

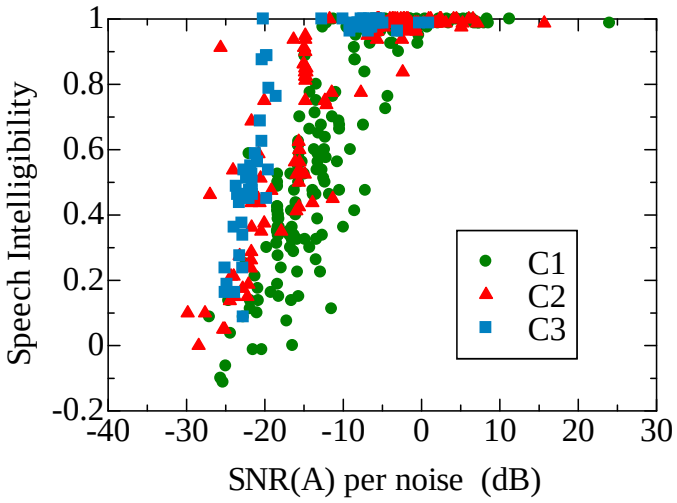


Fig. 6.1: Intelligibility vs. SNR(A) by cluster (Test set)

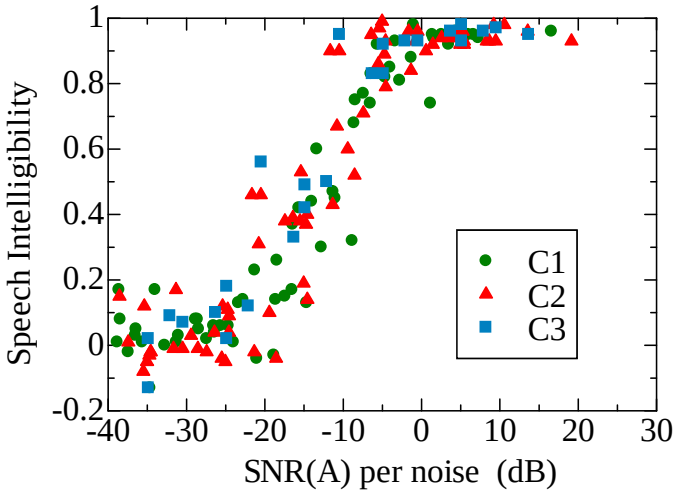


Fig. 6.2: Intelligibility vs. SNR(A) by cluster (Training set)

6.4 推定実験

オープンテストの主観評価結果をトレーニングセットで作成した推定関数から予測し，各推定関数の汎化性能を比較する．

6.4.1 推定条件

主観評価による了解度と 4 章と 5 章で作成した推定関数を比較する．比較に用いるのは 4 章で作成した 5 種の客観指標と交差検定で決定した回帰係数を用いた推定関数，5 章で検討した特徴量に cbSNRseg を用いた場合の最良のハイパーパラメータによる学習で作成した SVR，または L1 回帰の推定関数，この他にランダムフォレスト法（以下，RF）[179] による cbSNRseg を用いた回帰の 11 種を比較する．RF は，決定木を弱学習器とする集団学習アルゴリズムであり，単純な推定式を多数作成し，その平均値を予測値とする機械学習による回帰手法で，交差検定等による汎化性能の考慮をしなくてもそれなりの精度を出す手法である．RF の最大の特徴は学習時間の短さにあり，単純な回帰式を並列に多数用意しその平均を求める．本論文では SVR による推定式作成の時間をほとんど考慮していないため，計算時間の短い回帰手法の例として比較対象とする．RF を 5 章で比較しなかったのは，標準的な RF では交差検定を考慮しないことが多いこと[3]を考慮した．RF はトレーニングに用いた cbSNRseg のクローズテストの RMSE が最小になるように作成する決定木の数を調整した[4]．

オープンテストに用いた評価信号の各音質値および cbSNRseg を用いて全ての推定式の推定値を求め，主観評価値と推定値との式 (5.2) の RMSE と，式 (6.2) のピアソンの積率相関係数で比較する．式 (6.2) の $Sub(n)$ と $Est(n)$ は了解度の主観評価値と推定値，$\overline{Sub}$ と $\overline{Est}$ はそれぞれの平均値を示す．RMSE が同程度であれば，相関係数の高い方が良い回帰手法である．

$$r = \frac{\sum (Sub(n) - \overline{Sub})(Est(n) - \overline{Est})}{\sqrt{\sum (Sub(n) - \overline{Sub})^2} \sqrt{\sum (Est(n) - \overline{Est})^2}} \tag{6.2}$$

6.4.2 推定結果

各推定関数を用いて求めた推定了解度との RMSE を Table 6.3 に，ピアソンの積率相関係数を Table 6.4 に示す．Weighted sum は Table 6.1 の Test set の percentage を用いた重み付和である．

パラメトリック回帰 5 種の中では fwSNRseg(C) が最も RMSE が小さく，Weighted sum で 0.195 と最も小さい．ノンパラメトリックな多変量回帰の中では SVR(RBF) が，Weighted sum で 0.158 と 10 方式の中で最も RMSE が小さい．SVR(RBF) は，RMSE の重み付平均が fwSNRseg(C) に対して 0.77 倍まで減少している．相関係数は fwSNRseg(C) に対して僅かに低いものの，RMSE 差と異なり同一の値とみなせる程度である．以上の結果より，本論文で提案する提案する cbSNRseg を特徴量とした SVR による了解度推定は非常に有効である．汎化性能を要素ごとに考察する．

[3] Out of Bag 誤差などで同様の比較は行えなくはないが，厳密には交差検定と異なる．
[4] この様なチューニングでも過学習が起こりにくいことがランダムフォレストの特徴である．

Table 6.3: RMSE of open test

regression method	C1	C2	C3	Weighted sum	Multi
SNRseg	0.342	0.313	0.312	0.327	0.294
fwSNRseg(A)	0.218	0.207	0.310	0.228	0.241
fwSNRseg(C)	0.185	0.194	0.224	0.195	0.203
fwSNRseg(S)	0.235	0.218	0.291	0.238	0.248
AIseg	0.227	0.190	0.195	0.209	0.228
Ridge	0.203	0.224	0.226	0.214	0.215
Lasso	0.202	0.223	0.209	0.211	0.215
Kernel	0.440	0.529	0.183	0.431	0.295
SVR(linear)	0.218	0.224	0.224	0.223	0.230
SVR(RBF)	0.153	0.140	0.162	0.150	0.158
RF	0.162	0.172	0.150	0.164	0.187
MCI	0.0835	0.0683	0.695	0.0781	0.0781

Table 6.4: Pearson correlation(r) between subjective intelligibility and estimated intelligibility in open test

regression method	C1	C2	C3	Multi
SNRseg	0.746	0.546	0.719	0.675
fwSNRseg(A)	0.875	0.875	0.847	0.824
fwSNRseg(C)	0.898	0.915	0.892	0.888
fwSNRseg(S)	0.847	0.880	0.864	0.836
AIseg	0.818	0.871	0.884	0.771
Ridge	0.828	0.819	0.862	0.814
Lasso	0.828	0.819	0.870	0.814
Kernel	0.375	0.335	0.785	0.571
SVR(linear)	0.828	0.829	0.866	0.814
SVR(RBF)	0.892	0.914	0.899	0.874
RF	0.888	0.855	0.879	0.812

6.4.3　オープンテスト推定の考察

回帰手法の選択について

全体で最も推定性能が高かったのは SVR(RBF) で，次いで RF，fwSNRseg(C) の順であった．この結果は，ノンパラメトリック回帰を用いるときは RBF カーネルを用いてきちんとハイパーパラメータをチューニングし，汎化性能の高い推定関数を作ることができれば了解度推定性能は非常に高い物が作れるが，チューニングを十分にできないときはランダムフォレスト法や fwSNRseg(C) を説明変数に用いたパラメトリック回帰で十分であることを示唆する．つまり，予備実験等で信号処理アルゴリズム間の順位がわかれば良い時などの「ほどほどの推定精度」で良ければ，推定式を作成するのに時間のかかる SVR を用いる必要は無く，高速で学習可能な RF や fwSNRseg(C) によるパラメトリック回帰を用いれば比較的少ない計算コストで了解度推定ができる．

Fig. 6.3 に SVR(RBF) と fwSNRseg(C) の推定精度の比較をクラスタごとに示す．どちらも同程度の相関係数であるため，どの図もおおむね同傾向であるが，SVR(RBF) の方が等価線に近づいており，RMSE が小さくなっていることがわかる．他の回帰手法での推定精度は D に掲載する．

しかしながら，SVR(RBF) の RMSE でもいまだ MCI より大きいことを考えれば，より推定性能を向上させることができる特徴量がある可能性がある．また，交差検定でハイパーパラメータと同様に求めた正規化範囲の上限値と下限値も特徴量の次元ごとに行うことで推定性能が向上すると考えられる．

ランダムフォレスト法との比較について

ランダムフォレスト法は集団学習による弱回帰器式[5]を用いた高速な回帰・分類手法であり，明確な回帰基準を求める SVR とは異なる発想による機械学習の手法である．ランダムフォレスト法は SVM または SVR と比較されることが多く，多くの場合で SVM または SVR よりも若干精度が劣る．本論文でも Table 6.3 の Weighted sum では SVR に劣るものの，C3 に関しては 0.0150 で，SVR の 0.164 よりも高精度である．これは C3 のサンプル数が少なく，サンプル数が回帰精度に影響するカーネル法を用いた SVR よりもランダムフォレスト法の決定木間で重複を可とするランダムサンプリングによる弱回帰式を用いることに効果があったためであると推測される．よって特徴量の次元数よりもサンプル数が少ない場合にはランダムフォレスト法が有効である．

騒音クラスタリングの有効性について

Fig. 6.1 の主観評価結果ではクラスタ間差が明確にみられ，Table 6.3 の結果からも SVR(RBF) を用いた推定関数の RMSE は Weighted sum が Multi をほとんどの条件で下回る．特に Kernel は C1 と C2 で RMSE が他の回帰手法の 2 倍以上であり，相関も極端に低く過学習を起こしている．C3 については SVR(RBF) の次に RMSE が小さいが，相関係数は最も低く推定に用いる回帰手法に適さない．この他に SNRseg を用いたシグモイド回帰では Multi の方が RMSE が小さい．これら二つの推定法は他と比べて明らかに推定性能が低く，騒音クラスタリングの効果も見られなくなっている．一方でこの他の 8 種では騒音クラスタリングの効果があったと考えられる．しかし，

[5]推定精度の低い回帰式.

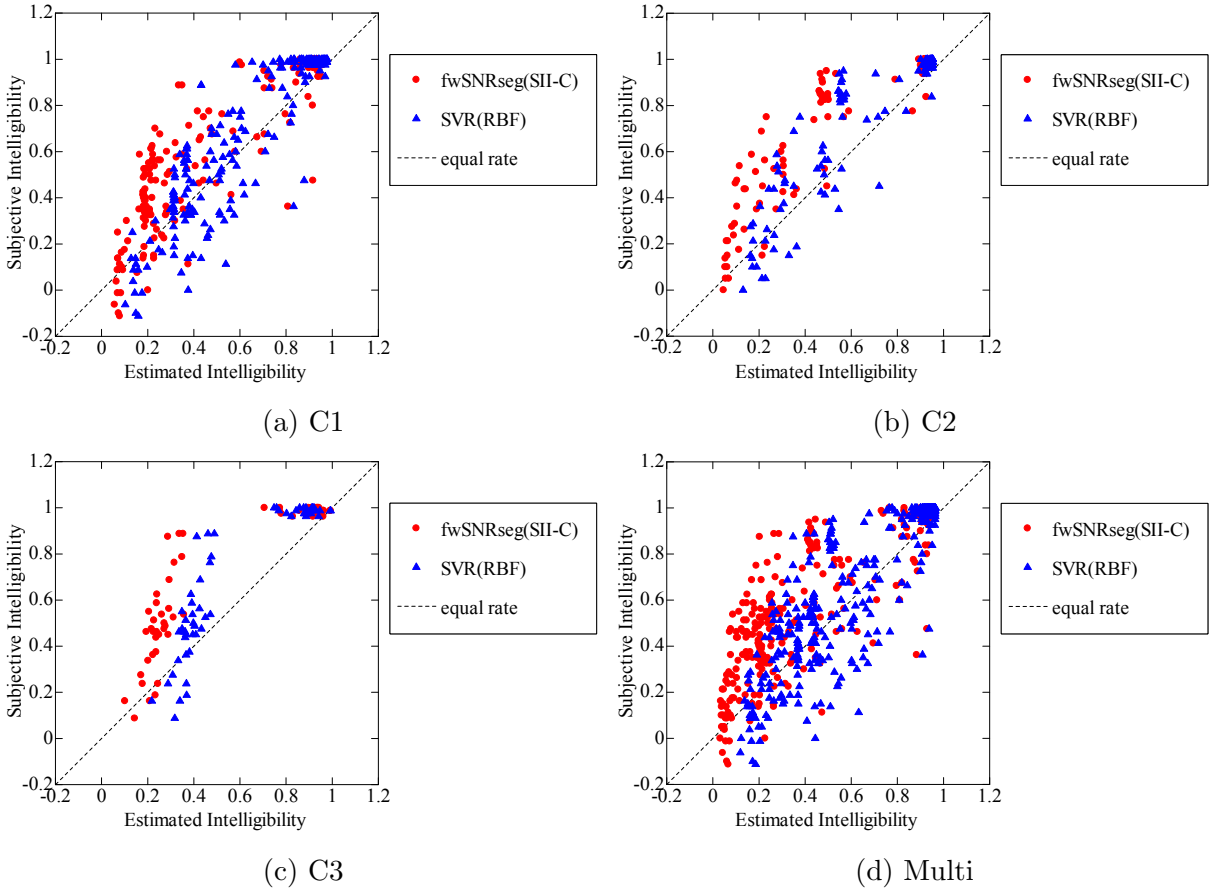


Fig. 6.3: Relationship between subjective and estimated intellugubility using SVR(RBF) and fwSNRseg(C)

最良な回帰手法である SVR(RBF) でその差は 0.008 と 1%未満であり，騒音クラスタリングの効果はほとんどなく，AIseg では了解度で 2%程度の RMSE 改善があるため，単独で用いた場合の推定性能がやや低い場合に騒音クラスタごとの推定関数が有効であったことを示唆している．提案法である SVR と cbSNRseg を用いた推定では，推定関数自体の汎化性能が高く，騒音クラスタリングの効果がほとんど見られなかったと考えられる．

MCI との比較について

Table 6.3 の結果より，オープンテストで最も RMSE が小さかった SVR(RBF) であっても RMSE は MCI よりも大きい．MCI は被験者による回答の分散を考慮した指標であるため，SNR の違いといったサンプル間の分散がある程度小さい場合には有効な指標である，しかし本論文の結果では騒音 LF による影響も均一ではないことを考慮する必要がある．Table 6.2 の結果でも Test set の MCI は Training set の MCI よりも小さくなっている．被験者数が 10 人ないし 8 人の比較的少

ない実験であるが，極端にSNRが悪くなるサンプルを含んでいないことを考えるとこの数値は妥当である．よって，被験者の分散によるMCIでは推定性能の指標としては十分ではない．また，交差検定誤差を考慮した5.3.5項の比較でもMCIはRMSEより大きく，未知データを想定ないし用いた推定にはMCIとの比較は妥当ではない．

高橋らによる文献[158]のMCIは,「ディジタル電話の受聴品質の推定」という非常に閉じた系の性能推定である．文献[158]が対象としたディジタル電話網の品質は系雑音はすでにほとんどなく，コーデックによる音声のひずみ感や，エコーキャンセラーの性能による劣化が主要因であった．このため，実験系による分散よりも，MOS評価による被験者の分散が大きい実験である．このため，電話系の受聴品質推定に対して，MCIは有効な指標であったと推測される．了解度の推定に関しても，電話系などの比較的条件が統制される場合にはMCIとの比較は妥当であると考えられるが，本論文が対象とする屋外での騒音を対象とした実験の目標指数にMCI以外の指標を検討する必要がある．

SVRの有効性について

5.2.1項で取り上げたSVRの汎化性能を保障する技術が了解度推定に有効であった理由を考察する．5.2.1項で述べたSVRの汎化性能が高い理由として挙げた下記の三点(a)〜(c)の他に非線形回帰の特徴として(d)，カーネル法の特徴として(e)を追加する．

(a) 正則化を用いることで（実質的に）変数選択を行う

(b) ϵ-不感応関数を用いるロバストな回帰

(c) 回帰関数とサポートベクトルとのマージンを最大化する明確な基準

(d) 交差検定によるハイパーパラメータの決定

(e) カーネル法を用いるノンパラメトリック回帰

(a)について，入力特徴量の次元数をモデル作成時に最適化することと等価なため，聴覚実験を基に設定したな細かい帯域分割による特徴量であっても了解度推定向けに最適化が可能となるのがメリットである．特に線形重回帰といった単純な実装と比べて，少数サンプルであってもモデルが作成できるのは大きなメリットである．また，騒音クラスタごとに最適な聴取帯域より細かく広い帯域分割法を用いることになるとしても正則化を用いる回帰手法であれば推定関数の最適化が行えるため，了解度推定のための多変量回帰問題において，正則化を用いることはもはや必須であるといえる．

(b)について，了解度は心理値であるため，十分な数の被験者で多くの実験パラメータを含めた主観評価による結果が得られる場合は，回帰したい関数に近いサンプルが得られる．このためパラメトリックな回帰でも十分推定が可能である．しかし，Fig. 6.1とFig. 6.2からも明らかなように8〜10人程度の主観評価では分散が完全に収束しているとは言い難く，回帰したい関数の大まかな形状は見えてくるものの，サンプルの分散は収束しているとは言い難い．MCIは，主観評価による大局的な音質変化の構造を明らかにし，微細な構造は計測誤差未満であるとすることである．よって，回帰したい推定関数近傍の誤差項を0とするϵ-不感応関数を用いることは，主観評

価などで計測限界未満にあたる微細な分散による影響を考慮することとなる．これは被験者の経験や心理実験の解答法による分散が比較的少ない安定した主観評価法であるほど回帰したい関数との誤差が小さくなるため，言語を用いた音質評価など比較的主観評価の統制がとりやすい実験系における推定に適していると考えられる．

(c) と (d) について，SVR(RBF) が最良の選択肢であることがわかっても，最良な推定関数を作成するためのハイパーパラメータの探索にかかる時間[6]がかかる欠点がある．しかし，誤差逆伝搬法によるニューラルネットを用いた回帰と異なり，必ず大域最適解が求まるメリットがある．このため事前の推定関数のチューニングに十分な時間を取れる場合は SVR(RBF) は最良の選択肢であるが，そうでない場合は Table 6.3 の結果より，RMSE 差が 0.014 程度の RF の使用を考慮する必要がある．

(e) についてはカーネル法共通の問題であるが，リプレゼンター定理を考慮するとサンプル数が増えるほど複雑な回帰を再現できる．これは膨大な数の騒音サンプルを主観評価するのが困難であるため了解度推定を行いたいという本論文の命題に対して矛盾している．しかしながら，本論文で扱ったサンプル数の場合に限ったとしても，他の L1 回帰法よりも汎化性能は高かったため，同一条件であればカーネル法の中でも過学習を抑制しやすい SVR を用いるのが有効である．

6.5 まとめ

本論文で提案した騒音クラスタリングと SVR を用いた了解度推定について，5 種の既存尺度を用いたシグモイドカーブフィッティングによる推定関数，SVR 以外のの多変量回帰手法 3 種との比較をオープンテストで行った．その結果，以下を明らかにした．

- ランダムサンプリングで作成したテストセットのクラスタリングであっても，騒音クラスタごとの同一の SNR(A) の値での了解度の序列はトレーニングセットと等しい．しかし，クラスタ境界に含まれるサンプルはトレーニングセットよりも多い．
- 上記を考慮すると，本論文の騒音クラスタリングだけでは完全に同傾向の騒音を完全に分類しきれていないと考えられる．
- 交差検定の RMSE ではほとんど差がみられない SVR と他の L1 正則化を用いたノンパラメトリック回帰と各種重みを用いた fwSNRseg によるパラメトリック回帰の RMSE もオープンテストで評価した場合に汎化性能は大きく異なり，交差検定の RMSE の序列だけで決定できない．
- SVR(RBF) の汎化性能が最も高く，同じく機械学習による RF がそれに次ぐ．fwSNRseg(c) を説明変数としたシグモイドカーブへのパラメトリック回帰は SVR(RBF) と RF を除く正則化を考慮した L1 回帰や他の重みによる fwSNRseg よりも汎化性能が高い．
- SVR は比較的少数のサンプルからでも汎化性能が高い推定関数を作成できるため，了解度推定の様に教師データが主観評価値のため多数のサンプルを集められないような場合に有効である．

[6]本論文の cbSNRseg を用いる場合の探索は正規化 SNR の組み合わせごとに数スレッドに処理を分割しても 3〜4 週間程度かかった．

- SVR はサンプル数の増加によって回帰式の学習に用いる時間が増加するため，多少精度は落ちるがランダムフォレスト法を用いる場合が良い場合もあり，特に学習サンプルが特徴量次元よりも少ない場合に有効である．
- しかし，オープンテストで最も RMSE が小さかった SVR(RBF) であっても RMSE は MCI よりも大きいため，本論文で提案した騒音クラスタリングと SVR を用いた推定について改良の余地はある．

以上の結果より，本論文では，cbSNRseg と RBF カーネルを用いた SVR による騒音クラスタ別の了解度推定は，騒音下での Sustention 了解度推定において，改良の余地はまだあるものの，比較した方式の中では最も最適な手法であると結論付ける．

第7章　結論

本章では，本論文を総括し，今後の検討課題について具体的に述べる．

7.1　総括

本論文では，騒音環境下における音声了解度を直接推定する手法に，機械学習の手法を適応することについて提案した．提案法は騒音の音色情報を用いたクラスタリングを行い，騒音クラスタごとに汎化性能の高い非線形回帰である SVR を用いて了解度推定関数を作成する．本論文では既存の尺度による了解度推定の問題点と，提案法の汎化性能を検証した．以下，本論文の各章を要約する．

1 章では，まず，本研究の背景として，今後より広く利用されるであろう屋外の騒音下における音声システムの利用の問題点についてを述べ，その中で筆者の考えを提示した．次に，音声品質の主観評価法と客観評価法の技術開発の経緯について受聴品質，ラウドネス，明瞭度，了解度について述べた．既存の研究では，単語を用いた了解度の予測の目安になるインデックス値を求める STI，SII 等があるものの，了解度値を直接推定する方式は無い．これより，本研究の目的を，騒音環境下における主観評価によって求める了解度値と対応する推定了解度値を求めることとした．

2 章では，予備実験としてバイノーラル音声システムを用いた騒音による妨害がある了解度主観評価を JDRT を用いて行い，既存尺度 16 方式を用いて了解度推定を行った．その結果，以下の 2 点が明らかとなった．

(a) Sustention は騒音種の影響を受けやすい．

(b) 既存尺度を用いた推定では何らかの聴覚重みを用いた fwSNRseg が良い．

3 章では，上記の課題を解決するために以下の 2 点を機械学習を用いて実装した了解度推定を提案した．

(α) 騒音種による最適な推定関数の選択

(β) 推定に最適な聴覚重みの作成

解決法 (α) は課題 (a) に対応し，騒音種の影響が同傾向の評価条件ごとに了解度推定関数を切り替えることを目的とする．このために騒音信号の音色特徴を求め，音色特徴のみを用いた教師なし学習である騒音クラスタリングを提案した．解決法 (β) は課題 (b) に対応し，特定の周波数帯域ごとに帯域分割した SNRseg に最適な回帰係数を求めることとした．このため，主観評価結果を教師信号とする教師あり学習の SVR を用いた了解度推定関数を騒音クラスタごとの作成を提案した．

4 章では，騒音クラスタリングの詳細について述べた．電子協騒音データベースのダイジェスト版を 3 sec ごとの LF に分割し，個々を 1 騒音として 15 次元の音色特徴を求め，x-means 方でク

ラスタリングした．その結果，605 個の LF が 3 つの騒音クラスタに分割された．各クラスタから代表騒音 32 種を求め，SNR を 5 段階に設定し，主観評価を行ったところ，クラスタ間に有意差がみられた．また，2 章の検討結果を用いて既存の品質評価尺度による推定関数を作成した．その結果，聴覚重み付けの無い SNRseg では交差検定の RMSE が大きく，何らかの聴覚重みが必要であることが明らかになった．

5 章では，騒音クラスタごとに機械学習を用いたノンパラメトリックな回帰による推定関数を作成した．提案法である SVR の他に，リッジ回帰，Lasso を用いた回帰，L1 正則化を用いたカーネル回帰を比較した．用いる特徴量にはクリティカルバンドによる 25 帯域への分割と 1/3 オクターブバンドによる 18 帯域への分割の 2 種類の正規化セグメンタル SNR を 4 章の主観評価結果を教師信号とした交差検定で比較した．その結果，全ての回帰手法でクリティカルバンドを用いる方が良いことを明らかにしたが，回帰手法の比較は交差検定だけでは不可能だった．

6 章では，4 章で作成した既存の客観音質を用いたシグモイドカーブフィッティングによるパラメトリック回帰による推定関数 5 種，正規化クリティカルバンドセグメンタル SNR を用いた 5 種の回帰式およびランダムフォレスト法の 11 種の推定関数について，推定関数作成に用いなかったオープンデータを用いて比較し，SVR に RBF カーネルを用いた推定は RMSE が 0.015 と最も性能が高いことを示した．同様に機械学習によるランダムフォレスト法による推定式は RMSE が 0.164 と若干精度は落ちるものの，学習時間差を考慮すれば十分な性能である．また従来法として高精度だった，fwSNRseg(C) を用いたシグモイドカーブフィッティングによる推定と比較すると RMSE は 0.045（0.77 倍）小さく，SVR による推定は非常に高い汎化性能を持っていることを明らかにした．

以上の特に 3 章から 6 章の検討から，騒音クラスタリングと SVR を用いた提案推定法は，既存の手法よりも高い汎化性能を持つ推定法であることを示した．本提案法を用いることにより，屋外で用いる音声システム設計に置ける新たな指針となり，より利便性の高い音声システムが開発に利用されることを期待する．

7.2 今後の展望

本論文で提案した，了解度推定法はいまだ完全なものとは言い難い．ここでは，提案法の中長期的な改良を中心とした今後の展望を述べる．

人工騒音と自然騒音

本論文では 2 章で検討した結果より，騒音による影響が顕著にみられた Sustention に絞って 3 章以降の主観評価と推定を行った．Fig. 2.35–(c) と Fig. 4.7–(c) を比べると，2 章の騒音による特異な影響は，白色雑音の様な自然騒音ではない人工音に対してのみだった可能性がある．付録 B 節に示した騒音ごとのスペクトログラムを見ても，自然騒音は低域から高域にかけてなだらかに減衰していくスペクトル構造を持つため，高域成分によるマスキングが Sustention の騒音傾向差であったとも考えられる．このため自然騒音と人工騒音の比較を検討する必要がある．

他の子音特徴の評価

また，2 章の結果を用いた推定実験であったため，3 章以降は Sustention 以外の子音特徴を評価していない．6 章の結論では，RBF カーネルを用いた SVR はは fwSNRseg(S) によるカーブフィッティングよりも高精度であった．しかし，SII の子音重みは全ての子音の平均的な重みとして設計されているため，本論文の比較だけでは最適な比較とも言い難い．よって，子音特徴ごとの SVR による推定を行い，子音特徴ごとに fwSNRseg による回帰と比較する必要がある．また子音特徴ごとの推定結果を統合した平均了解度の推定精度も必要である．

ファジィクラスタリング

4 章で検討した騒音クラスタリングでは，その後の主観評価値選択の明確な基準を求めるため，非階層クラスタリングでハードクラスタリングの x-means アルゴリズムを用いた．しかし，Fig. 6.1 より，オープンテストではクラスタ間の重なりが見られる．そこで重なりを許容するファジィクラスタリングの考慮が必要である．

MIR 特徴量の最適化

4.2.2 項で検討した MIR 特徴量は，提案した特徴全ての検定結果が有意な傾向 ($p \simeq 0.05$) にあったため，そのまま全て用いた．しかし下位検定の結果では一部の特徴量間で有意差が無かったため，性能向上のためには主成分分析等を用いて特徴量の最適化が考えられる．4.2.2 項の検討時には，特徴量と主観値の傾向が未知であったため，特徴量選択を行わなかった．6.1 節でオープンテストのためにランダムサンプリングした 400 個の騒音 LF は，フルセットの統計量と比べ各特徴量ともに分布を良く近似しており，騒音クラスタリングの特徴量最適化に用いるのに最適であると考えられる．ファジィクラスタリングの検討と共に，騒音クラスタリングを最適化した場合に，SVR による推定関数の汎化性能の高さが騒音クラスタリングよりも高くなるか検証することが必要である．

SVR ハイパーパラメータ探索の高速化

現在の SVR ハイパーパラメータの探索はメッシュ法による総当たりであるため，非常に時間がかかる．特に RBF カーネルでは，C と γ だけで SNRseg の上下限値の組み合わせ 1 つにつき 1 スレッドで計算した場合に約 1 日かかる．本論文の予備実験では，粗いメッシュと細かいメッシュを用いる二重探索法も検討したが，現在の cbSNRseg と RBF カーネルの組み合わせの値は粗いメッシュで確認できるピークではなく，最適な探索法ではなかった．今後，バンドごとに特徴量の上下限値の比較を行うと，組み合わせ数が膨大となる．このため同一のメッシュを用いる場合には，分散計算を用いるなどの実装上のテクニックを用いるとと共に，伊藤らによる MCV 正則化 [180] 等の効率の良いハイパーパラメータ探索を検討する．現状では，RMSE が二番目に小さいランダムフォレスト法は計算時間が圧倒的に早いメリットがあるため，パラメータ探索の高速化を考慮した SVR と計算時間等も含めた総合比較を行う必要がある．

MIR 特徴量の SVR への利用

5 章では，SNRseg の最適な回帰係数を用いるという命題に基づき，SVR の特徴量は帯域ごとの SNRseg とした．これは fwSNRseg の導出過程を模擬したためだが，SVR は異なる次元の物理量の結合にも用いることができる．現在の推定では騒音クラスタリングの結果であるクラスタ番号を SVR 推定関数の切り替えにしか用いていない．クラスタリングに用いた MIR 特徴量を SVR に用いれば，騒音クラスタリングモデルを SVR 推定関数の外ではなく，推定関数内で特徴量の結合として持つことが可能となる．一般に，カーネル回帰は特徴量数よりもサンプル数の増加の方が汎化性能が高くなるため，特徴量数の次元を不要に増加させるのは望ましくない．しかし，次元数の少ない obSNRseg の方が汎化性能が低かったことを考慮すると，素性の良い特徴量であれば汎化性能が向上する可能性がある．

ノンリファレンス特徴量による推定

上記の MIR 特徴量の利用と同様に，3SQM（ITU–T P.563）等と同様のノンリファレンス特徴量の結合による了解度推定を検討する．MIR 特徴量は騒音の音色特徴であり，3SQM では雑音抑圧を用いて推定した音声とのセグメンタル SNR 等を用いているため，特徴量の重み付き結合に SVR を用いれば，本論文で検討した obSNRseg ／ cbSNRseg と異なり，原音が入手できない評価系でも了解度を推定できる．1 章で考察した実社会における騒音下での音声システムの利用では，STI の様に原音を自ら提示しながら評価するには手間がかかるため，ノンリファレンス特徴量による推定が必要である．

実空間音源の推定

本論文の主観評価では，2 章の一部を除き，全て電子協騒音データベースを用いた．これは検聴表に録音環境の詳細が記載されているため扱いやすい．しかし，提案モデルの真のオープンテストを考慮した場合，完全に未知な騒音を用いる必要がある．特に，3 章以降の検討では伝達関数等を考慮せず，単純な騒音と音声の加算で実験を行っている．このため，民生品での利用を考えると，録音条件の厳密な統制は不可能であり，残響等の伝達特性も考慮しなければならない．このため，提案の性能評価に，民生品のスマートフォン等で録音された騒音等を用いて，より現実的な分析を行う必要がある．

7.3 結び

以上の様に，本論文で提案した了解度推定法には解決するべき課題は多いものの，従来の STI, SII 等によるインデックス値予測と異なり，了解度の推定値を直接求めることが可能になる．また，既存の品質評価を用いた回帰や，他の L1 正則化を用いた回帰よりも汎化性能が高い．提案法がより良い音声システム設計の一指標となることを願い，本論文の結びとする．

謝 辞

本論文をまとめるにあたり，多くの方々のご指導及びご協力を頂きました．この場を借りて，以下に名前をあげられなかった方々も含め皆様に感謝します．

本研究を行うに当たって，山形大学大学院理工学研究科 近藤和弘准教授には主指導教員として大学院への進学を受け入れていただき，学部 4 年生の研究室配属以後のべ 6 年にわたり熱心にご指導を頂きました．近藤先生からは，音声工学に対する知識の教授だけでなく，研究活動の基礎から，物事の捉え方，プログラミング，プレゼンテーション技法，国際会議発表論文をはじめとした多くの論文添削，海外旅行のマナーなど，この欄だけでは挙げきれないほど様々なことを学ばせて頂きました．ここに深く感謝します．

本論文を作成するにあたり，山形大学大学院理工学研究科 小坂哲夫教授には博士前期課程から副査として論文の審査において貴重な御意見をいただき，音声言語処理特論では，大変有益な文献をご紹介頂きました．ここに深く感謝します．田村安孝教授，大槻恭士准教授には博士後期課程から副査として論文の審査において貴重な御意見をいただきました．ここに深く感謝します．

山形大学大学院理工学研究科 中川清司元教授，高野勝美准教授，堺三洋技術専門職員には，学部ゼミや研究室・グループゼミの場において貴重なご意見，ご指導を頂き，研究室生活においても大変お世話になりました．ここに深く感謝します．古閑敏夫元教授には博士後期課程の画像伝送工学を受け持って頂き，研究活動に対してもご助言頂きました．ここに深く感謝します．

日本電気通信システム株式会社 渋谷徹博士には，父子ほどの年齢差がありながら，研究活動に対し貴重なご意見を頂きました．さらに社会経験のない私に，企業における研究開発の話から，電子メールの作法まで非常に多くのことを学ばせて頂きました．ここに深く感謝します．近藤研究室の三浦正範氏には，先輩後輩として機械システム分野出身であり異なる視点からご意見をいただきました．ここに深く感謝します．この他，近藤研究室に所属した学生の皆様には，膨大な数の主観評価にご協力頂きました．本論文を書き上げることができたのも一緒に研究室生活を送った，多くの先輩，同輩，後輩の皆様おかげです．ここに深く感謝します．

東北学院大学 岩谷幸雄教授には東北大学在籍中に共同研究プロジェクトの担当として貴重なご意見をいただきました．ここに深く感謝します．独立行政法人産業技術総合研究所健康工学研究部門 中川誠司主任研究員，籠宮隆之特別研究員には学会発表等において貴重なご意見をいただきました．ここに深く感謝します．この他，学会等でご助言いただいた多くの研究者の皆様，論文査読者の皆様に深く感謝します．

最後に，9 年に渡る長い大学生活を支えてくれた両親と婚約者に感謝します．

参 考 文 献

[1] 大石 康智，後藤 真孝，伊藤 克亘，武田一哉，“スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別 (音楽情報,< 特集 > 情報処理技術のフロンティア),” 情報処理学会論文誌，vol.47，no.6，pp.1822–1830，2006.

[2] Ronald T. Azuma, “A Survey of Augmented Reality,” Teleoperators and Virtual Environments, vol.6, no.4, pp.355–385, Aug. 1997.

[3] 鈴木 陽一，西村 竜一，“超臨場感音響の展開,” 電子情報通信学会誌，vol.93，no.5，pp.392–396，May. 2010.

[4] 安藤 彰男，“音響の高臨場感技術,” 映像情報メディア学会誌，vol.66，pp.671–672，2012.

[5] 小澤 賢司，小坂 直敏，山内 勝也，高田 正幸，藤沢望，音色の感性学，岩宮 眞一郎（編），コロナ社，2010.

[6] ITU–T, “Methods for subjective determination of transmission quality,” Recommendation P.800, International Telecommunication Union, Aug. 1996.

[7] ITU–T, “Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm,” Recommendation P.835, International Telecommunication Union, Nov. 2003.

[8] ITU–T, “Modulated noise reference unit (MNRU),” Recommendation P.810, International Telecommunication Union, Feb. 1996.

[9] 伊藤 憲三，北脇 信彦，筧一彦，“音声のディジタル波形符号化方式の客観的品質評価尺度の検討,” 電子通信学会論文誌（A），vol.66，no.3，pp.274–281，1983.

[10] N. Kitawaki, M. Honda, and K. Itoh, “Speech quality assessment methods for speech coding systems,” IEEE Comm. Magazine., vol.22, pp.26–33, 1984.

[11] J.H.L. Hansen, and B. L. Pellom, “AN EFFECTIVE QUALITY EVALUATION PROTOCOL FOR SPEECH ENHANCEMENT ALGORITHMS,” Proc. International Conference on Spoken Language Processing 98, 1998.

[12] 山田 武志，牧野 昭二，北脇信彦，“雑音抑圧音声の主観・客観品質評価法,” 日本音響学会誌，vol.67，no.10，pp.476–481，Aug. 2011.

[13] B.S. Atal, “The history of linear prediction,” Signal Processing Magazine, IEEE, vol.23, no.2, pp.154–161, 2006.

[14] ITU–T, “Objective quality measurement of telephoneband (300—3400 Hz) speech codecs,” Recommendation P.861, International Telecommunication Union, Aug. 1996.

[15] ITU–T, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” Recommendation P.862, International Telecommunication Union, Feb. 2001.

[16] ITU–T, “Mapping function for transforming P.862 raw result scores to MOS–LQO,” Recommendation P.862.1, International Telecommunication Union, Nov. 2003.

[17] ITU–T, “Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs,” Recommendation P.862.2, International Telecommunication Union, Nov. 2007.

[18] ITU–T, “Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2,” Recommendation P.862.3, International Telecommunication Union, Nov. 2007.

[19] 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.93，no.9，pp.583–593，2010.

[20] T. Yamada, M. Kumakura, and N. Kitawaki, “Performance estimation of speech recognition system under noise conditions using objective quality measures and artificial voice,” IEEE Trans. Audio, Sp. & Lang. Process., vol.14, no.6, pp.2006 – 2013, Nov. 2006.

[21] K. Kondo, Subjective Quality Measurement of Speech – Its Evaluation, Estimation and Applications, Springer–Verlag, Mar. 2012.

[22] J.G. Beerends, E. Larsen, N. Iyer, and J.M. van Vugt, “Measurement of speech intelligibility based on the PESQ approach,” Proc. On-line Workshop Meas. Sp. Audio Quality Netw. ’04, Jun. 2004.

[23] T. Yamada, M. Kumakura, and N. Kitawaki, “Objective Estimation of Word Intelligibility for Noise–Reduced Speech,” IEICE Trans. Commun., vol.E91-B, no.12, pp.4075–4077, Dec. 2008.

[24] ITU–T, “Perceptual objective listening quality assessment,” Recommendation P.863, International Telecommunication Union, Jan. 2011.

[25] ITU–T, “Single–ended method for objective speech quality assessment in narrow-band telephony applications,” Recommendation P.563, International Telecommunication Union, May. 2004.

[26] ITU–R, “A guide to ITU-R Recommendations for subjective assessment of sound quality,” Recommendation BS.1116–1, International Telecommunication Union, Dec. 2003.

[27] ITU–R, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” Recommendation BS.1116–1, International Telecommunication Union, Oct. 1997.

[28] ITU–R, “General methods for the subjective assessment of sound quality,” Recommendation BS.1284–1, International Telecommunication Union, Oct. 1997.

[29] ITU–R, “Pre–selection methods for the subjective assessment of small impairments in audio systems,” Recommendation BS.1285, International Telecommunication Union, Oct. 1997.

[30] ITU–R, “Methods for the subjective assessment of audio systems with accompanying picture,” Recommendation BS.1286, International Telecommunication Union, Oct. 1997.

[31] ITU–R, “Method for the subjective assessment of intermediate quality levels of coding systems,” Recommendation BS.1534–1, International Telecommunication Union, Jan. 2003.

[32] ITU–R, “Method for objective measurements of perceived audio quality,” Recommendation BS.1387–1, International Telecommunication Union, Nov. 2001.

[33] ISO/IEC, “Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio,” Technical Report 11172, International Organization for Standardization / International Electrotechnical Commission, 1993.

[34] ISO/IEC, “Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC),” Technical Report 13818–7, International Organization for Standardization / International Electrotechnical Commission, Oct. 2004.

[35] ISO/IEC, “Information technology – Coding of audio–visual objects – Part 3: Audio,” Technical Report 14496–3, International Organization for Standardization / International Electrotechnical Commission, Sep. 2009.

[36] R. Huber, and B. Kollmeier, “Pemo–Q –A new Method for Objective Audio Quality Assessment using a Model of Auditory Perception,” IEEE Trans. Audio, Sp. & Lang. Process., vol.14, no.6, pp.1902–1911, 2006.

[37] 三浦 種敏, “音質評価の基本問題,” 日本音響学会誌, vol.20, no.3, pp.145–146, 1964.

[38] 中山 剛, 越川 常治, 三浦 種敏, “音質評価法の基本的考察,” 日本音響学会誌, vol.21, no.4, pp.209–215, 1965.

[39] 中山 剛, 三浦 種敏, “音質評価の方法論について,” 日本音響学会誌, vol.22, no.6, pp.319–331, 1966.

[40] 中山 剛, 宮川 陸男, 三浦 種敏, “音質の総合評価,” 日本音響学会誌, vol.22, no.6, pp.332–339, 1966.

[41] ITU–R, “The E–model: a computational model for use in transmission planning,” Recommendation G.107, International Telecommunication Union, Dec. 2011.

[42] H. Fletcher, and W. A. Munson, “Loudness, Its Definition, Mesurement and Calculation,” J. Acoust. Soc. Am., vol.5, pp.82–108, 1933.

[43] D. W. Robinson, and R. S. Dadson, “A redetermination of the equal–loudness relations for pure tones,” Br. J. Appl. Phys., vol.7, pp.166–181, 1956.

[44] ISO, “Normal equal–loudness–level contours,” Technical Report 226:2003, International Organization for Standardization, 2003.

[45] IEC, “Electroacoustics – Sound level meters – Part 2: Pattern evaluation tests,” Technical Report 61672:2003, International Electrotechnical Commission, Apl. 2003.

[46] ITU–T, “Determination of loudness ratings; fundamental principles,” Recommendation P.76, International Telecommunication Union, Nov. 1988.

[47] ISO, “Methods for calculating loudness level,” Technical Report 532B, International Organization for Standardization, 1975.

[48] ITU–R, “Algorithms to measure audio programme loudness and true-peak audio level,” Recommendation BS.1770–3, International Telecommunication Union, Aug. 2012.

[49] ITU–R, “Requirements for loudness and true–peak indicating meters,” Recommendation BS.1771–1, International Telecommunication Union, Jan. 2012.

[50] ITU–R, “Operational practices for loudness in the international exchange of digital television programmes,” Recommendation BS.1864, International Telecommunication Union, May. 2010.

[51] 山下 公一，松平 登志正，“語音聴力検査,” Audiology Japan，vol.51，pp.167–176，2008.

[52] H. Fletcher, and J. C. Steinberg, “Articulation testing methods,” Bell System Technical Journal, vol.8, pp.806–854, 1929.

[53] 落合 宣之，“空間中における音韻の明瞭性,” 心理学研究，vol.13，pp.285–287，1938.

[54] 日本音響学会明瞭度委員会，“明瞭度試験法の基準,” Technical report，日本音響学会，1957.

[55] 飯田 茂隆，“建築分野の明瞭度試験 (< 小特集 > 試験用音声の標準化),” 日本音響学会誌，vol.41，no.10，pp.704–708，1985.

[56] N. R. French, and J. C. Steinberg, “Factors Governing the Intelligibility of Speech Sounds,” J. Acoust. Soc. Am., vol.19, no.1, pp.90–119, 1947.

[57] 文部省科学研究費聴力測定法の規準班，“聴力測定法の基準,” Technical report，文部省科学研究費総合研究報告書，1956.

[58] 日本聴覚医学会，“補聴器適合検査の指針（2010）,” Audiology Japan，vol.53，pp.708–726，2010.

[59] 粕谷 英樹，“合成音声の品質評価法 (< 特集 > 音声合成技術の動向と今後),” 日本音響学会誌，vol.49，no.12，pp.866–870，1993.

[60] L. C. W. Pols, Quality assessment of text–to–speech synthesis by rule, S. Furui, and M. M. Sondhi, eds., Marcel Deckker Inc., New York, 1991.

[61] J.P. Egan, Psycho–Acoustics Laboratory Report, OSRD, 1944.

[62] G. Fairbanks, “Test of phonetic differentiation: The rhyme test,” J. Acoust. Soc. Am., vol.30, pp.596–600, 1958.

[63] A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, “Articulation testing methods : Consonantal differentiation with a closed–response set,” J. Acoust. Soc. Am., vol.37, pp.158–166, 1965.

[64] W. D. Voiers, Diagnostic Evaluation of Speech Intelligibility, in Speech Intelligibility and Speaker Recognition, M. E. Hawley, ed., Dowden, Hutchinson & Ross, PA, 1977.

[65] W. D. Voiers, “Evaluating Processed Speech using the Diagnostic Rhyme Test,” Speech Tech., vol.1, pp.30–39, 1983.

[66] ANSI, “Method For Measuring The Intelligibility Of Speech Over Communication Systems,” Technical Report S3.2–2009, American National Standards Institute, 1989, reaffirmed 1995, 1999 and 2009.

[67] R. Jakobson, C.G.M. Fant, and M. Halle, Preliminaries to speech analysis; the distinctive features and their correlates, A. Laboratory, ed., MIT, 1952.

[68] M. Oldman, “The intelligibility of speech and the judgment of meaning,” Journal of Sound and Vibration, vol.47, no.3, pp.323–331, 1976.

[69] D.N. Kalikow, K.N. Stevens, and L.L. Elliott, “Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,” J. Acoust. Soc. Am., vol.61, no.5, pp.1337–1351, 1977.

[70] M. Nilssonm, “Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am., vol.95, pp.1085–1099, 1994.

[71] M. Blue, C. A. Ntuen, and T. Letowski, “Speech Intelligibility of the Callsign Acquisition Test in a Quiet Environment,” International Journal of Occupational Safety and Ergonomic, vol.10, pp.179–189, 2004.

[72] 幸田 彰，久我 新一，“講演室の明瞭度試験の性能について,” 日本建築學會研究報告，vol.22，pp.247-248，may 1953.

[73] 飯田 茂隆，“明瞭度テストにおける問題,” 日本音響学会建築研究会資料，vol.AA 74–13，1974.

[74] 戸井田 義徳，“エコーとノイズが文章了解度に及ぼす影響について：野外拡声装置の明瞭度改善に関する研究その 2,”日本建築学会論文報告集，no.346，pp.112–123，1984.

[75] 小川 有子，“無意味三連音節による明瞭度試験,”日本音響学会第 1 回シンポジウム「試験用音声の標準化」，pp.10–16，1985.

[76] 佐藤 洋，長友 宗重，吉野 博，佐藤隆，“文章音表を用いた残響・騒音の音声聴取に及ぼす影響の評価に関する研究,”日本建築学会計画系論文集，vol.495，no.495，pp.15–20，1997.

[77] 田中 美郷，“補聴器適合評価機器の試作に関する研究,”昭和 63 年度科学研究費補助金研究成果報告書，1989.

[78] 米本 清，“補聴器適合評価用 CD (TY-89) 及び 57-S 語表の単音節明瞭度と音圧,” Audiology Japan，vol.32，no.5，pp.429–430，1989.

[79] 日本人工内耳研究会，“人工内耳装用のための語音聴取評価検査 CI–2004（試案）,”エスコアール，2004.

[80] 井脇 貴子，城間 将江，久保 武，SOLI Sigfrid，“HINT–Japanese 雑音下における語音聴取閾値検査の開発,” Audiology Japan，vol.42，no.5，pp.421–422，1999.

[81] 井脇 貴子，城間 将江，久保 武，Soli Sigfrid，“HINT–Japanese(雑音下における語音聴力検査) Norming Study,” Audiology Japan，vol.43，no.5，pp.499–500，2000.

[82] 井脇 貴子，城間 将江，久保 武，Soli Sigfrid，“HINT–Japanese(雑音下における語音聴力検査) Norming Study 2,” Audiology Japan，vol.44，no.5，pp.561–562，2001.

[83] 井脇 貴子，城間 将江，久保 武，Soli Sigfrid，“Japanese–HINT（雑音下における語音聴力検査）Norming Study 3–無響室における検査–,” Audiology Japan，vol.44，no.5，pp.561–562，2001.

[84] Soli Sigflid，城間 将江，井脇 貴子，久保武，“多言語による語音聴力検査 HINT(Hearing in Noise Test) 開発の背景 : Japanese HINT の意義,” Audiology Japan，vol.45，no.5，pp.497–498，2002.

[85] 天野 成昭，近藤 公久，日本語の語彙特性 1　単語親密度，三省堂，1999.

[86] 近藤 公久，天野 成昭，“「日本語の語彙特性」データベース：有効性と問題点,”電子情報通信学会技術研究報告，vol.TL200–14，no.335，pp.1–8，2000.

[87] 天野 成昭，近藤公久，“NTT データベースシリーズ「日本語の語彙特性」について (< 特集 > 音声研究関連データベースの動向),”音声研究，vol.4，no.2，pp.44–50，2000.

[88] 金田一 京助，山田 明雄，柴田 武，山田忠雄（編），新明解国語辞典　第四版，三省堂，1999.

[89] 加藤 和美，天野 成昭，近藤公久，“雑音を付加した音声の単語了解度に対する親密度の影響,”日本音響学会聴覚研究会資料，vol.H–99–8，1999.

[90] 坂本 修一，鈴木 陽一，天野 成昭，小澤 賢司，近藤 公久，曽根敏夫，“親密度と音韻バランスを考慮した単語了解度試験用リストの構築,”日本音響学会誌，vol.54，no.12，pp.842–849，Dec. 1998.

[91] 坂本 修一，天野 成昭，鈴木 陽一，近藤 公久，小澤 賢司，曽根敏夫，“単語了解度試験におけるモーラ同定に対する親密度の影響,”日本音響学会誌，vol.60，no.7，pp.351–357，2004.

[92] 近藤 公久，坂本 修一，天野 成昭，鈴木陽一，“親密度別単語了解度試験用音声データセット(FW03)に収録された単音節音声の雑音下における認知閾,”日本音響学会誌，vol.66，no.3，pp.105–111，2010.

[93] 佐藤 洋，佐藤 逸人，吉野 博，鈴木 陽一，天野 成昭，近藤 公久，長友宗重，“単語親密度と加齢による聴力損失が残響及び騒音下における単語了解度に及ぼす影響,”日本音響学会誌，vol.58，no.6，pp.346–354，2002.

[94] 近藤 公久，天野 成昭，坂本 修一，鈴木陽一，“親密度別単語了解度試験用音声データセット2007(FW07)の作成,”電子情報通信学会技術研究報告，vol.SP2007–157，no.432，pp.43–48，2008.

[95] H. Sato, S. Sakamoto, T. Kondo, S. Amano, and Y. Suzuki, “A new method for measurement of word intelligibility under noisy and reverberant environment in a short duration,” Proc. Inter–Noise 2011, 2011.

[96] 阿部 純一，桃内 佳雄，金子 康朗，李 光五，人間の言語情報処理–言語理解の認知科学–，サイエンス社，1994.

[97] 大槻 恭士，坂本 修一，牧野正三，“単語親密度を考慮した単語了解度の予測法,”日本音響学会聴覚研究会資料，vol.40，no.6，pp.483–488，2010.

[98] Seiji Nakagawa, Chika Fujiyuki, and Takayuki Kagomiya, “Development of Bone-Conducted Ultrasonic Hearing Aid for the Profoundly Deaf: Assessments of the Modulation Type with Regard to Intelligibility and Sound Quality,” Japanese Journal of Applied Physics, vol.51, p.07GF22, 2012.

[99] 翁長 博，“残響音場の音声了解度に対応する物理指標の提案,”日本音響学会誌，vol.66，no.3，pp.97–104，2010.

[100] YAMADA Takeshi, KUMAKURA Masakazu, and KITAWAKI Nobuhiko, “Objective Estimation of Word Intelligibility for Noise-Reduced Speech,” IEICE transactions on communications, vol.91, no.12, pp.4075–4077, 2008.

[101] 中貝 順一，小澤賢司，“音の再生方式と高能率符号化が競合話者存在下での単語了解度に及ぼす影響(電気音響, 音響一般),”電子情報通信学会論文誌（A），vol.88，no.9，pp.1026–1034，2005.

[102] M. Morimoto, H. Sato, and M. Kobayashi, “Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces,” J. Acoust. Soc. Am., vol.116, no.3, pp.1607–1613, 2004.

[103] 西本 卓也，狩谷 幸香，渡辺隆行，“早口音声の聴取訓練における単語親密度の影響,” 電子情報通信学会技術研究報告，vol.SP2007–116，no.406，pp.119–124，2007.

[104] 近藤 和弘，泉 良，藤森 雅也，加賀 類，中川清司，“二者択一型日本語音声了解度試験方法の検討,” 日本音響学会誌，vol.63，no.4，pp.195–205，Apl. 2007.

[105] 近藤 和弘，泉 良，中川清司，“新しい日本語了解度試験方法の評価,” 電子情報通信学会技術研究報告，vol.SP2000–163，no.726，pp.25–32，2001.

[106] 藤森 雅也，近藤 和弘，高野 勝美，中川清司，“二者択一式日本語了解度試験方法の評定用リストの再検討,” 電子情報通信学会技術研究報告，vol.SP2005–180，no.685，pp.103–108，2006.

[107] 佐藤 正幸，生山 雅人，錦戸 暖，豊島 広紀，坂田 聡，上田裕市，“人工内耳システムのための W–SPEAK 方式の提案と評価シミュレータの設計 (肢体障害・聴覚障害,HCG シンポジウム),” 電子情報通信学会技術研究報告，vol.WIT2007–105，no.555，pp.85-90，2008.

[108] Y. Kitashima, K. Kondo, H. Terada, T. Chiba, and K. Nakagawa, “Intelligibility of read Japanese words with competing noise in virtual acoustic space,” Acoust. Sci. & Tech., vol.29, no.1, pp.74–81, Jan. 2008.

[109] Y. Kobayashi, K. Kondo, and K. Nakagawa, “Intelligibility of HE-AAC coded Japanese words with various stereo coding modes in virtual 3D audio space,” Auditory Display, vol.5954, pp.219–238, May. 2010.

[110] K. Kondo, T. Kanda, Y. Kobayashi, and H. Yagyu, “Speech Intelligibility of Diagonally Localized Speech with Competing Noise Using Bone-Conduction Headphones,” Proc. Interspeech 2010, pp.1213–1216, 2010.

[111] K. Kondo, T. Chiba, Y. Kitashima, and N. Yano, “Intelligibility comparison of Japanese speech with competing noise spatialized in real and virtual acoustic environments,” Acoust. Sci. & Tech., vol.31, no.3, pp.231–238, May. 2010.

[112] 渋谷 徹，渡邉 瞳，小林 洋介，近藤和弘，“音声の伸長・短縮の了解度への影響と適応話速変換方法の提案,” 映像情報メディア学会誌，vol.66，no.10，pp.J377–J384，Oct. 2012.

[113] Y. Hu, and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” IEEE Trans. Audio, Sp. & Lang. Process., vol.16, no.1, pp.229–238, Jan. 2008.

[114] J.R.Deller,Jr., J.G. Proakis, and J.H.L. Hansen, Discrete–Time Processing of Speech Signals, Macmillan, 1993.

[115] B.J. McDermott, “Multidimensional Analyses of Circuit Quality Judgments,” J. Acoust. Soc. Am., vol.45, no.3, pp.774–781, 1969.

[116] B.H. Juang, “On Using the Itakura-Saito Measures for Speech Coder Performance Evaluation”, AT&T Bell,” AT&T Bell Laboratories Technical Journal, vol.63, no.8, pp.1477–1498, Oct. 1984.

[117] F. Itakura, and S. Saito, “Analysis–synthesis telephone based on the maximum–likelihood method,” Proc. 6th ICA, pp.C17–C20, 1968.

[118] T.P. Barnwell, and W.D. Voiers, “An analysis ofobjective measures for user acceptance of voice communication systems,” DCA Final Tecniqual Report, vol.DCA100-78-C-0003, 1979.

[119] D.H. Klatt, “Prediction of perceived phonetic distance from critical band spectra : a first step,” Proc. ICASSP, 1982.

[120] 村田 晴美，荻原 昭夫，岩田 基，汐崎陽，“音楽電子透かしにおける埋込多重化に対する直流成分を用いた音質改善,” 電子情報通信学会論文誌 (A)，vol.93，no.3，pp.171–180，2010.

[121] B.S. Atal, “Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification,” J. Acoust. Soc. Am., vol.55, no.6, pp.1304–1312, 1974.

[122] J.M. Tribolet, P. Nolt, B.J. McDermott, and R.E. Cochiere, “A STUDY OF COMPLEXITY AND QUALITY OF SPEECH WAVEFORM CODERS,” Proc. ICASSP, 1978.

[123] J. Ma, Y. Hu, and P.C. Loizou, “Objective measures for predicting speech intelligibility in noisy conditions based on new band–importance functions,” J. Acoust. Soc. Am., vol.125, no.5, pp.3387–3405, 2009.

[124] H. Fletcher, and R.H. Galt, “The Perception of Speech and Its Relation to Telephony,” J. Acoust. Soc. Am., vol.22, no.2, pp.89–151, 1950.

[125] K.D. Kryter, “Methods for the Calculation and Use of the Articulation Index,” J. Acoust. Soc. Am, vol.34, no.11, pp.1689–1697, 1962.

[126] ANSI, “Methods for the calculation of the articulation index,” Technical Report S3.5–1969, American National Standards Institute, 1969.

[127] 勝木 保次，村田 計一，吉田 登美男，亀田 和夫，越川 常治，三浦種敏，山口 善治，中田 和男，新版聴覚と音声，三浦種敏（編），電子情報通信学会，1980.

[128] 三浦 種敏，“日本語に対する定量的な伝送品質と伝送特性との関係,” 日本電信電話公社，研究実用化報告，vol.3，no.4，pp.527–534，1954.

[129] 佐藤 洋，長友 宗重，吉野博，“単音節明瞭度試験法による帯域騒音及び聴力損失が音声情報伝達に及ぼす影響の評価について,” 日本建築学会計画系論文集，vol.494，no.494，pp.1–6，1997.

[130] 島原 正男，“室内およびその音響装置の明瞭度予測,” 日本音響学会建築研究会資料，vol.AA 74–7，1974.

[131] 植松 道治，曽根 敏夫，二村忠元，“ランダム変動騒音下の音声明瞭度と了解度に関する基礎実験 : 変動騒音の言語聴取妨害に関する研究 その 1,” 日本音響学会誌，vol.34，no.9，pp.516–521，1978.

[132] 曽根 敏夫，植松 道治，金指 久則，二村忠元，“ランダム変動騒音下の音声明瞭度の予測 : 変動騒音の言語聴取妨害に関する研究その 2,” 日本音響学会誌，vol.35，no.2，pp.58–62，1979.

[133] H.G. Latham, “The signal–to–noise ratio for speech intelligibility — An auditorium acoustics design index,” Applied Acoustics, vol.12, pp.253–320, 1979.

[134] H.J.M Steeneken, and T. Houtgast, “A physical method for measureing speech transmission quality,” J. Acoust. Soc. Am., vol.67, no.1, pp.318–326, Jan. 1980.

[135] ISO, “Ergonomics – Assessment of speech communication,” Technical Report 9921:2003, International Organization for Standardization, 2003.

[136] IEC, “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index,” Technical Report 60268–16 ED. 4.0 B:2011, International Electrotechnical Commission, Jan. 2011.

[137] IEC, “The objective rating of speech intelligibility in auditoria by the ”RASTI” method,” Technical Report 268–16, International Electrotechnical Commission, 1988.

[138] 小椋 靖夫，浜田 晴夫，三浦種敏，“音場における音声伝送品質のための MTF と STI について,” 日本音響学会誌，vol.40，no.3，pp.181–191，1984.

[139] 小椋 靖夫，浜田 晴夫，三浦種敏，“音場における音声伝送品質のための MTF と STI について,” 日本音響学会誌，vol.40，no.3，pp.181–191，1984.

[140] H. Haas, “Uber den Einfluss eines Einfachechos auf die Horsamkeit von Sprache,” Acustica, vol.1, pp.49–58, 1951.

[141] Sander J. van Wijngaarden, and Rob Drullman, “Binaural intelligibility prediction based on the speech transmission index,” J. Acoust. Soc. Am., vol.123, no.6, pp.4514–4523, Jun. 2008.

[142] F.F. Lia, and T.J. Coxb, “A neural network model for speech intelligibility quantification,” Applied Soft Computing, vol.7, pp.145–155, 2007.

[143] ANSI, “Methods for calculation of the speech intelligibility index,” Technical Report S3.5–1997, American National Standards Institute, 1997.

[144] Jianfen Ma, and Philipos C. Loizou, “SNR loss: A new objective measure for predicting the intelligibility of noise–suppressed speech,” Speech Communication, vol.53, pp.340–354, 2011.

[145] J. H. Friedman, “Multivariate Adaptive Regression Splines,” Annals of Statistics, vol.19, no.1, pp.1–67, 1991.

[146] J. Kates, and K. Arehart, “Coherence and the speech intelligibility index,” J. Acoust. Soc. Am., vol.117, pp.2224–2237, 2005.

[147] I. Hollube, and K. Kollmeier, “Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model,” J. Acoust. Soc. Am., vol.100, pp.1703–1715, 1996.

[148] J. Li, L. Yang, J. Zhang, T. Yan, Y. Hu, M. Akagi, and P. Loizou, “Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese, and English,” J. Acoust. Soc. Am., vol.129, pp.3291–3301, 2011.

[149] K. Kondo, “Estimation of Speech Intelligibility Using Objective Measures,” Applied Acoustics, vol.74, pp.63–70, 2012.

[150] W.M. Liu, K.A. Jellyman, N.W. Evans, and J.S.D. Mason, “Assessment of Objective Quality Measures for Speech Intelligibility,” Proc. Interspeech 2008, pp.699–702, Sept. 2008.

[151] 加藤 宏明，津崎 実，匂坂芳典，“聴知覚特性を考慮した音韻長制御規則の客観評価モデル,” 日本音響学会誌，vol.55，no.11，pp.752–760，nov 1999.

[152] 田中良和，白石君男，“補聴器適合検査のための印象評価による雑音および環境騒音の分類,” Audiology Japan，vol.54，no.2，pp.130–137，2011.

[153] EnShuo Tsau, Sachin Chachada, and C.-C. Jay Kuo, “Content/Context-Adaptive Feature Selection for Environmental Sound Recognition,” Proc. APSIPA 2012, 2012.

[154] V. Vapnik, The Nature of Statistical Learning Theory; Statistics for Engineering and Information Science, Springer, 1995.

[155] C. Bergmeir, I. Triguero, D. Molina, J. L. Aznarte, and J. M. Benitez, “Time Series Modeling and Forecasting Using Memetic Algorithms for Regime–Switching Models,” IEEE Trans. on Neural Networks and Learning Systems, vol.23, pp.1841–1847, Nov. 2012.

[156] B. Gardner, and K. Martin, HRTF Measurements of a KEMAR Dummy–Head Microphone, MIT Media Lab., ed., Perceptual Computing – Technical Report, 1994.

[157] 板橋 秀一，“騒音データベースと日本語共通音声データ DAT 版,” 日本音響学会誌，vol.47，no.12，pp.951–953，1991.

[158] 高橋 玲，北脇 信彦，“符号化音声品質客観評価尺度の性能評価,” 電子情報通信学会論文誌（B），vol.80，no.6，pp.480–487，1997.

[159] MG Kendall, “A New Measure of Rank Correlation,” Biometrika, vol.30, pp.81–89, 1938.

[160] J. McQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” Proc. 5th Berkeley Symp. Math. Statistics and Probability, 1967.

[161] D. Pelleg, and A. W. Moore, “X-means: Extending K-means with Efficient Estimation of the Number of Clusters,” Prc. 7th Inter’l Conf. on Machine Learning, pp.727–734, 2000.

[162] S. Jianbo, and M. Jitendra, “Normalized Cuts and Image Segmentation,” IEEE Trans. on Pattern analysis and machine intelligence, vol.22, no.8, pp.888–905, 2000.

[163] C. Ding, C. H. Q. Ding, X. He, H. Zha, M. Gu, and H. Simon, “A Min–max Cut Algorithm for Graph Partitioning and Data Clustering,” Proc. IEEE International Conference on Data Mining, 2001.

[164] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, 1981.

[165] T. Hofmann, “Probabilistic Latent Semantic Indexing,” Proc. of the 32 Annual International SIGIR Conference on Research and Development in Information Retrieval, 1999.

[166] Wei Xu, Xin Liu, and Yihong Gong, “Document clustering based on non-negative matrix factorization,” Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 2003.

[167] Chin–Chung Chang, and Chin–Jen Lin, “LIBSVM: a library for support vector machines,” , 2001.

[168] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, “A practical guide to support vector classification,” , 2003.

[169] R Core Team, R: A Language and Environment for Statistical Computing, , R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0.

[170] R. Ihaka, and R. Gentleman, “R: A language for data analysis and graphics,” Journal of Computational and Graphical Statistics, vol.5, no.3, pp.299–314, 1996.

[171] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, e1071: Misc Functions of the Department of Statistics (e1071), TU Wien, , 2012, R package version 1.6–1.

[172] G. Seymour, Predictive Inference, Chapman and Hall, 1993.

[173] O. Lartillot, P. Toiviainen, and T. Eerola, “MIRtoolbox,” , 2007.

[174] H.I. Witten, and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques (3rd Edition), Motgan Kaufmann, 2005.

[175] A. E. Hoerl, and R. W. Kennard, “Ridge regression: biased estimation for nonorthogonal problems,” Technometrics, vol.12, pp.55–68, 1970.

[176] R. Tibshirani, “The lasso method for variable selection in the cox model,” Statistics in Medicine, vol.16, pp.385–395, 1997.

[177] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization Paths for Generalized Linear Models via Coordinate Descent,” Journal of Statistical Software, vol.33, no.1, pp.1–22, 2010.

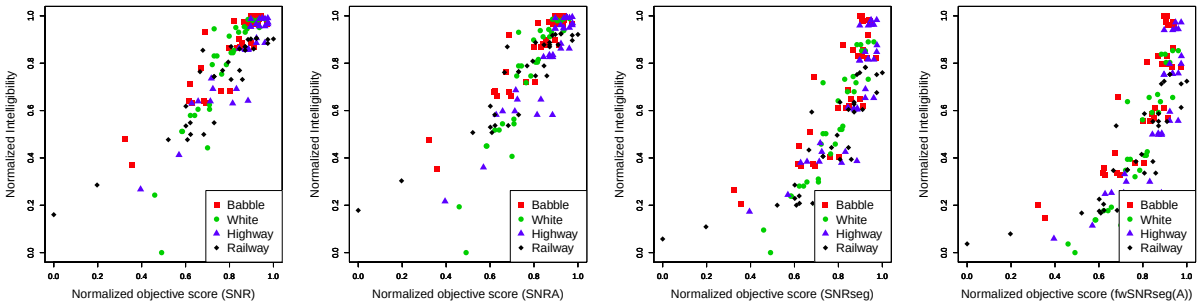
[178] 赤穂昭太郎，カーネル多変量解析—非線形データ解析の新しい展開，岩波書店，2008.

[179] L. Breiman, “Random Forests,” Machine Learning, vol.45, no.1, pp.5–32, 2001.

[180] 伊藤 健太郎，中野良平，“サポートベクトル回帰におけるハイパーパラメータの最適化法,” 電子情報通信学会技術研究報告. NC, ニューロコンピューティング，vol.102，no.508，pp.7–12，dec 2002.

付 録A　子音特徴ごとの客観音質値

2 章の子音特徴別音質値の散布図を Fig. A.1〜Fig. A.5 に示す．Fig. 2.24 と同様に，図中では騒音種ごとにプロットを変えてあるが，相関係数は全ての騒音種を混合して求めてあり，正規化に用いる最大値，最小値は全評価値から求めた値を用いた．

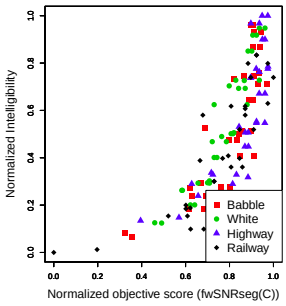


(a) SNR

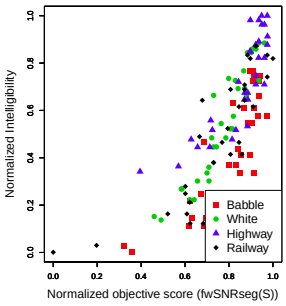
(b) SNRA

(c) SNRseg

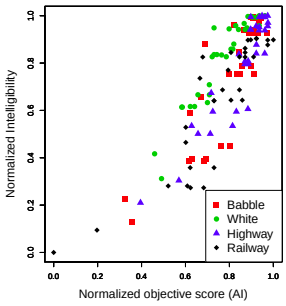
(d) fwSNRseg(A)



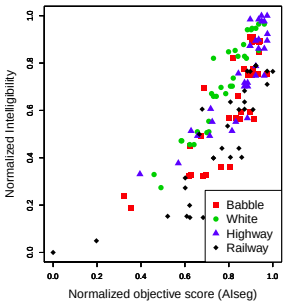
(e) fwSNRseg(C)



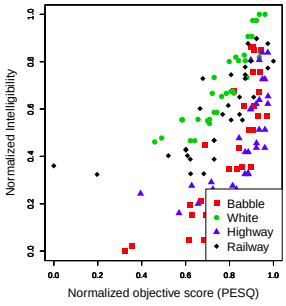
(f) fwSNRseg(S)



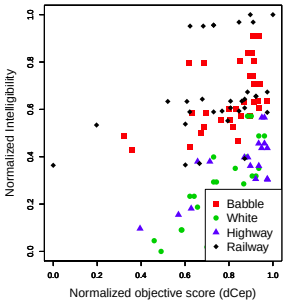
(g) AI



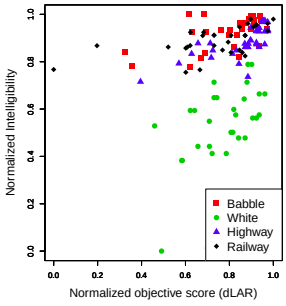
(h) AIseg



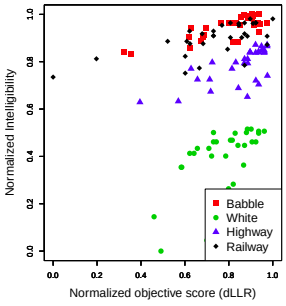
(i) PESQ



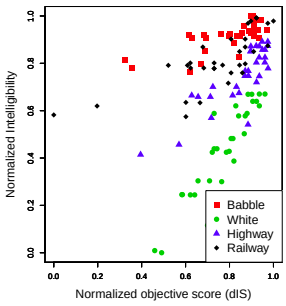
(j) d_{Cep}



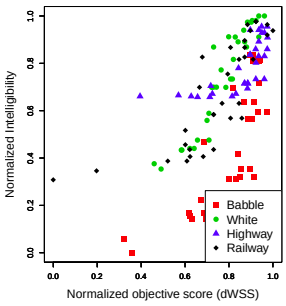
(k) d_{LAR}



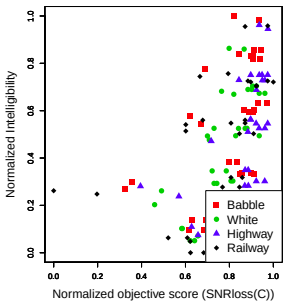
(l) d_{LLR}



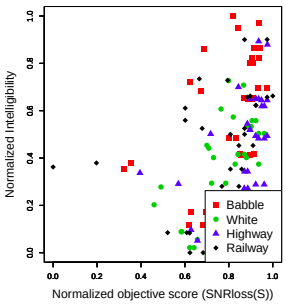
(m) d_{IS}



(n) d_{WSS}

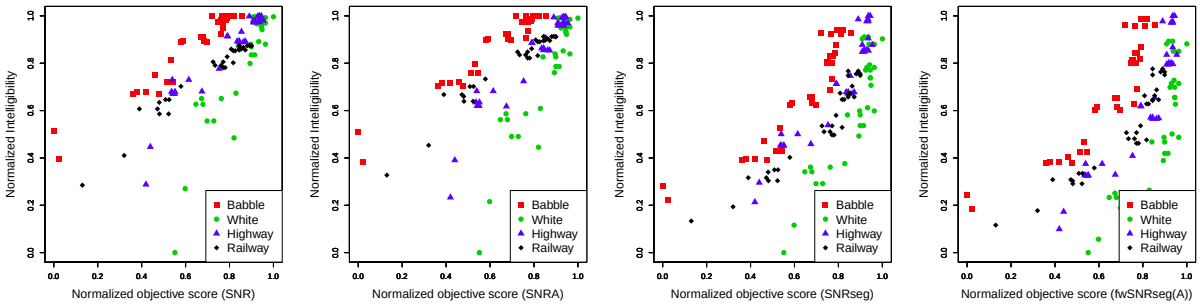


(o) SNRloss(C)

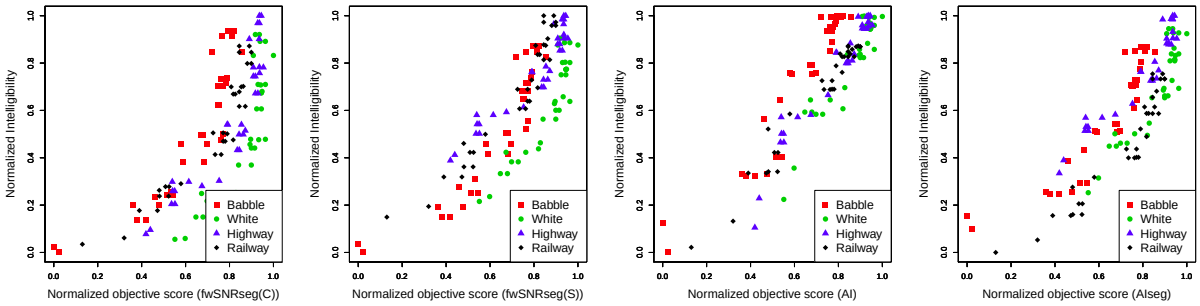


(p) SNRloss(S)

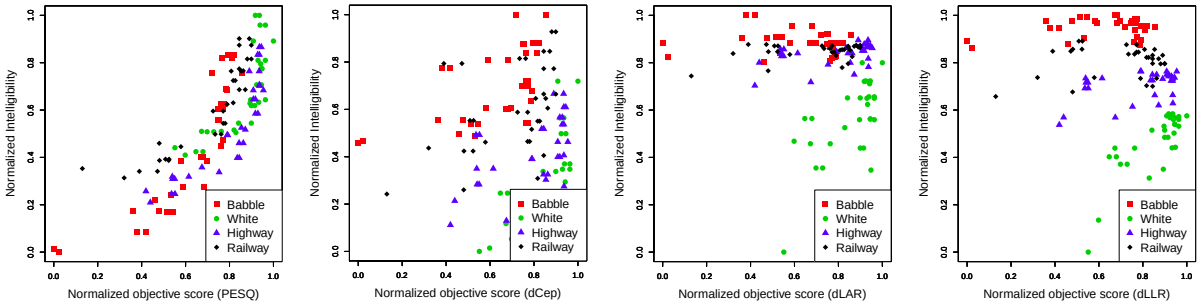
Fig. A.1: Comparison between normalized intelligibility(voicing) score and normalized objective speech quality score



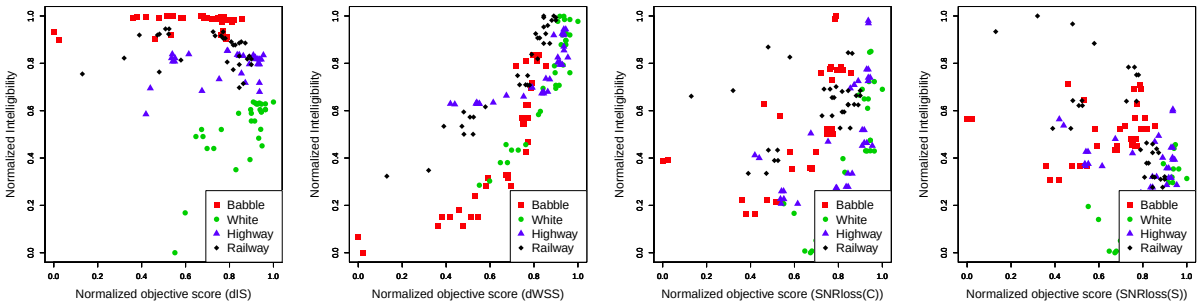
(a) SNR (b) SNRA (c) SNRseg (d) fwSNRseg(A)



(e) fwSNRseg(C) (f) fwSNRseg(S) (g) AI (h) AIseg

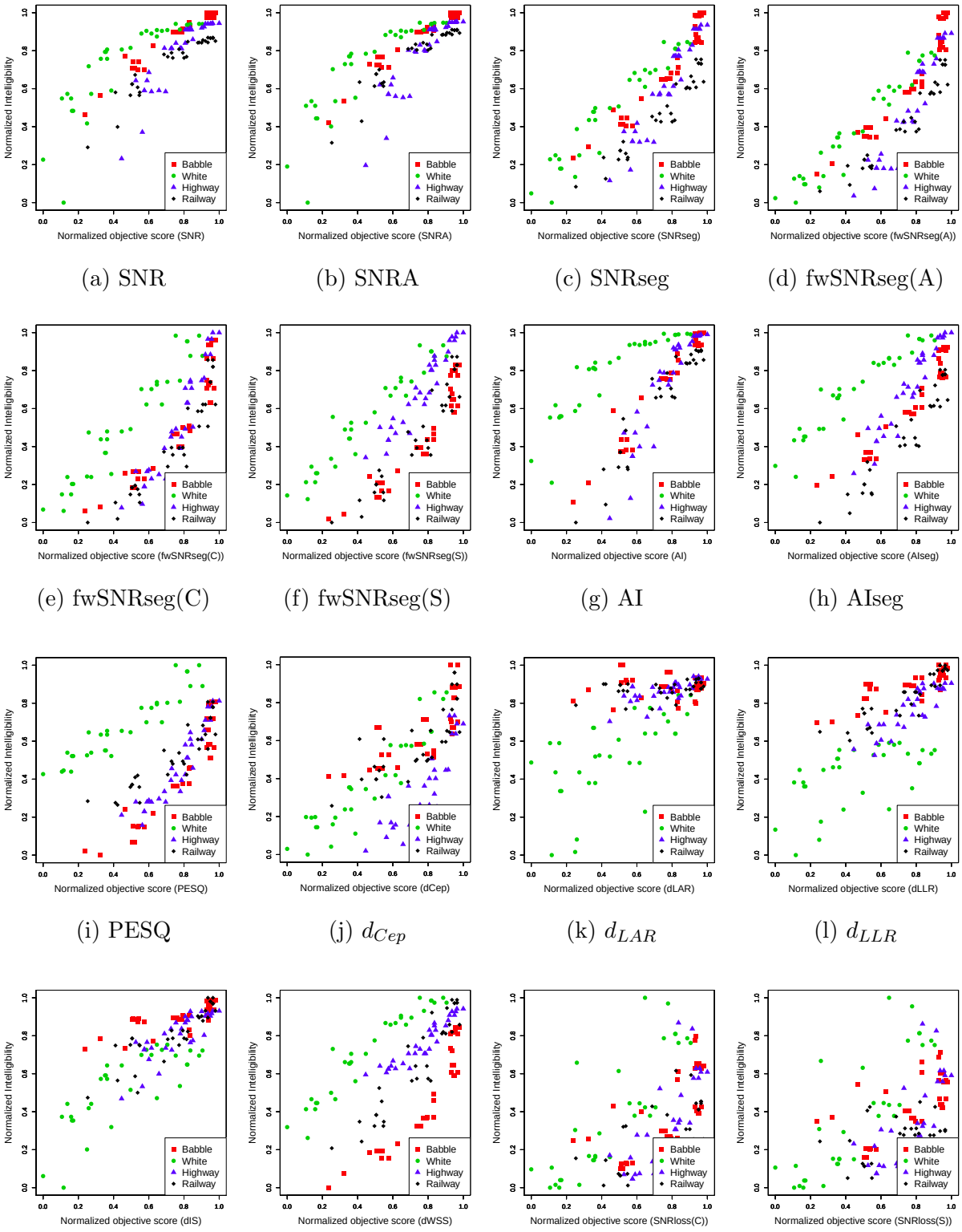


(i) PESQ (j) d_{Cep} (k) d_{LAR} (l) d_{LLR}



(m) d_{IS} (n) d_{WSS} (o) SNRloss(C) (p) SNRloss(S)

Fig. A.2: Comparison between normalized intelligibility(nasality) score and normalized objective speech quality score



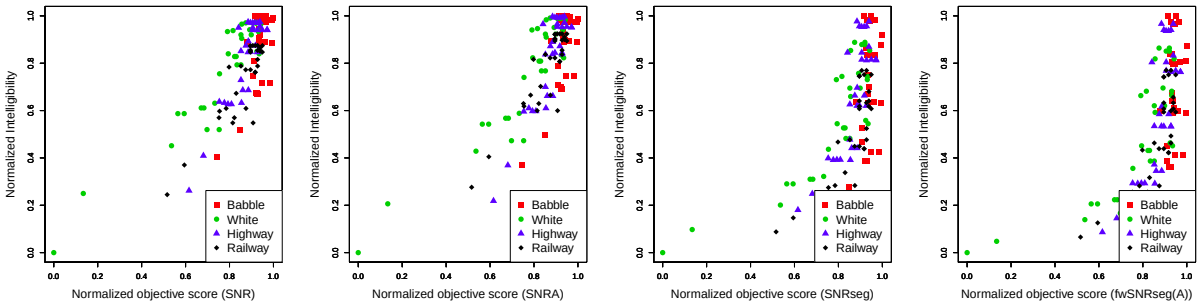
(a) SNR (b) SNRA (c) SNRseg (d) fwSNRseg(A)

(e) fwSNRseg(C) (f) fwSNRseg(S) (g) AI (h) AIseg

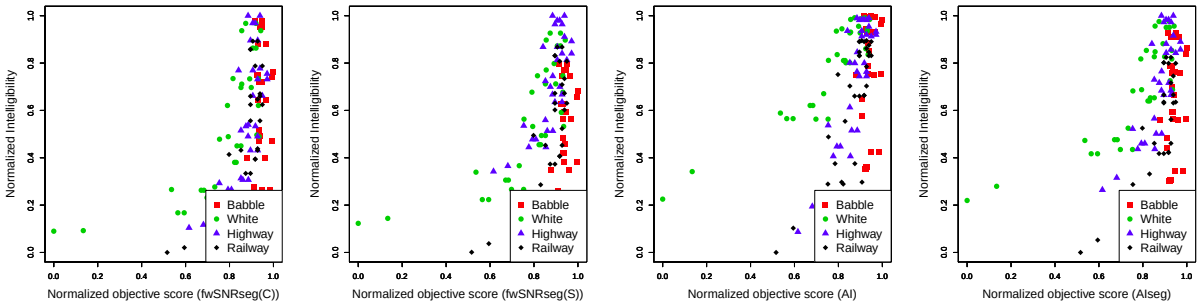
(i) PESQ (j) d_{Cep} (k) d_{LAR} (l) d_{LLR}

(m) d_{IS} (n) d_{WSS} (o) SNRloss(C) (p) SNRloss(S)

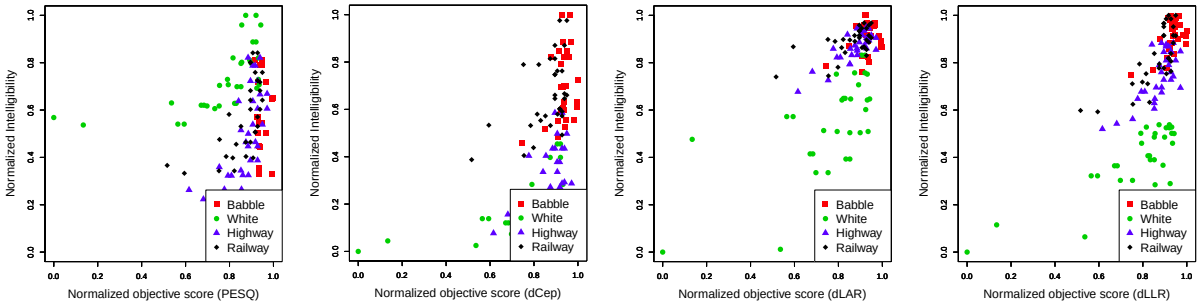
Fig. A.3: Comparison between normalized intelligibility(sustention) score and normalized objective speech quality score



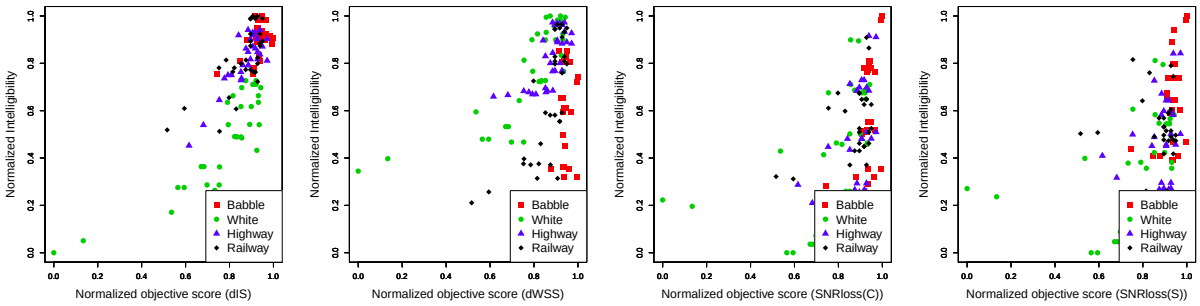
(a) SNR (b) SNRA (c) SNRseg (d) fwSNRseg(A)



(e) fwSNRseg(C) (f) fwSNRseg(S) (g) AI (h) AIseg

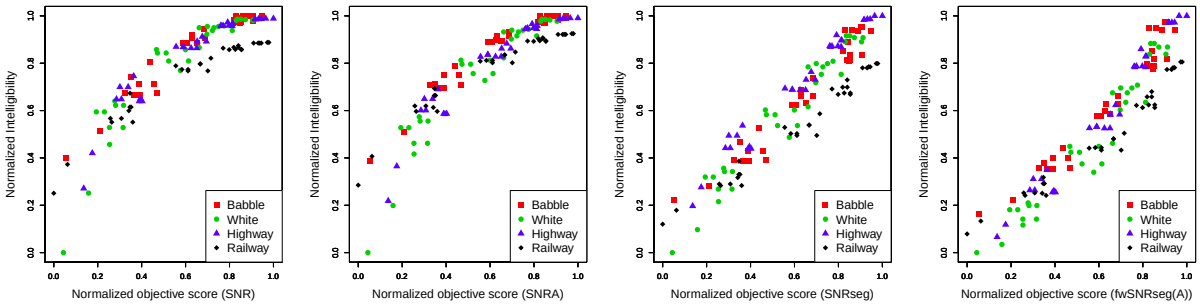


(i) PESQ (j) d_{Cep} (k) d_{LAR} (l) d_{LLR}

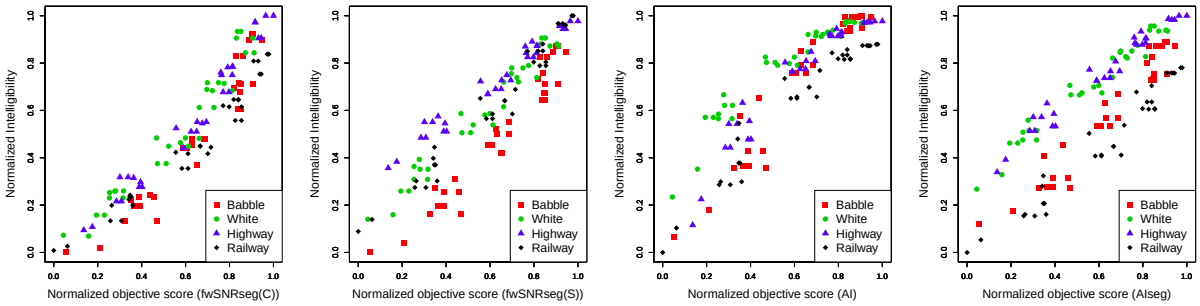


(m) d_{IS} (n) d_{WSS} (o) SNRloss(C) (p) SNRloss(S)

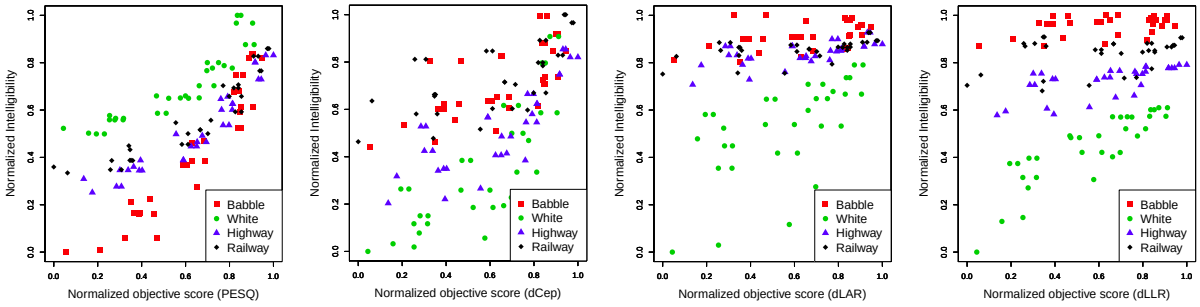
Fig. A.4: Comparison between normalized intelligibility(sibilation) score and normalized objective speech quality score



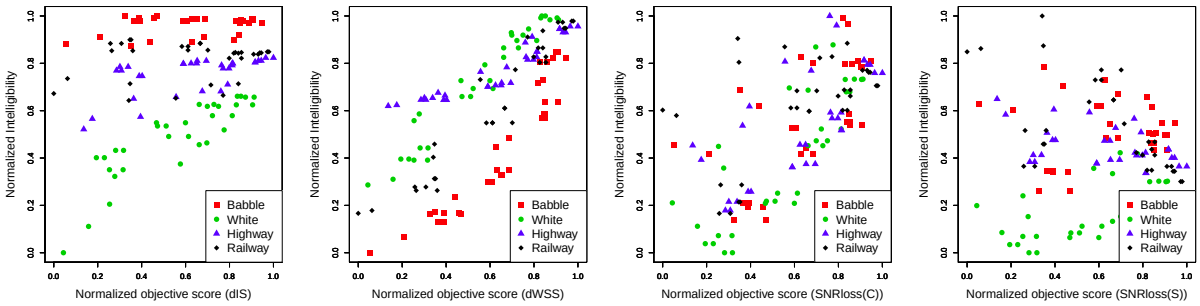
(a) SNR (b) SNRA (c) SNRseg (d) fwSNRseg(A)



(e) fwSNRseg(C) (f) fwSNRseg(S) (g) AI (h) AIseg

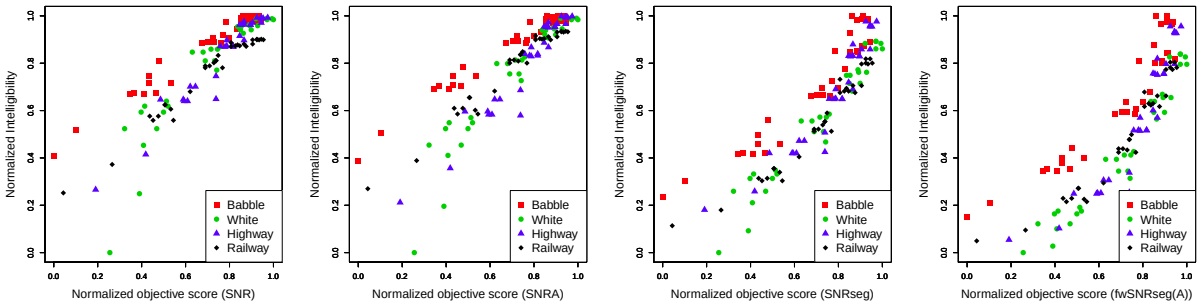


(i) PESQ (j) d_{Cep} (k) d_{LAR} (l) d_{LLR}

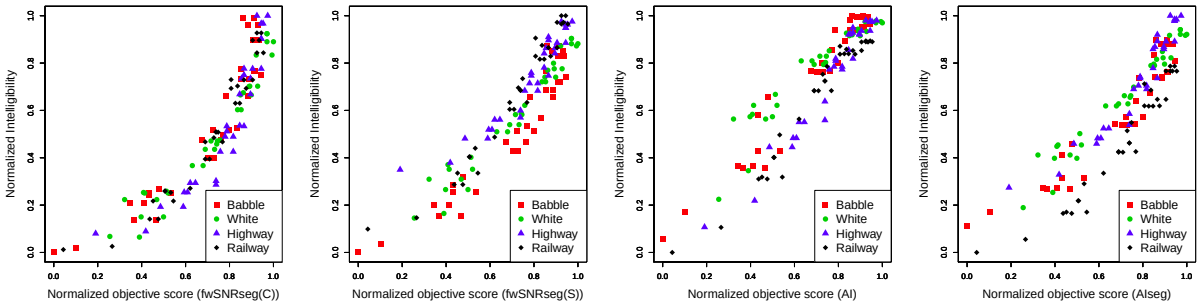


(m) d_{IS} (n) d_{WSS} (o) SNRloss(C) (p) SNRloss(S)

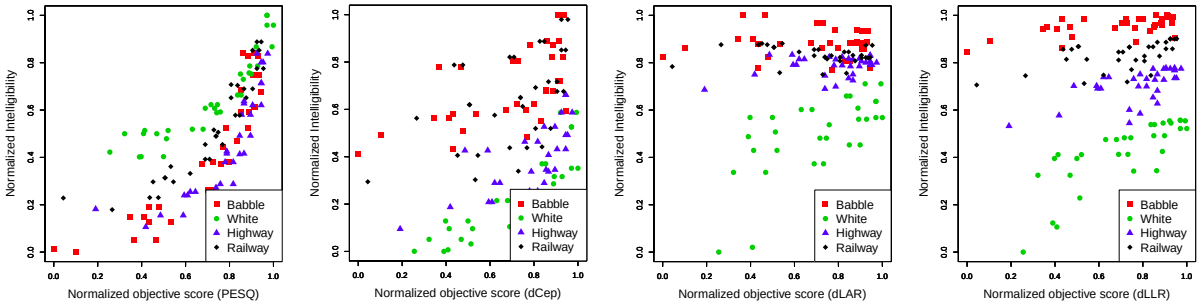
Fig. A.5: Comparison between normalized intelligibility(graveness) score and normalized objective speech quality score



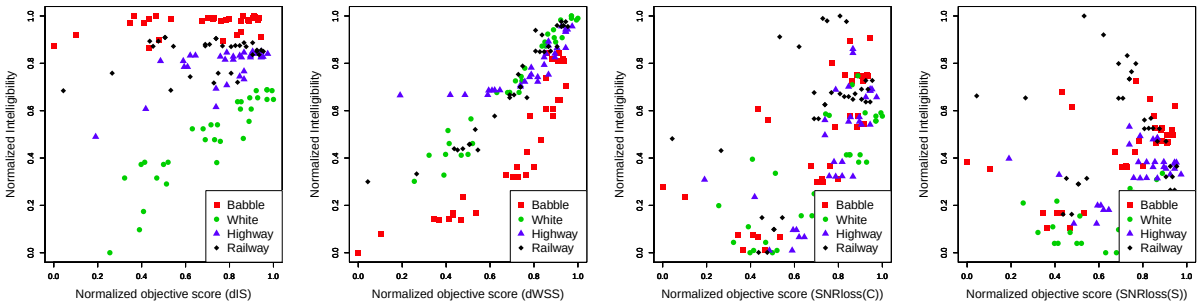
(a) SNR (b) SNRA (c) SNRseg (d) fwSNRseg(A)



(e) fwSNRseg(C) (f) fwSNRseg(S) (g) AI (h) AIseg



(i) PESQ (j) d_{Cep} (k) d_{LAR} (l) d_{LLR}

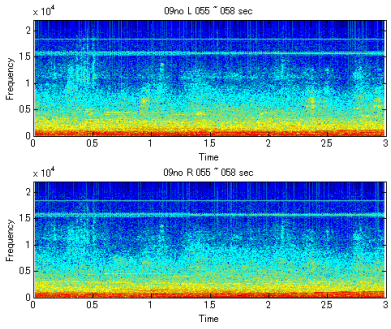


(m) d_{IS} (n) d_{WSS} (o) SNRloss(C) (p) SNRloss(S)

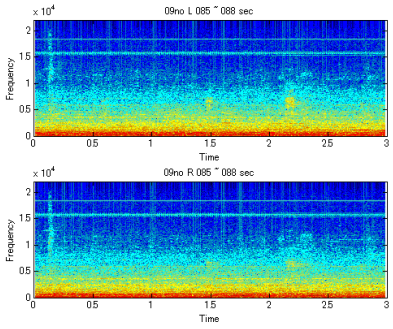
Fig. A.6: Comparison between normalized intelligibility(compactness) score and normalized objective speech quality score

付 録B　騒音LFのスペクトログラム

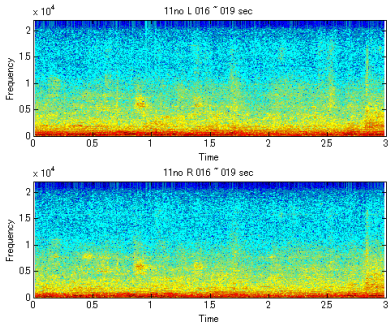
4 章の主観評価に用いた騒音 LF のスペクトログラムを示す．全てのスペクトログラム間のレベル統制をしている．作図に利用したのは主観評価で 0 dB に相当する音圧を用いた．



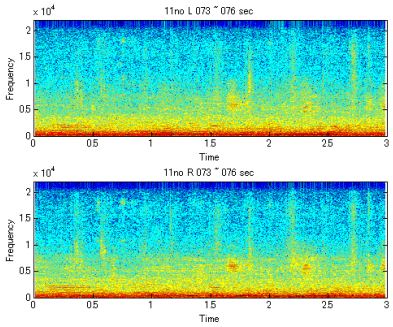
exhibition booth 1 (C2)



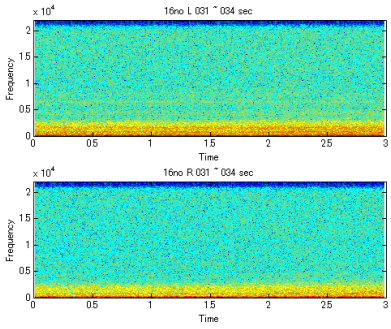
exhibition booth 1 (C3)



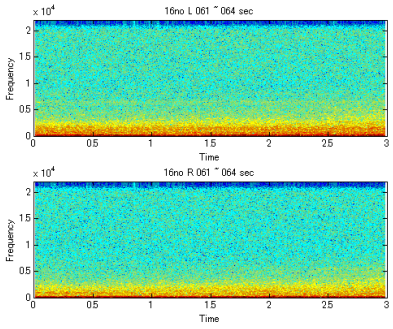
exhibition booth 2 (C2)



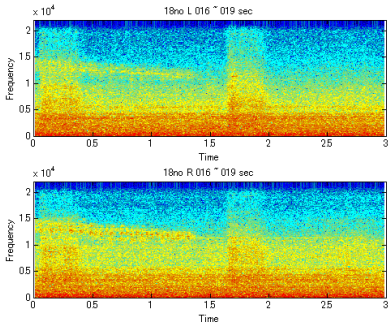
exhibition booth 2 (C3)



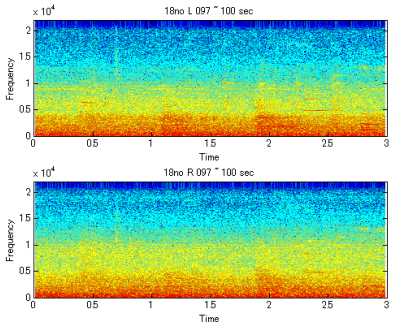
telephone booth (C1)



telephone booth (C2)

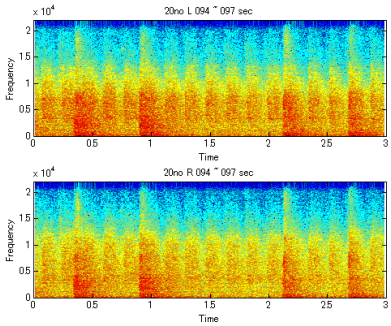


factory 1 (C1)

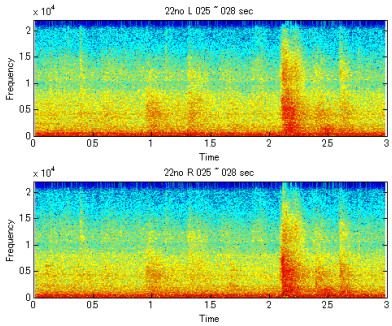


factory 1 (C2)

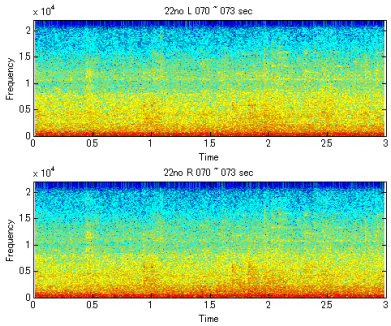
Fig. B.1: Spectrogram of various noise



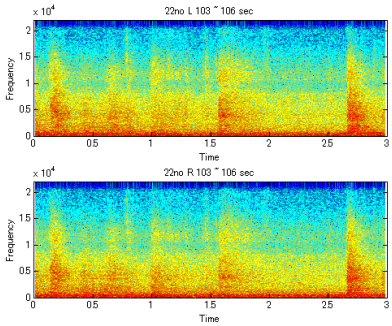
factory 2 (C1)



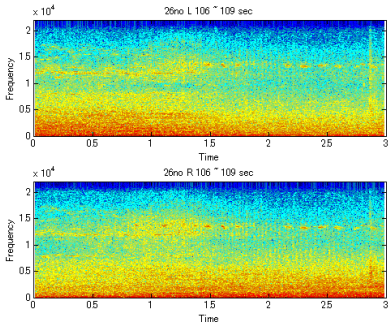
sorting facility (C1)



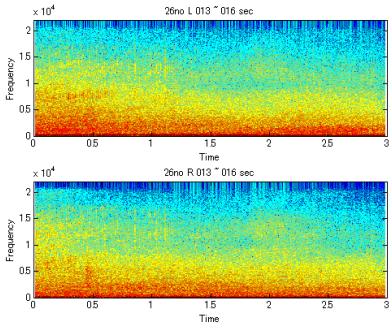
sorting facility (C2)



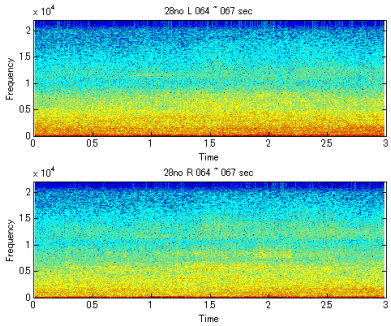
highway 1 (C1)



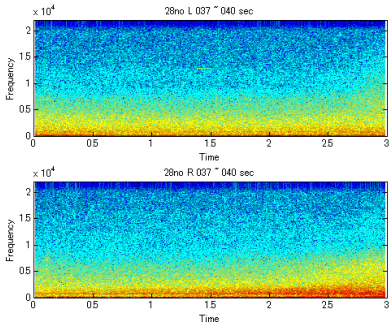
highway 2 (C1)



highway 2 (C2)

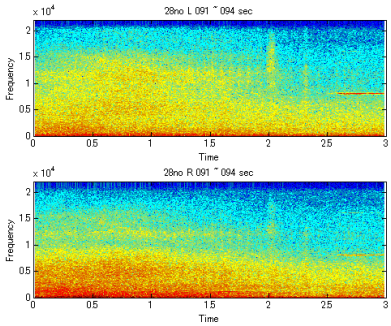


crossing (C1)

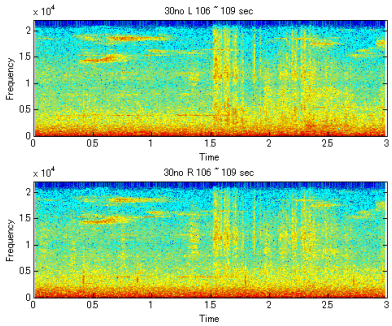


crossing (C2)

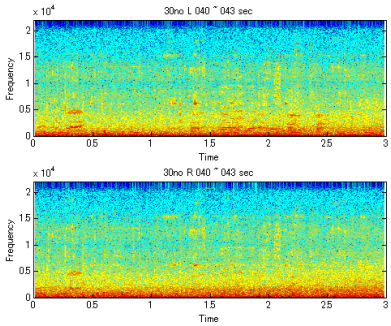
Fig. B.1 Spectrogram of various noise(cont'd)



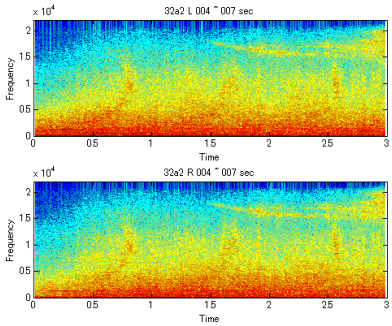
crossing (C3)



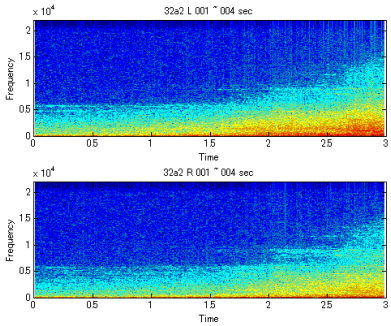
crowd (C1)



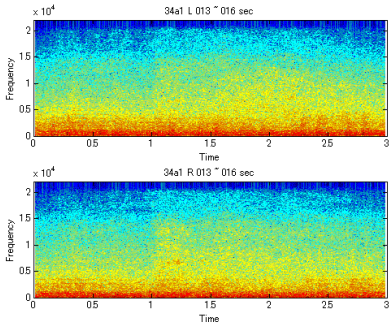
crowd (C2)



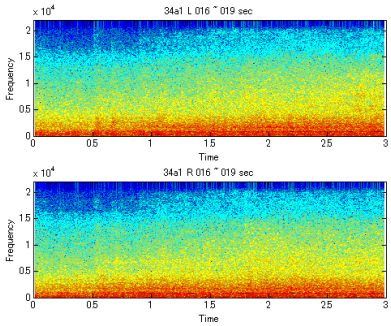
bullet train (C1)



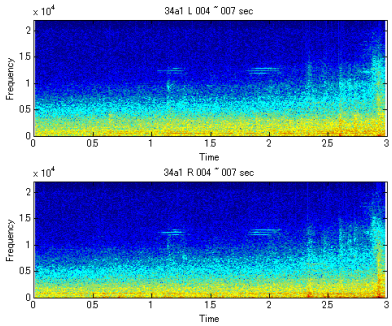
bullet train (C2)



train (C1)

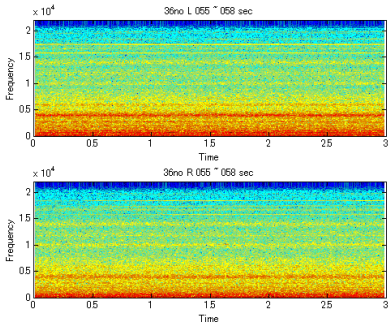


train (C2)

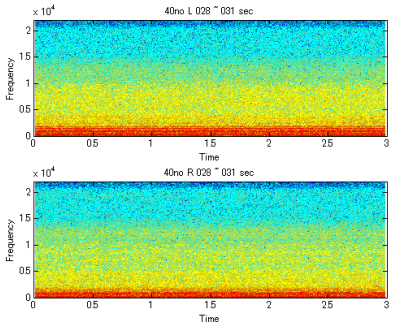


train (C3)

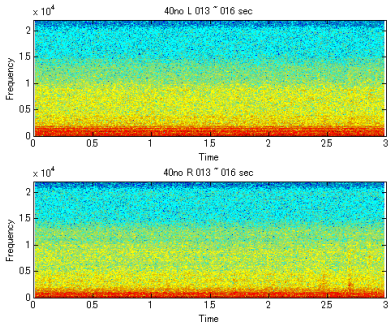
Fig. B.1 Spectrogram of various noise(cont'd)



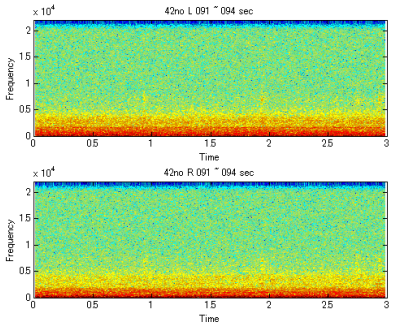
computer room (C1)



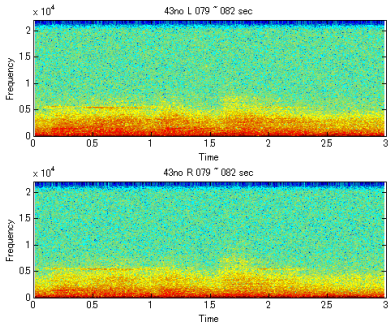
air conditioner 1(C2)



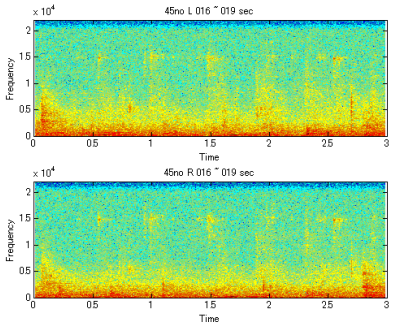
air conditioner 1 (C3)



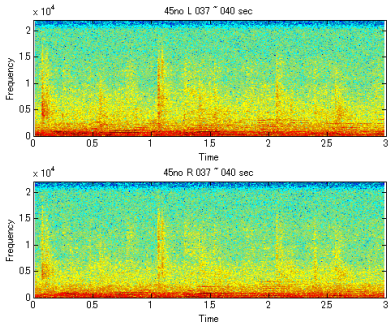
air conditioner 2 (C2)



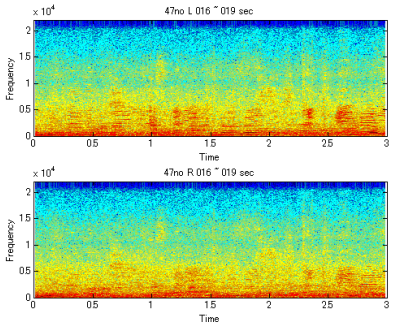
air duct(C2)



elevator hall 1(C1)



elevator hall 1 (C2)



elevator hall 2 (C1)

Fig. B.1 Spectrogram of various noise(cont'd)

付 録C　採用したハイパーパラメータ

5 章の結果で，cbSNRseg を用いた SVR で選択したハイパーパラメータの一覧を Table C.1. と Table C.2. に示す．提案法で RMSE が最小になる ϵ は設定した値の最小値であり，前後の値を比較したところ RMSE の差は 0.001 未満とほとんど見られなかった．このため，10^{-3} を提案法の ϵ の値とすることとした．C と γ に関しては，設定した範囲内に RMSE を最小とする値がみられたため，その値を採用した．max SNRseg と min SNRseg の値は特徴量の正規化に用いる値であり，回帰に用いるハイパーパラメータではないが，提案法ではハイパーパラメータと同様に最適値探索と並行して最適組み合わせ探索を行い，最も交差検定 RMSE が小さくなる組み合わせを採用した．

Table C.1: Selected parameter in the linear kernel

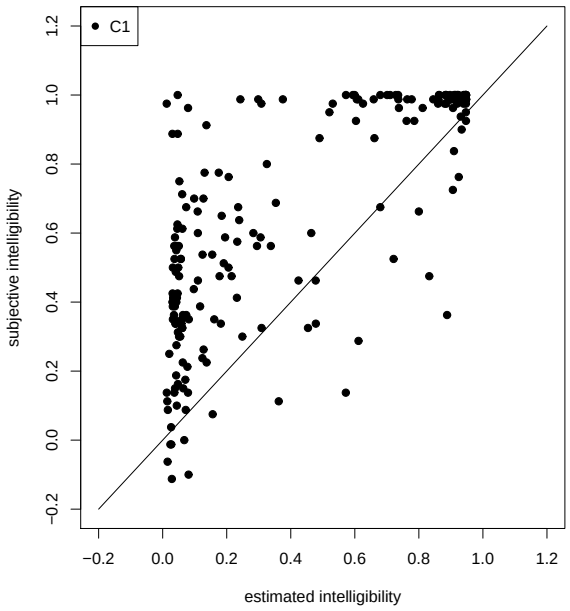
feature	name	C1	C2	C3
obSNRseg	ϵ	10^{-3}	10^{-3}	10^{-3}
	C	1.28	0.468	0.0935
	max SNRseg (dB)	5	5	5
	min SNRseg (dB)	-10	-10	-40
cbSNRseg	ϵ	10^{-3}	10^{-3}	10^{-3}
	C	1.28	0.468	0.0935
	max SNRseg (dB)	0	0	0
	min SNRseg (dB)	-10	-10	-10

Table C.2: Selected parameter in the RBF kernel

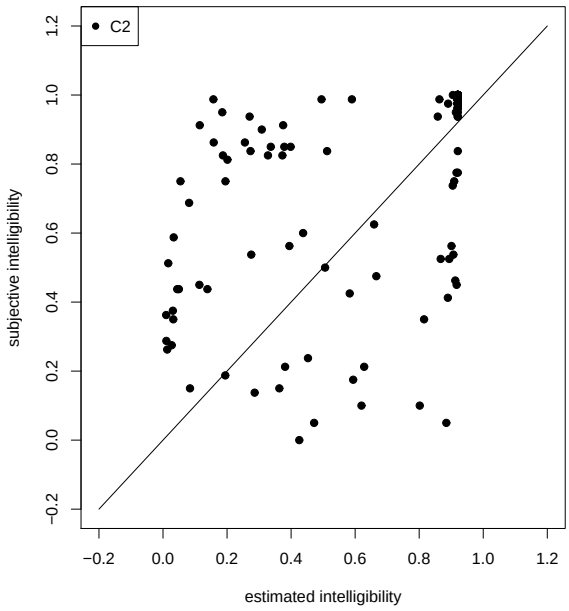
name	name	C1	C2	C3
obSNRseg	ϵ	10^{-3}	10^{-3}	10^{-3}
	C	21.4	6.40	21.4
	γ	0.00328	0.0743	0.00116
	max SNRseg (dB)	5	5	5
	min SNRseg (dB)	-10	-10	-10
cbSNRseg	ϵ	10^{-3}	10^{-3}	10^{-3}
	C	21.4	6.40	21.4
	γ	0.00328	0.0743	0.00116
	max SNRseg (dB)	0	0	0
	min SNRseg (dB)	-10	-10	-10

付 録D　オープンテストの推定精度

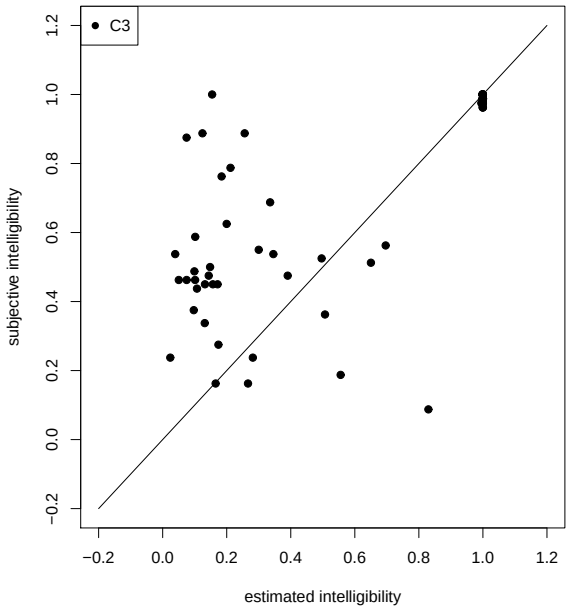
オープンテストにおける主観評価による了解度と推定了解度の関係を回帰手法ごとに Fig. D.1 〜Fig. D.11 に示す.



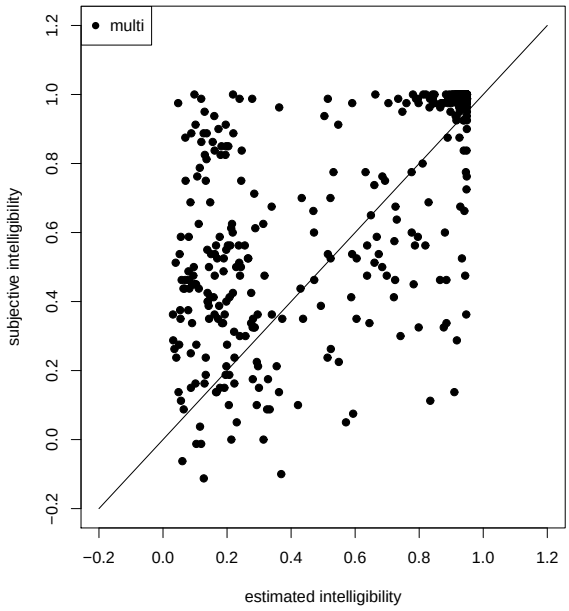
(a) C1



(b) C2

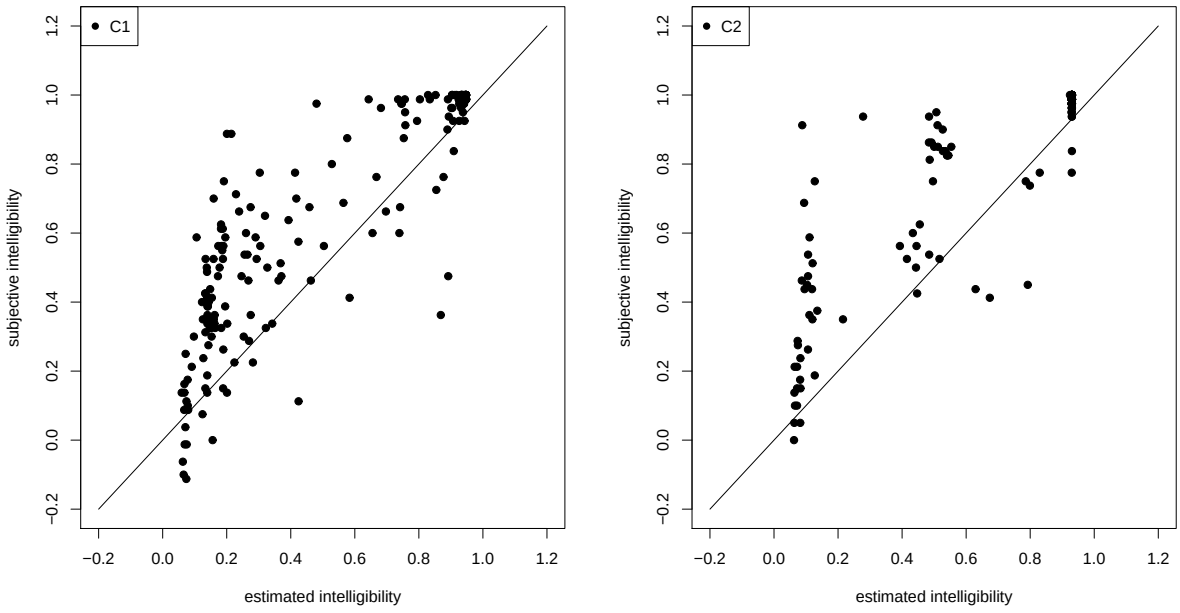


(c) C3



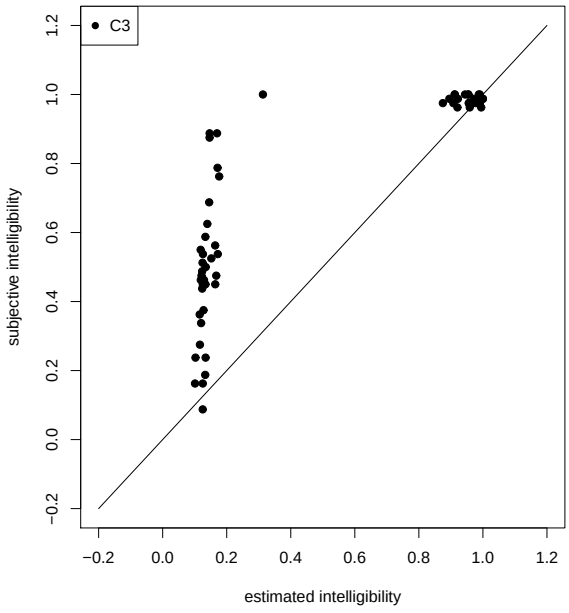
(d) Multi

Fig. D.1: Comparison of open test subjective intelligibility and estimated intelligibility with sigmoid function(SNRseg)

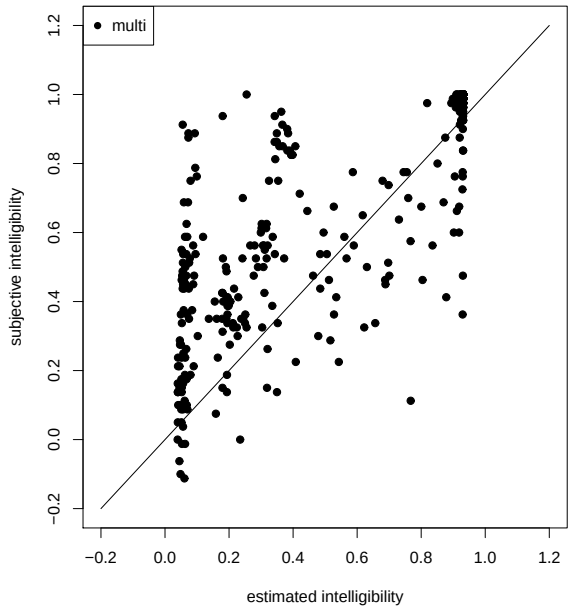


(a) C1

(b) C2

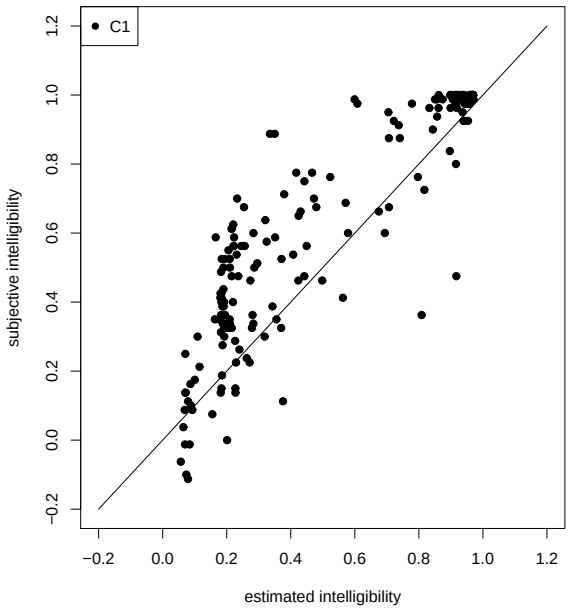


(c) C3

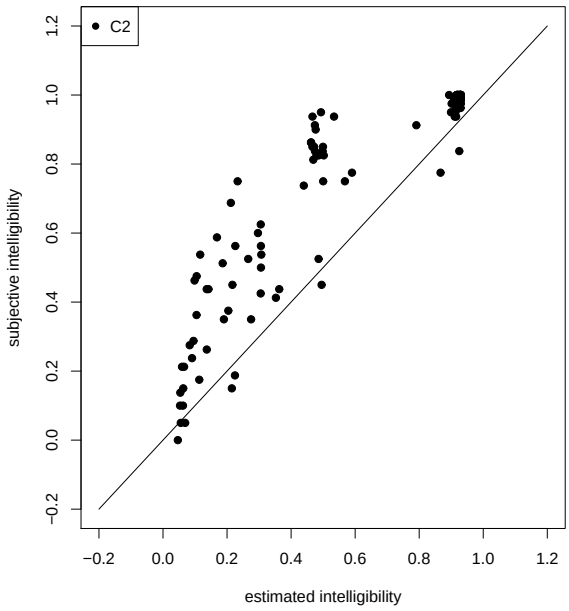


(d) Multi

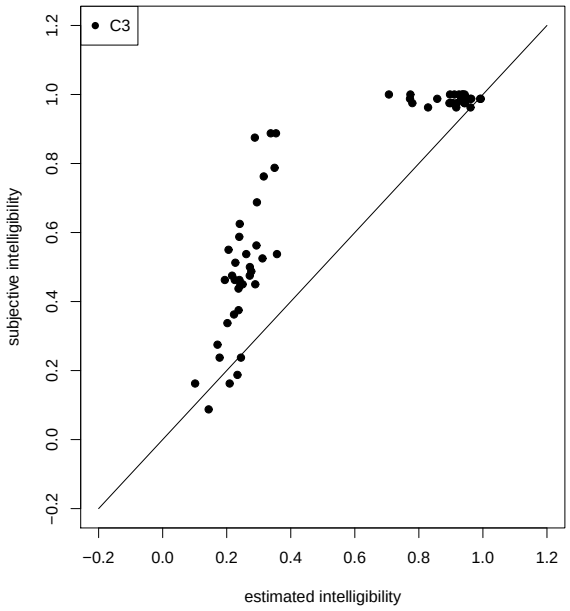
Fig. D.2: Comparison of open test subjective intelligibility and estimated intelligibility with sigmoid function(fwSNRseg(A))



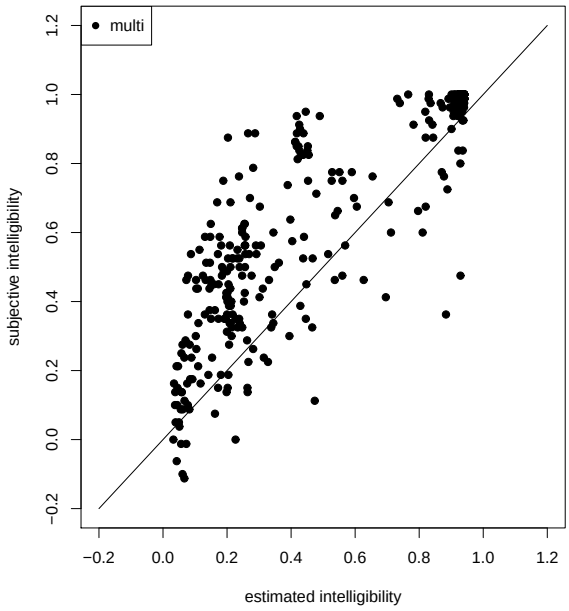
(a) C1



(b) C2

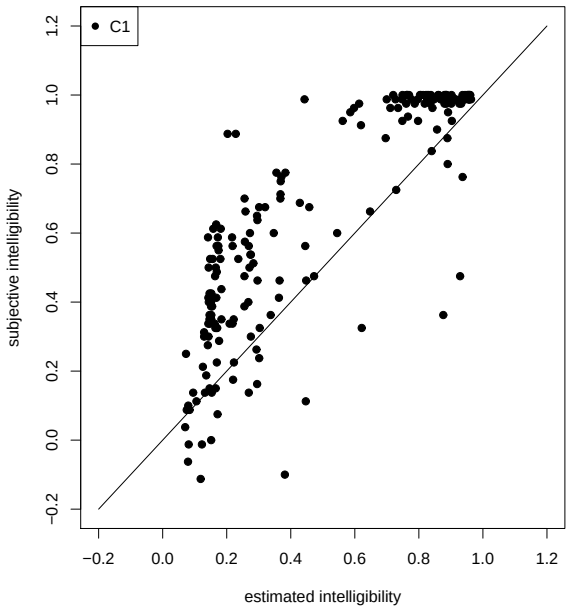


(c) C3

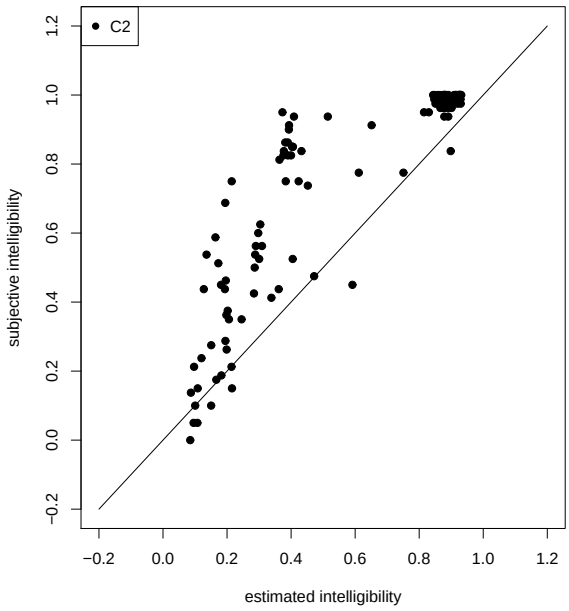


(d) Multi

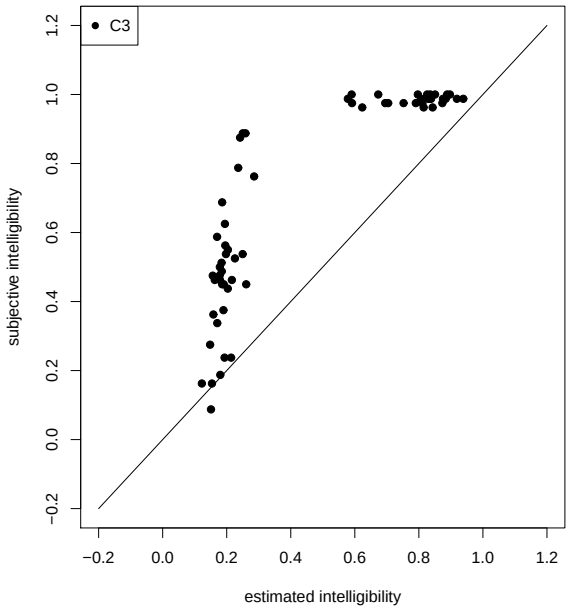
Fig. D.3: Comparison of open test subjective intelligibility and estimated intelligibility with sigmoid function(fwSNRseg(C))



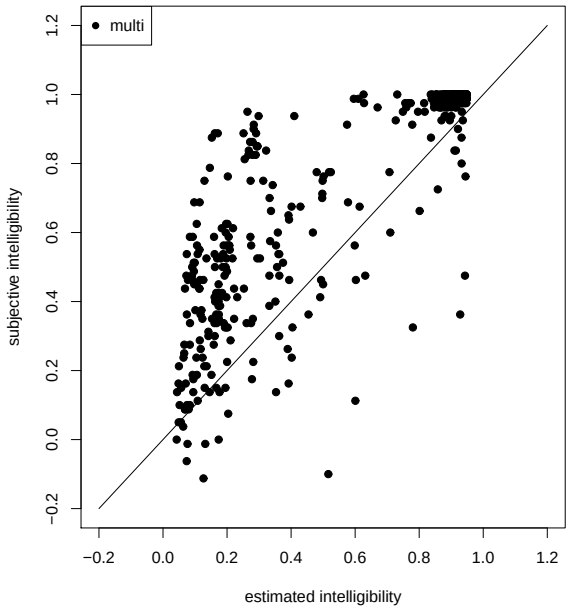
(a) C1



(b) C2

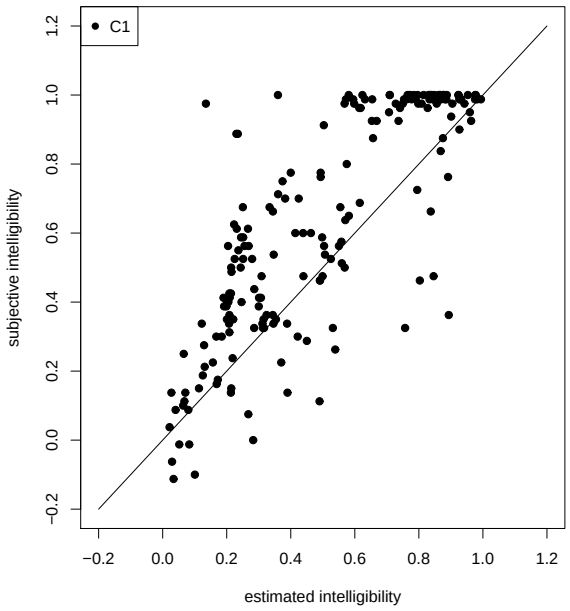


(c) C3

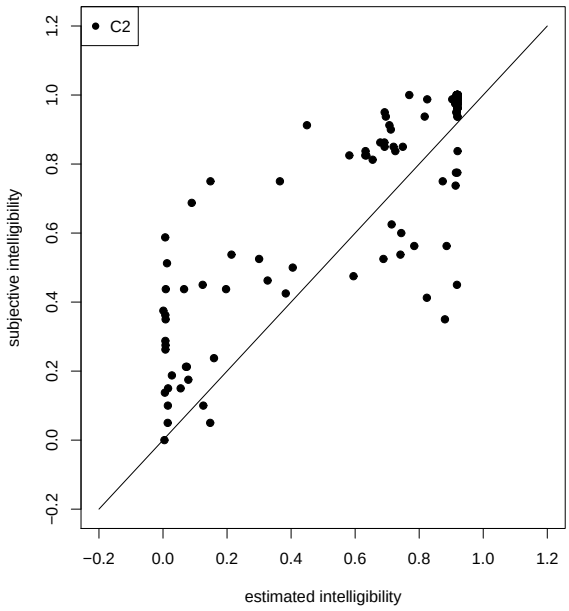


(d) Multi

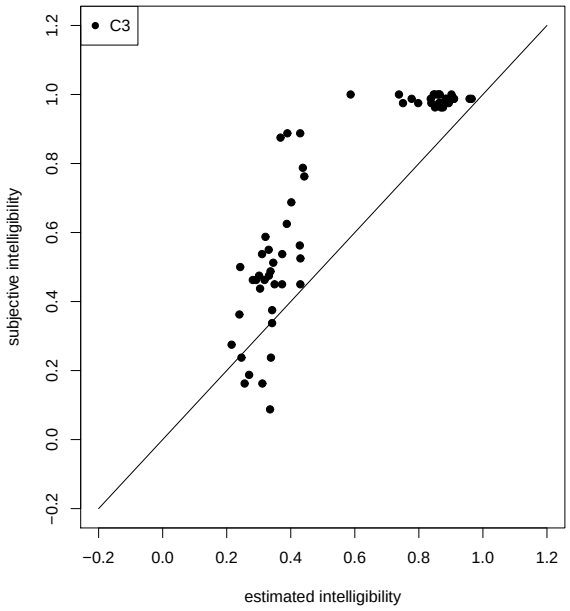
Fig. D.4: Comparison of open test subjective intelligibility and estimated intelligibility with sigmoid function(fwSNRseg(S))



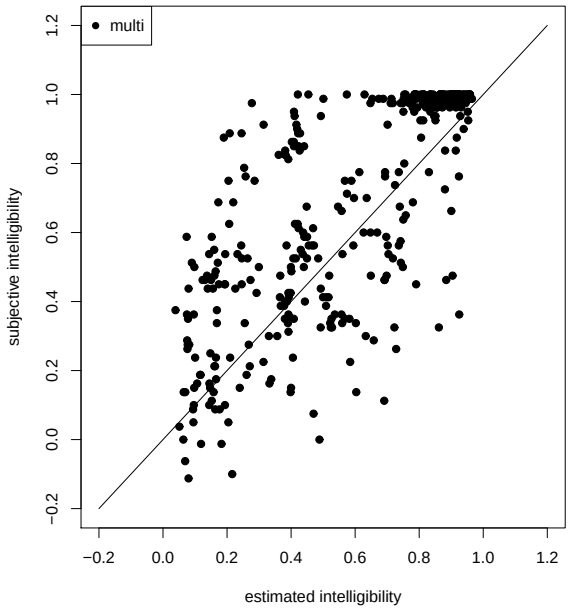
(a) C1



(b) C2

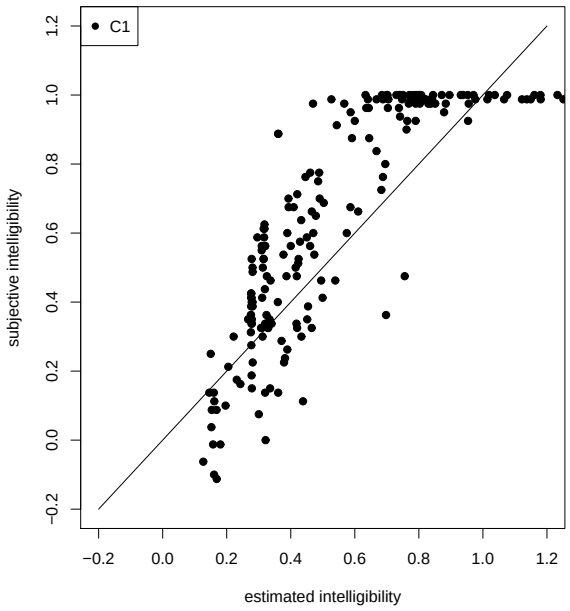


(c) C3

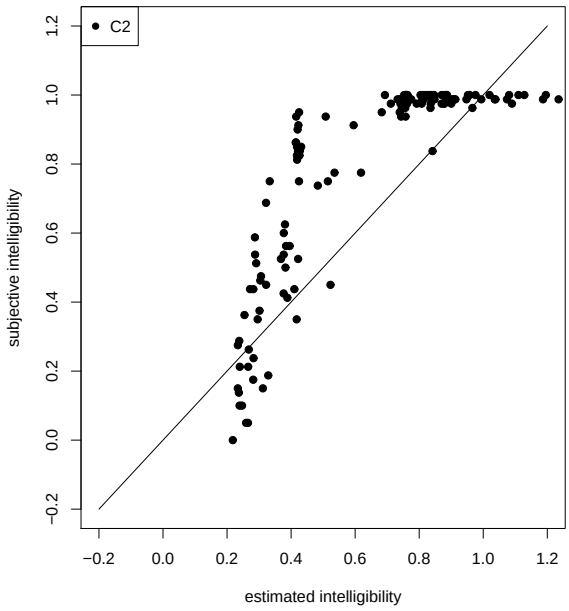


(d) Multi

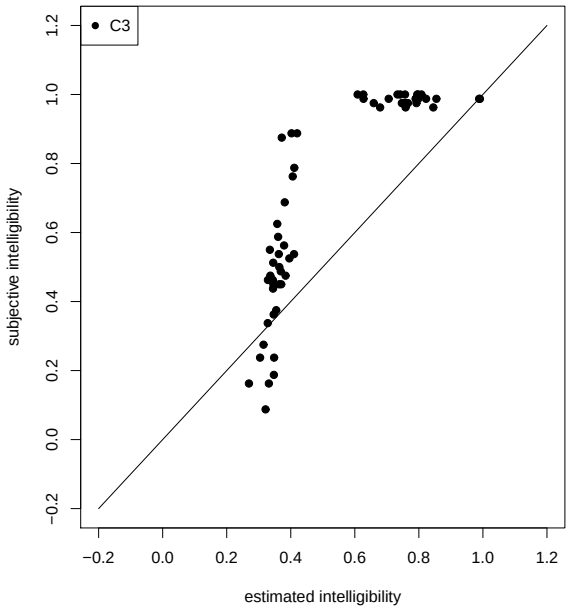
Fig. D.5: Comparison of open test subjective intelligibility and estimated intelligibility with sigmoid function(AIseg)



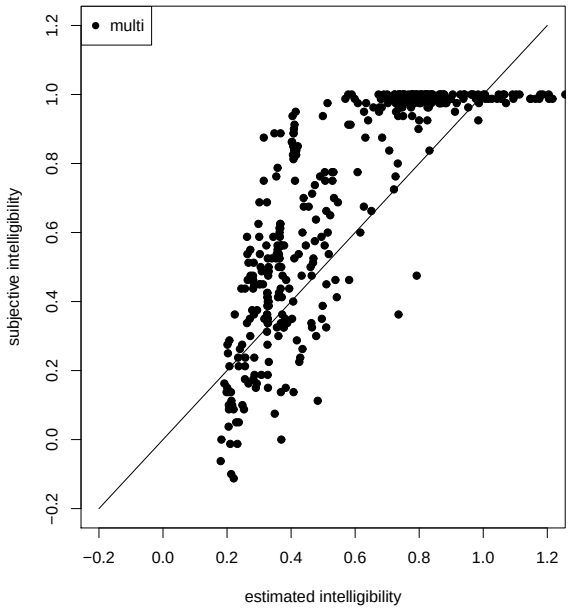
(a) C1



(b) C2

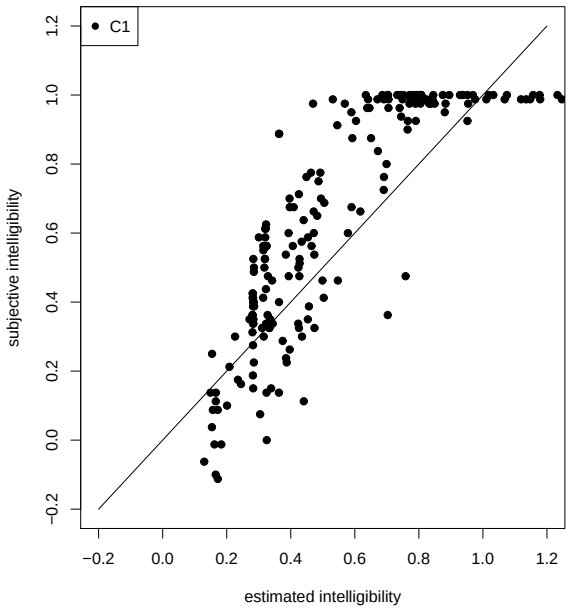


(c) C3

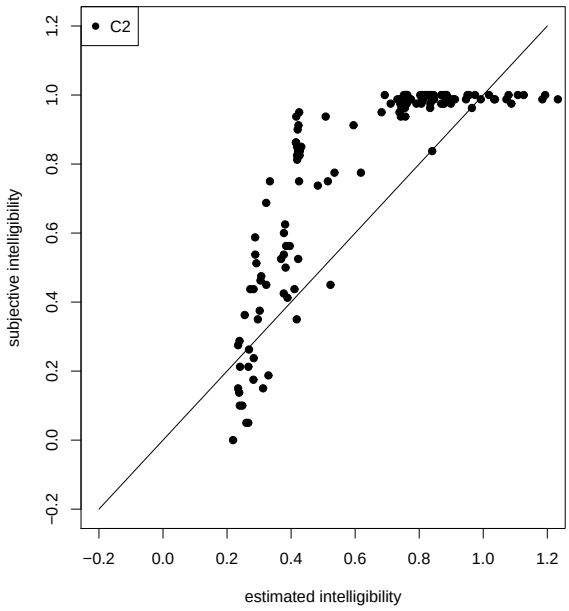


(d) Multi

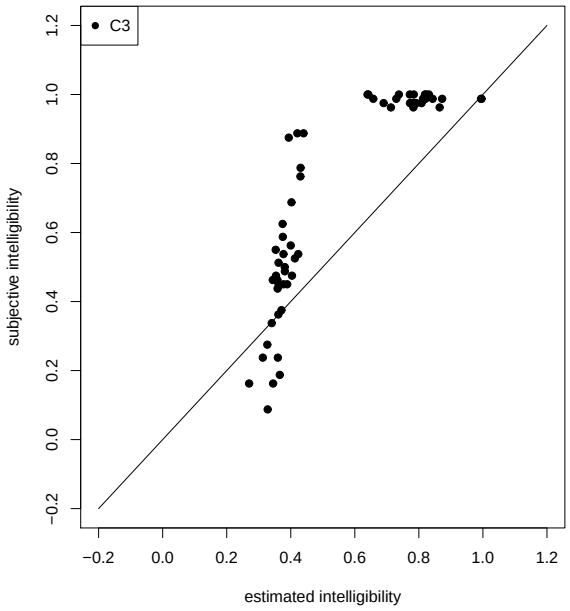
Fig. D.6: Comparison of open test subjective intelligibility and estimated intelligibility with Ridge regression(cbSNRseg)



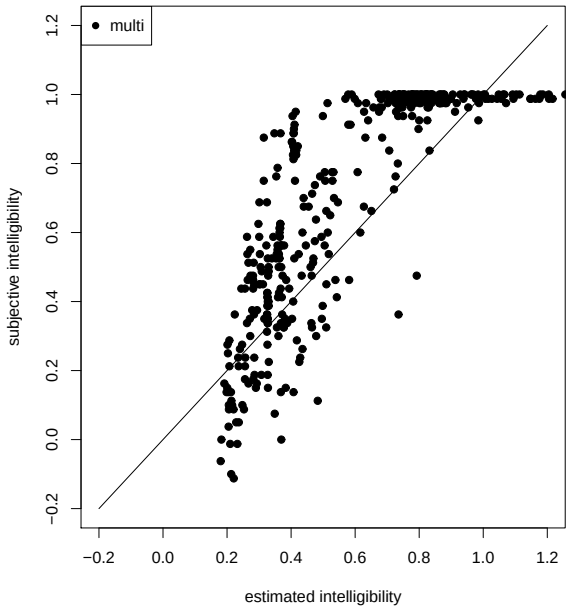
(a) C1



(b) C2

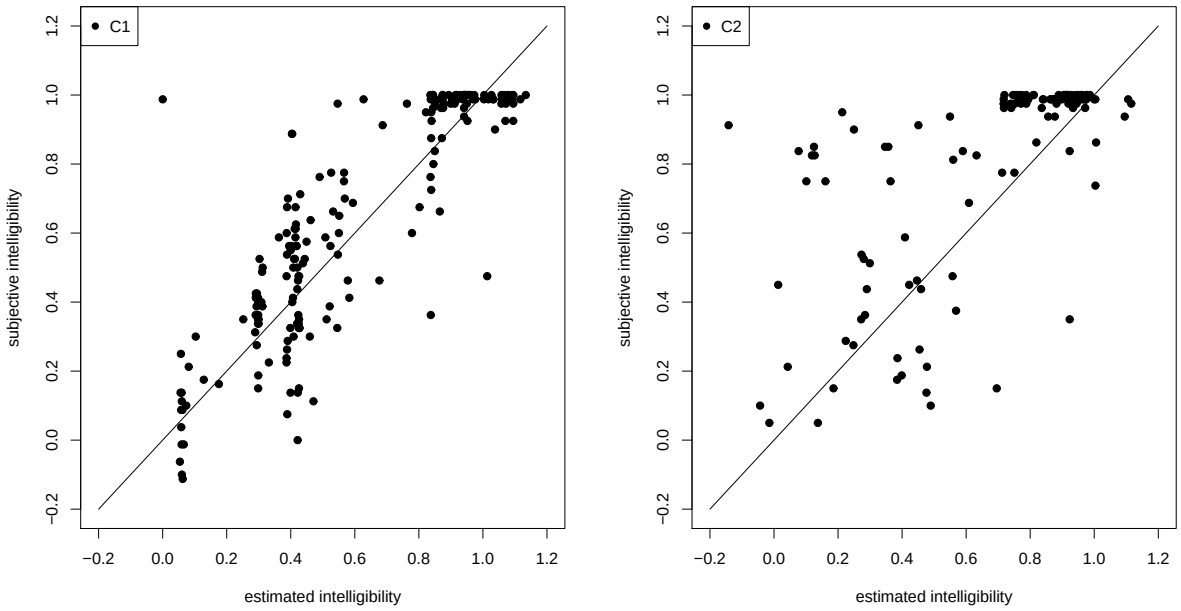


(c) C3



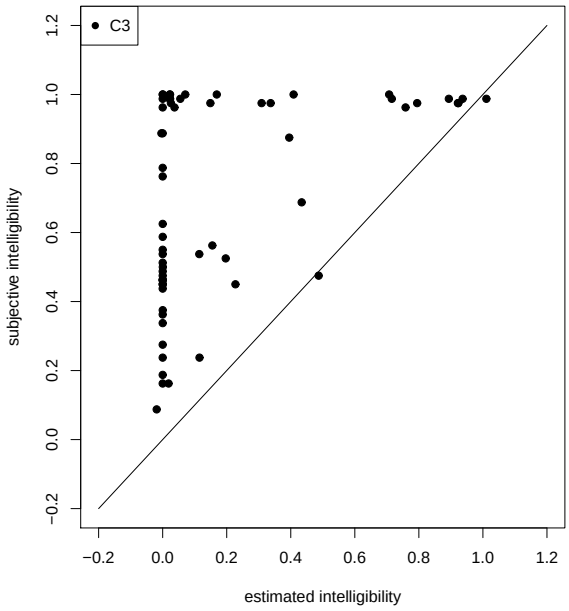
(d) Multi

Fig. D.7: Comparison of open test subjective intelligibility and estimated intelligibility with Lasso regression(cbSNRseg)

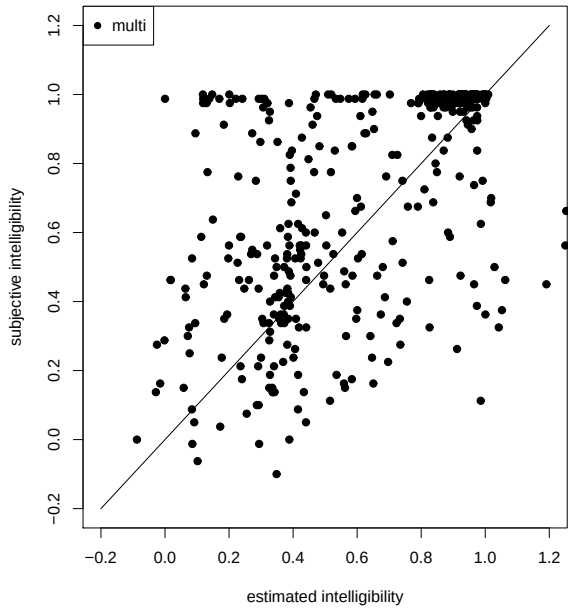


(a) C1

(b) C2

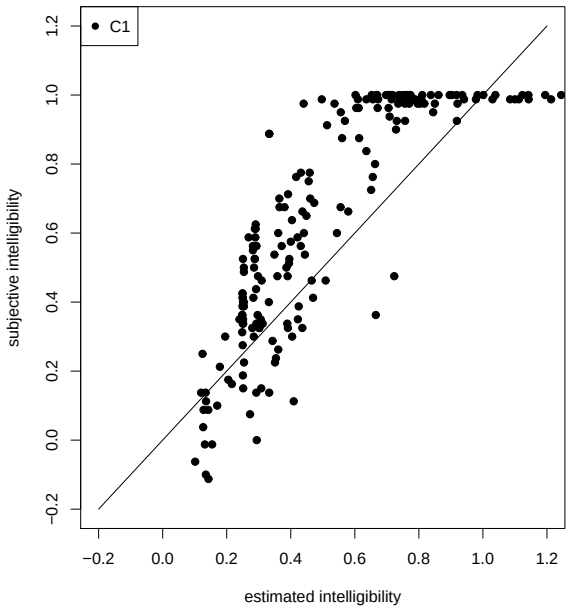


(c) C3

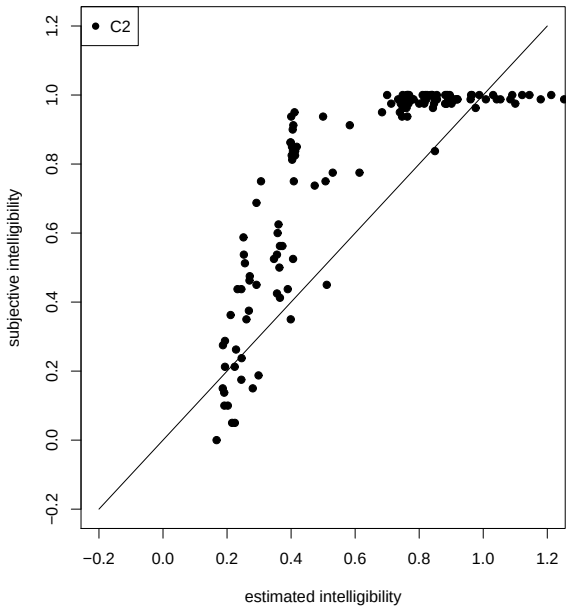


(d) Multi

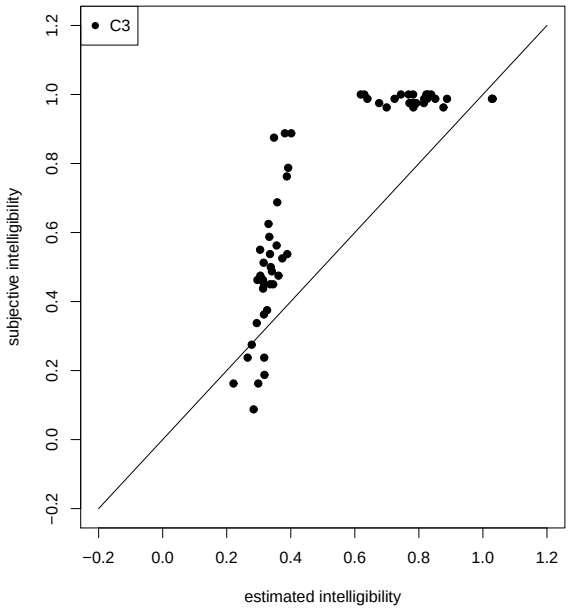
Fig. D.8: Comparison of open test subjective intelligibility and estimated intelligibility with L1 kernel regression(RBF, cbSNRseg)



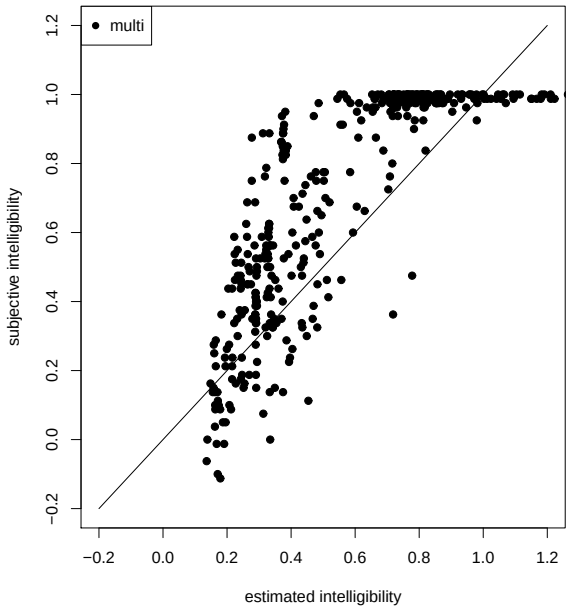
(a) C1



(b) C2

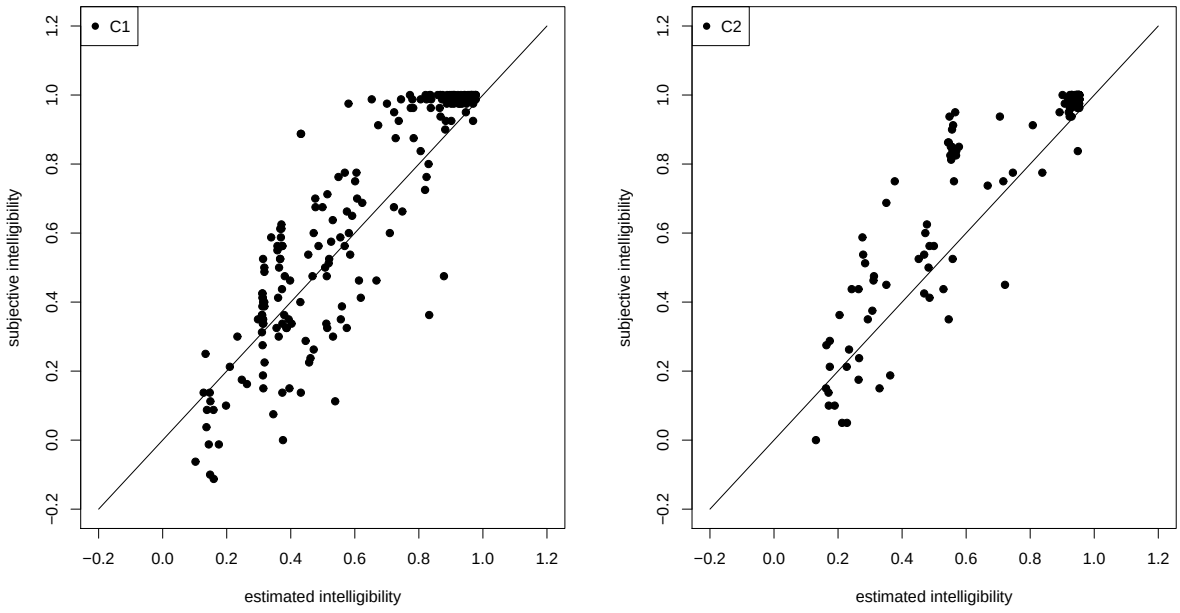


(c) C3



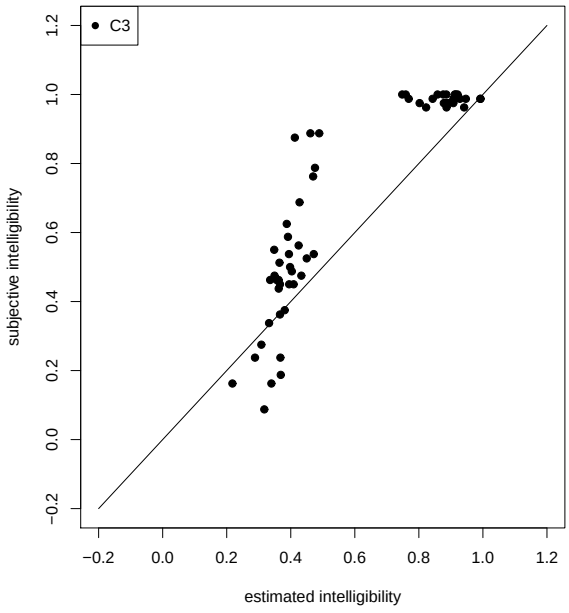
(d) Multi

Fig. D.9: Comparison of open test subjective intelligibility and estimated intelligibility with SVR(linear, cbSNRseg)

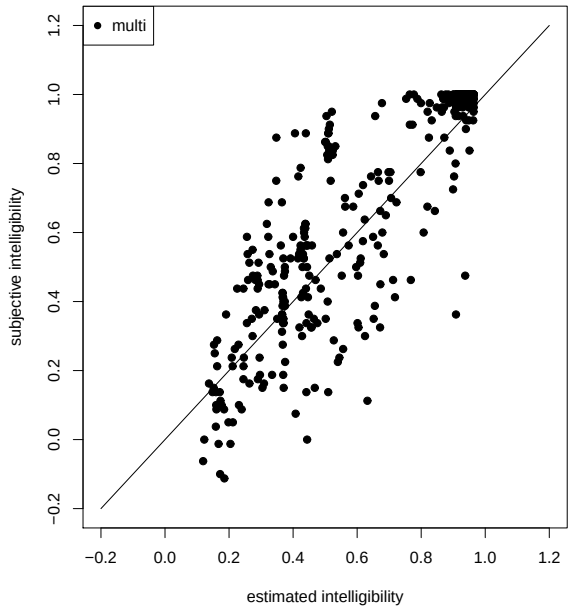


(a) C1

(b) C2

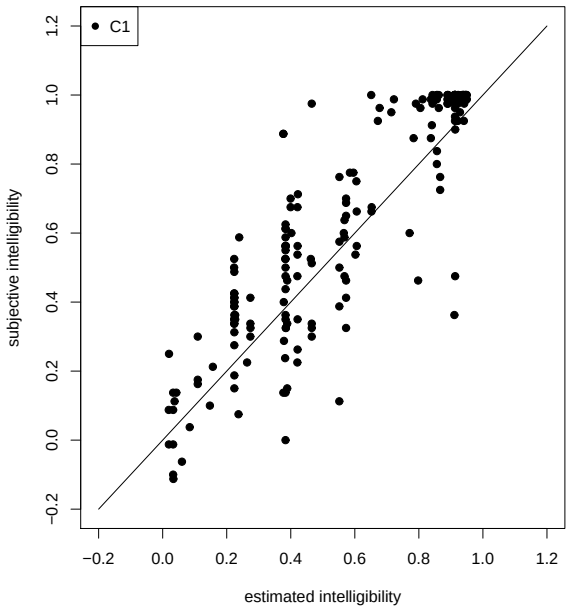


(c) C3

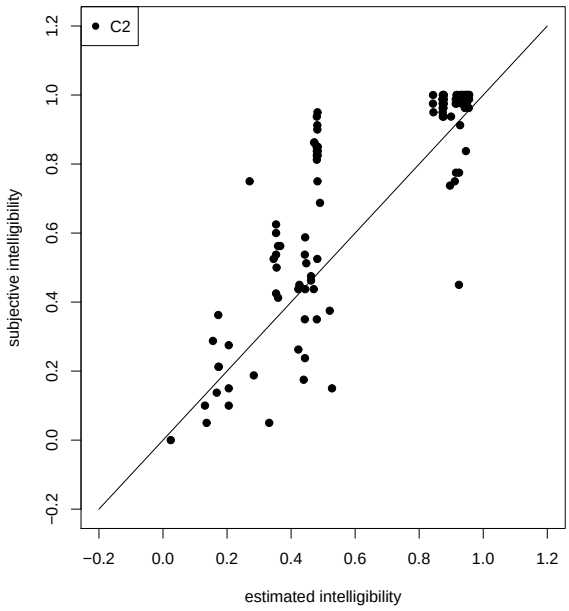


(d) Multi

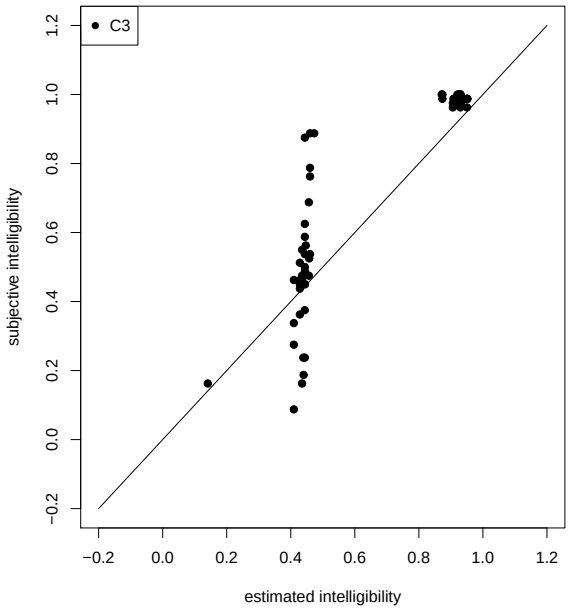
Fig. D.10: Comparison of open test subjective intelligibility and estimated intelligibility with SVR(RBF, cbSNRseg)



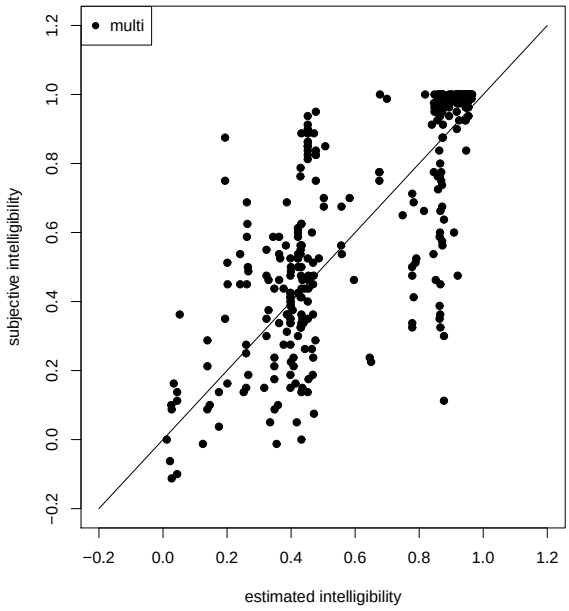
(a) C1



(b) C2



(c) C3



(d) Multi

Fig. D.11: Comparison of open test subjective intelligibility and estimated intelligibility with RF(cbSNRseg)

付 録E　発表論文

†は本論文の内容に関係するものである.

学術雑誌・専門書等への投稿（筆頭著者）

(1) Yosuke Kobayashi, Kazuhiro Kondo and Kiyoshi Nakagawa, “Intelligibility of HE–AAC coded Japanese words with various stereo coding modes in virtual 3D audio space,” Solvi Ystad, Mitsuko Aramaki, Richard Kronland–Martinet, Kristoffer Jensen (Eds.) in Lecture Notes in Computer Science ,vol. 5954, Springer–Verlag, Heidelberg, pp.219–238, (May. 2010)

†(2) 小林洋介，近藤和弘, “音声了解度の客観推定に用いる騒音クラスタリングとその性能評価,” 電気学会論文誌 C，Vol. 133, No. 2, Sec. C，pp.380–387

†(3) 小林洋介，近藤和弘, “帯域別セグメンタル SNR とサポートベクトル回帰を用いた騒音下音声了解度推定,” 電気学会論文誌 C　投稿中

査読付き国際会議発表論文（筆頭著者）

(4) Yousuke Kobayashi, Kazuhiro Kondo, and Kiyoshi Nakagawa, “Intelligibility of low bit rate MPEG–coded Japanese speech in virtual 3D audio space,” Proc. The 15th International Conference on Auditory Display (ICAD2009), pp. 99–102, Copenhagen, Denmark, (May 18–22, 2009)

(5) Yosuke Kobayashi, Kazuhiro Kondo, and Kiyoshi Nakagawa, “Influence of Various Stereo Coding Modes on the Encoded Japanese Speech Intelligibility with Competing Noise ”, Proc. International Workshop on the Principles and Applications of Spatial Hearing (IWPASH), P19(Poster), Sendai, Japan, Nov. 11–13 2009.

(6) Yosuke Kobayashi, Kazuhiro Kondo, Kiyoshi Nakagawa and Yukio Iwaya, “ON THE INFLUENCE CODING METHOD ON JAPANESE SPEECH INTELLIGIBILITY IN VIRTUAL 3D AUDIO SPACE”, Proc. AES The 40th International Conference, P–12, Tokyo, Japan, Oct. 8–10 2010.

†(7) Yosuke Kobayashi, Kazuhiro Kondo, ”On distortion measures effective for the estimation of Japanese speech intelligibility of localized speech with competing noise in virtual acoustic

space," Proc. 40th International Congress and Exhibition on Noise Control Engineering (Inter–Noise), Mon–P–13, Osaka, Japan, Sept. 4–7, 2011.

国内学会・研究会発表予稿論文（筆頭著者）

電子情報通信学会 応用音響研究会（日本音響学会　電気音響研究会）

(8) 小林洋介，近藤和弘，中川清司，高野勝美，“ステレオ符号化が仮想 3 次元空間音声の音像定位精度に与える影響,” 電子情報通信学会技術研究報告，vol. 108, no. 179, EA2008–56, pp. 65–70，仙台市，(2008 年 8 月 4 –5 日)

(9) 小林洋介，近藤和弘，中川清司　“ステレオ符号化が仮想 3 次元空間音声の音声了解度へ与える影響,” 電子情報通信学会技術研究報告，vol. 110, no. 71, EA2010–23, pp. 7–12 札幌市，(2010 年 6 月 10 –11 日)

日本音響学会研究発表会

(10) 小林洋介，矢野式安，近藤和弘，中川清司，“仮想 3 次元空間音声の伝送符号化方式が音声了解度に与える影響,” 日本音響学会 2008 年秋季研究発表会, 1–P–8(Poster), pp. 733–734, 福岡市, (2008 年 9 月 10–12 日)

(11) 小林洋介，近藤和弘，中川清司，“低ビットレート MPEG 音響符号化方式が仮想 3 次元音声に与える影響,” 日本音響学会 2009 年春季研究発表会, 1–9–21, pp. 1509–1512, 東京都目黒区, (2009 年 3 月 17–19 日)

(12) 小林洋介，近藤和弘，中川清司，“HE–AAC におけるステレオ符号化方式が日本語音声了解度に与える影響,” 日本音響学会 2009 年秋季研究発表会, 3–Q–28(Poster), pp. 803–806, 郡山市, (2009 年 9 月 15–17 日)

(13) 小林洋介, 井上脩平, 近藤和弘, 中川清司 “仮想 3 次元音響空間における競合立体妨害音の仰角が日本語音声了解度に与える影響,” 日本音響学会 2010 年春季研究大会, 3–P–24(Poster),pp.921–924, 調布市, (2010 年 3 月 8–10 日)

†(14) 小林洋介，近藤和弘，“仮想 3 次元空間における音声了解度推定の検討,” 日本音響学会 2010 年秋季研究大会,3–Q–1,pp.735–738 , 吹田市, (2010 年 9 月 14–16 日)

†(15) 小林洋介，近藤和弘，“バイノーラル音声の了解度推定に用いる学習条件の検討,” 日本音響学会 2011 年秋季研究大会, 1–Q–25(Poster), pp.593–596, 松江市, (2011 年 9 月 20–23 日)

†(16) 小林洋介，近藤和弘，“サポートベクトル回帰を用いたバイノーラル音声了解度推定の詳細検討,” 日本音響学会 2012 年春季研究大会, 3–Q–5(Poster), pp.601–604, 横浜市, (2012 年 3 月 13–15 日)

†(17) 小林洋介，近藤和弘，“音声了解度推定のための騒音クラスタリングの検討,” 日本音響学会 2012 年秋季研究大会, 2–Q–a1(Poster), pp.507–510, 長野市, (2012 年 9 月 19–21 日)

†(18) 小林洋介，近藤和弘，“帯域別 SNR とサポートベクトル回帰を用いた音声了解度推定,”日本音響学会 2012 年秋季研究大会, 2–Q–a2(Poster), pp.511–514, 長野市, (2012 年 9 月 19–21 日)

†(19) 小林洋介，近藤和弘，“音声了解度推定のための騒音クラスタリングの詳細評価,”日本音響学会 2013 年春季研究大会, 2–Q–19(Poster), pp.641–644, 八王子市, (2013 年 3 月 13–15 日)

情報科学技術フォーラム (FIT)

†(20) 小林洋介，近藤和弘, “仮想音響空間内の音声了解度推定に用いるひずみ尺度の検討,”第 10 回情報科学技術フォーラム (FIT2011)，E–020，函館市，(2011 年 9 月 9 日)

情報処理学会東北支部研究会

(21) 小林洋介，近藤和弘，高野勝美，中川清司，“仮想 3 次元空間音声の伝送符号化方式が音像定位精度に与える影響,”平成 19 年度第 6 回情報処理学会東北支部研究会，7–6–B–1–2，米沢市，(2008 年 3 月 11 日).

(22) 小林洋介，近藤和弘，中川清司，“HE–AAC 符号化におけるステレオ方式が仮想 3 次元音声に与える影響 ”，平成 20 年度第 6 回情報処理学会東北支部研究会，8–6–A–1–3，米沢市，(2009 年 3 月 9 日).

(23) 小林洋介, 近藤和弘, 中川清司 “HE–AAC 符号化が立体音声の音質と音声了解度へ与える影響,”平成 21 年度第 6 回情報処理学会東北支部研究会，9–6–B–2–1，米沢市，(2010 年 3 月 5 日)

†(24) 小林洋介，近藤和弘, “仮想音響空間内の音声了解度推定に用いるひずみ尺度の検討,”平成 22 年度第 6 回情報処理学会東北支部研究会，10–6–B–2–1，米沢市，(2011 年 3 月 14 日)

その他の発表

(25) AES ジャパンコンファレンス・仙台 2012, AES 学生支部イベント学生による研究室紹介, 山形大近藤研究室, AES ジャパンコンファレンス・仙台 2012，仙台市, (2012 年 10 月 10 日)

学術雑誌・専門書等への投稿（共著）

(26) 渋谷 徹，渡邉 瞳，小林洋介，近藤和弘, “音声の伸長・短縮の了解度への影響と適応話速変換方法の提案,”映像情報メディア学会誌, Vol. 66, No.10, pp.J377–J384, (Oct. 2012)

査読付き国際会議発表論文（共著）

(27) Takayuki Kanda, Hiroyuki Yagyu, Yosuke Kobayashi, Kazuhiro Kondo and Kiyoshi Nakagawa, “Comparison of localized speech intelligibility with competing noise using regular and bone–conduction stereo headphones”, Proc. International Workshop on the Principles and Applications of Spatial Hearing (IWPASH), P21(Poster) , Sendai, Japan (November 11–13 2009)

(28) Kazuhiro Kondo, Takayuki Kanda, Yosuke Kobayashi and Hiroyuki Yagyu, “Speech Intelligibility of Diagnoally Localized Speech with Competing Noise Using Bone–Conduction Headphones ”, Proc. Inter–Speech 2010, pp1213–1216, Chiba, Japan, (September 26–30 2010)

(29) Toru Shibuya, Yousuke Kobayashi, Kazuhiro Kondo, “Differences in the effect of speech rate on intelligibility in artificially speed–altered speech by phonetic feature,” Proc. 40th International Congress and Exhibition on Noise Control Engineering, Mon–P–26 , Osaka, Japan, (Sept. 2011)

(30) Naoya Anazawa, Yosuke Kobayashi, Hiroyuki Yagyu, Takayuki Kanda, Kazuhiro Kondo, “Evaluation of localized speech intelligibility from bone–conduction headphones with competing noise for augmented audio reality,” Proc. 40th International Congress and Exhibition on Noise Control Engineering, Mon–P–15 , Osaka, Japan, (Sept. 2011)

(31) Toru Shibuya, Yosuke Kobayashi, Hitomi Watanabe and Kazuhiro Kondo, “DIFFERENCES IN THE EFFECT OF TIME–EXPANDED AND TIME–CONTRACTED SPEECH ON INTELLIGIBILITY BY PHONETIC FEATURE,” Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), SP–P6.10, Kyoto, Japan, (Mar. 2012)

国内学会発表予稿論文（共著）

日本音響学会研究発表会

(32) 阪野俊洋，小林洋介, 近藤和弘 “原音を用いない特徴量による音声了解度推定, ” 日本音響学会 2013 年春季研究大会, 2–Q–20(Poster), pp.645–646 八王子市, (2013 年 3 月 13–15 日)

電気関係学会東北支部連合大会

(33) 神田敬幸, 近藤和弘, 小林洋介, 柳生寛幸，“空気伝導と骨伝導ヘッドホンを用いた空間定位音声了解度の比較, “平成 22 年度電気関係学会東北支部連合大会，2F06 ，八戸市，(2010 年 8 月 26–27 日) 若手研究者優秀論文賞受賞

(34) 阪野俊洋，小林洋介, 近藤和弘，“原音を必要としない特徴量に対し Neural Net を適用した日本語音声了解度推定, ” 平成 24 年度電気関係学会東北支部連合大会

東北地区若手研究者発表会

(35) 渡邉瞳，渋谷徹，小林洋介，近藤和弘，“ 語頭子音特徴による音声の伸長・短縮の単語了解度への影響,” 平成 24 年東北地区若手研究者研究発表会，YS10–A15, pp.29–31, 仙台市，(2012 年 3 月 9 日) 優秀賞受賞

情報処理学会東北支部研究会

(36) 狩野二人, 三浦正範, 小林洋介, 近藤和弘, 中川清司，“指向性の強いパラメトリック・スピーカーを並行配置した音声了解度の評価,” 平成 20 年度情報処理学会東北支部研究会，9–6–B–2–3, 米沢市，(2009 年 3 月 9 日)

(37) 神田敬幸, 近藤和弘, 小林洋介, 中川清司, 柳生寛幸, 岩谷幸雄，“空気伝導と骨伝導ヘッドホンを用いた空間定位音声了解度の比較,” 平成 20 年度情報処理学会東北支部研究会， 9–6–B–2–4, 米沢市，(2009 年 3 月 9 日)

(38) 井上脩平, 小林洋介 , 近藤和弘, 中川清司, 岩谷幸雄，“仮想音響空間内の妨害雑音の仰角が音声了解度に及ぼす影響,” 平成 20 年度情報処理学会東北支部研究会， 9–6–B–2–5，米沢市，(2009 年 3 月 9 日)

(38) 渋谷徹, 小林洋介, 近藤和弘，“話速変換による単語了解度の音韻的特徴の検討” 平成 22 年度第 6 回情報処理学会東北支部研究会，10–6–B–2–2，米沢市，(2010 年 3 月 14 日)