

室内近距離音場の時間・周波数特性
と音の知覚に関する研究

Sound Perception and
Temporal-Spectral Characteristics
in Sound Field near Sound Source

2014年 1月

工学院大学大学院工学研究科
情報学専攻 計算機構成研究室

原 佳史

博士学位論文

氏名(本籍)	原 佳史 (東京都)
学位の種類	博士 (情報学)
学位記番号	博甲第126号
学位授与年月日	平成 26年 1月 27日
学位授与の要件	学位規則第4条第1項
学位論文題目	室内近距離音場の時間・周波数特性と 音の知覚に関する研究

論文審査委員

主査 三好 和憲

副査 管村 昇

” 馬場 則男

” 清水 寧

” 東山 三樹夫

概要

本論文は最適な音コミュニケーション環境構築のために、音源に近接する音場特性がどのように音の知覚に寄与するかを明らかにし、音、音場表現に富んだ収録・再生環境の最適化について論じるものである。その結果、直接音のエネルギー大きい近距離音場における音の好ましさや音の大きさ感、音の聴き易さに関する主観的な知覚変化ならびに、音のマスキング効果を音の時間・周波数特徴に着目した分析・評価をすることによって、音の心理的・物理的要因を明らかにした。

室内を伝わる音の特徴は、音源から受音点に達する直接音と受音点に達する反射音群の重なりによって表現される。音の知覚または音場の主観的印象は、主として直接音によるものと考えられる音の方向定位と反射音による音場の広がり感に関するものに大別される。室内音場における直接音領域の範囲(Coherent領域)は、すでに波動理論に基づいてMorseとBolt(1944)によって理論的に推定されている。しかし、Lyon(1983)は直接音距離を伝達関数の極・零分布の視点から考察し、直接音領域の伝達特性が最小位相特性の零点分布によって表現されると推論した。これらの考察をもとに高橋等(2006)は残響室内音場で実測したインパルス応答から位相周波数特性を帯域別に分析することによって、Morse等による直接音距離の推定値を実証することに成功した。

以上の研究経緯を踏まえ、本研究では直接音領域から残響音場において収録された音源・受音点間距離(Sound Source Distance: SSD(m))に伴う音の心理的・物理的要因の変化に注目した。残響室ならびに無響室においてSSD(m)を

変えて測定したインパルス応答を用いて主観評価実験を行なうとともに、主観値変化に対応した物理的分析手法を検討した。

近距離音場における知覚変化の尺度化には一対比較法 (Thurstone CaseV) を用いた。距離別音楽信号を用い、音の知覚において最も重要とされる音の好ましき (主観的 Preference) と大きさ感 (主観的 Loudness) に関する受聴実験を行なった。その結果、Preference と Loudness の両者が共に最大となる收音点が直接音領域内に存在することを確認した。

直接音のエネルギーが大きい音源近傍での收音においては、初期反射音が音楽の音質に最も影響を与える。安藤等 (1985) によると、反射音の好ましい遅れ時間は音源信号の自己相関関数 (Auto-Correlation Function: ACF) の継続時間と対応するとされている。また主観的に好ましい最適な遅れ時間は ACF の値が 0.1 になる遅れ時間によって特徴づけられると考えられている。

聴覚モデルにおいて音響信号は蝸牛の周波数選択性から狭帯域信号に分割される。分割された狭帯域信号は有毛細胞モデルにより半波整流後にローパスフィルターを施した狭帯域包絡線を得ると考えられている。Meddis 等 (1997) はこの聴覚モデルに基づき、狭帯域包絡線の自己相関関数を用いた Summary Auto-Correlation Function (SACF) による聴覚脳幹の周期性検出に基づく Pitch 知覚モデルを提案した。そこで本論文では、ACF の初期部の変化が直接音の増強効果を表していると考え、短時間観測区間にわたる ACF を示すことで近距離音場における主観評価との対応を示した。さらに残響音が付加された楽音波形のパワースペクトルから際立ったピークを形成する周波数成分を抽出した後、ピーク周波数列の基本周波数と高調波成分を推定することによって、音の調波構造を可視化する時間・周波数分析 (SPEC² 分析) を試みた。

室内空間における人々の会話において、音声の明瞭性が重要である。Houtgast (1973) は狭帯域包絡線の変調度によって、室内の反射音や雑音による了解度評価が可能であることを報告している。一方、反射音のエネルギーが小さい

近距離音場においては、残響による音声マスキングが起こらず、音声情報の損失がないため、文章了解度評価や音声明瞭度指数 (Speech Transmission Index : STI) 分析を実施することが困難である。そこで音声の聞きやすさに関する主観的な明瞭性評価実験を行なった。また狭帯域信号を用いた聴覚モデル分析を取り入れた時間・周波数特徴の可視化によって近距離音場での SSD(m) によって変化する音の微細な物理的变化と明瞭性に関する知覚評価結果との対応を明らかにした。

共有空間における音コミュニケーションにおいては、上述した SSD(m) による音の変化の他に、受聴者には周囲雑音や会話音声などの外部からの音が到来する。共有性の高い空間における音環境設計では、会話情報の秘話性が重要になる。近年、他人の会話が邪魔にならない空間を目指した Speech Privacy が重視されている。そこで本論文ではこれまで検討されてこなかった会話音声を効率良く遮る情報マスキング効果の評価方法を開拓した。白色雑音による定常信号と情報マスキング効果が期待できる逆転再生音声をマスキング信号とし、単語了解度評価と時間・周波数分析によるマスキング効率評価をおこなった。また、音声波形の特徴を加味することによって、効率の良い情報マスキング効果が期待出来るマスキングの設計、さらには環境特徴を取り入れたマスキングシステムの設計方法を開拓した。音声の了解性に深く関わっている時間波形の帯域別包絡線に加えて、音声の調波構造が示す時間変化にもまたマスキング効果を評価する上で重要な要因があることが明らかとなった。

第1章では、序論として本研究の背景及び目的・意義について述べる。

第2章では、残響室ならびに無響室において音源・受音点間距離 (Sound Source Distance: SSD(m)) を変化させたときの室内伝達関数の測定方法、並びに測定結果を示す。測定点は音源近傍の 0.01(m) から 2.56(m) までの残響室において推測された直接音領域から残響音場領域までを設定した。各測定点における時間波形ならびに振幅・周波数特性を図示することによって SSD(m) の増大に

よってスペクトルの山谷の分散が増えていくことを確認した。そこで振幅・周波数特性の標準偏差 (dB) を計算することによって、SSD(m) に比例するスペクトル成分の増大を示した。しかしこれまで用いられてきた残響音場での知覚に対応するとされる音場評価指標の一つである D 値、ならびに時間重心値では、近距離音場の Coherent 領域の物理的な変化を表すことができなかった。

第3章では、第2章で測定した残響インパルス応答を音楽信号に畳み込むことによって作成した仮想距離別音楽信号を用いた主観評価実験を示す。音の知覚評価には、Thurstone の一対比較法 (Case V) を用いた。距離が異なった対の刺激音を評価することで SSD(m) の増大によって変化する音の大きさ感を尺度化した。残響室の評価結果において、第1章にて示した従来音場評価指数が一定となる Coherent 領域内において、主観値が最大となる結果が示された。

第4章ではまず、音の最適な時間遅れと対応するとされる自己相関関数 (Auto-Correlation Function: ACF) に着目する。初期反射音は音源の直接音エネルギーを増強し、知覚変化にとって重要とされる。このことから、距離別伝達関数を音楽信号に畳み込むことによって仮想的に作成した距離別音楽信号の自己相関関数 (Auto-Correlation Function: ACF) の 30(ms) 以内の初期部の変化を観測することによって微細な周波数特徴の変化を示した。また、最も音源からの距離が近い SSD:0.01(m) の信号を参照とした短時間 ACF のエネルギー比では、Coherent 領域内で緩やかに変化する一方、Coherent 領域を超えると急激にエネルギー比が変化し残響付加によって、原信号が崩れていく様子が確認できた。

さらに残響音が付加された楽音波形のパワースペクトルから際立ったピークを形成する周波数成分を選出 (Peak-Picking) した後、ピーク周波数列の基本周波数と高調波成分を推定することによって、調波構造の時間・周波数分析 (SPEC² 分析) を試みた。

第5章では音声の主観的明瞭性に関する評価実験ならびに、音声の主要帯域における包絡線スペクトル分析に着目した。会話音声の明瞭度には狭帯域包絡

線特徴が重要であるとされている。近距離音場における音声の狭帯域包絡線スペクトルエネルギーを分析することにより、主観的明瞭性変化との対応が示された。

第6章では、共有空間の Speech Privacy を求めた音環境設計において、会話音声の情報を聞こえにくくするマスキング音に関する単語理解度評価実験ならびに時間・周波数特徴に着目したマスキング効率の評価方法を検討した。エネルギーマスキングに代表される白色雑音と、音声理解度を減少させる効果が期待できる逆転音声を用い、音声単文章の書き取り評価を行うことで、マスキング音による理解度の差を示した。また Target となる音声にマスキングを重畳した信号の基本周波数分析を行なうことにより、Target の理解度評価との対応を明らかにした。

付録 A では包絡線特徴を取り入れた情報マスキング効率の高いマスキングの設計方法と、室内空間を伝わる音声をマスクするマスキングシステムに適したマスキング作成・評価方法に関する基本原理とアルゴリズムを開拓した。

第7章では、本論文の総括と今後の課題を述べる。

以上、要するに個人性が追求される現代の近距離音場での音コミュニケーション環境において、音の知覚と音場の物理的変化の関係を主観評価実験による尺度化と音の時間・周波数特徴に着目した分析をすることで明らかにした。

Abstract

This thesis describes the sound perception and the temporal-spectral characteristics for the sound field near sound source in order to design a comfortable communication environment.

The contents of this thesis are as follows;

Chapter 1 is the introduction that presents the background and the objectives of this thesis. Chapter 2 introduces the measuring method of the transfer function in the reverberation room and the anechoic room, and indicates its results as sound source distance (SSD) is varied. Chapter 3 presents a framework of the subjective evaluation test by Thurstone's case V with SSD-dependent musical signals that are artificially created by the convolution between musical signals and the impulse responses measured in the reverberation room and the anechoic room in Chapter 2. Chapter 4 evaluates the early reflection of the SSD-dependent musical sounds by the short-time auto-correlation function (ACF). Additionally, the harmonic structures of the SSD-dependent musical sounds with the reverberation are visualized through the temporal-spectral analysis referred to as SPEC². This is led by general fact that the intelligibility of the direct sound in a reverberation room is greatly dependent on the amount of reverberation. Chapter 5 shows the evaluation test for the subjective speech intelligibility (SI), and the effect of the direct sound in the sound field near sound source by investigating the spectral change for the narrow-

band envelope of speech. Chapter 6 explores a new approach for the speech intelligibility test using informational maskers. Japanese short sentences with white noise as the energy maskers and reversed speech as the informational maskers, are used for speech intelligibility tests. The dissimilarity between the signal and the masked signal in terms of the frame-based correlation coefficient by SPEC² is evaluated so that the efficacy of the maskers including the informational masking effects on SI can be verified. Chapter 7 concludes this thesis, and then Appendix A proposes a design method for the effective informational masker, and the fundamental principle and algorithm for a masking system including room environmental characteristics.

目次

第1章 序論	17
1.1 本研究の意義と目的	17
1.2 本研究分野に関する背景	18
1.2.1 近距離音場の伝達関数に関する研究	18
1.2.2 音の基本周波数特徴と狭帯域包絡線特徴に着目した音の 時間・周波数分析	19
1.2.3 Speech Privacy における情報マスクング効果	20
1.3 本論文の構成	21
第2章 室内における近距離音場のインパルス応答測定と評価	23
2.1 はじめに	23
2.2 室内伝達関数の測定	24
2.3 インパルス応答とエネルギー減衰	26
2.4 Coherent 領域	33
2.5 残響インパルス応答による室内音場評価	33
2.5.1 直接音と残響音のエネルギー比 D_{30} 値	34
2.5.2 時間重心	35
2.5.3 振幅周波数特性の標準偏差	36
2.5.4 初期反射成分の分析	38
2.6 まとめ	38

第3章	近距離音場における主観評価実験	41
3.1	はじめに	41
3.2	Thurstone Case Vによる一対比較評価方法	42
3.3	距離別音楽モチーフによる主観的 Preference 評価実験	46
3.3.1	実験方法	46
3.3.2	評価結果	48
3.4	距離別音楽モチーフによる主観的 Loudness 評価実験	50
3.4.1	実験方法	50
3.4.2	評価結果	51
3.5	まとめ	53
第4章	基本周波数分析 SPEC ² 用いた調波構造の可視化	55
4.1	はじめに	55
4.2	短時間自己相関分析	56
4.3	スペクトルのピーク抽出による基本周波数分析 SPEC ²	59
4.3.1	Spectral Peak-Picking による信号表現	59
4.3.2	SPEC ² による基本周波数分析	61
4.4	まとめ	64
第5章	包絡線スペクトル特徴に着目した音声の主観的明瞭性評価	67
5.1	はじめに	67
5.2	近距離音場における距離別音声の明瞭性評価実験	68
5.2.1	実験方法	69
5.2.2	実験結果	69
5.3	距離別音声信号の狭帯域包絡線スペクトル分析	70
5.4	まとめ	71

第 6 章 情報マスク効果の単語理解度評価と時間・周波数特徴の類似性分析	77
6.1 はじめに	77
6.2 単語理解度試験による情報マスク評価	78
6.2.1 実験方法	79
6.2.2 実験結果	81
6.3 時間・周波数分析によるマスク効果の評価	81
6.3.1 帯域別狭帯域包絡線相関	81
6.3.2 SPEC ² 相関によるスペクトル類似性分析	83
6.4 まとめ	86
第 7 章 総括	89
付 録 A 包絡線特徴に着目した情報マスク音の設計と評価方法	93
A.1 はじめに	93
A.2 包絡線マスキング (EHM) 設計方法	93
A.3 単語理解度評価試験	95
A.3.1 実験結果	96
A.4 室内マスクシステム的设计	96
A.5 室内マスクシステムの評価	97
A.5.1 SPEC ² 相関によるマスク効果の評価	97
A.6 まとめ	100
研究業績	101
謝辞	105
参考文献	107

目 次

2.1	(a) 残響室, (b) 無響室におけるインパルス応答の測定環境例 . . .	25
2.2	インパルス応答測定機器の構成	26
2.3	残響室内における距離別インパルス応答の時間波形	27
2.4	残響室内における距離別インパルス応答の振幅・周波数特性 . . .	28
2.5	無響室内における距離別インパルス応答の時間波形	29
2.6	無響室内における距離別インパルス応答の振幅・周波数特性 . . .	30
2.7	残響室内における距離別インパルス応答の残響曲線	32
2.8	残響室内における残響時間 RT_{60} (s)	34
2.9	距離別インパルス応答の D_{30} (dB)	35
2.10	距離別インパルス応答の時間重心値	36
2.11	距離別インパルス応答の振幅・周波数特性の標準偏差	37
2.12	初期反射成分の振幅・周波数特性	39
2.13	初期反射成分の振幅・周波数特性の標準偏差	40
3.1	Preference 試験音作成手順	47
3.2	Preference 試験音構成	47
3.3	Preference 評価結果 (a):Water Music Suite No.6, (b):Symphony No.4	49
3.4	Loudness 試験音作成手順	50
3.5	Loudness 試験音構成	51

3.6 (a) 残響室ならびに (b) 無響室における距離別音楽モチーフによる主観的 Loudness 評価結果	52
4.1 残響室における距離別音楽モチーフの短時間自己相関関数例	57
4.2 SSD(m) と短時間自己相関エネルギー比 ΔACF (dB)	58
4.3 Spectral Peak-Picking による線スペクトル抽出手順	60
4.4 Spectral Peak-Picking 例	62
4.5 SPEC ² 分析方法	63
4.6 音楽モチーフの SPEC ² 分析例	64
4.7 距離別音楽モチーフの SPEC ² 分析例	65
5.1 一対比較実験による明瞭性評価結果	69
5.2 狭帯域包絡線スペクトル分析方法	70
5.3 距離別狭帯域包絡線スペクトルの 3次元図 (残響室)	72
5.4 距離別狭帯域包絡線スペクトルの 3次元図 (無響室)	73
5.5 狭帯域包絡線スペクトルエネルギー比	74
5.6 狭帯域包絡線スペクトルエネルギー (EME) 計算方法	75
5.7 距離別音声の狭帯域包絡線スペクトルエネルギー (EME)	76
6.1 マスカーに埋もれた発話音声の単語了解度試験音構成 (a) と試験条件 (b)	80
6.2 単語了解度試験結果	81
6.3 TM 比 (dB) ごとの狭帯域包絡線相関係数 (EC)	83
6.4 狭帯域包絡線相関係数 $\rho_E(l, i)$ と単語了解度 (%)	84
6.5 TM 比 (dB) ごとの SPEC ² 相関係数 (SC)	85
6.6 SPEC ² 相関係数 $\rho_F(l, fr)$ と単語了解度 (%)	86
A.1 残響音による包絡線の変化	94

A.2	音声包絡線を部分的に保存した音声変調雑音マスク (EHM) 作成方法	94
A.3	Target と TM 比 (dB) ごとのマスクの時間波形	95
A.4	単語理解度試験結果	96
A.5	室内マスクシステムにおける Target	97
A.6	環境マスクを取り入れた室内マスクシステム構成と作成アルゴリズム	98
A.7	(a) 基本マスクならびに (b) 環境マスクに対する SPEC ² 相関分析	99

第1章 序論

1.1 本研究の意義と目的

本論文は音のコミュニケーションにおいて最適な環境づくりのために、音場特性がどのように知覚に起因するか明らかにし、音、音場表現に富んだ収録・再生環境の最適化について論じるものである。人々は生活の中で空間を伝わる音を介してコミュニケーションを行っている [1]。空間内を振動によって伝わる音波を聴いたり、触れたり、見たりすることにより、音響事象を知覚し音の働きを理解する。この音響事象を複数の人間、機械、自然が共有することによって音のコミュニケーションが生まれる。

現代では、CPU や DSP の小型化や通信効率の向上によって PC やモバイルデバイスをインターフェースとした音のコミュニケーションが行なわれる機会が多くなってきた。デジタル通信技術では逆フィルタ処理やエコーキャンセリングによって環境騒音や室内の響きを信号処理によって抑制し、音声の明瞭度向上を行っている。しかし本来の音のコミュニケーションにおいて、利便性追求のための技術により音場特性を排除し音の響きの持つ豊かさまでも制限されるべきではない。これまでに室内音場特性を評価する指標として、知覚に対応する様々な物理パラメータが規準化される一方、室内の近距離音場の評価指標に関しては検討されてこなかった。そこで本論文では室内音場において音源から放射された音波が球面波で伝わりとされる領域 (Coherent 領域) に着目した。Coherent 領域を含む近距離音場において、音源・受音点間距離 (Sound Source Distance: SSD(m)) の増大に伴う音の知覚変化を受聴実験によって明

らかにするとともに、音の時間・周波数特徴変化の可視化・評価方法を開拓する。また近年、コミュニケーション空間の音環境設計において会話情報の保護という視点から重視される Speech Privacy に着目し、空間を伝わる会話情報を伝え難くする背景音(情報マスキング音)の設計とマスキング効率の評価方法を提案する。

1.2 本研究分野に関する背景

1.2.1 近距離音場の伝達関数に関する研究

室内を伝わる音の特徴は、音源から受音点に達する直接音と周囲の壁面による反射をくり返した後に受音点に達する反射音群の重なりによって表現される。その結果、音の知覚または音場の主観印象は、主として直接音によるものと考えられる音の方向定位と反射音による音場の広がり感に関するものに大別される。室内音場に関する音の主観印象には、直接音にわずかな反射音が重なりあって作られる近距離音場での伝達特性が重要である。

波動理論によると、室内音場において音源から放射された音波が球面波で伝わりとされる領域 (Coherent 領域) の存在が Morse と Bolt によって予想されている [2]。近年、高橋等 [3][4][5] は実測した残響室インパルス応答から最小位相・位相周波数特性の局所的変化の回帰分析によって、音源からの距離に伴う伝達位相変化を実証し、その結果伝達位相変化の観測できる範囲が Coherent 領域と概ね一致することを確認した。最小位相零点の周波数軸上の移動によって生じる伝達位相変化によって音源・受音点間距離情報が含まれている事が、位相特性解析により実証されている。伝達位相変化とその分散に着目することで、音源からの距離情報を含む伝達関数を音場の詳細な境界条件を必要とすること無く雑音信号から生成する手法が検討されている。

そこで本論文では、残響室ならびに無響室において音源・受音点間距離 (Sound Source Distance: SSD(m)) を変えて測定したインパルス応答を用いた主観評価実験, および主観値の変化に対応した物理量の変化を明らかにするためにインパルス応答の分析を行なった。

1.2.2 音の基本周波数特徴と狭帯域包絡線特徴に着目した音の時間・周波数分析

信号波形における狭帯域包絡線は音声情報を得る大きな手掛かりとして知られており, 音声の明瞭性にとって重要なファクターである。Drullman 等 [6] は 100-6.4k(Hz) にわたる 24 帯域の 1/4 オクターブ帯域包絡線とそれぞれの周波数帯域に対応する帯域雑音から了解性のある音声を合成できることを示した。また Shannon 等 [7] は音声帯域を概ね 4 帯域に分割した帯域包絡線を用いて了解性のある音声を実現できることを示した。Houtgast[8][9] は狭帯域包絡線の変調度によって, 室内の反射音や雑音による了解度評価が可能であることを報告している。

聴覚モデルにおいて音響信号は蝸牛の周波数選択性から狭帯域信号に分割される。分割された狭帯域信号は有毛細胞モデルにより半波整流後にローパスフィルターを施した狭帯域包絡線を得ると考えられている。Meddis 等 [13] は聴覚モデルに基づき, 狭帯域包絡線の自己相関関数を用いた Summary Auto-Correlation Function (SACF) による聴覚脳幹の周期性検出に基づく Pitch 知覚モデルを提案した。近年では風間等 [14] が音声の了解性には位相特性の及ぼす効果が大きいことを明らかにした。音声の分析・合成における振幅・位相特性いずれかの優位性は分析時間長に依存するものとなり, 概ね 10(ms) 以下あるいは 200(ms) 以上の分析時間長ではいずれも位相特性が振幅特性に比べて重要な要因となる。本研究では, 上述した聴覚モデルを踏まえた時間・周波数分

析を行なうことで近距離音場での音の変化と知覚評価結果との対応を明らかにする。

1.2.3 Speech Privacy における情報マスキング効果

これまでに音のマスキング効果に着目した公共空間の設計指針の検討がされてきた [15]。近年、共有空間の音環境設計においては他人の会話が邪魔にならない空間を目指した Speech Privacy, さらに個人情報の保護を目的とした、会話の秘話性を追求する Speech Security が重視されている。

共有空間内での受聴者周囲における会話相手の音声 (Target) の音圧レベルは、天井や家具、照明設備、壁、窓等の設備によって変化する。一方、空間内の受聴点によっては空調設備等から発せられる暗騒音が Target となる音声のマスキングに寄与する。ASTM E1130 では、共有空間内の受聴者の耳の周囲の音圧レベル SNR(dB) を測定し、標準明瞭度指数 (AI: Articulation Index) を用いて Speech Privacy を評価している [16]。つまり、これまでの Speech Privacy 評価を取り入れた音環境設計においては会話の秘話性ではなく、オフィス等の音環境評価を目的とされている。会話音声を Target とすると、マスキングには受聴点での音圧レベルに依存するエネルギーマスキング効果と Target の会話情報を覆う情報マスキング効果にわけられる [17][18][19]。Speech Privacy, Speech Security において効率良く会話情報を保護するためには、情報マスキング効果の考慮が重要である。そこで本論文では Speech Privacy を制御するための情報マスキングシステムの設計ならびに、時間とともに変化する Target とマスキング音の時間・周波数特徴に着目したマスキング音のマスキング効率評価方法を開拓する。

1.3 本論文の構成

第1章では、序論として本研究の背景及び目的・意義について述べる。

第2章では、残響室ならびに無響室において音源・受音点間距離 (Sound Source Distance: SSD(m)) を変えたときの室内伝達関数の測定方法、並びに測定結果を示す。測定点は音源近傍の 0.01(m) から 2.56(m) までの残響室において推測された直接音領域から残響音場領域までとし、測定をおこなった。そして各測定点における、時間波形並びに振幅・周波数特性を図示することによって SSD(m) の増大によるスペクトルの山谷の変化を確認する。さらに振幅・周波数特性の標準偏差 (dB) を計算し、SSD(m) の増大によるスペクトル変化との対応を明らかにする。また、残響音場での音の知覚に対応するとされる音場評価指標の一つである D 値、ならびに時間重心値に着目し、近距離音場の Coherent 領域での従来音場評価を示す。

第3章では、第2章で測定した残響インパルス応答を音楽信号に畳み込むことによって作成した仮想距離別音楽信号を用いた主観評価実験を示す。音の知覚評価には、Thurstone の一対比較法 (Case V) を用いる。距離が異なった対の刺激音を評価することで SSD(m) の増大によって変化する音の大きさ感を尺度化する。

第4章ではまず、音の最適な時間遅れと対応するとされる自己相関関数 (Auto-Correlation Function: ACF) に着目する。第3章における主観評価実験で用いた距離別音楽信号の短時間フレーム処理を取り入れた短時間自己相関分析を行ない、SSD(m) の増大に伴う ACF の変化を視覚化・評価する。

また残響音が付加された楽音波形のパワースペクトルから際立ったピークを形成する周波数成分を選出 (Peak-Picking) し、ピーク周波数列の基本周波数と高調波成分を推定することによって、音の微細な時間・周波数分析 (SPEC² 分析) を試みる。

第5章では音声の主観的明瞭性に関する評価実験ならびに、音声の主要帯域における包絡線スペクトル分析に着目する。近距離音場における音声の主観的明瞭性と、主要帯域における狭帯域包絡線スペクトルエネルギーとの対応を示す。

第6章では、共有空間の Speech Privacy を求めた音環境設計において、会話音声の情報を聞こえ難くするマスキング音に関する音声明瞭度評価実験ならびに時間・周波数特徴に着目したマスキング効率の評価方法を検討する。

会話音声の情報を遮る現象に情報マスキング効果がある [20][21]。白色雑音と、音声了解度を減少させる情報マスキング効果が期待できる逆転音声を用い、音声単文章の書き取りによる単語了解度評価実験をおこなう。実験では Target とマスキング音の信号のエネルギー比を TM 比 (dB) とし、エネルギー変化とマスキング音の違いによる了解度の差を示す。また、TM 比 (dB) ごとに Target とマスキング音を合成した信号の基本周波数分析を行なうことにより、Target の原信号と TM 比 (dB) ごとの合成信号のスペクトルの類似性を評価することによって、了解度評価との対応を明らかにする。

また付録 A では室内空間を伝わる音声をマスクするマスキングシステムに適した音声の狭帯域包絡線特徴を含むマスキング音の作成とその評価方法に関する基本原理、アルゴリズムを開拓する。また環境特徴を取り入れた、マスキングシステム設計の検討をおこなう。

第7章では、本論文の総括と今後の課題を述べる。

第2章 室内における近距離音場の インパルス応答測定と評価

2.1 はじめに

音コミュニケーション空間において、人々は意識せずとも音場特徴を理解し共有しながら目的となる音 (Target) を受け取り、会話をおこなっている。室内共有空間内では同一の室内においても話者間の距離が変わると Target の音の好ましきや大きき感、明瞭さなど受聴者の知覚に変化が起こる。本章では、Target からの距離が変わった時に受聴者が知覚する音場特徴の物理的变化を明らかにするために、室内音場において音源・受音点間距離 (Sound Source Distance: SSD(m)) を変えたインパルス応答の測定とその物理的分析を行なった。

室内に設置された音源が音を発生すると、室内で観測される音の大ききはやがて一定となって定常状態に達する。そこで音源を停止すると、室内で観測される音の大ききは時間とともに減少しやがて消滅する。この音源を停止した後に観測される音の減衰過程を室内残響過程という。室内残響過程は音源から衝撃音 (インパルスと呼ぶ) を発したとき、室内で観測される音のエネルギーの時間変化で表すことができる。インパルス音源信号に対する室内音圧の時間変化を室内音場の (音圧) インパルス応答という。室内音場のインパルス応答を確率過程とみなすとき、室内音場の統計論的取り扱いが可能となる。

波動理論によると、室内音場において音源から放射された音波が球面波で伝わるとされる領域 (Coherent 領域) の存在が Morse と Bolt によって予想されている [2]。近年、高橋等 [3][4][5] によって実測した残響室インパルス応答か

ら最小位相・位相周波数特性の局所的変化の回帰分析によって、音源からの距離に伴う伝達位相変化を実証し、その結果伝達位相変化の観測できる範囲が Coherent 領域と概ね一致することを確認した。上述した結果を踏まえ、本論文では音源近傍から Coherent 領域外になるまで SSD(m) を変えて残響インパルス応答を測定した。

一般的に SSD(m) が変化することによって、様々な知覚が変化する。残響音場において、音源からの距離が遠い録音では残響音のエネルギーが高く、音場特徴が多く含まれる一方で音の明瞭性が損なわれる。反対に音源に最も近接した録音は直接音のエネルギーが高く音の明瞭度が高くなる一方、音場特徴は損なわれる。SSD(m) 変化にともなう知覚の変化では直接音と反射音のエネルギー比 (D/R) が重要である。そこで本章では、直接音と反射音のエネルギー比を表す (D 値) や時間重心に着目し、従来用いられてきた主観量に対応するとされる音場評価指標を用いて物理的評価を行なった。

室内残響音場における約 30ms 以内の初期反射音は音源の直接音エネルギーを増強し音声の明瞭性向上に貢献し音質にも影響を与える [25][26]。直接音のエネルギーが大きい音源近傍での収録においては、初期反射音が音楽の音質に最も影響を与える。安藤等 [27] は単一の初期反射音の最適な遅れ時間と音源信号の自己相関関数 (Auto-Correlation Function: ACF) の継続時間との対応を明らかにしている。そこで本章では複数の反射音からなる初期反射成分の特徴が重要であると考え、室内で実測した残響インパルス応答の初期反射区間の振幅・周波数分析を行い、SSD(m) の増大に伴ったスペクトル変化を示した。

2.2 室内伝達関数の測定

高橋等 [3][4] は実測した残響室インパルス応答から最小位相・位相周波数特性の局所的変化の回帰分析によって、音源からの距離に伴う伝達位相変化を実証している。本論文では、音源特性を含めた主観評価実験および伝達系の評価

を行うために、残響室に加えて無響室内においても SSD(m) を変えてインパルス応答を測定した。測定は工学院大学八王子校舎の9号館内の残響室ならびに無響室にておこなった。図 2.1 は実測した際の測定機器の設置例である。イン

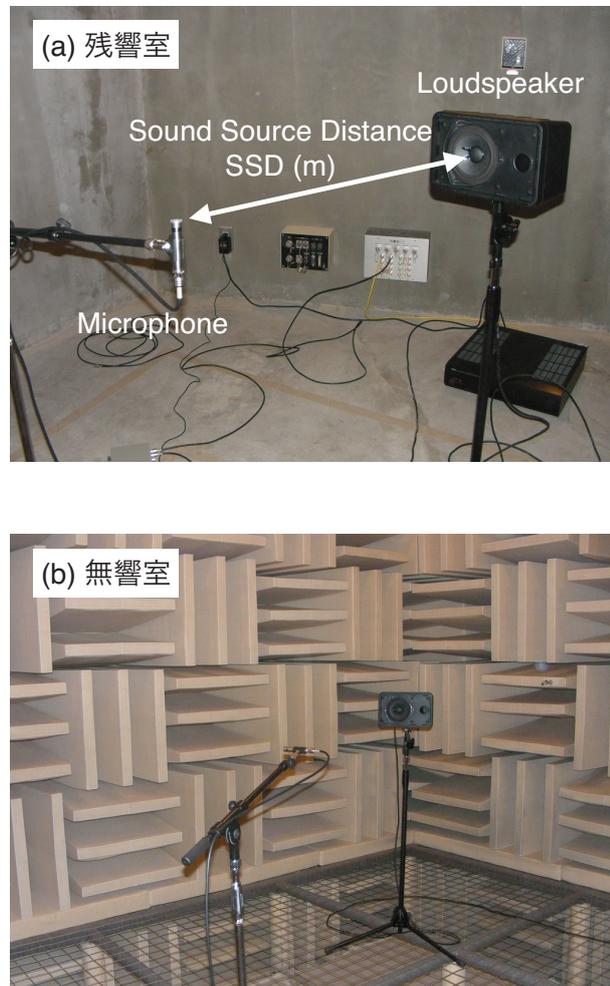


図 2.1: (a) 残響室, (b) 無響室におけるインパルス応答の測定環境例

パルス応答測定機器の構成は図 2.2 に示す。

図 2.3 は残響室にて測定した SSD(m) ごとのインパルス応答の時間波形である。振幅は最大値で正規化して表示してある。SSD(m) が増大するにつれ残響のエネルギーが上昇することがわかる。2.56(m) でのインパルス応答では実応

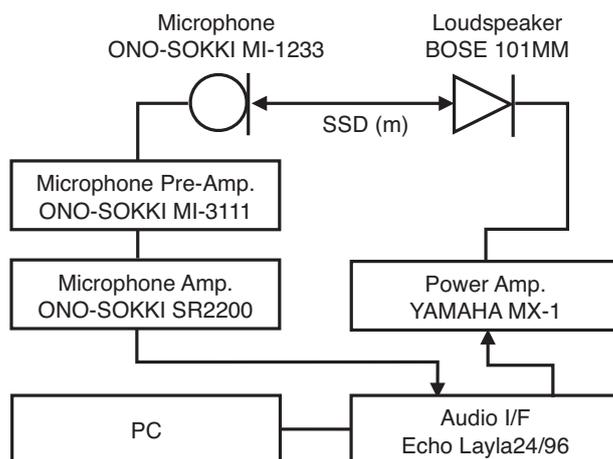


図 2.2: インパルス応答測定機器の構成

答の時間波形から、壁面からの反射音が含まれていることがわかる。また本論文で測定したインパルス応答は後続する雑音の影響を受けずに十分な SN 比 (dB) を確保するため、全て指数窓によって 1(s) に減衰させている。

図 2.4 は残響室にて測定した SSD(m) ごとのインパルス応答の振幅・周波数特性である。SSD(m) が増大するにつれスペクトルの山谷の分散が大きくなることからわかる。

図 2.5 は無響室にて測定した SSD(m) ごとのインパルス応答の時間波形である。残響室内に比べて SSD(m) の増大にともなった時間波形の変化は僅かである。また図 2.6 は無響室にて測定した SSD(m) ごとのインパルス応答の振幅・周波数特性である。残響室内のインパルス応答に比べ残響のエネルギーの付加がなく、SSD(m) の増大によるスペクトルの差は見られない。

2.3 インパルス応答とエネルギー減衰

室内に白色雑音を発生する音源があるとする。時刻 $n = 0$ において音源が停止したとき、音場内の 1 点で観測される受信信号 $y(n)$ は

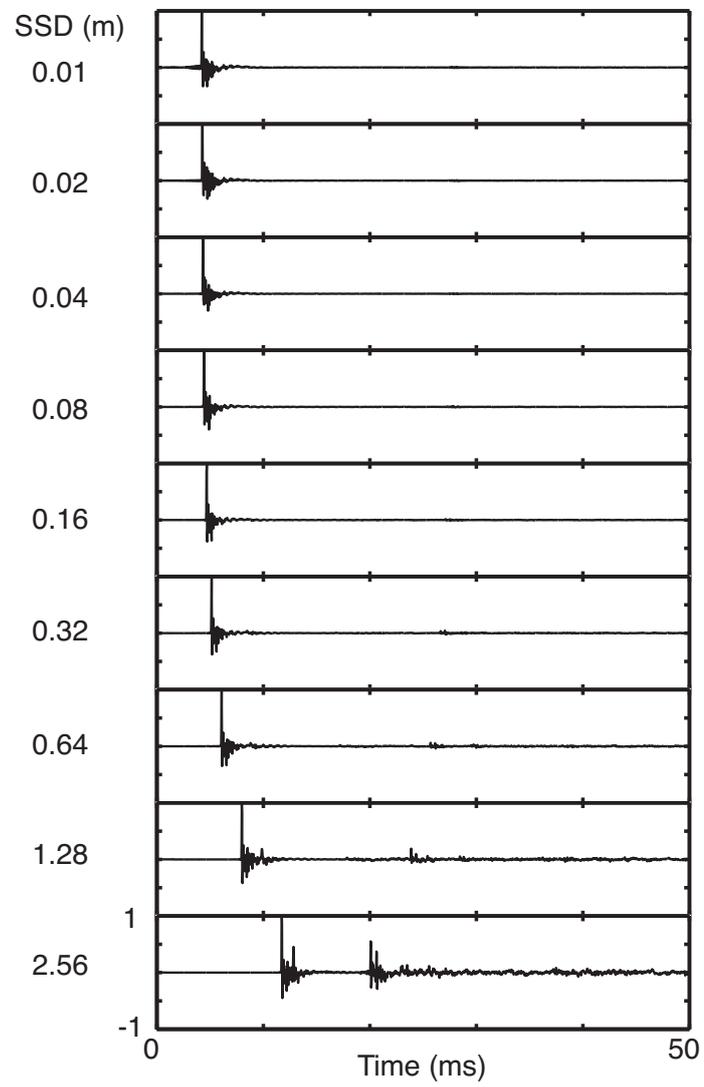


図 2.3: 残響室内における距離別インパルス応答の時間波形

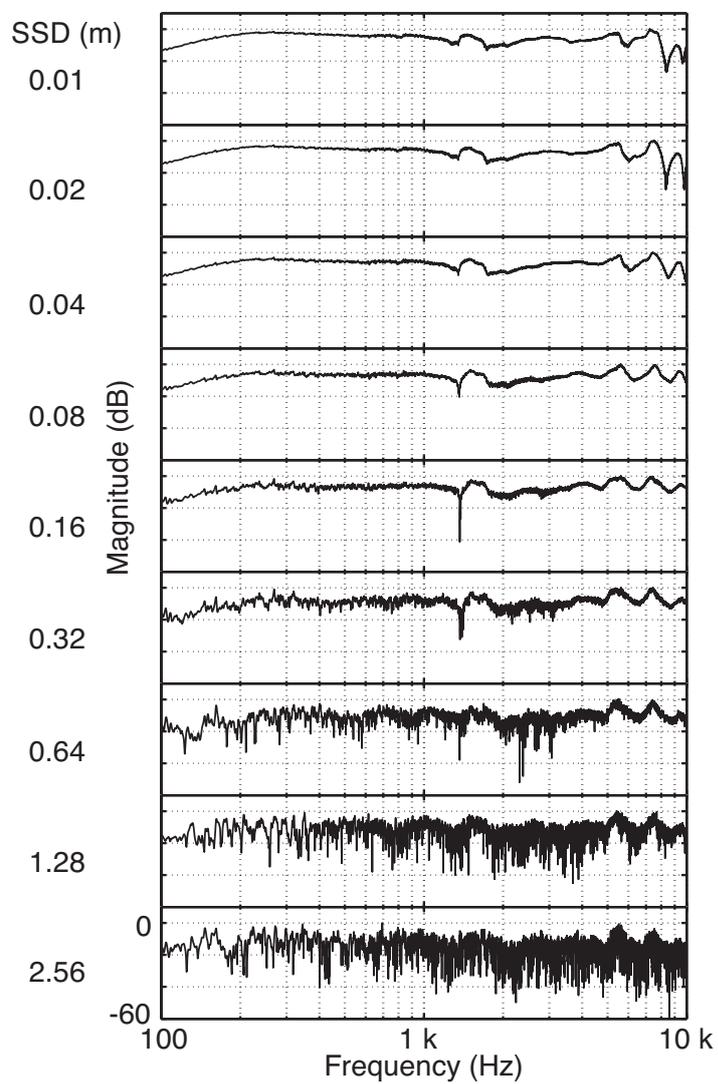


図 2.4: 残響室内における距離別インパルス応答の振幅・周波数特性

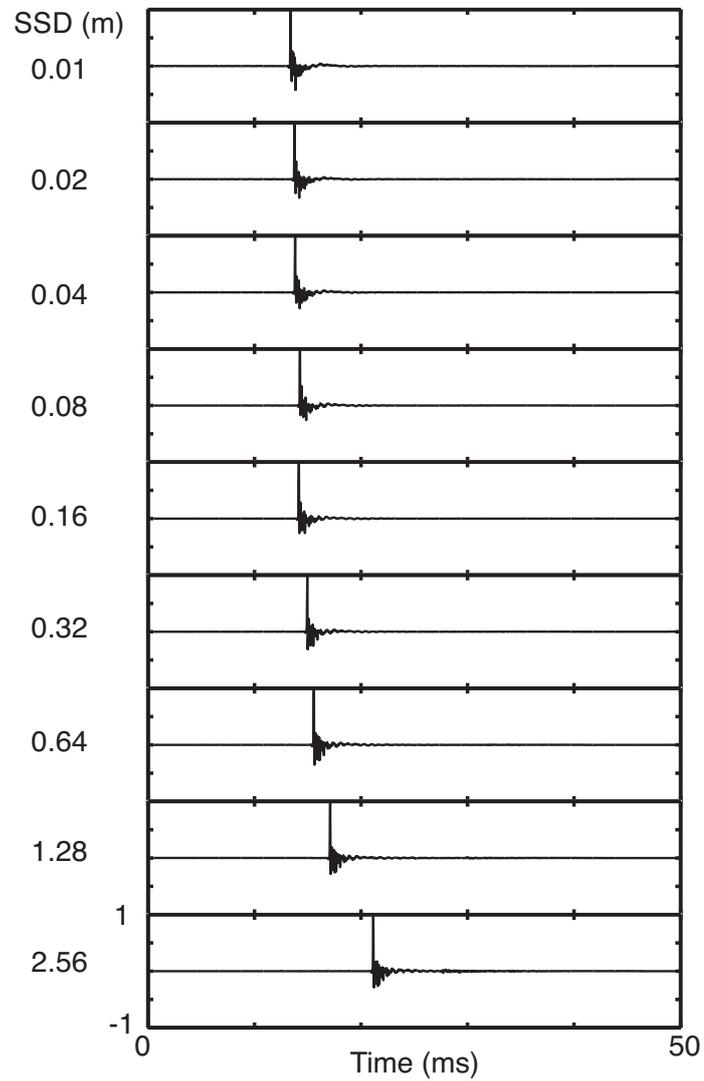


図 2.5: 無響室内における距離別インパルス応答の時間波形

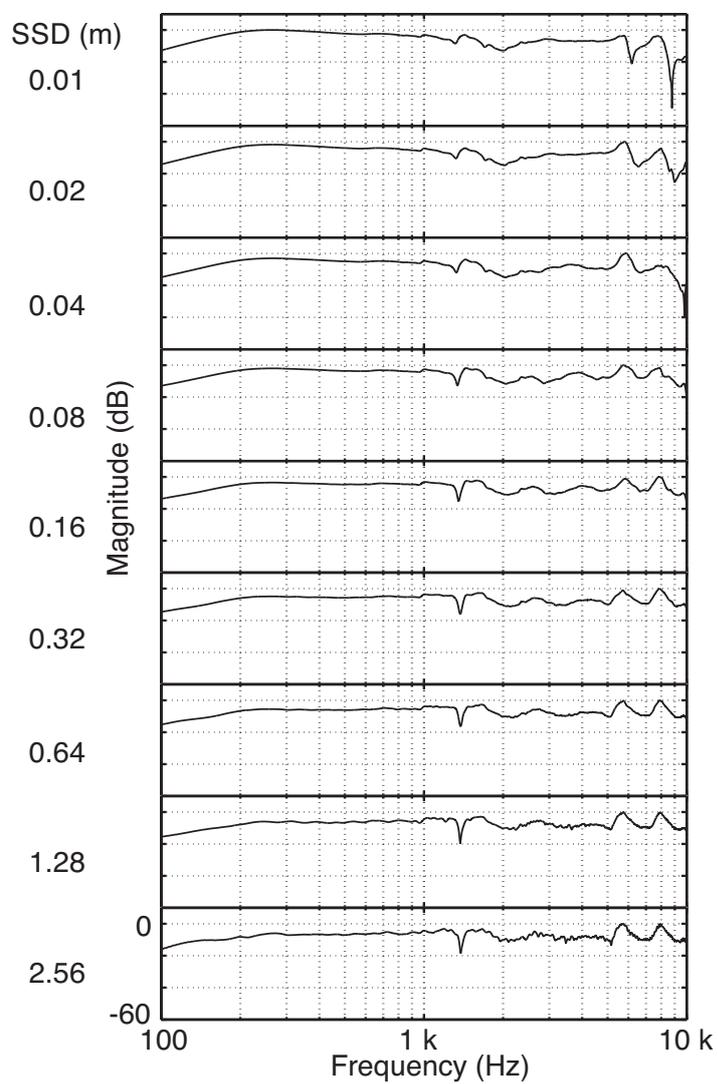


図 2.6: 無響室内における距離別インパルス応答の振幅・周波数特性

$$y(n) = x(n) * h(n) = \sum_{m=n}^{N-1} x(n-m)h(m) \quad (2.1)$$

のとおり畳み込みで表される。ここで $h(n)$ は長さ N を持つ音源・受音点間のインパルス応答である。

音源信号が白色雑音信号であることを考慮して上記の出力信号 $y(n)$ の自乗期待値を求めれば

$$\begin{aligned} E[y^2(n)] &= \sum_{l=n}^{N-1} \sum_{m=n}^{N-1} E[x(n-m)x(n-l)]h(m)h(l) \\ &= K \sum_{i=n}^{N-1} h^2(i) \end{aligned} \quad (2.2)$$

が得られる。ここで K は白色雑音の自乗期待値である。不規則信号 $x(n)$ が

$$E[x(m)x(l)] = \begin{cases} 0 & m \neq l \\ K & m = l \end{cases} \quad (2.3)$$

を満足するとき、信号 $x(n)$ を白色雑音という。 $E[X]$ は確率変数 X の期待値を表す。白色雑音を放射する音源の停止後に受音点にて観測されるエネルギー減衰過程の期待値は、音源・受音点間のインパルス応答の自乗積分で表される [22][23]

図 2.7 は、残響室内における距離別インパルス応答の減衰過程を表した残響曲線 (Schroeder Curve) である。縦軸は音圧 (dB)、横軸は時間 (ms) である。SSD(m) が増大するにつれ、残響曲線の傾斜が緩やかになり、残響のエネルギーが多くなる様子が確認できる。

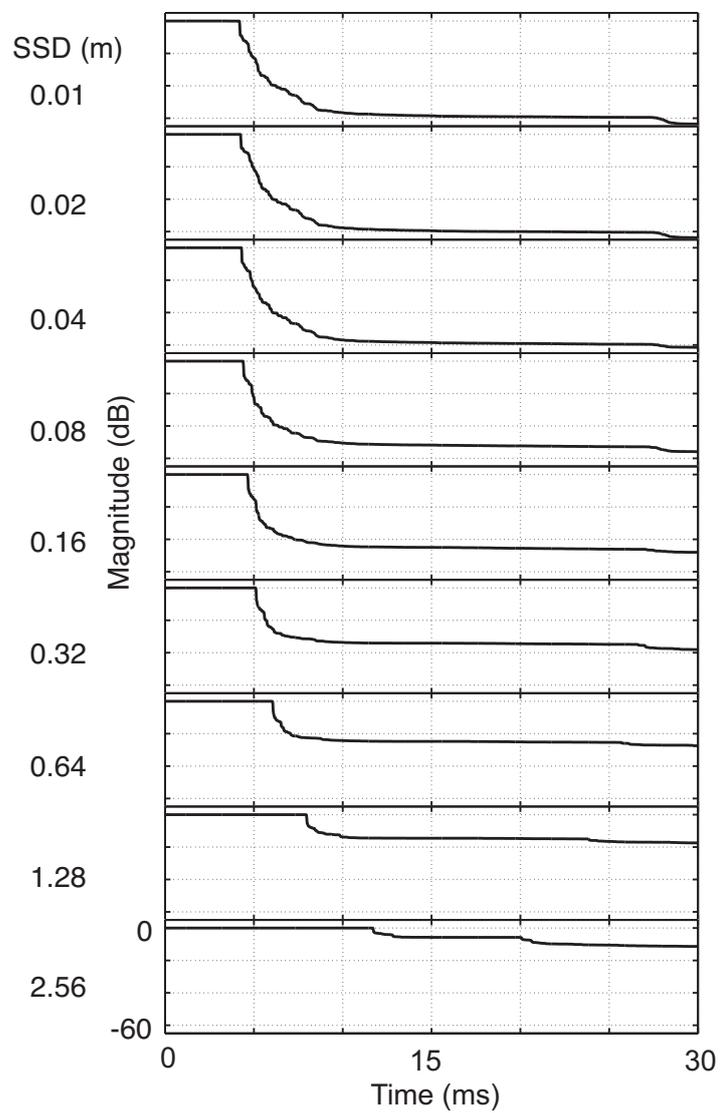


図 2.7: 残響室内における距離別インパルス応答の残響曲線

2.4 Coherent 領域

受音点を音源の近くに置くと、音源近傍の音場は自由音場とほぼ同様であると考えることができる。音場において球面波の伝達が観測可能な範囲は Coherent 領域 (直接音領域) と呼ばれている。Coherent 領域の存在は 1944 年に Morse と Bolt によって予想された [2][24]。しかしその存在は、60 年にわたって実証されることは無かった。

直接音に従う Coherent 領域の条件である音源からの距離 R_c は

$$R_c = \sqrt{\frac{V\alpha}{4\pi}} \cong \sqrt{\frac{V\beta_0}{8}} \cong \sqrt{\frac{6.8V}{8cT_R}} = \sqrt{\frac{A}{64}} \quad (2.4)$$

のように計算される [28]。ただし、 $A \cong 0.161V/T_R$ (m^2) である。

ここで本残響室における Coherent 領域を算出する。本測定条件は残響時間 1(s)、室容量 $V = 183(\text{m}^3)$ であり、上記の Coherent 領域の限界は

$$R_c = \sqrt{\frac{0.161 * 183}{64 * 1}} \cong 0.68 \quad (\text{m}) \quad (2.5)$$

となる。

2.5 残響インパルス応答による室内音場評価

室内の伝達関数は音場の特徴を含んでいる。測定したインパルス応答を分析することによって音場特徴を抽出することができる。

一般に室内の残響の長さを知る基本的指標として、残響時間が用いられる。音源停止後、受音点信号エネルギーが 60(dB) 減衰するに要する時間を残響時間 ($RT_{60}(\text{s})$) と定義される。図 2.8 は SSD(m) ごとの残響室内のインパルス応答の残響曲線において、60(dB) 減衰した時間から求めた残響時間 $RT_{60}(\text{s})$ である。残響室内の距離別インパルス応答を 1/4oct. のフィルターバンクを用いて分析し、125-8k(Hz) の範囲で 1/1oct. バンド間隔で表示してある。SSD(m)

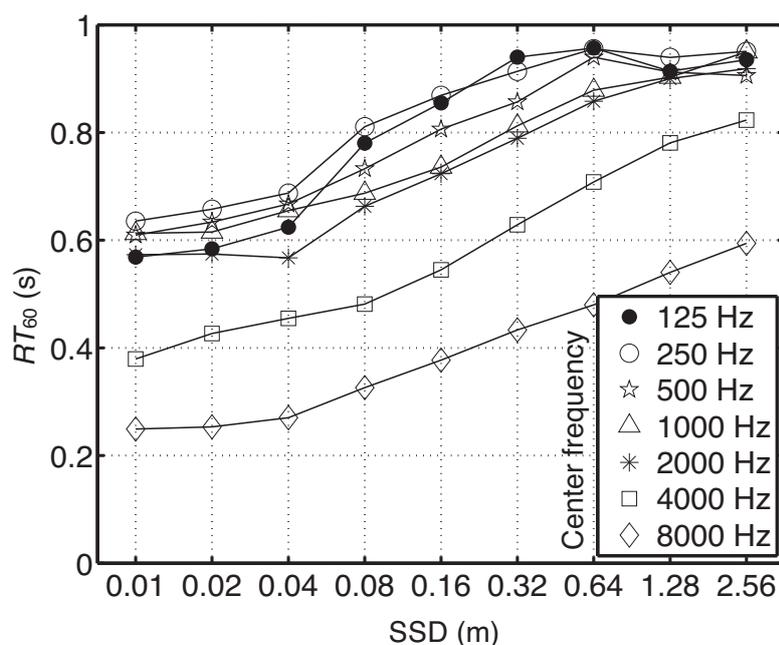


図 2.8: 残響室内における残響時間 RT_{60} (s)

の増加に伴って、残響時間も上昇していき、125–2k(Hz) では Coherent 領域の 0.68(m) を超えると一定になっていく様子が確認できる。

2.5.1 直接音と残響音のエネルギー比 D_{30} 値

次に主観量と対応するとされる従来音場評価方法として、直接音と残響音のエネルギー比 (D/R) に着目する。一般的に直接音と残響音のエネルギー比は、会話上で Target となる音声の明瞭度を評価する尺度として用いられる [30][31]。SSD(m) に伴った音場知覚には D 値に代表される直接音と残響音のエネルギー比が Cue の一つであるとされている [32]。室内音場設計指標の一つとして、受聴点において直接音から遅れて 30–50(ms) に到達する初期反射音の仮定を基にした定義 (D 値 (Deutlichkeit)) が用いられる。一般的にコンサートホールの設計指標において、 D_{50} が用いられるが、近距離音場においては、初期反射音の

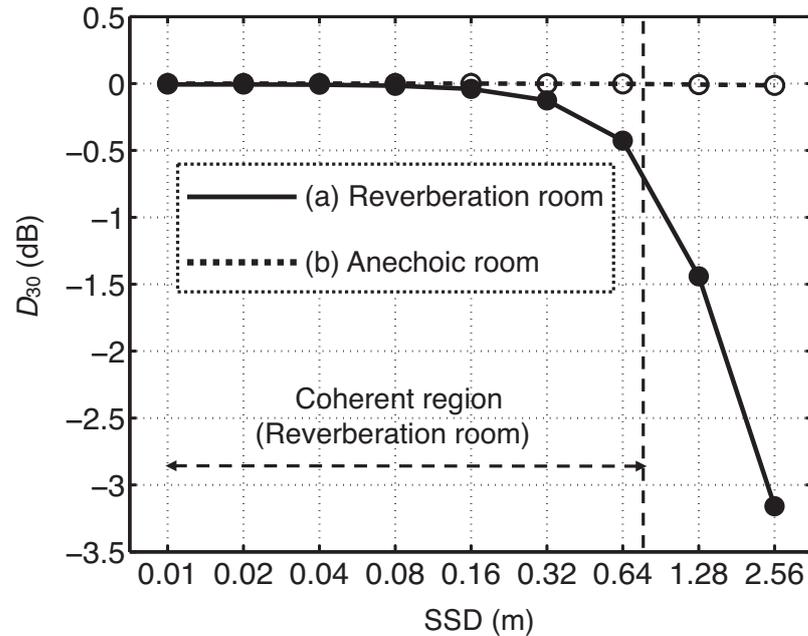


図 2.9: 距離別インパルス応答の D_{30} (dB)

エネルギーがより短い 30(ms) までに集中していると推測し, D_{30} を示すことにした. 時間を t , 測定した残響インパルス応答を $h(t)$ とすると D_{30} (dB) は

$$D_{30} = 10 \log_{10} \frac{\sum_{t=0}^{30(\text{ms})} h^2(t)}{\sum_{t=0}^{\infty} h^2(t)} \quad (\text{dB}) \quad (2.6)$$

として表される. 図 2.9 は距離別インパルス応答の D_{30} (dB) である [29]. 残響室内では D_{30} は直接音領域内の 0.01–0.64(m) ではほぼ一定となり, Coherent 領域を超えると下降する様子がわかる.

2.5.2 時間重心

D_{30} (dB) に加え, Cremer[33] が提案した空間を評価する音響物理指標である時間重心を挙げる. 時間重心値は, 主観的に捉えられる残響の量と対応すると

されている。時間重心 T_s (s) は

$$T_s = \frac{\sum_{t=0}^{\infty} t \cdot h^2(t)}{\sum_{t=0}^{\infty} h^2(t)} \quad (\text{s}) \quad (2.7)$$

として表される。図 2.10 は SSD の増大に伴う各インパルス応答の時間重心値

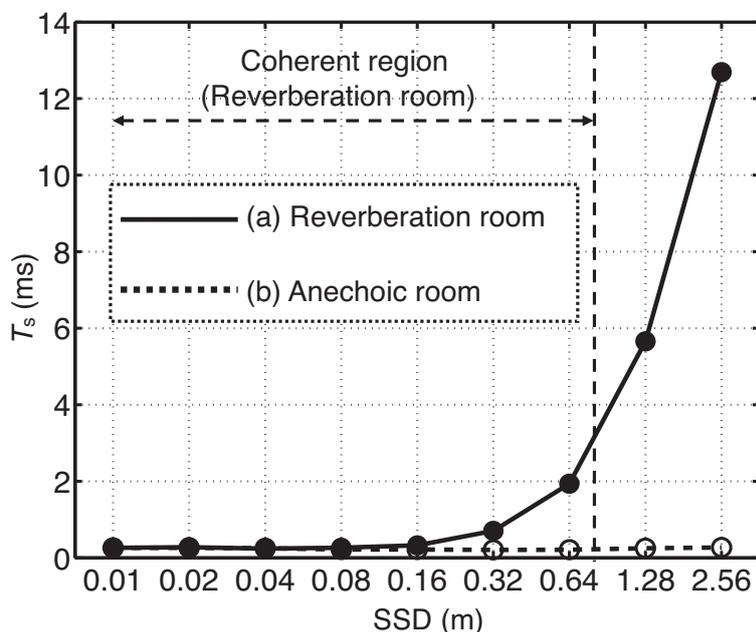


図 2.10: 距離別インパルス応答の時間重心値

T_s (ms) の変化を示す。縦軸は対数表示した時間重心値 T_s (ms)、横軸は SSD(m) である。無響室内ではどの距離でもほぼ一定の値を示している。一方で残響室においては、SSD:0.16(m) 以降では時間重心値が増大していくことがわかる。

2.5.3 振幅周波数特性の標準偏差

図 2.4 に見られるように、SSD(m) の増大によってスペクトルの山谷の分散が大きくなる。そこで伝達系の振幅周波数特性の標準偏差 [34] に着目する。 H_r を SSD: r (m) におけるスペクトル列、 N を点数、 k は周波数 (100-8k(Hz)) とす

ると振幅周波数特性の標準偏差 $\sigma_r(k)$ (dB) は

$$\sigma_r(k) = \sqrt{\frac{1}{N} \sum_{k=1}^N \left\{ H_r(k) - \overline{H_r(k)} - H_1(k) \right\}^2} \quad (\text{dB}) \quad (2.8)$$

と表される。図 2.11 は SSD(m) の増大に伴った振幅周波数特性の標準偏差

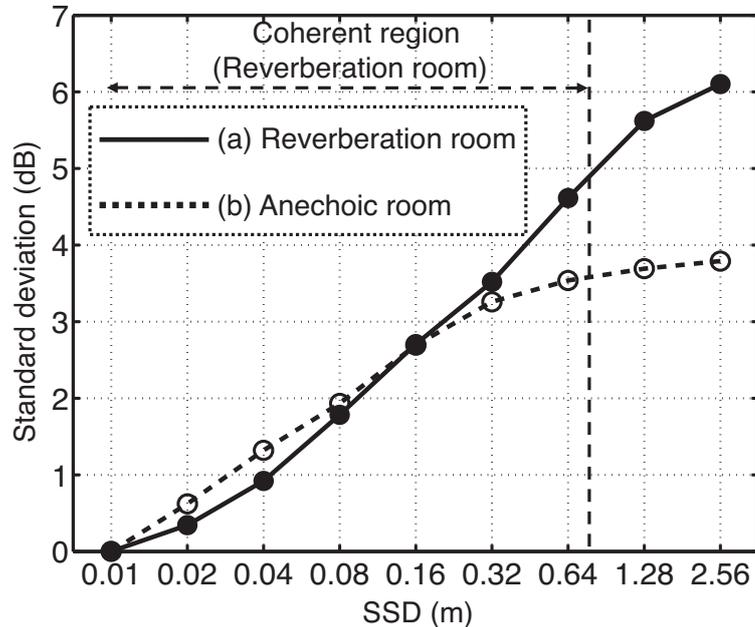


図 2.11: 距離別インパルス応答の振幅・周波数特性の標準偏差

$\sigma_r(k)$ (dB) である。残響音場，無響室ともに Coherent 領域内の SSD: 0.16(m) までは無響室での標準偏差値が 3(dB) まで上昇するが 3(dB) を超えると無響室ではほぼ一定となっていき，残響室における標準偏差値はさらに上昇していく様子がわかる。また，残響音場理論において残響室内での振幅周波数特性の標準偏差は 5.5(dB) で一定となるとされ，本残響室内の Coherent 領域外の結果から確認することができる [34].

2.5.4 初期反射成分の分析

直接音のエネルギーが大きい音源近傍での収録においては、初期反射音が音楽の音質に最も影響を与える。約 30(ms) 以内の初期反射音は音源の直接音エネルギーを増強し音声の明瞭性向上に貢献し音質にも影響を与える [25][26]。そこで本章では、測定した残響インパルス応答の 30(ms) までの区間に含まれる初期反射成分の分析をおこなう。

図 2.12 は 30(ms) までの初期反射区間における残響室内のインパルス応答の振幅・周波数特性である。破線は音源近傍の SSD: 0.01(m) における振幅・周波数特性であり、0.02(m) 以降は実線で表示した。初期反射成分であっても、SSD(m) の増大に伴ってスペクトルの山谷が増加する。

図 2.13 は図 2.11 と同様に残響室内のインパルス応答の 30(ms) までの初期反射成分に関して振幅・周波数特性の標準偏差 (dB) を計算した例である。SSD(m) が 0.16(m) までは増加していき、一度 Coherent 領域内で一定になった後、Coherent 領域を超えると増加することが示された。上述した事から Coherent 領域内の初期反射成分のスペクトル特徴は原音の構造を保ったまま僅かに豊かになっていき、Coherent 領域を超えて残響のエネルギーが付加されることによってスペクトル構造が崩れる事が明らかになった。

2.6 まとめ

本章では音源・受音点距離 (SSD(m)) が変わったことによる音場特徴の評価を行うために、残響室と無響室において SSD(m) を変えてインパルス応答を測定した。Coherent 領域内において SSD(m) に伴うインパルス応答から、従来の音場評価指標を用いて評価した。 D_{30} 値や時間重心のような一般的に用いられる主観量と対応するとされる音場評価指標では Coherent 領域内の物理特性の微細な変化を表す事ができなかった。一方、図 2.4 ならびに図 2.11 に見られ

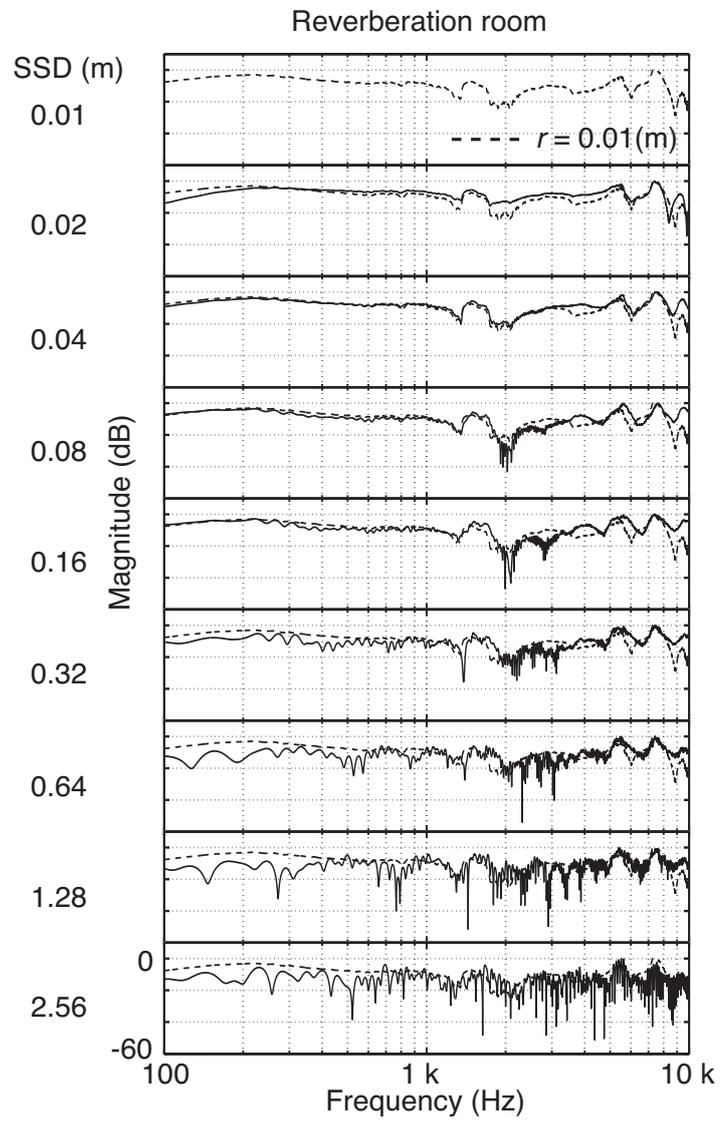


図 2.12: 初期反射成分の振幅・周波数特性

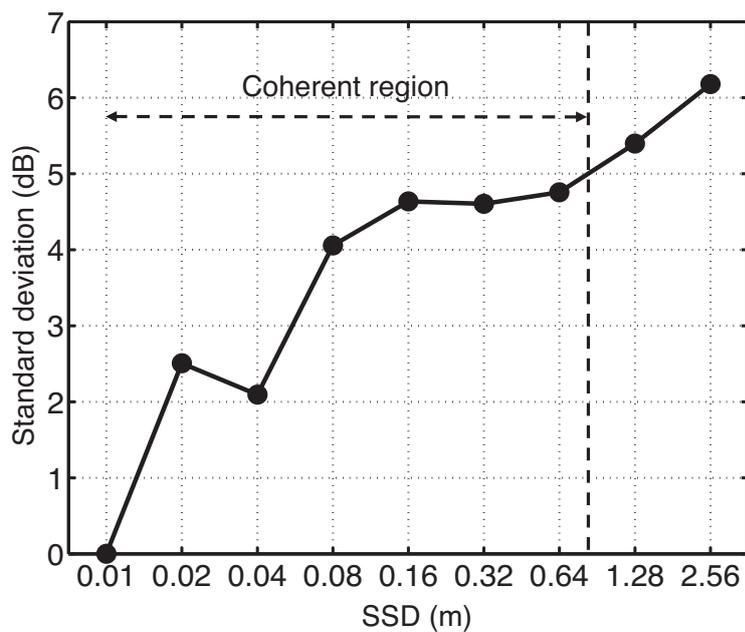


図 2.13: 初期反射成分の振幅・周波数特性の標準偏差

る通り、スペクトルの山谷は SSD(m) の増加に伴って大きくなり、分散も増えていくことがわかる。次章では Coherent 領域内における SSD(m) の増大に伴う知覚変化に着目する。そこで本章で測定・評価したインパルス応答を畳み込んだ音楽モチーフを用いて、主観的な音の好ましさと音の大きさ感の主観評価実験を行なう。

第3章 近距離音場における主観評価実験

3.1 はじめに

共有空間における音コミュニケーションにおいて、人は Target となる音を快適に受聴できるように常に意識が働いていることを経験的に知っている。たとえば、室内において話者に耳を傾けたり、Target に近づくことによって音の情報を効率良く受け取っている。そこで本論文では室内の近距離音場において音源・受音点間距離 (SSD(m)) によって変化する音の物理的変化と心理的変化の関係に着目した。

好ましい受聴環境を設定するには、室内近距離音場での最適な SSD(m) を見つける必要がある。そこで本論文では残響室内における音源近傍の Coherent 領域内にて録音した音楽モチーフを用いた主観評価実験を行った。Zahorik[35] は仮想的に設定した距離別信号を用いて、距離感知覚に関する評価を行なっている。安藤等 [36] は単一反射音場において音楽モチーフを用いた反射音の最適な遅れ時間と Preference の関係を受聴実験によって明らかにしている。そこで本論文では実測した残響インパルス応答を用いて、SSD(m) の増大に伴った音の好ましさ (主観的 Preference) 知覚の変化を一対比較法により評価することにより尺度化し、近距離音場の音コミュニケーションにとって最適な距離を明らかにする。被験者数はヘッドホン (Diotic) 受聴による一対評価実験により SSD(m) が異なった試験信号対を評価する。評価方法は Thurstone による比較判断の法則のにおける Case V を用い、主観評価結果を尺度化した [39][40]。

音の大きさ感は音の知覚にとって明白に違いを聞き分けることのできる音響属性である。しかし音の大きさ感を量的に表すことは、これまでに数多く検討され規格化が行われてきたが[50][51]、全ての音響信号に対し統一的に一つのパラメータとして尺度化を行うことは困難である。そこで本論文では音の知覚に重要な近距離音場における音の大きさ感 (Loudness)[52] の主観的な変化に着目した。近年では、録音作品製作における Loudness War と呼ばれる音の大きさ感とダイナミックレンジの問題[44] や放送における音量差問題を解決するための Loudness Meter とよばれる物理的 Loudness パラメータが国際的規格化[45][46][47] され、日本国内の放送業界においても運用が開始されており[48]、Loudness の主観的な違いと定量性評価について関心が高まっている。本論文では、近距離音場における音の大きさ感 (主観的 Loudness) 変化の尺度化とそれに対応する時間・周波数分析による評価方法の検討を試みた。本章では主観的 Preference 評価と同様に残響音場の Coherent 領域内で收音された音楽モチーフを Diotic 受聴することによって、Coherent 領域内における主観的 Loudness を一対比較法により評価する。測定したインパルス応答には、音場特性とともに音源としたスピーカ特性も含まれる。そこで残響室と同様に無響室内で距離別に測定したインパルス応答を同様の音楽モチーフに畳み込み、主観的 Loudness を評価した。

3.2 Thurstone Case V による一対比較評価方法

比較判断法の Thurstone Case V を基にした一対比較方法[37][38] は、一般に画像の視覚評価の尺度化などに多く用いられ、観測者の感知する心理的刺激量の定量化が試みられている。本論文では、SSD(m) の異なる距離別音の評価に一対比較法の Thurstone Case V を用いることによって心理的評価の尺度化をおこなった。

表 3.1: 一対比較法における比較行列 P

	S_a	S_b	S_c	\dots	S_j	\dots	S_n
S_a	$p_a > a$	$p_b > a$	$p_c > a$	\dots	$p_j > a$	\dots	$p_n > a$
S_b	$p_a > b$	$p_b > b$	$p_c > b$	\dots	$p_j > b$	\dots	$p_n > b$
S_c	$p_a > c$	$p_b > c$	$p_c > c$	\dots	$p_j > c$	\dots	$p_n > c$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
S_j	$p_a > j$	$p_b > j$	$p_c > j$	\dots	$p_j > j$	\dots	$p_n > j$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
S_n	$p_a > n$	$p_b > n$	$p_c > n$	\dots	$p_j > n$	\dots	$p_n > n$

一対比較法の典型的な場合では、心理的尺度によってその値を決めようとするすべての刺激は、初めそのあらゆる可能な組み合わせによる対の形で、観測者 O に提示される。 O はある一定の観点から、対のうちの一方が他方よりも量的に大きいかどうかを判断する。その判断は2範疇で行われ、観測者の思った通りに判断することが要求される。刺激には性質の似通ったものが用いられる。観測者の回答は、表面上では比較判断である。同一の観測者がすべての刺激対を、様々な場合において多数回判断したり、多数の似通った観測者が全ての刺激対をそれぞれただ1回だけ判断することもある。前者では、場合行列 (Occation Matrix) が作られ、後者では個人行列 (Indivisidual Matrix) が作られる。いずれの場合にも数量的な結果として各々の刺激について、それを他の各々の刺激と比較してその尺度上で上位にあると判断された回数と、その比率が求められる。これをもとにして表3.1に示されるような比率行列 P が作られる。比較行列の一般的形式、上欄の各刺激 S_j が左側の各刺激 S_k よりも大きいと判断された回数の比率 $P_{j < k}$ を示している。

尺度化の問題は行列 P から始まる。つまりこの知識に基づいて各刺激に1次元の尺度上の1個の値を与える。この尺度は、間隔尺度 (Interval Scale) 特性を備えたものであることが望ましい。

次にこのようなデータから尺度化を行う場合の原理を述べる。比較判断から

始まり、尺度化を行う場合の手がかりを Thurstone の比較判断の法則に見いだすことができる。刺激 $S_j(S_k)$ とそれぞれ特有のものとして与えられた心理学的平均値を $R_j(R_k)$ とおき、単位正規分布の平均からの標準速度距離 (偏差率) を z_{jk} 、分布 $R_{hj}(R_{hk})$ の標準偏差を $\delta_j(\delta_k)$ 、各分布間の相関係数を r_{jk} とすると

$$R_j - R_k = z_{jk} \sqrt{\delta_j^2 + \delta_k^2 - 2\gamma_{jk}\delta_j\delta_k} \quad (3.1)$$

と表される。根号項は、差 $R_{hj} - R_{hk}$ の標準偏差であり、根号中の各項の関数として表される。つまり、 $R_j - R_k$ それぞれの距離を表す尺度の単位である。上式の右辺がわかれば、各 $R_j - R_k$ の距離の大きさを決定することができる。 z_{jk} は実測比率 $p_{j>k}$ から求められる。

Thurstone は比較判断の法則の適用に関して、次の5つの”Case”を区別している。

- Case I.

3.1式のままの完全な形で、ただ一人の観測者の反復判断に法則を適用する場合。これは法則が、最初に考えられた理想的なケースである。この場合には未知の助変数に関する実測値を得ることを常に必要とする。

- Case II.

集団の場面に法則を適用する場合。この場合には多数の観測者が各刺激対に1回ずつの判断を行う。これにも Case I と同様な完全データを必要とする。

- Case III.

$\gamma_{jk} = 0$ と仮定する場合。どの刺激対に対する回答の間にもまったく相関がないと仮定する。この仮定は刺激間の一致を認めることができない場合に、特に許されるものであり、刺激が明らかにただ1つの面だけで変化するとすれば、刺激間の交互作用はほとんどないと考えられる。

刺激の相互相関が0となることによって、3.1式は次式のような形の簡略化された法則となる。

$$R_j - R_k = z_{jk} \sqrt{\gamma_j^2 + \gamma_k^2} \quad (3.2)$$

ここで、 γ_j^2 と γ_k^2 の相対的な値を推定することは可能であり、 z_{jk} は実験結果からこの式を解くことが可能となる。

- Case IV.

弁別のちらばりがほぼ等しいと仮定する場合である。この場合には Thurstone が示しているように、法則は次式のような形になる。

$$R_j - R_k = .707 z_{jk} \gamma_j + \gamma_k \quad (3.3)$$

どの刺激も同じ程度に、容易に尺度上に位置づけることができる場合には、弁別のちらばりは等しくなる。しかし評価対象が様々な特徴を含む刺激は多くの基準によって判断されることが可能になり、他のものよりも判断は困難である。視覚的に知覚された線分のような比較的客観的な刺激では、当面の問題に関連のある少数の特性を用意に分離することができるため、等しいちらばりを持つものと考えられる。また集団の判断の間では、単一個人の判断の中でも、等しくないちらばりを生じる可能性が多い。

Case IV を適用する場合にも、Case III の場合と同様に標準偏差を推定する。正確性の点では Case III が合理的であるとされる。

- Case V.

すべての解法のうちでもっとも簡単な解法をおこなうためには、さらにもう1つの仮定を追加しなければならない。その仮定は弁別のちらばりがすべて等しいということである。 $\gamma_j = \gamma_k$ とし、 γ_j で両方を表せば、法則は次のように簡約化される。

$$R_j - R_k = z_{jk} \gamma_j \sqrt{2} \quad (3.4)$$

γ_j を尺度の単位とすれば, 1 に等しくなり

$$R_j - R_k = z_{jk} \sqrt{2} \quad (3.5)$$

尺度の単位を $\gamma_j \sqrt{2}$ とすれば, その法則は

$$R_j - R_k = z_{jk} \quad (3.6)$$

のように表すことができる.

3.3 距離別音楽モチーフによる主観的Preference評価実験

本章では Coherent 領域内における異なる SSD(m) において録音された音楽モチーフの好ましき (Preference) の変化を一对評価法 (Thurstone V) によって尺度化する.

3.3.1 実験方法

図 3.1 は試験音の試験音作成手順である. まず, 9つの SSD(m) ごとのインパルス応答を音楽モチーフに畳み込み, 仮想距離別楽音を作成した. インパルス応答と音楽モチーフは, それぞれの信号のエネルギーによって正規化されている. また雑音を加えることによって知覚に対する遅い残響効果が取り除かれると予想し, SN 比: 10(dB) で SSD: 2.56(m) の残響インパルス応答を畳み込んだ白色雑音を付加した.

図 3.2 は試験音の構成である. 一つの試験信号には無音区間を隔てて, 2つの距離の異なる音楽モチーフが順に再生される. 音楽モチーフには無響室にて録音された楽曲 [49] の一部を使用し, バンドパスフィルタを通過させ 100-10k(Hz)

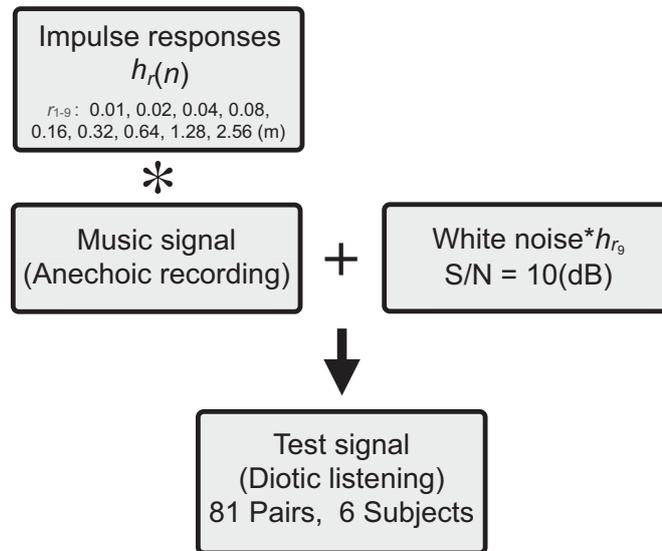


図 3.1: Preference 試験音作成手順

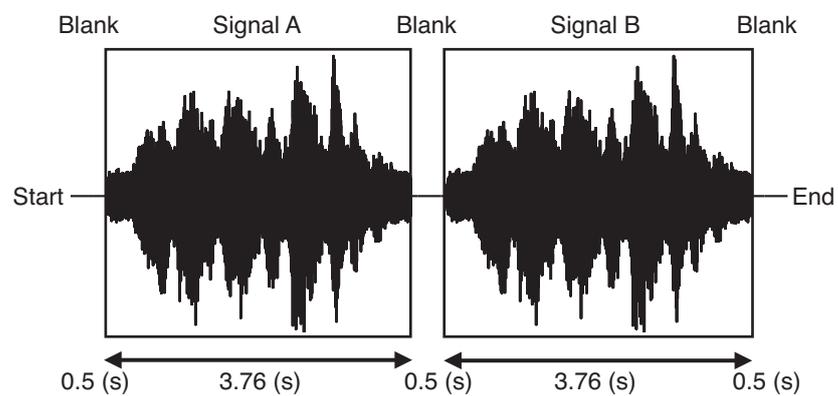


図 3.2: Preference 試験音構成

の帯域制限を行なった信号を用いた。試験音は全 81 対で、評価結果には含まない SSD(m) が同一の対も含まれている。試験音は 0.5 秒の無音区間を隔てて再生される音楽モチーフを受聴する。信号の立ち上(下)がり部はハニング窓の立ち上(下)がりを用いて、滑らかに再生されるようにした。被験者はすべて同様の再生装置、ヘッドホン (AKG K240 Studio) を通した Diotic 受聴によってどちらが好ましい音かを判断する。被験者数は特に訓練していない学生の 6 人で 20 から 40 代までである。

3.3.2 評価結果

図 3.3 は 2 種類の音楽モチーフを用いた Preference の評価結果 ((a):Water Music Suite No.6, (b):Symphony No. 4) である。横軸は SSD(m), 縦軸は主観値を表す心理尺度値 (Thurstone V) である [39]。Thurstone V による心理尺度化には信頼性評価に関する方法が定義されていない。実験結果の信頼区間推定は検討されているが、より多くの被験者数を得ることが必要となる [41][42]。そこで本論文における心理尺度の誤差評価には、モンテカルロシミュレーションによって被験者数と刺激数を関数とした実験式を用いて図中のエラーバーを推定した [43]。エラーバーに関する実験式は n を刺激数, N を被験者数とすると

$$\sigma_{obs} = b_1(n - b_2)^{b_3}(N - b_4)^{b_5} \quad (3.7)$$

ただし

$$b_1 = 1.76, b_2 = -3.08, b_3 = -0.613, b_4 = 2.55, b_5 = -0.491$$

として表される。図中破線は残響室内の Coherent 領域を表している。図 3.3(a) は音源近傍の Coherent 領域内においても Preference が変化することがわかる。特に SSD: 0.16(m) において最も高い主観値が得られた。また Coherent 領域を超えると急速に下降する。以上のことから、第 2 章、図 2.9, 図 2.10 で示した

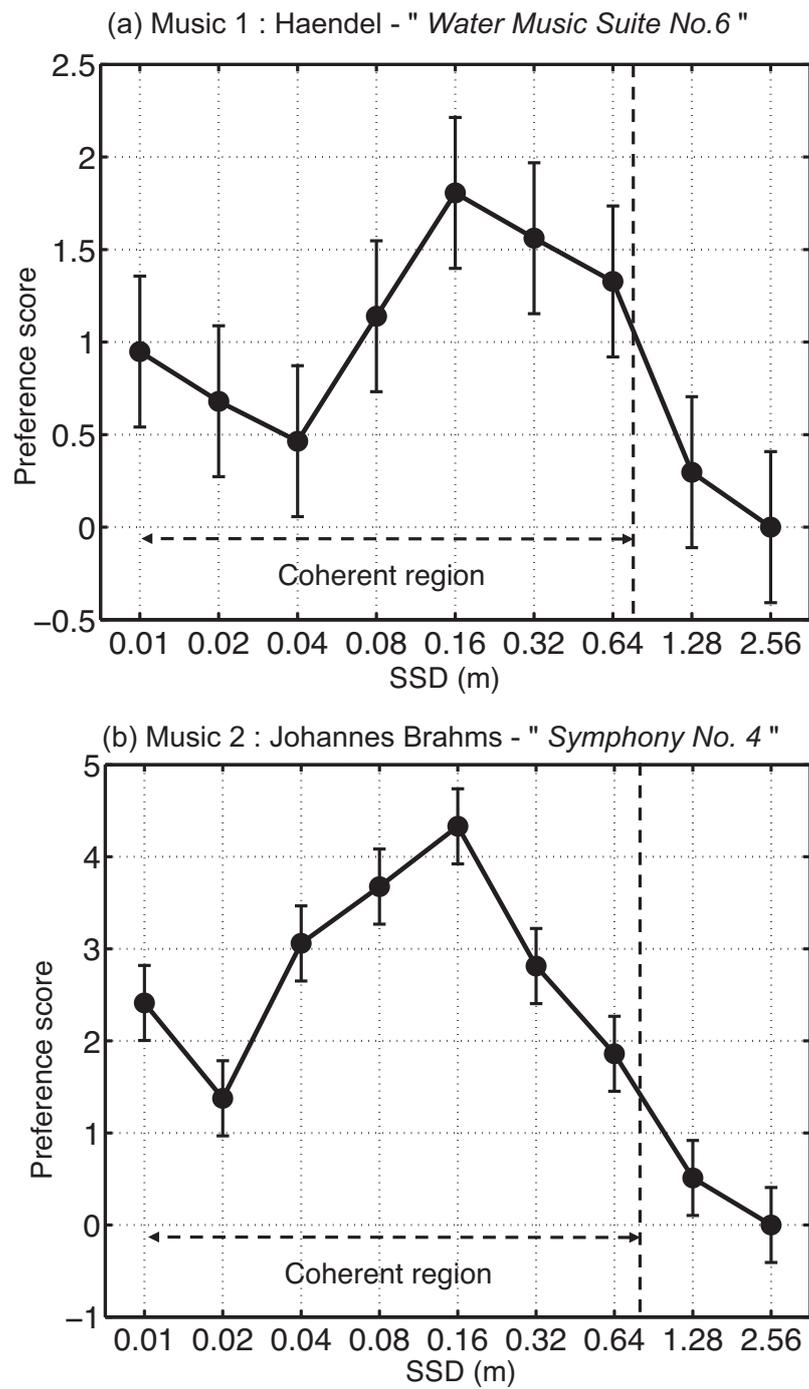


図 3.3: Preference 評価結果 (a):Water Music Suite No.6, (b):Symphony No.4

ような従来用いられてきた物理評価では Coherent 領域内での知覚変化を表すことができないことが明らかとなった。

3.4 距離別音楽モチーフによる主観的 Loudness 評価実験

Preference 評価と同様の一対比較による音の大きさ感 (主観的 Loudness) 評価試験を行った。

3.4.1 実験方法

図 3.4 は試験音を作成するための試験音作成手順、図 3.5 は試験音の構成

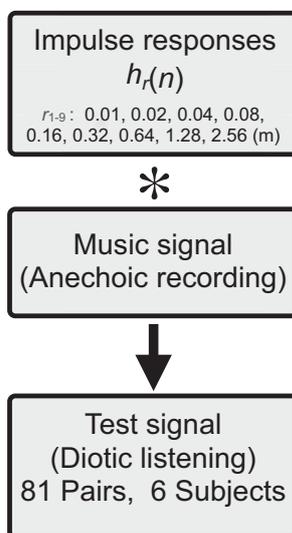


図 3.4: Loudness 試験音作成手順

である。前章における Preference 試験と同様に音楽モチーフには無響室録音された Haendel: Water Music Suite No.6 の演奏の一部を使用した。試験には SSD(m) の異なった 2 つの音楽モチーフの 81 組の信号を使用した。また本章で

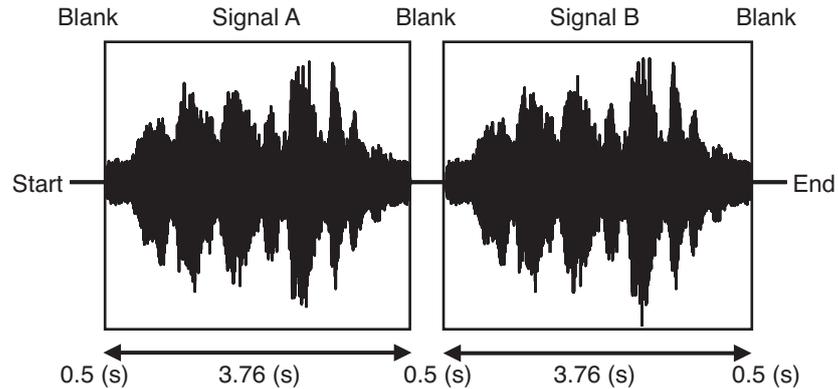


図 3.5: Loudness 試験音構成

は音源の特性を含めた評価を行うため、同様の音楽モチーフに無響室内で測定したインパルス応答を畳み込んだ試験信号を作成し、試聴実験を行った。

3.4.2 評価結果

図 3.6 は距離別音楽モチーフ ((a): 残響室, (b): 無響室) を用いた主観的 Loudness 評価結果である。(a) の残響室における Loudness は Preference 評価と同様に, Coherent 領域内の SSD: 0.16(m) において主観値が最大となり, 領域を超えると主観値が下降していくことが明らかになった。一方で, (b) の無響室においては SSD(m) の増大に比例して主観値も上昇していくことがわかる。したがって, Coherent 領域内の主観値の上昇はスピーカの物理的特徴等の音源特性変化による直接音成分の増強効果であると推測できる。また, Coherent 領域を超えた残響音場領域では残響エネルギーの付加により直接音成分が不明瞭になり, 主観値が低下していくことがわかった。

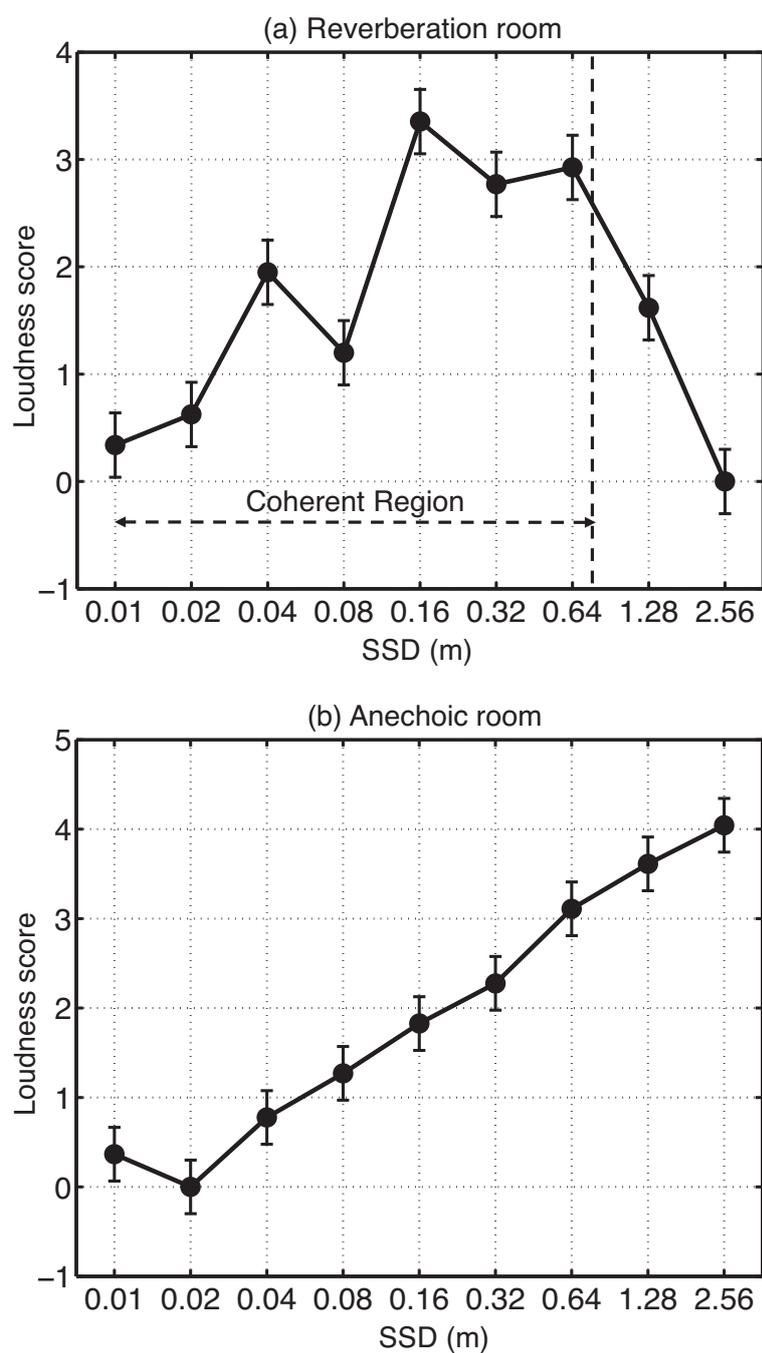


図 3.6: (a) 残響室ならびに (b) 無響室における距離別音楽モチーフによる主観的 Loudness 評価結果

3.5 まとめ

本章では音源・受音点間距離の変化に伴った知覚の変化を音の好ましきならびに音の大きさ感に関して、一対比較法を用いた受聴実験により尺度化した。音楽録音を考慮し、残響音場内における音源近傍の Coherent 領域の SSD(m) が極めて小さい場所から、残響エネルギーが増大する Coherent 領域外までの音楽モチーフに対する主観値の変化を明らかにした。Preference は Coherent 領域内で最大値まで上昇し、Coherent 領域を超えると SSD(m) の増大に伴って主観値が下降した。極めて音源近傍の 0.02(m) や 0.04(m) 付近の主観結果では、Preference が低くなることがわかる。これはスピーカ特性により主観評価に影響を与えたものだと推測できる。直接音のエネルギーが大きい Coherent 領域内では、僅かな音の物理的特徴変化によって Preference が変化する。

また音の知覚に重要である大きさ感に関しても同様の試聴実験を行い主観的 Loudness 評価を行った。残響室内での主観的 Loudness は Preference と同様に Coherent 領域内において主観値が最大となり、領域を超えると急激に下降する。一方、残響が付加されない無響室録音での評価は SSD(m) の増大に伴って上昇した。したがって従来音場評価指標で表せない Coherent 領域内の物理特性変化は、主として音源特性 (スピーカ構造に起因する特性) により、試験信号が増強され主観値が上昇するものだと考えられる。

次章以降では、知覚変化に影響を与える物理的な変化を距離別録音信号の時間・周波数特徴に着目した評価方法、可視化の検討を行なう。

第4章 基本周波数分析SPEC²用いた調波構造の可視化

4.1 はじめに

本論文では、聴覚器官にとって信号の短時間内に含まれる調波構造(音情報)変化の認識が重要であると仮定する。Meddis等[10][11][12][13]は聴覚モデル分析に短時間ACFを取り入れてピッチを推定する方法を提案した。安藤等によって反射音の最適遅れ時間についての研究がなされている[27]。主観的好ましき(Preference)と対応する最適遅れ時間は音源信号の自己相関関数(Auto-Correlation Function :ACF)の有効継続時間によるとされている。その結果、最適な遅れ時間は自己相関係数が0.1になる遅れ時間によって特徴づけられている。この最適な遅れ時間と主観量変化との対応に加え、複数の反射音からなる初期反射音の特徴についても伝達系の周波数特性によって明らかにされるべきである。信号の短時間ACFに着目する事によって、聴覚的に重要である帯域のスペクトル変化を表すことができる。一般に録音現場では録音条件が変わったときの伝達系の特性は知らされない。残響が含まれた録音音源から音の知覚に影響を与える特徴を見つける事が望ましい。そこで本章では残響インパルス応答が付加された音楽モチーフのACFの初期部を分析することで、知覚に寄与するPreference変化を視覚化する。

さらに本章ではSSD(m)の変化に伴って残響が付加された距離別音楽モチーフの音の時間・周波数の視覚化を行なう。雑音中からの音声抽出のモデルの一つに、パワースペクトルのピークを選び出す波形検出方式がある[53]。このス

ペクトルピーク抽出 (Spectral Peak-Picking) は雑音中の音声抽出に限らず、音声・音楽の基本周波数分析にも有効であることが知られている。そこで本章では残響音が付加された楽音波形のパワースペクトルから際立ったピークを形成する周波数成分を抽出した後、ピーク周波数列の基本周波数と高調波成分を推定することによって、音の微細な時間・周波数分析 (SPEC² 分析) を試みることにした。

4.2 短時間自己相関分析

ACF の初期部のエネルギーの増加は直接音を増強し、音の知覚に影響を与えるとみなされる。そこで短時間観測区間 30(ms) にわたる短時間 ACF のエネルギーを求める。分析区間 N の音源信号を時間関数 $s(t)$ 、遅れ時間を τ とした自己相関関数 $\phi(\tau)$ は

$$\Phi(\tau) = \frac{1}{N/2} \sum_{t=0}^{N/2} s(t)s(t+\tau) \quad (4.1)$$

$$\phi(\tau) = \frac{\Phi(\tau)}{\Phi(0)} \quad (4.2)$$

と定義される。

図 4.1 は残響インパルス応答が畳み込まれた音楽信号の短時間観測区間 30(ms) にわたる短時間自己相関関数例である。室内において残響のエネルギーが増すと、音楽信号の自己相関関数は無相関に近づき原信号の ACF の形状が崩れ始める。最も近傍の SSD: 0.01(m) から前章の主観評価結果において、主観値が最大となった 0.16(m) までの ACF は原信号の特徴を保存しつつ、山谷が微細になってきている。これはスペクトル成分が豊かになり、30(ms) に含まれる直接音のエネルギーが増強されていると推測できる。一方 0.32(m) 以降になると原信号の ACF の形は崩れることが明らかになった。

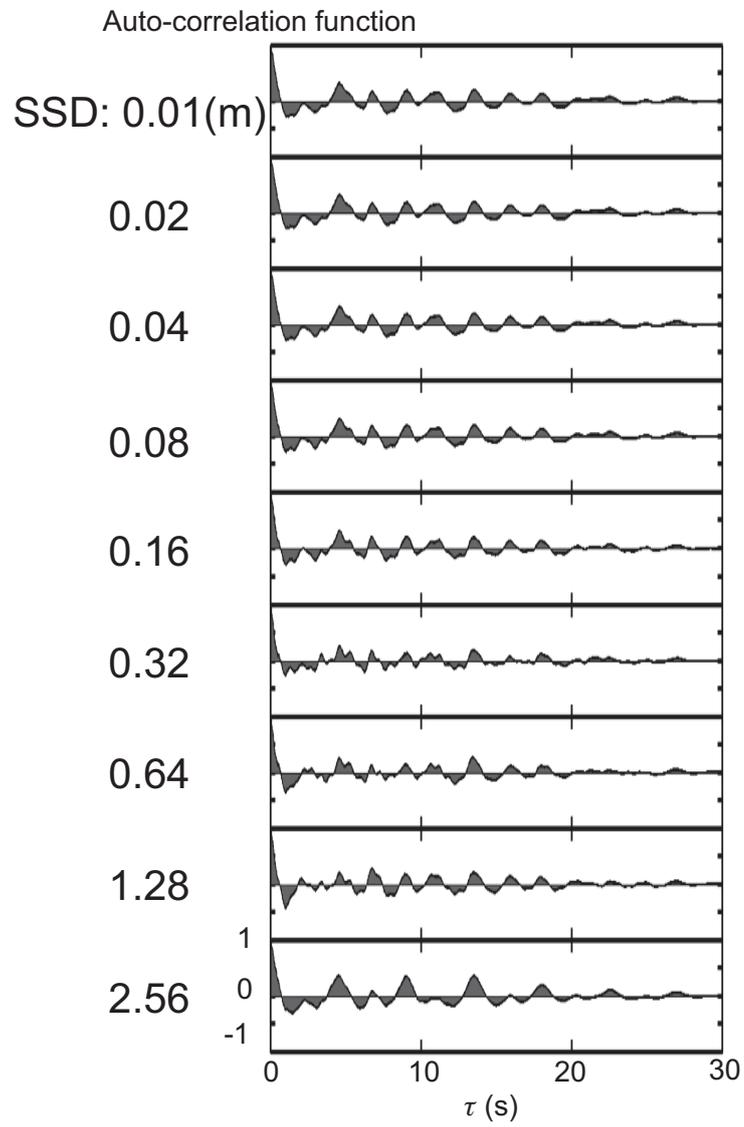


図 4.1: 残響室における距離別音楽モチーフの短時間自己相関関数例

短時間の分析区間 N' までの短時間自己相関関数のエネルギーを A とすると

$$A = \frac{1}{N'} \sum_{\tau=0}^{N'} |\phi(\tau)(N' - \tau)| \quad (4.3)$$

と表される。したがって SSD:0.01(m) における音楽モチーフと異なる SSD(m) の短時間 ACF のエネルギー比 ΔF (dB) は

$$\Delta ACF = 20 \log_{10} \frac{A_{SSD(m)}}{A_{0.01(m)}} \quad (\text{dB}) \quad (4.4)$$

と表すことができる。以下の図 4.2 は SSD にともなった短時間自己相関エネ

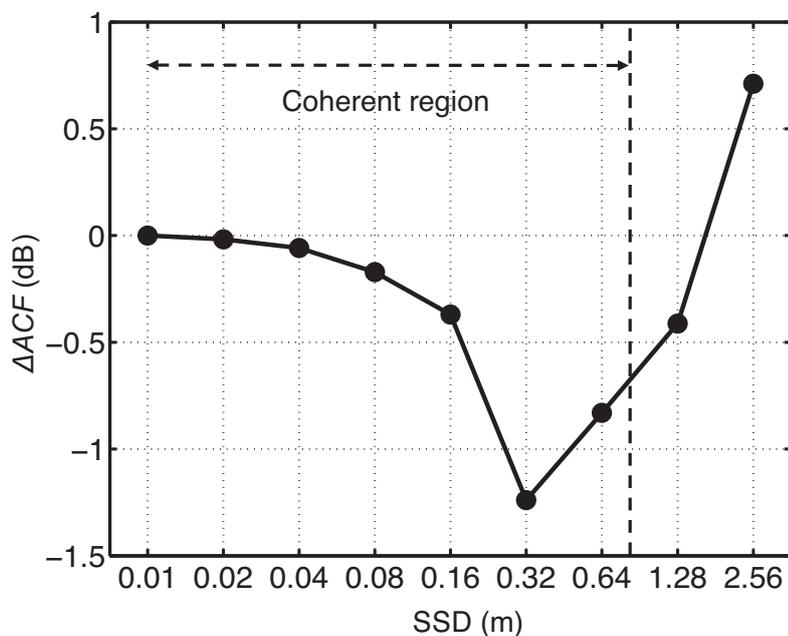


図 4.2: SSD(m) と短時間自己相関エネルギー比 ΔACF (dB)

ルギー比 ΔACF (dB) の変化である。SSD: 0.16(m) までは緩やかに ΔACF が変化し、0.32(m) 以降は図 4.1 に見られるように、SSD(m) が増大するにつれ残響のエネルギーにより大きく原信号が崩れていく様子がわかる。

4.3 スペクトルのピーク抽出による基本周波数分析 SPEC²

音の時間・周波数分析を行なうために、本論文では Signal Period Counting by Spectral Peak Correlation (SPEC²) 分析を提案する。

4.3.1 Spectral Peak-Picking による信号表現

雑音中からの音声抽出のモデルの一つに、パワースペクトルのピークを選び出す波形検出方式がある [53]。スペクトル振幅に観測されるピークのみを抽出する手法を Spectral Peak-Picking と呼ぶ。このスペクトルピークの抽出は雑音中の音声抽出に限らず、音声・音楽の基本周波数分析にも有効であることが知られている。信号観測区間長によらず Spectral Peak-Picking によって信号を構成する正弦波成分を推定することができる。図 4.3 は Spectral Peak-Picking による線スペクトル抽出手順である。音響信号がいくつかの正弦波信号の重畳

$$x_a(n) = \sum_{k=1}^K A(k)e^{j2\pi f(k)n} + \epsilon_K(n) \quad (4.5)$$

で表されているとする。ここで $A(k)$, $f(k)$ は k 番目の正弦波成分の位相を含む複素数振幅, 周波数, K は重畳される正弦波の数, $\epsilon_K(n)$ は正弦波成分で表現しきれない残差成分を表すものである。以下の手順によって、上記信号を構成する正弦波成分を逐次的に推定することができる。

Step 1: 零を付加した M 点 DFT

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x_a(n)e^{-j\frac{2\pi k}{M}n} \quad (4.6)$$

を行う。観測信号のデータ長を N とする。

Step 2: 振幅が最大となる周波数成分 $|X(k_p)|$ を抽出

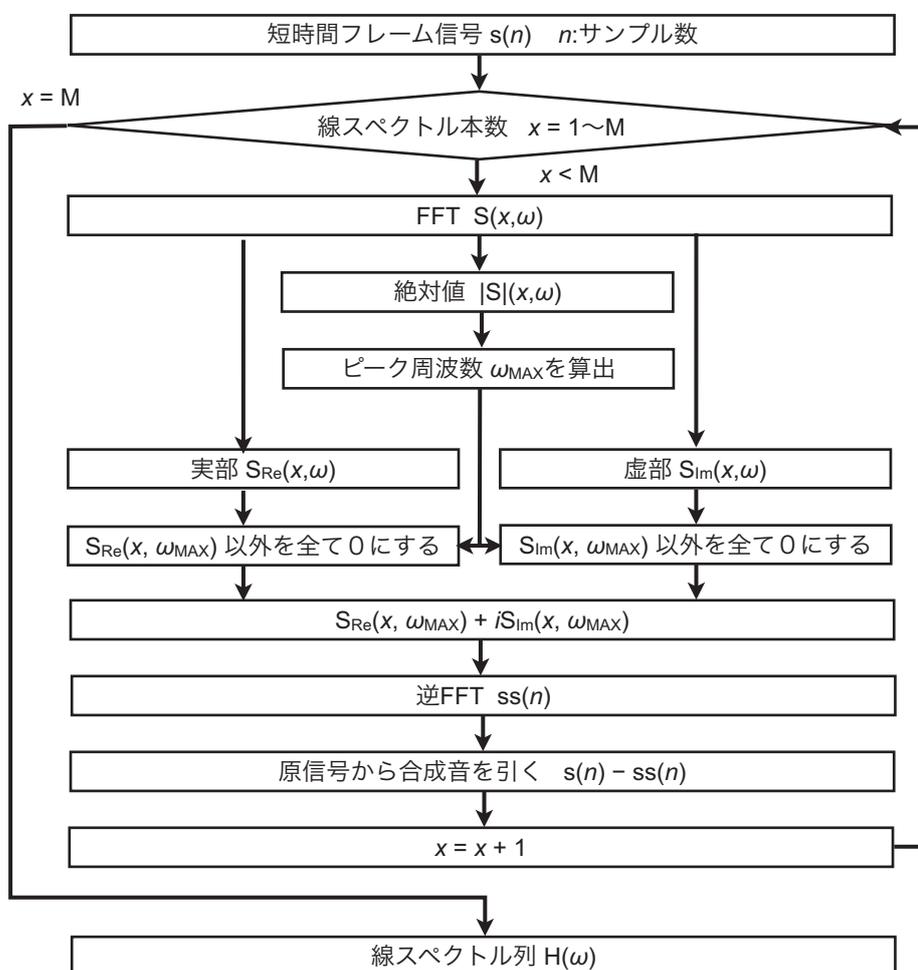


図 4.3: Spectral Peak-Picking による線スペクトル抽出手順

Step 3: 最大振幅周波数成分を有する正弦波信号 $X(k_p)e^{j\frac{2\pi k_p n}{M}}$ を

$$e_a(n) = x_a(n) - X(k_p)e^{j\frac{2\pi k_p n}{M}} \quad n = 0, 1, \dots, N - 1 \quad (4.7)$$

のとおり信号波形から引き去る。この引き算は窓関数 $w(n)$ によって決定される。有限長観測信号に含まれる真のスペクトル以外の周波数成分(スプリアス: $W(k)$)を含めて引き去ることによって、周波数領域で行うこともできる。式 4.6 の例では $w(n)$ は矩形窓である。

Step 4: 改めて

$$x_a(n) \leftarrow e_a(n) \quad n = 0, 1, 2, \dots, N - 1 \quad (4.8)$$

として、Step 1 から Step 3 を

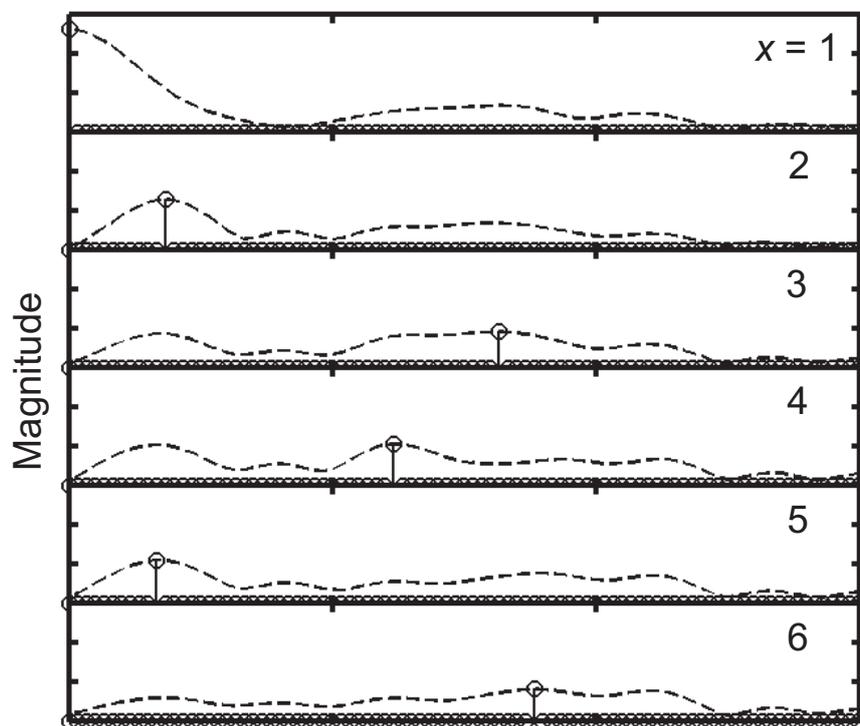
$$\sum_{n=0}^{N-1} |e_a(n)|^2 < E \quad (4.9)$$

となるまで繰り返す。ここで E は許容する残差エネルギーである。図 4.4 は下記の手続きを 6 回繰り返して得られたパワースペクトルのピークを検出し、残差成分から線スペクトルを抽出する例である。残差成分はほぼ雑音のような微小信号である。

4.3.2 SPEC² による基本周波数分析

図 4.5 は SPEC² と呼ぶ音声の基本周波数分析方法の原理を示すものである。図 4.3 によって抽出した線スペクトル列から自己相関関数を求めることによって基本周波数成分の変化を視覚的に判断することができる。図 4.6 は第 3 章の Preference 評価実験において用いた音楽モチーフ ((a):Water Music Suite No.6)[49] の一部を SPEC² 表示した例である。抽出された基本周波数成分が図中の赤部分で表示されている。縦軸は 100-1k(Hz) の周波数を示し、横軸は時

(a) パワースペクトル



(b) 線スペクトル列

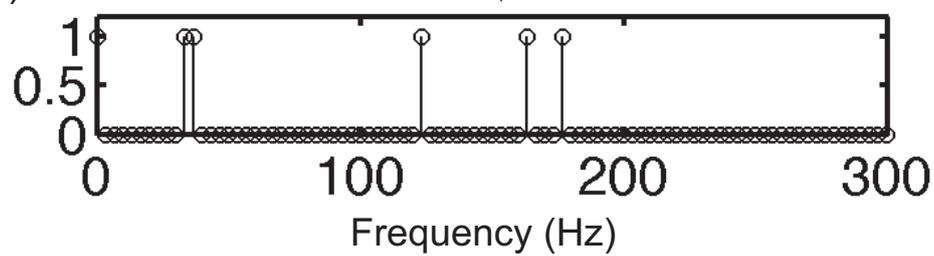
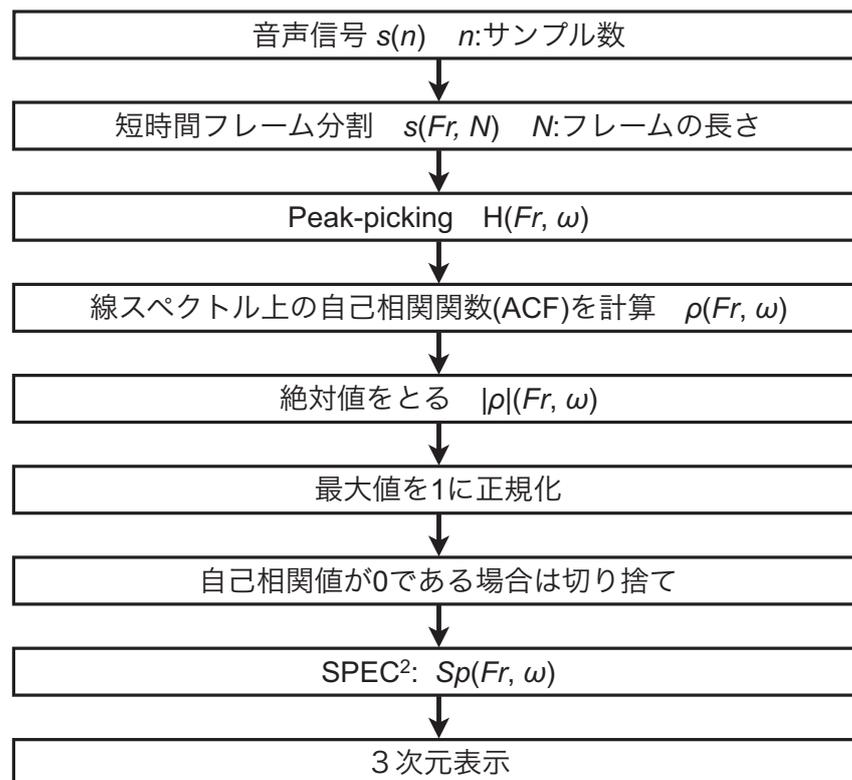


図 4.4: Spectral Peak-Picking 例

図 4.5: SPEC² 分析方法

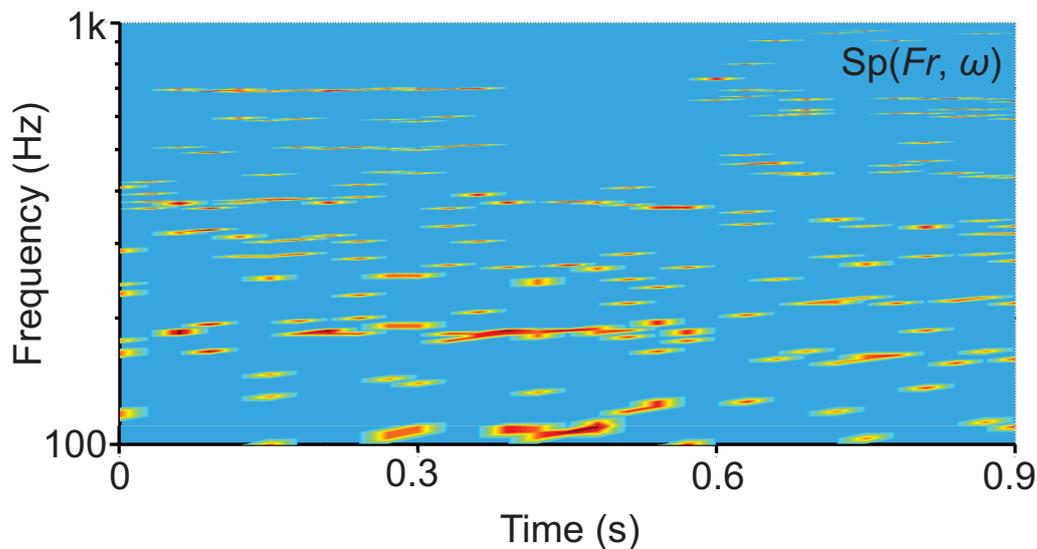


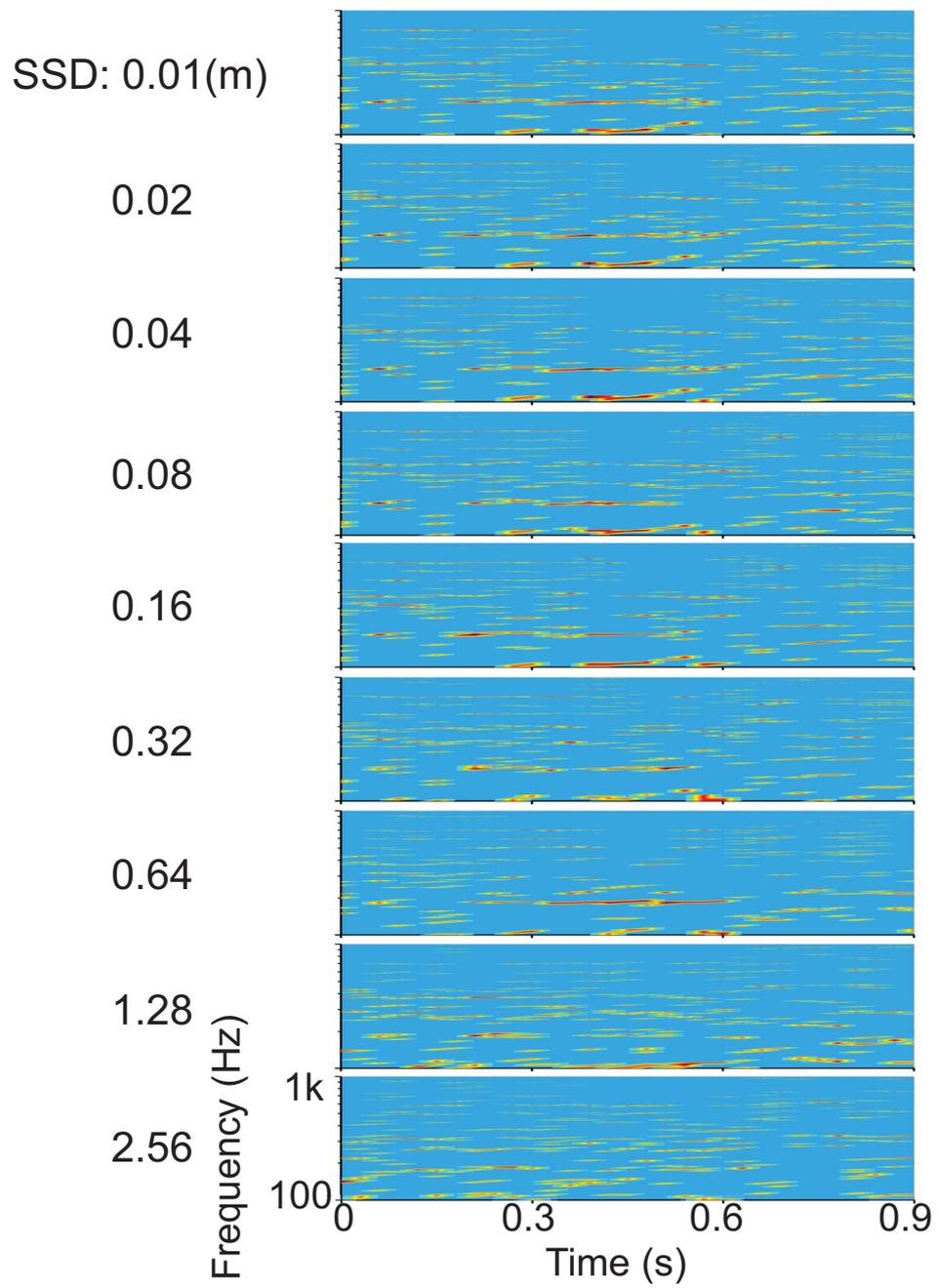
図 4.6: 音楽モチーフの SPEC² 分析例

間である。分析区間のフレーム長は 30(ms) を用い、音楽モチーフの 0.9(s) 分の SPEC² 分析結果を表示してある。

図 4.7 は図 4.1 と同様の距離別に録音された音楽モチーフを用いた、SPEC² 分析例である。音源近傍の 0.01(m) の分析例と比較すると SSD(m) が増加するにつれ、基本周波数成分が横軸の時間方向に伸びて抽出されており、残響のエネルギーの付加により直接音成分の減衰が長くなっていることが確認できる。一方で Coherent 領域の 0.68(m) を超えると、0.01(m) で確認できる基本周波数成分が無くなり、調波構造が乱れて不明瞭になっていく様子が確認できる。

4.4 まとめ

本章では、残響が含まれた録音音源から音の知覚変化に寄与する直接音のスペクトル特徴を明らかにするために、自己相関関数 ACF を用いた分析を行った。SSD: 0.01(m) における音楽モチーフを参照音とすると Coherent 領域

図 4.7: 距離別音楽モチーフの SPEC² 分析例

内の 0.16(m) までわずかに ACF が変化し、領域を超えると残響のエネルギーによって ACF が崩れていくことが明らかになった。さらに、音の時間・周波数分析を行なうために、SPEC² 分析を提案し、残響を含んだ音源から直接音成分の時間・周波数特徴を可視化することができた。このことから、主観量の変化に影響を与えるとされる物理的特徴を SPEC² 分析によって視覚化した。残響エネルギーや目的となる音 (Target) 以外の音が付加されたときに、SPEC² 分析をおこなう事で、Target に対する調波構造の保存度と強調効果を判断することができる。

本章までに距離別音楽モチーフを用いた受聴実験による心理的評価、ならびに短時間 ACF 分析と SPEC² 分析による物理的評価をおこなった。共有空間における快適な音コミュニケーション環境の追求には、音楽の視聴だけではなく話者同士の会話情報の明瞭性が重要となる。そこで次章では音声の主観的な明瞭性に着目し、Preference, Loudness と同様の評価実験によって尺度化するとともに、音声の明瞭性にとって重要とされる包絡線特徴に着目し、音の時間的变化を周波数的に視覚化することを試みる。

第5章 包絡線スペクトル特徴に着目した音声の主観的明瞭性評価

5.1 はじめに

第2章で述べたように、これまで音場の主観評価の研究では、主としてコンサートホールの初期反射音を対象として評価が進められてきた [27][29][33][54]. 第3章では室内音場における音楽の最適な録音距離を主観評価実験によって明らかにした. 本章では共有空間での人々の会話に重要である、会話音声の明瞭性(聴き易さ)変化について着目する.

信号波形における狭帯域包絡線は音声情報を得る大きな手掛かりとして知られており、音声の明瞭性にとって重要なファクターである. Drullman 等 [6] は 100–6.4k(Hz) にわたる 24 帯域の 1/4 オクターブ帯域包絡線とそれぞれの周波数帯域に対応する帯域雑音から了解性のある音声を合成できることを示した. また Shannon 等 [7] は音声帯域を概ね 4 帯域に分割した帯域包絡線を用いて了解性のある音声を実現できることを示した. Houtgast[8][9] は狭帯域包絡線の変調度によって、室内の反射音や雑音による了解度評価が可能であることを報告している. 一方、反射音のエネルギーが小さい近距離音場においては、残響による音声マスキングが起こらず、音声情報の損失がないため、文章了解度評価や音声明瞭度指数 (Speech Transmission Index : STI) 分析 [55] を実施することが困難である.

聴覚モデルにおいて音響信号は蝸牛の周波数選択性から狭帯域信号に分割される。分割された狭帯域信号は有毛細胞モデルにより半波整流後にローパスフィルターを施した狭帯域包絡線を得ると考えられている。Meddis 等 [13] は聴覚モデルに基づき、狭帯域包絡線の自己相関関数を用いた Summary Auto-Correlation Function (SACF) による聴覚脳幹の周期性検出に基づく Pitch 知覚モデルを提案した。近年では風間等 [14] が音声の了解性には位相特性の及ぼす効果が大きいことを明らかにした。音声の分析・合成における振幅・位相特性いずれかの優位性は分析時間長に依存するものとなり、概ね 10(ms) 以下あるいは 200(ms) 以上の分析時間長ではいずれも位相特性が振幅特性に比べて重要な要因となる。

そこで本章では、近距離音場において音源・受音点間距離と共に変化する音声の主観的明瞭性に関する評価実験ならびに、音声の主要帯域における包絡線スペクトル分析に着目した。上述した狭帯域分析に着目した時間・周波数特徴の可視化によって近距離音場での SSD(m) によって変化する音の微細な物理的变化と明瞭性に関する知覚評価結果との対応を明らかにする。その結果から近接音場における音声の明瞭さと、狭帯域包絡線スペクトルエネルギーとの対応を明らかにした。

5.2 近距離音場における距離別音声の明瞭性評価実験

反射音のエネルギーが小さい近距離音場では、音声が残響によってマスクングされず音声情報の損失がないため、Speech Transmission Index (STI)[55] に代表されるこれまでの文章了解度評価を実施することが困難である。そこで本章では、音声の聞きやすさに関して一対比較法 (Thurstone V) による主観的な明瞭性評価実験を行なった。

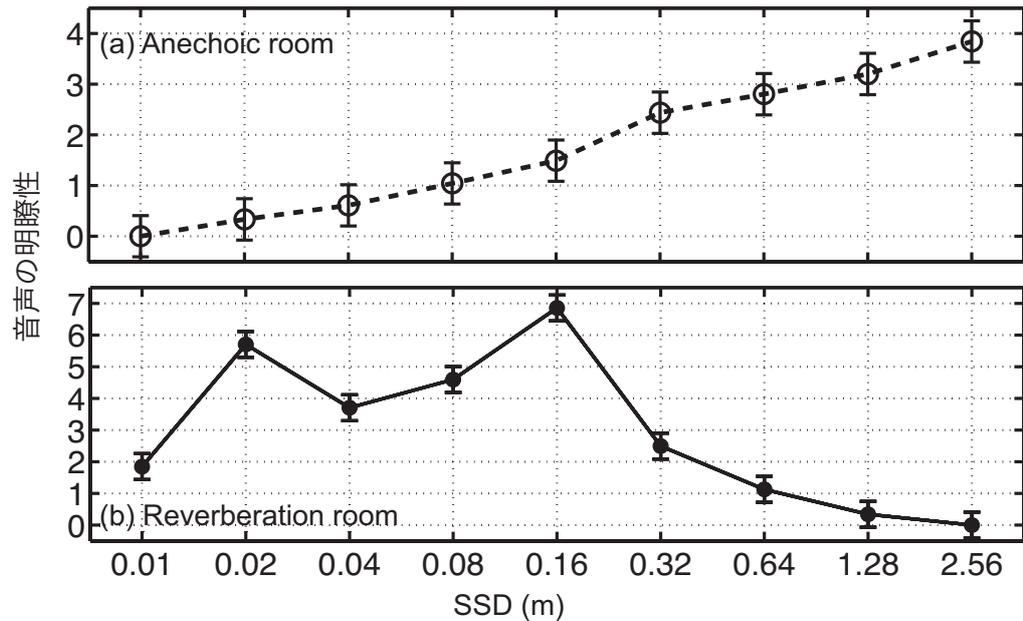


図 5.1: 一対比較実験による明瞭性評価結果

5.2.1 実験方法

試験音は第3章における Preference 試験方法と同様の構成で各残響インパルス応答を約 2(s) の女声サンプルに畳み込み、バンドパスフィルタにより 100–10k(Hz) まで帯域制限した。また遅い残響効果を取り除かれると予想して、S/N 比: 6(dB) で白色雑音を付加した。被験者は特別に訓練をしていない 20 代から 30 代の学生 6 名である。2 つの距離別音声が含まれる 36 対の試験音を受聴し、どちらの音声か明瞭であるかを回答する。

5.2.2 実験結果

図 5.1 は距離別音声の明瞭性に関する一対比較評価結果 ((a):無響室, (b):残響室) である。横軸は物理尺度 SSD(m), 縦軸は明瞭性に対応する心理尺度値である。図中のエラーバーは刺激数と被験者数を関数として求めた信頼区間

である [43]. 図 5.1(a) において SSD(m) の増大に伴って明瞭性が高くなる. 図 5.1(b) では直接音領域内の 0.16(m) までは図 (a) と同様に明瞭性が上昇する一方, 0.16(m) を超えると急激に下降する. 直接音領域内の明瞭性向上は, 主としてスピーカ特性による音源特徴の影響と考えられる.

5.3 距離別音声信号の狭帯域包絡線スペクトル分析

図 2.9 に示された通り, D_{30} は直接音領域内の 0.01–0.64(m) ではほぼ一定となり, 反射音のエネルギーが低い領域での音質評価は明瞭性変化との対応が困難であることを表している.

直接音の明瞭性にとって包絡線特徴の変化が重要である [8][9]. そこで図 5.2

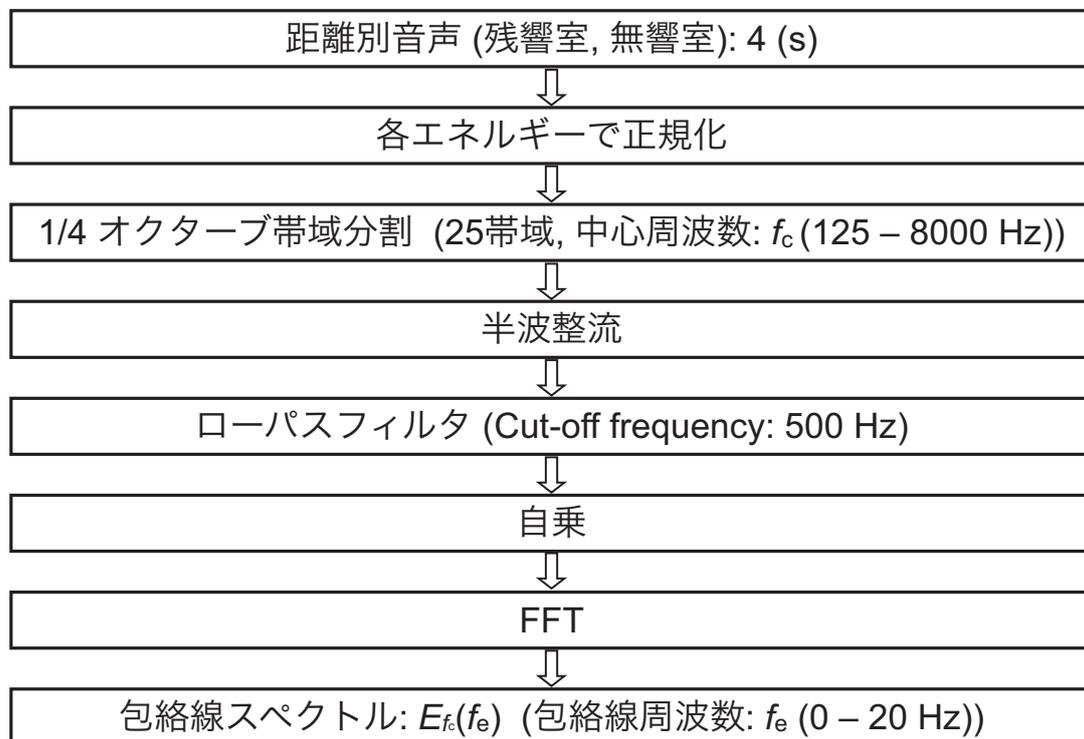


図 5.2: 狭帯域包絡線スペクトル分析方法

の手順に従い、インパルス応答が畳み込まれた距離別音声进行分析し狭帯域包絡スペクトル $E_{f_c}(f_e)$ を分析した。

図 5.3 ならびに図 5.4 は残響室・無響室内で録音された距離別音声(女声)の狭帯域包絡線スペクトル構造の 3次元表示例である。SSD(m)が増大するにつれ、スペクトルの山谷が崩れていく様子がわかる。また音声の主要な包絡線スペクトルが $f_c : 297(\text{Hz}), 707(\text{Hz})$ に表れている。そこで図 5.5 に仮想距離別音声の主要帯域に関して

$$\Delta E_{f_c}(r_n) = 10 \log_{10} \frac{\sum_{f_e=0}^{20(\text{Hz})} E_{f_c}(f_e, r_n)}{\sum_{f_e=0}^{20(\text{Hz})} E_{f_c}(f_e, r_1)} \quad (\text{dB}) \quad (5.1)$$

として $r_1: 0.01(\text{m})$ における音声の狭帯域包絡線スペクトルを参照とする狭帯域包絡線スペクトルエネルギー比 $\Delta E_{f_c}(r_n)$ (dB) を示す。図 5.5(b) $f_c : 707(\text{Hz})$ では図 5.1(b) の明瞭性評価が最大となった $0.16(\text{m})$ まで、SSD(m) の増大に従って包絡線スペクトルエネルギーが上昇することがわかる。

図 5.6 はさらに、図 5.3・図 5.4 の狭帯域包絡線スペクトル上で最大値を正規化した狭帯域包絡線スペクトルエネルギー (EME) の計算方法である。図 5.7 は距離別音声の距離別音声の狭帯域包絡線スペクトルエネルギー EME(dB) である。残響室では EME(dB) は Coherent 領域内の SSD: $0.16(\text{m})$ まで上昇し、 $0.32(\text{m})$ 以降は下降する様子がわかる。一方で、無響室では EME(dB) は SSD(m) と比例して増大していく。したがって、EME(dB) の傾向は図 5.1 に示された明瞭性評価に対応することが明らかになった。

5.4 まとめ

本章では、従来の明瞭度評価指標では表せない近距離音場の音声の明瞭性を一対比較法によって尺度化するとともに狭帯域包絡線分析によって視覚化した。無響室での包絡線変化は音源特性によるものと推測されるが、残響室内では直接音領域を超えると反射音により音声の包絡線特徴が崩れ、急激に包絡線

(a) Reverberation room

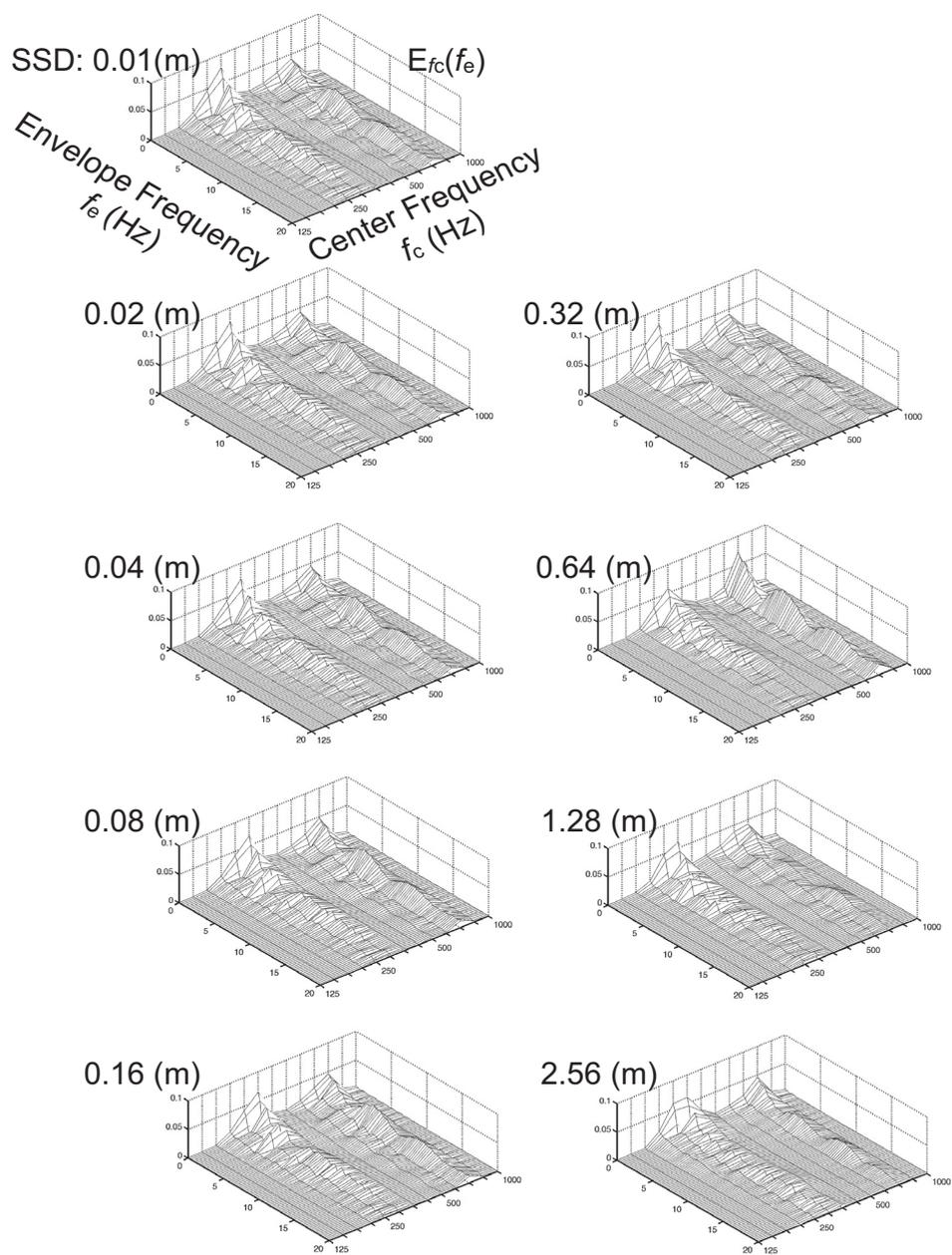


図 5.3: 距離別狭帯域包絡線スペクトルの3次元図(残響室)

(b) Anechoic room

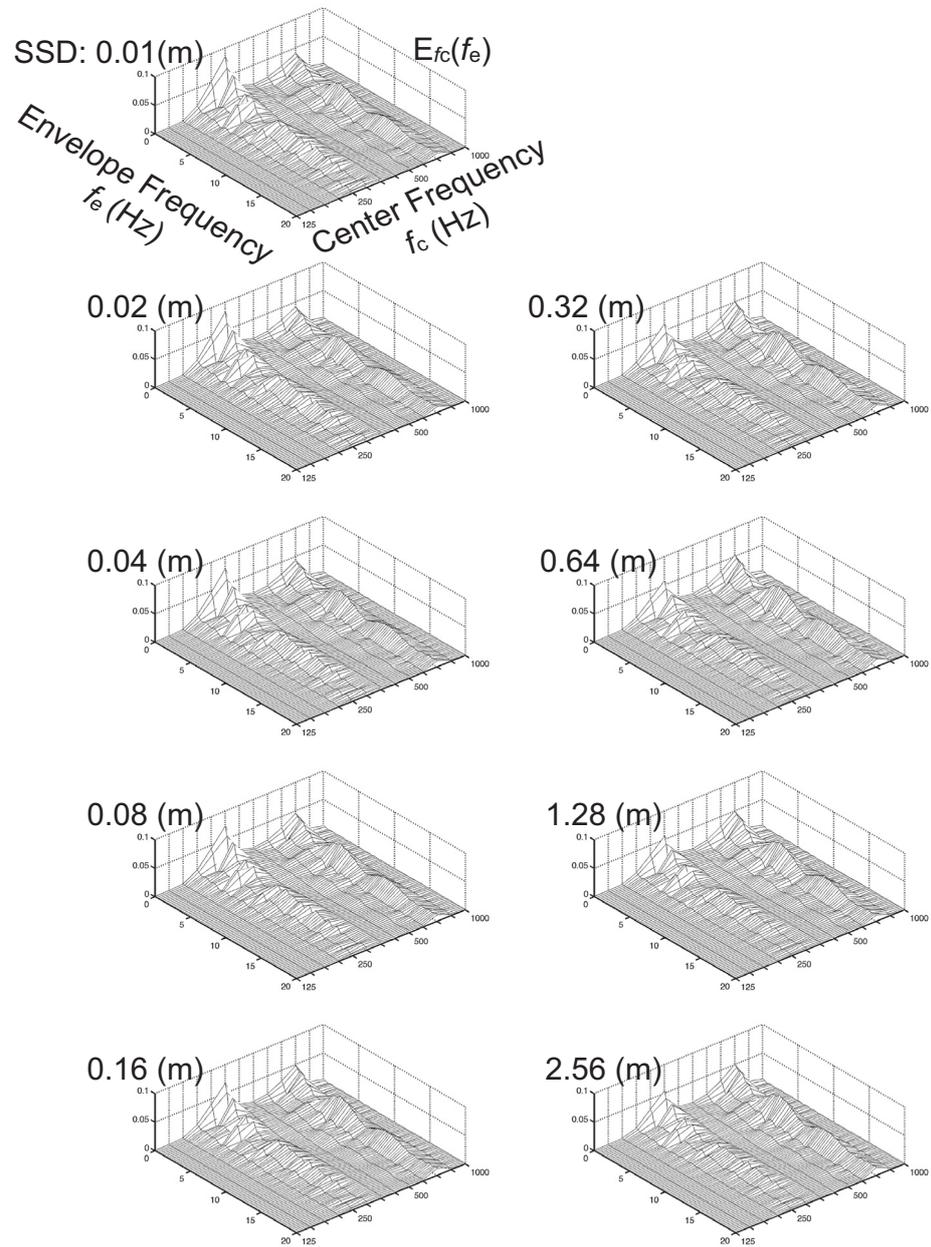


図 5.4: 距離別狭帯域包絡線スペクトルの3次元図(無響室)

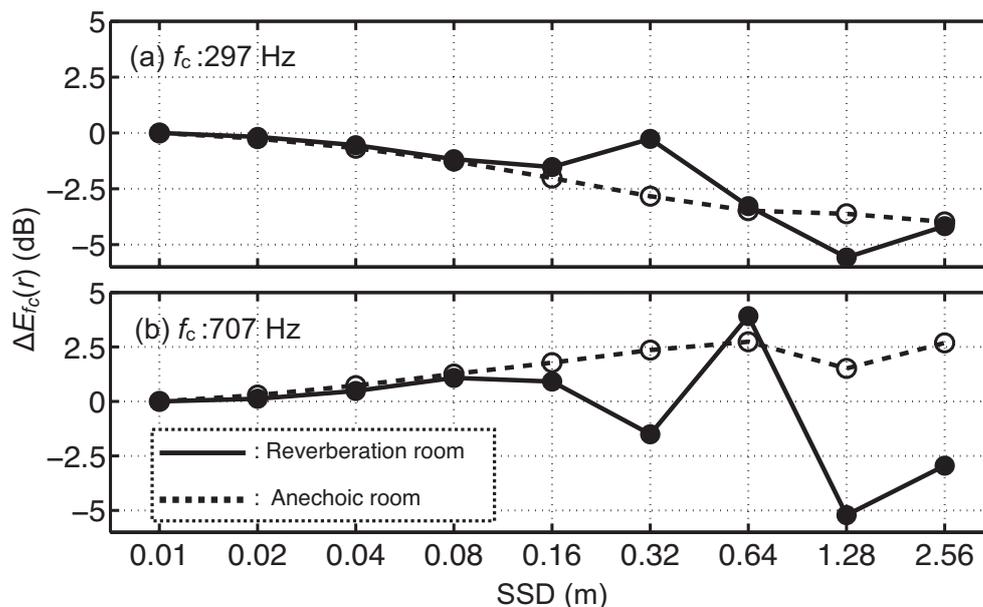


図 5.5: 狭帯域包絡線スペクトルエネルギー比

スペクトルエネルギーが低下する。主要帯域の包絡線スペクトル特徴が豊かになると直接音が補強され、明瞭性が向上していると考えられる。本論文ではこれまで、残響音場内での評価を行ってきた。共有空間での室内コミュニケーションにおいては、残響以外にも暗騒音や会話相手以外からの音が到来する。基本周波数分析 SPEC² ならびに狭帯域包絡線分析は、残響以外の騒音が付加された環境においても特徴を抽出することができる。次章では外部からの音の到来によって受聴点での会話音声の了解度が低下する情報マスキング効果に着目し、マスキング音が重畳されたときの明瞭度評価実験ならびに時間・周波数分析による物理的評価を行なった。

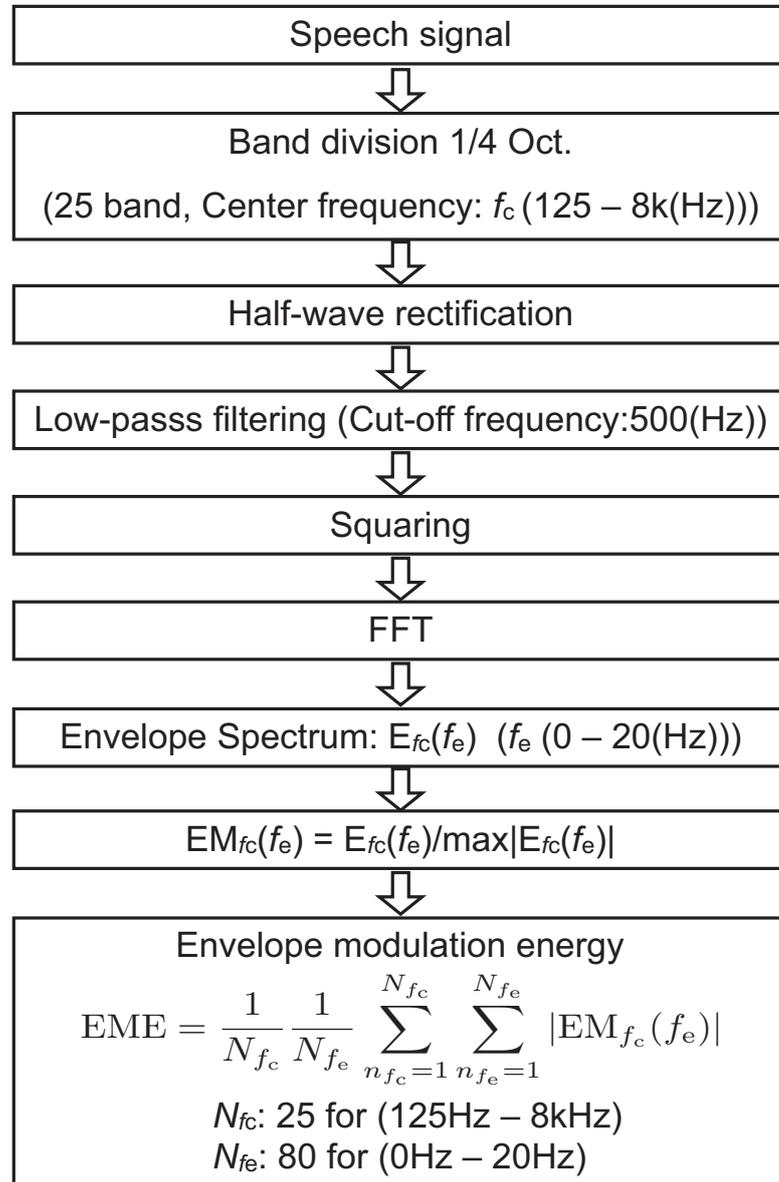


図 5.6: 狭帯域包絡線スペクトルエネルギー (EME) 計算方法

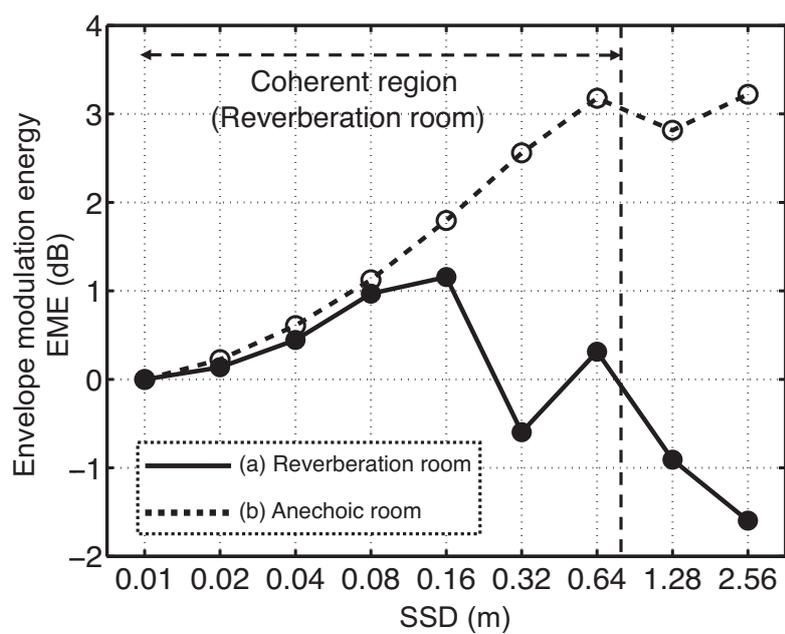


図 5.7: 距離別音声の狭帯域包絡線スペクトルエネルギー (EME)

第6章 情報マスキング効果の単語 了解度評価と時間・周波数 特徴の類似性分析

6.1 はじめに

本論文ではこれまで、室内環境における音コミュニケーション設計において重要となる近距離音場において変化する音の知覚変化と音場の物理的特徴を評価ならびに視覚化することによって明らかにした。一方で、公共性が存在する空間では残響を含む音場特性の他に、意図せず外部からの会話情報や環境設備からの暗騒音が到来し、コミュニケーション情報を覆うマスキング現象が起こる。目的とする会話音声を Target とすると、マスキングには受聴点での音圧レベルに依存するエネルギーマスキング効果と Target の会話情報を覆う情報マスキング効果にわけられる [17][18][19].

音のマスキング効果に着目した公共空間の設計指針には長い歴史がある [15]. 近年の室内音環境においては暗騒音レベルの低下や遮音性能の向上により、従来環境と比較すると静寂になる一方、人の音コミュニケーションにおける会話音声に関する評価が重要である。共有空間の音環境設計においては他人の会話が邪魔にならない空間を目指した Speech Privacy、さらに個人情報の保護を目的とした、会話の秘話生を追求する Speech Security が重視されている [57][58]. 清水等 [59][60][61] は、Speech Privacy に着目し、オフィスや医療施設などの執務空間における最適な音環境の評価・制御・設計手法について報告している。

共有空間内での受聴者周囲における会話相手の音声 (Target) の音圧レベルは、天井や家具、照明設備、壁、窓等の設備によって変化する。さらに、空間内の受聴点によっては空調設備等から発せられる暗騒音がターゲットとなる音声のマスクングに寄与する。ASTM E1130[16]では、共有空間内の受聴者の耳の周囲の音圧レベル SNR(dB) を測定し、標準明瞭度指数 (AI: Articulation Index) を用いて Speech Privacy を評価している。つまり、これまでの Speech Privacy 評価を取り入れた音環境設計においては会話の秘話性ではなく、主にオフィス等の音環境評価を目的とされている。本論文では Speech Privacy, Speech Security において効率良く会話情報を保護するためには、情報マスクング効果の考慮が重要であると考えた。

そこで本章ではマスキングの音圧レベルの付加によってマスクングするエネルギーマスクング効果に代わって、マスキングの波形に依存する情報マスクング効果に着目する。エネルギーマスクング効果を引き起す白色雑音と高い情報マスクング効果がある逆転音声を用いて、単語理解度試験を行なう。またマスキングに埋もれた会話音声の狭帯域包絡線分析と SPEC² を行い、マスクング効率評価方法を開拓する。

6.2 単語理解度試験による情報マスクング評価

音声のマスクング効果は単文章の書き取り試験 (単語理解度試験) によって評価することができる。情報マスクング効果をもたらす代表的なマスキングの一つに時間軸を反転した逆転音声がある。逆転音声には Target に対して時間的特徴の一時的な相違とスペクトル類似性を含んでおり、効率良く会話情報をマスクングすることが期待できる [62][63][64]。Target 音声に逆転音声を重ねれば、音声の了解性の低下が考えられる。マスクング効果を利用した公共空間設計においてはこれまでに、定常雑音や暗騒音のエネルギーマスクング効果や遮音壁の設置等が検討されてきた。一方、Speech Privacy を考慮した室

内設計においては秘話性が高く、効率良く会話情報をマスキングする信号の検討が必要である。エネルギーマスキング効果によって生じる雑音に埋もれた音声に対する了解度の劣化は、音声波形の帯域別包絡線の保存度によって予測できることが知られている [65]。しかし情報マスキング効果が含まれる音声了解度の劣化予測は、今もって研究課題である。

マスキング効果はエネルギーマスキング効果と情報マスキング効果の合計として解釈される [66]。そこで本論文では、白色雑音と逆転音声を用い、Target とマスキャーのエネルギー比 (TM 比 (dB)) を変化させた時の単語了解度試験を行なうことで、エネルギーマスキング効果とマスキャーの種類による情報マスキング効果の両者を含んだ条件によってマスキング効果の違いを明らかにする。

6.2.1 実験方法

一般にマスキング効果の実験では Target となる信号を予め設定することが行われる。しかし日常に起こるマスキング現象では、Target となる信号はそれぞれの場合において変化する。そこで本論文では Target を固定することなく混合信号から聞き取れる音声信号を被験者に自由に書き取ってもらうこととした。図 6.1 (a) は含まれる発話文章の了解度 (%) の試験音構成ならびに (b) は試験条件である。試験に用いた音声は、いずれも形容詞と名詞からなる意味のある簡単な言葉とマスキャーとの混合音とした。試験音声に含まれる音声の話者はマスキャーも含めて全て同一話者である。被験者は試験音声と同様の構造となる形容詞と名詞の例をサンプル音声として試聴した後、試験混合音声から聞き取れた言葉を書き取る。形容詞部あるいは名詞部のいずれかが書き取ればその音声信号に対する正解点数を 0.5 点、同様に両部分を聞き取れた場合には正解点を 1 点として集計した。

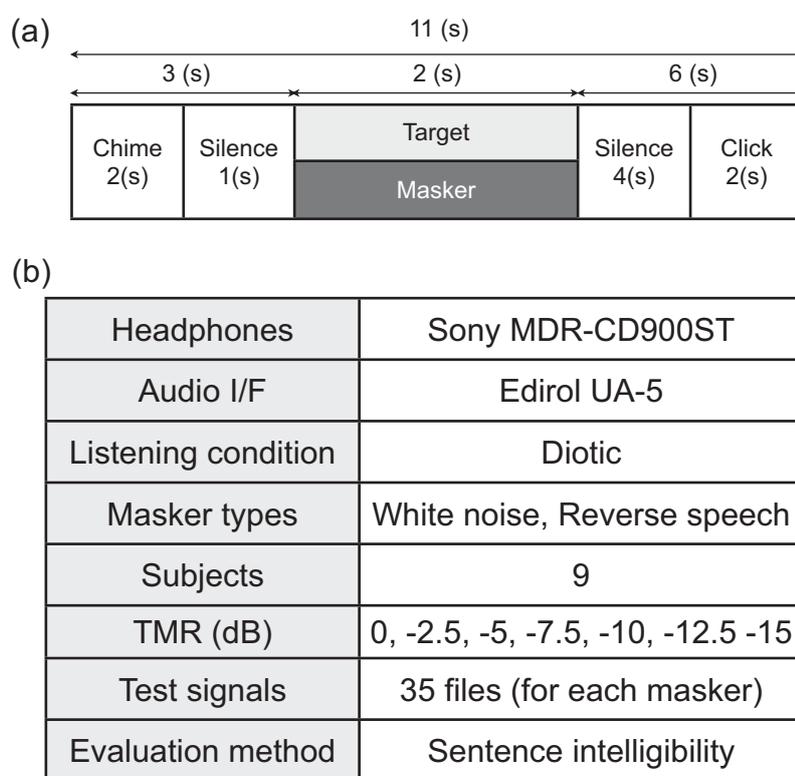


図 6.1: マスカーに埋もれた発話音声の単語了解度試験音構成 (a) と試験条件 (b)

6.2.2 実験結果

図 6.2 は白色雑音 (実線) と逆転音声 (破線) 中に含まれる発話文章の了解度

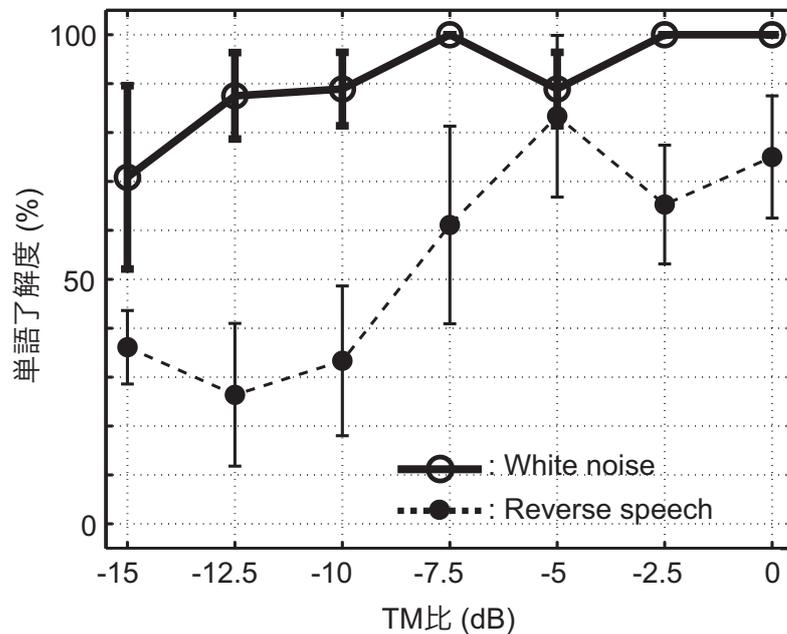


図 6.2: 単語了解度試験結果

(%)の試験例である。横軸に示す TM 比 (Target 音声対マスキング音のエネルギー比)(dB) の下降とともに了解度はマスキング効果によって低下する。特に逆転音声によるマスキング効果が白色雑音に比べて顕著である。

6.3 時間・周波数分析によるマスキング効果の評価

6.3.1 帯域別狭帯域包絡線相関

上述したようにマスキング効果は音声の了解度試験によって、概ね評価することができるであろう。しかし、マスキング設計においては受聴実験によらずマスキング効果を予測・評価できることが望ましい。そこで音声了解度の重要

な要因であった音声の帯域別包絡線を，マスキング中の音声波形と原音声との間の包絡線相互相関係数で比較してみることにした。

ここで，短文章番号 l ，狭帯域番号を i ，サンプル数を n ，自乗包絡線を e としたときの Target(T) と Target にマスキング (T+M) が付加された信号の短時間フレーム間の相関係数 $\rho_E(l, i)$ は

$$\rho_E(l, i) = \frac{\widehat{e}_T(l, i, n)\widehat{e}_{T+M}(l, i, n)}{\sqrt{\widehat{E}_T(l, i)\widehat{E}_{T+M}(l, i)}}, \quad (6.1)$$

ただし，

$$\widehat{e}_T(l, i, n) = e_T(l, i, n) - \overline{e_T(l, i, n)}, \quad (6.2)$$

$$\widehat{e}_{T+M}(l, i, n) = e_{T+M}(l, i, n) - \overline{e_{T+M}(l, i, n)}, \quad (6.3)$$

$$\widehat{E}_T(l, i) = \overline{\widehat{e}_T(l, i, n)^2}, \quad (6.4)$$

$$\widehat{E}_{T+M}(l, i) = \overline{\widehat{e}_{T+M}(l, i, n)^2}. \quad (6.5)$$

さらに式 6.1 を短文章と帯域平均をしたものを狭帯域包絡線相関係数 (EC)

$$EC = \overline{\overline{\rho_E(l, i)}} \quad (6.6)$$

として表すことにする。

図 6.3 は図 6.2 にて用いた TM 比 (dB) ごとの試験信号の狭帯域包絡線相関係数 (EC) の計算例である。白色雑音は EC の相関が高く，TM 比の低下による変化は大きくないが逆転音声は EC の変化が顕著であることが示された。従って，白色雑音のような定常的なマスキングが Target に重畳されたときには包絡線特徴を崩す事が難しい事がわかる。

また，図 6.4 は単語理解度に対する狭帯域包絡線相関係数 $\rho_E(l, i)$ の計算例である。図中破線は 2 次近似曲線である。狭帯域包絡線相関係数 $\rho_E(l, i)$ の分散が大きく，単語理解度との対応を示すことが出来ないことが明らかになった。したがってこの結果から，情報マスクング効果が従来のエネルギーマスクング

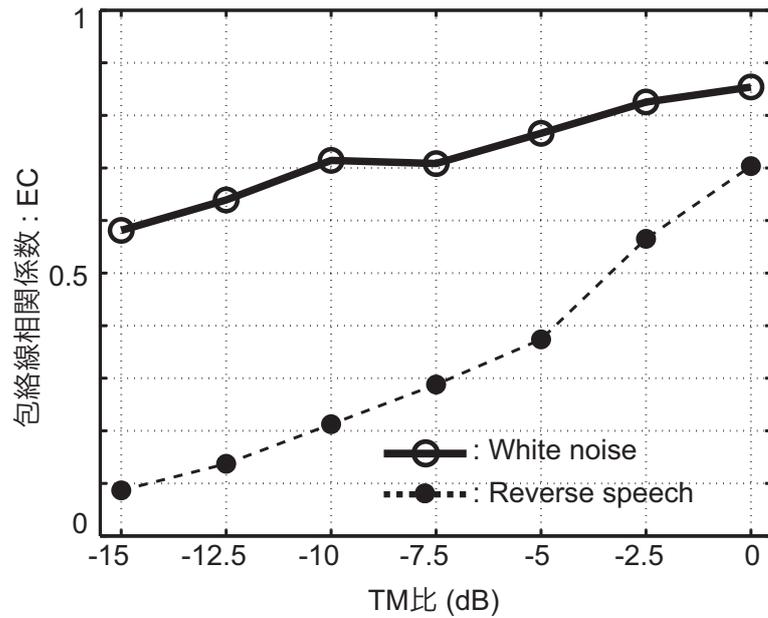


図 6.3: TM 比 (dB) ごとの狭帯域包絡線相関係数 (EC)

効果と異なり、包絡線保存度だけでは予測できない要因を含むことを示唆している。

6.3.2 SPEC² 相関によるスペクトル類似性分析

包絡線相関係数に代る指標として、第4章で述べた SPEC² 分析をおこない、両者の短時間フレーム間の相互相関係数 (SC) について TM 比による変化を分析した。

ここで、短文章番号 l 、短時間フレーム番号を fr 、周波数を k (Hz)、SPEC² 行列を p としたときの Target (T) と Target にマスキャーが付加された信号 (T+M) の短時間フレーム間の SPEC² 相関係数 $\rho_F(l, fr)$ は

$$\rho_F(l, fr) = \frac{\widehat{p}_T(l, fr, k)\widehat{p}_{T+M}(l, fr, k)}{\sqrt{\widehat{P}_T(l, fr)\widehat{P}_{T+M}(l, fr)}}, \quad (6.7)$$

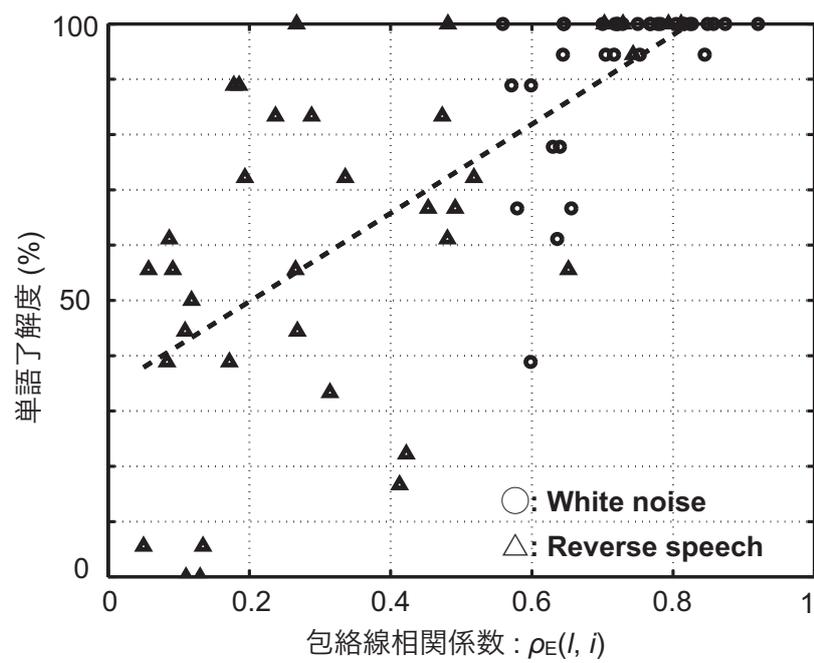


図 6.4: 狭帯域包絡線相関係数 $\rho_E(l, i)$ と単語理解度 (%)

ただし,

$$\widehat{p}_T(l, \text{fr}, k) = p_T(l, \text{fr}, k) - \overline{p_T(l, \text{fr}, k)}, \quad (6.8)$$

$$\widehat{p}_{T+M}(l, \text{fr}, k) = p_{T+M}(l, \text{fr}, k) - \overline{p_{T+M}(l, \text{fr}, k)}, \quad (6.9)$$

$$\widehat{P}_T(l, \text{fr}) = \overline{\widehat{p}_T(l, \text{fr}, k)^2}, \quad (6.10)$$

$$\widehat{P}_{T+M}(l, \text{fr}) = \overline{\widehat{p}_{T+M}(l, \text{fr}, k)^2}, \quad (6.11)$$

さらに式 6.7 を短文章と短時間フレームの平均をしたものを SPEC² 相関係数 (SC)

$$\text{SC} = \overline{\rho_F(l, \text{fr})} \quad (6.12)$$

として表すことにする.

図 6.5 は図 6.2 にて用いた TM 比 (dB) ごとの試験信号の SPEC² 相関係数

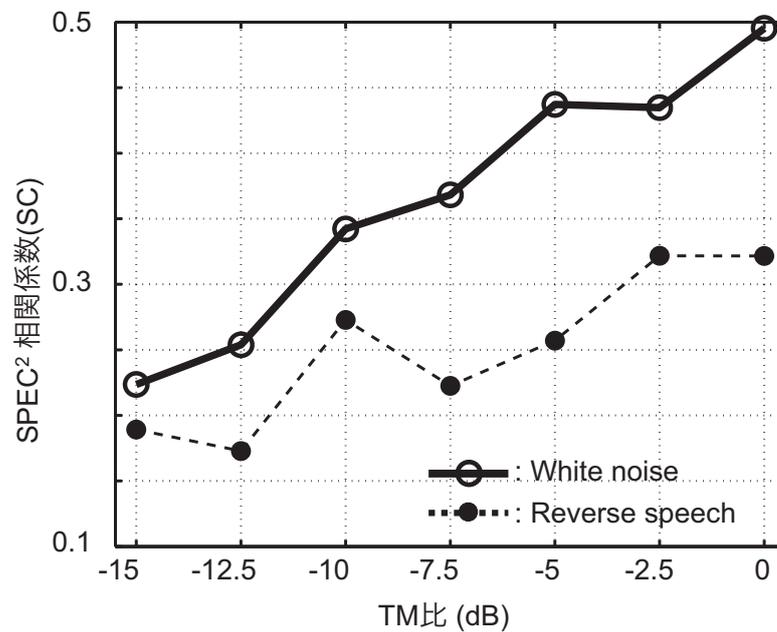


図 6.5: TM 比 (dB) ごとの SPEC² 相関係数 (SC)

(SC) の計算例である. 白色雑音は TM 比の低下と比例して SC も下降する一

方、逆転音声は相関値が大きく変化しない。このことから Target と逆転音声の調波構造の類似性が高いことが示された。

図 6.6 は SPEC² 相関係数 $\rho_F(l, fr)$ と単語理解度の関係を示す結果である。図

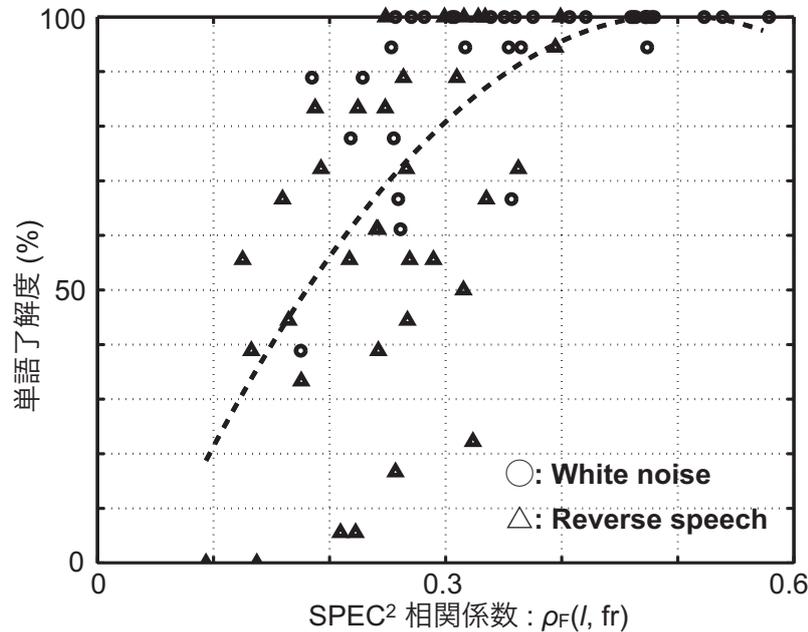


図 6.6: SPEC² 相関係数 $\rho_F(l, fr)$ と単語理解度 (%)

中破線は白色雑音と逆転音声の SPEC² 相関係数の結果の 2 次近似曲線である。この結果から、情報マスクング効果の変化は SPEC² 相関係数によって評価出来ることが示された。

6.4 まとめ

本章では、Speech Privacy 問題におけるエネルギーマスクング効果と情報マスクング効果を文章明瞭度試験と時間・周波数分析による信号の類似性評価によって明らかにした。白色雑音と逆転再生音をマスキングとして用い、マスキングの種類と TM 比 (dB) が変わった事によるマスクング効率の違いを受聴実験に

よって示した。マスキングに埋もれた音声の包絡線相関分析ならびに SPEC² 分析を用いて新たなマスキング評価方法を開拓した。情報マスキング効果が SPEC² 分析によるスペクトル類似性評価によって量的に表すことが可能であることが明らかになった。

マスキングシステム設計にあたって、会話音声と同じ特性を持った逆転再生音を生成しリアルタイム再生することは不可能である。そこで付録 A では包絡線特徴を保存した新たなマスキング信号の設計方法を検討し、さらに環境特徴を含めたマスキングシステムの設計・評価方法を検討している。

第7章 総括

本論文では室内音場における近距離音場に着目し、音源・受音点間距離 (Sound Source Distance: SSD(m)) の増大とともに変化する音の好ましさや音の大きさ感、音の明瞭性に関する主観的な音の知覚ならびに音のマスキング効果を尺度化するとともに、時間変化する信号の基本周波数特徴と狭帯域包絡線分析によって視覚化ならびに評価方法を開拓した。

本論文ではまず、室内音場特性を調査するために残響室と無響室において同一条件で SSD(m) を変えて残響インパルス応答を測定した。測定データから従来用いられてきた D_{30} 値や時間重心値等の主観量に対応するとされる音場特性指標を示した。しかし、残響のエネルギーが低い近距離音場においては従来評価方法では変化を表す事が困難であった。

音の好ましさ (Preference) に関する研究では、これまでに初期反射音が起因するとされ、録音された楽音の自己相関関数 (ACF) の初期部を観測することで求められる、最適遅れ時間と対応するとされてきた。そこで測定した残響室内で測定したインパルス応答を用いて仮想距離別音楽信号を作成し、受聴実験を行なうことで SSD(m) の増大に応じて変化する Preference を一対比較法 (Thurstone Case V) によって尺度化した。その結果、室内空間における近接音場の直接音領域 (Coherent 領域) 内で Preference に関する主観値が最大となり、領域を超えると急激に下降することが示された。また、音の知覚変化にとって音の大きさ感 (主観的 Loudness) は重要とされている。Preference 評価実験と同様にして、主観的 Loudness 変化に関しても一対比較法を用いた受聴

実験をおこなった。Preference 評価実験で示された Coherent 領域内で主観値が上昇する結果には、直接音として知覚される領域、つまり初期反射音や音源特徴が起因すると考え、Loudness 評価では無響室におけるインパルス応答も用いて同様の受聴実験を行なった。その結果、残響室内では Preference と同様に Coherent 領域内で主観値が最大となり、Coherent 領域を超えると急激に下降することが示された。一方で無響室での評価は SSD(m) の増大にともなって主観値も上昇していく結果となった。そこで従来音場評価では表せない、直接音成分の物理変化を明確にする必要があると考え、音の時間・周波数分析を検討した。

雑音中から音声抽出をするモデルであるスペクトルピーク抽出手法 (Spectral Peak-Picking) を用いてピーク周波数列 (線スペクトル列) を推定し、線スペクトル列の自己相関関数を短時間フレーム分析することで、主観評価に寄与すると考えられる直接音成分の基本周波数成分を推定し、視覚化する SPEC² 分析を試みた。主観評価に用いた、仮想距離別音楽信号を SPEC² 分析し、SSD(m) に伴って調波構造が保存・強調されたり、不明瞭になっていく様子を明らかにした。したがって本論文で提案した SPEC² 分析により基本周波数成分の時間的变化を視覚化することで、残響エネルギーの付加や外部から到来する暗騒音の影響による調波構造の変化を評価でき、音のコミュニケーションにおける対象音 (ターゲット) の特徴に特化した共有空間設計の一つの指標として活用することが可能となった。

従来の研究によって音の明瞭性には狭帯域包絡線特徴が重要であることが明らかにされている。明瞭な音の特徴の可視化は、音コミュニケーション環境構築において会話がしやすい環境づくりの目安となる事であろう。そこで本論文では Preference や主観的 Loudness 実験に加え、同様の条件での音声の明瞭性に関する主観量の尺度化を行なった。その結果、上述した残響室内における主観評価結果と同様、明瞭性も Coherent 領域内において最大となり領域を超え

ると急速に下降することが明らかとなった。無響室での評価もまた SSD(m) の増大に伴って上昇していく。さらに音声の時間的変化をスペクトル変化として表す、狭帯域包絡線スペクトルを用いて明瞭性の変化を評価することにした。残響室ならびに無響室での距離別音声を用い、狭帯域包絡線スペクトルの3次元表示をすることで明瞭性に関わる特徴を可視化することができた。実験音で使用した音声(女声)の主要帯域の狭帯域包絡線スペクトルのエネルギー比を音源から最も近い0.01(m)を参照として計算すると SSD(m) の増大に伴って、明瞭性に関する主観値と対応して変化していることがわかった。要するに、音声の狭帯域包絡線スペクトルを示し、主要帯域を保存または強調したコミュニケーション環境を設計すれば会話が明瞭な共有空間を作ることが可能であることが明らかとなった。

前述したように公共空間における音のコミュニケーション環境では、対象音(Target)の情報を得る以外にも空調などの室内設備からの暗騒音や Target とは違う話者の情報が受聴者に到来する。そこで本論文では共有空間での Speech Privacy に着目し、Target に加わるマスキング音の情報マスキング効果に関する文章了解度試験ならびに評価方法を開拓した。まず、エネルギーマスキングの代表とされる白色雑音と高い情報マスキング効果があると考えられる逆転音声を用いて、Target とマスキング音のエネルギー比(TM比(dB))を変えた状況での発話文章の書き取り試験によって音声了解度評価を行なった。その結果、エネルギーマスキングを引き起す白色雑音に比べ、逆転音声は顕著に音声了解度を低下させることが明らかになった。さらに従来のマスキング評価で用いられてきた包絡線保存度を示した。実験で用いた Target と TM 比ごとに付加されたマスキング音のサンプルを用いて明瞭度評価に重要となる狭帯域包絡線の相関係数を示すことで包絡線保存度と単語了解度との関係を示した。白色雑音は従来からの知見どおり相関値が1に近づくに伴い単語了解度も上昇しており、エネルギーマスキング効果を表すことが確認できた。一方、逆転音声では単語

了解度の変化に対応せず従来のマスキング包絡線保存度だけではマスキング効果が表せないことがわかった。そこで、SPEC²の相関を計算することによってターゲットの原音声とのスペクトルの類似性分析をおこなった。その結果、単語了解度とSPEC²相関の関連が強いことが示された。

情報マスキング効果に有効である逆転音声はマスキングシステム設計においてリアルタイムでは再生することができない。そこで本論文では白色雑音と逆転音声をマスキャーにしたときの単語了解度評価ならびに評価方法の検討からヒントを得て、新たなマスキングシステムの設計を目指した。音声の狭帯域包絡線特徴を保存しつつ定常的な特徴をもった包絡線マスキャー (EHM) を提案した。同様に TM 比 (dB) を変えて単語了解度試験を行なったところ、顕著に単語了解度を低下させることがわかった。さらに室内におけるマスキングシステムのモデルを提案し、上述したマスキャーにさらに環境特徴を含んだマスキング再生システムを開拓した。

以上、要するに個人性が追求される現代の近距離音場での音コミュニケーション環境において変化する、音の知覚と音場の物理的変化の関係を主観評価実験による尺度化と音の時間・周波数特徴に着目した分析をすることで明らかにした。

付録A 包絡線特徴に着目した情報マスキング音の設計と評価方法

A.1 はじめに

室内空間を伝わる音声をマスクするマスキングシステムに適したマスキング音の作成とその評価方法に関する基本原理とアルゴリズムを開拓する。これまでエネルギーマスキング効果の代表例として知られる白色雑音に音声波形の特徴を加味することによって、情報マスキング効果が期待出来るマスキング音の作成とその評価について検討した。その結果音声の了解性に深く関わっている時間波形の帯域別包絡線に加えて、音声の調波構造が示す時間変化にもまたマスキング効果を評価する上で重要な要因があることが明らかとなってきた。そこで音声波形の特徴に加えて空間を伝わる音の周波数特性、さらには暗騒音をも考慮したマスキング音の作成とその評価について検討を進める。

A.2 包絡線マスキング音 (EHM) 設計方法

音声に対するマスキング音の効果は基本的には音声の了解性の低下に表れる。妨害音の中に含まれる音声の了解性の劣化に関する研究では、図 A.1 から想像されるような、音声の狭帯域包絡線の変化が大きな劣化要因であることが明らかにされてきた。そこで、逆転音声にヒントを得た包絡線マスキング音を提案する。図 A.2 は音声の狭帯域包絡線を部分的に利用した音声変調雑音 (EHM) の作成

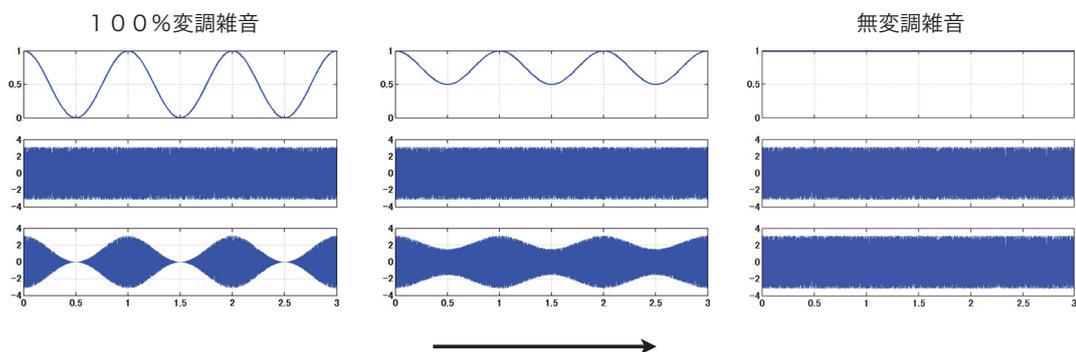


図 A.1: 残響音による包絡線の変化

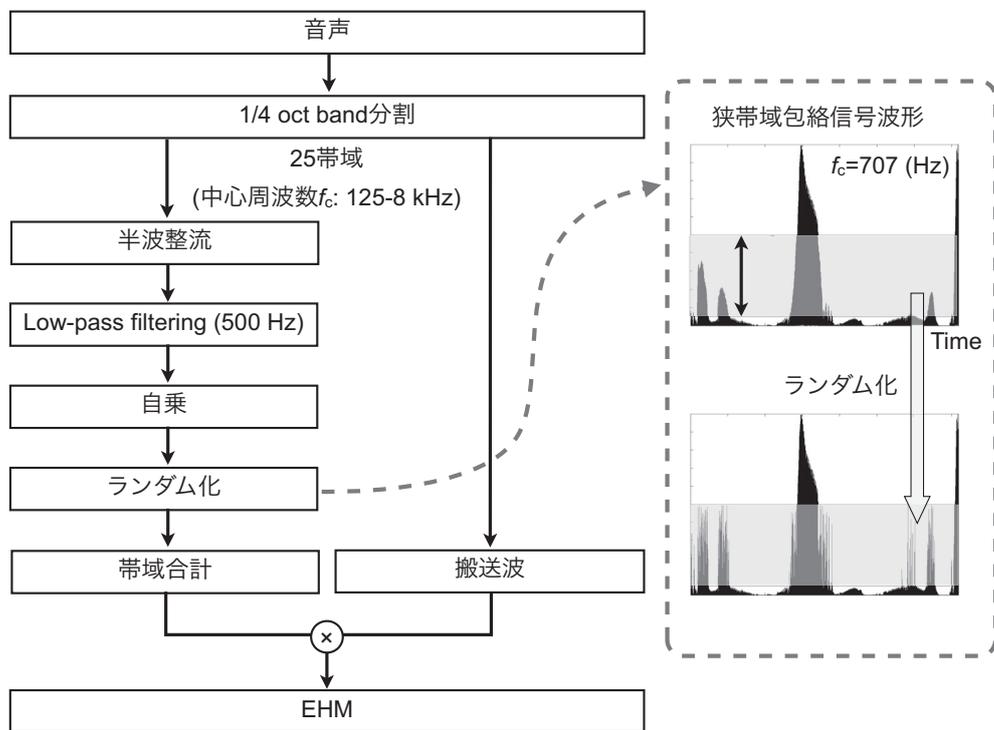


図 A.2: 音声包絡線を部分的に保存した音声変調雑音マスク (EHM) 作成方法

方法である。これは音声の包絡線に代表される非定期的特徴と広帯域雑音の定常性を混合した試験信号である。図に示すとおり音声 (Target とは異なる音声) のエネルギーの大きな部分と小さな部分の包絡線を保存し、その中間部分を包絡線のレベル分布を保存しながらランダム化したものである。音声合成に用いた搬送波はそれぞれの帯域における帯域雑音である。

図 A.3 は Target の音声信号の時間波形と、Target にマスキャーを TM 比 (dB)

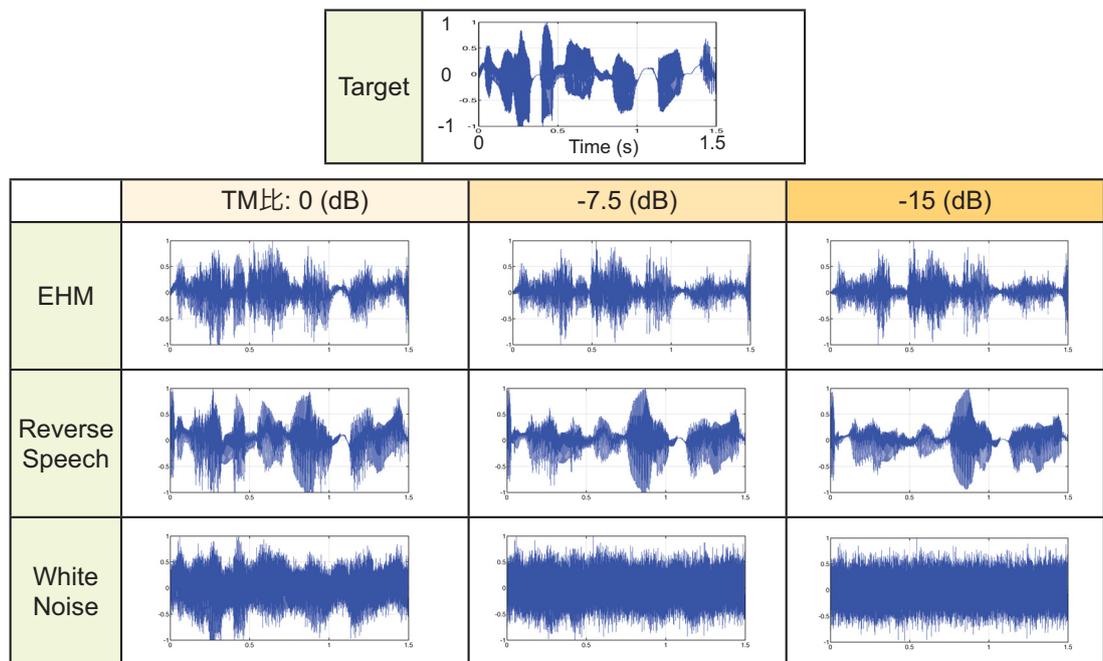


図 A.3: Target と TM 比 (dB) ごとのマスキャーの時間波形

ごとに重畳した、EHM を含む各マスキャーの時間波形である。EHM は音声の包絡線特徴を残しながら、定常性を持ったマスキャーであることが判る。

A.3 単語理解度評価試験

第 6 章、図 6.1 と同様の条件で EHM を含めた単語理解度評価試験をおこなった。

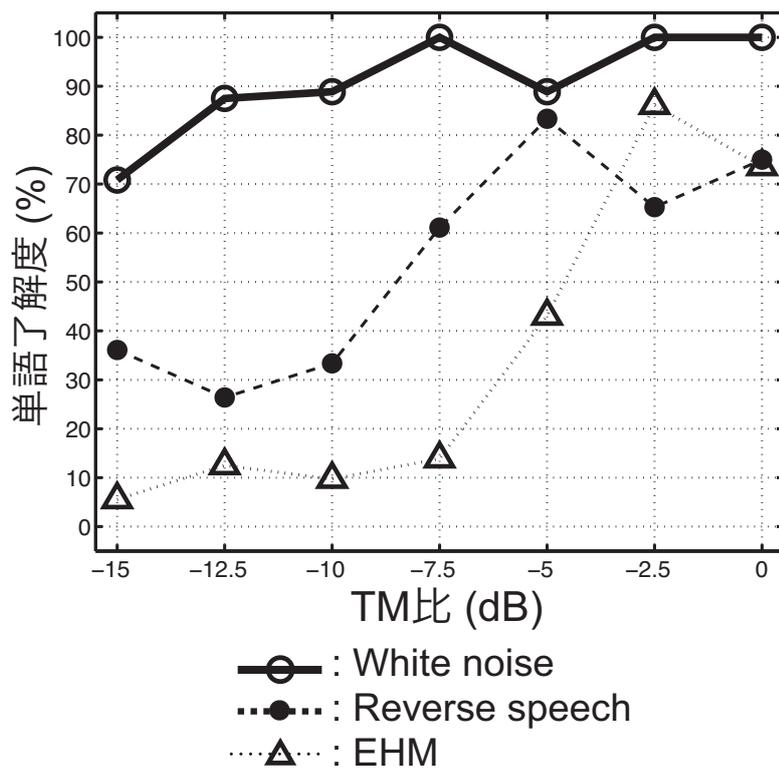


図 A.4: 単語了解度試験結果

A.3.1 実験結果

図 A.4 は単語了解度 (%) の結果である。TM 比が -10 (dB) 以下になると、音声の了解性が概ね失われることとなる。即ちマスクーのマスクング効果が顕著となる。しかしマスクーごとにマスクング効果をみれば、音声の帯域別包絡線に着目した音声変調雑音マスクーの効果が顕著であることが確認できる。

A.4 室内マスクングシステムの設計

マスクングシステムは一般に図 A.5 にモデル化される室内空間を伝わる音声 $y(n)$ を Target として設計される。そこでこれまで検討した基本マスクーを

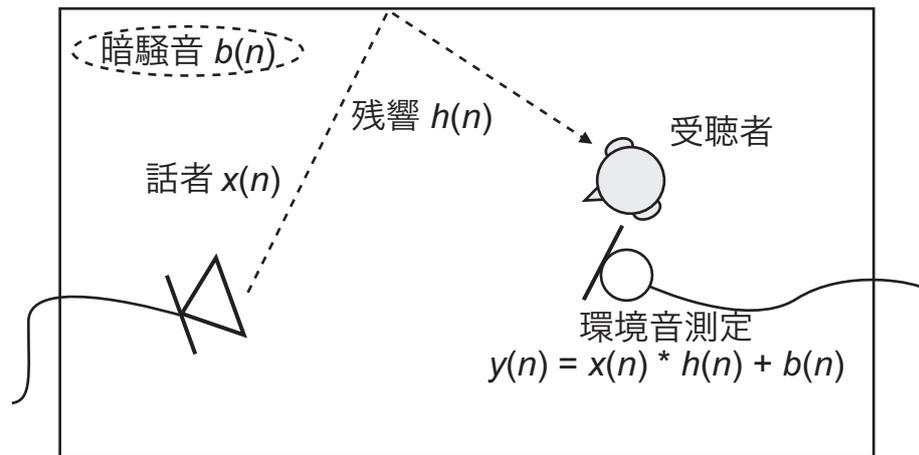


図 A.5: 室内マスキングシステムにおける Target

$y(n)$ の自己相関関数に畳み込みこみ，環境特徴を含んだスペクトル特徴を持たせることによって環境マスキングを作成した。

図 A.6 は環境音マスキングの室内マスキングシステム構成と生成アルゴリズムである。ただし，このシステムにおけるマスキング音の再生用スピーカの位置や環境音収音用マイクロホンの設置方法については，マスキング音が含まれた環境音を収音することによってハウリング現象を誘起することが考えられる。実環境における再生装置に関しては多点でスピーカを設置したり，Target 音声収音用のマイクロホンの設置などにも考慮が必要である。

A.5 室内マスキングシステムの評価

A.5.1 SPEC² 相関によるマスキング効果の評価

図 A.7 は基本マスキング (a) 並びに環境マスキング (b) に対する SPEC² 相関分析を図の横軸に示す TM 比において行ったものである。Target はいずれも図 A.5 に示した環境音声 $y(n)$ である。情報マスキング効果に重要な TM 比 0～7.5(dB) における SPEC² 相関の変化に着目すると，白色雑音に比べ，いずれの

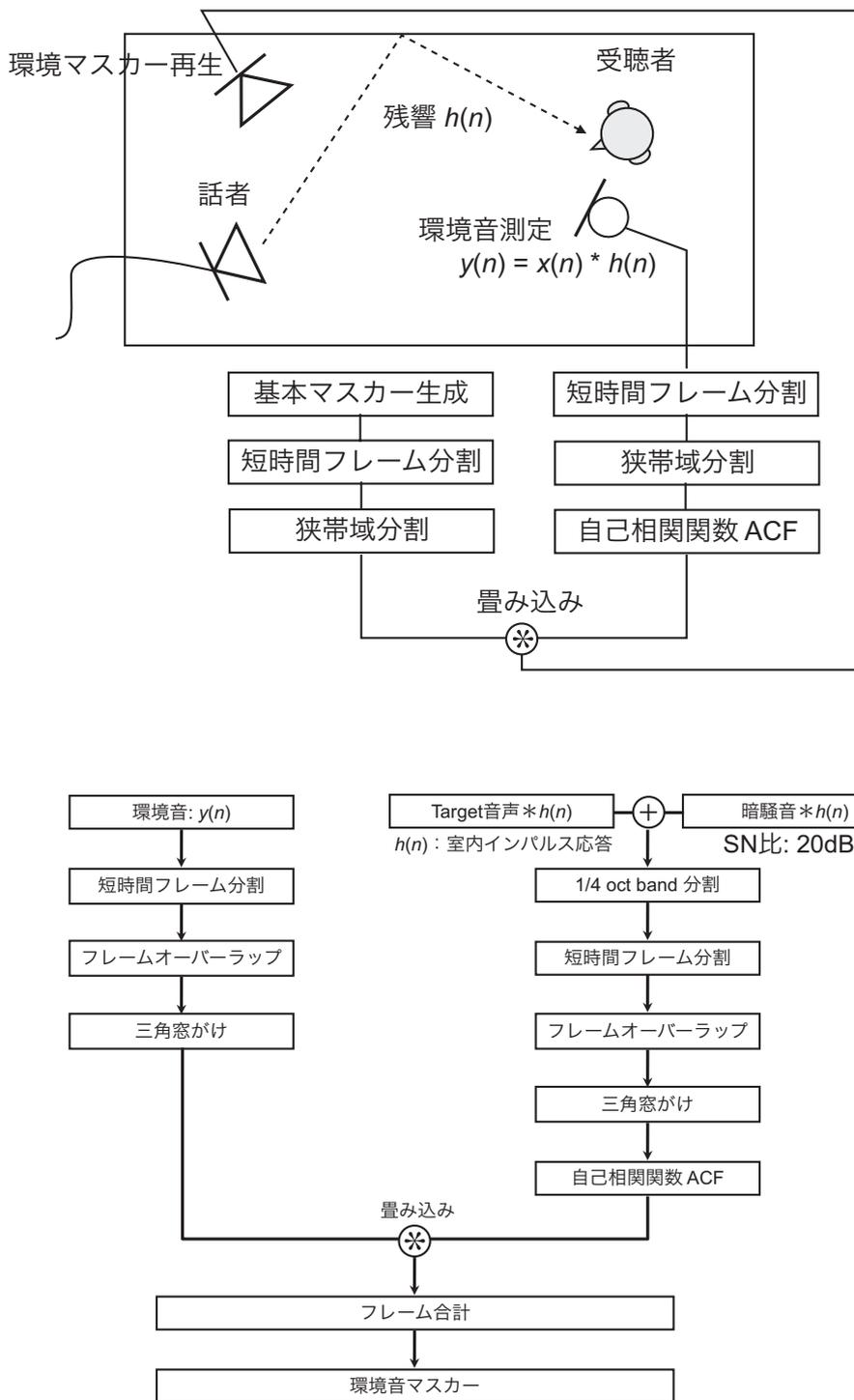


図 A.6: 環境マスキングを取り入れた室内マスクングシステム構成と作成アルゴリズム

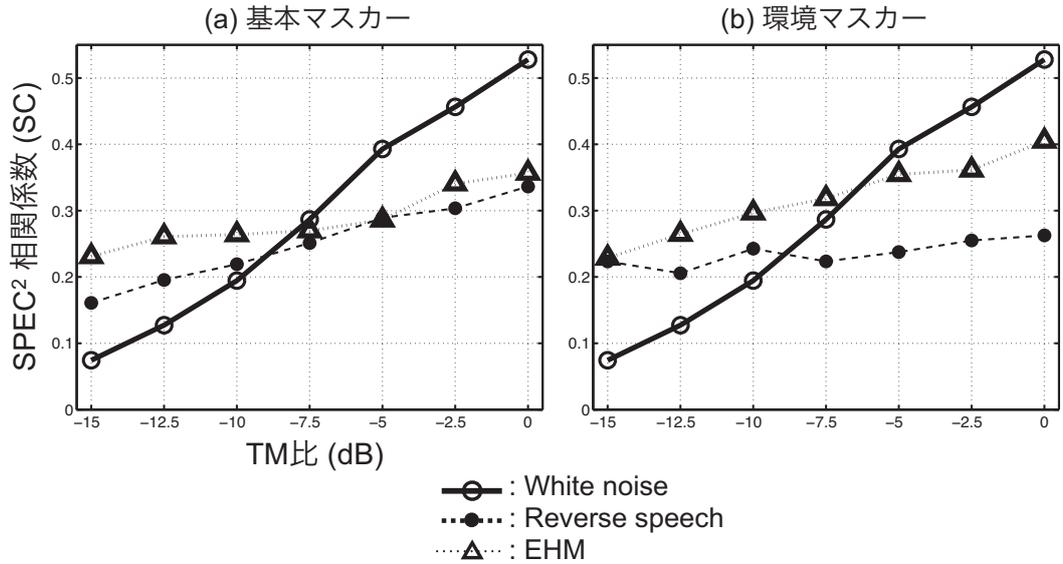


図 A.7: (a) 基本マスクーならびに (b) 環境マスクーに対する SPEC² 相関分析

マスクーにおいても SPEC² 相関が概ね低下することがわかる。また白色雑音では SPEC² 相関が TM 比の減少とともに低下するのに対して、いずれのマスクーにおいても SPEC² 相関が TM 比が変化しても大きく変化しないことも見てとることができる。これらの白色雑音との相違がマスクーに情報マスクング効果が期待できる要因である。

図 A.7(a)(b) の両図の差は基本マスクーと環境マスクーの間のマスクング効果の差を示している。環境音マスクー (EHM) は基本マスクーに比べて SPEC² 相関が若干高くなっていることがわかる。これは図 A.6 に示す生成アルゴリズムの通り、環境 Target 音声 $y(n)$ の自己相関関数 (ACF) が畳み込まれる事によって Target のスペクトル特徴がマスクーに含まれるようになり、Target 音声との相関が高くなると考えられる。一方、 $y(n)$ の逆転音声を用いた環境マスクーは、残響や暗騒音が含まれていない基本マスクーに比べ、SPEC² 相関が低くなっていることがわかる。

A.6 まとめ

本論文付録では，室内空間を伝わる音声をマスクすることに適したマスクの生成アルゴリズムを開拓した．白色雑音に代表されるようなTM比によって変化するエネルギーマスクング効果によるマスクングだけでなく，Target 音声の特徴を含んだマスクの生成方法を検討した．そこで音声波形の時間変化とスペクトル変化の特徴に加え，室内環境を伝わる音声の空間特徴を含んだ音声波形のスペクトル特性と暗騒音を含めたマスクを生成した．またエネルギーマスクング効果を表す帯域別包絡線に代る音声スペクトルの周波数軸上の自己相関関数 (SPEC²) 相関を求め，TM比によって変化するマスクング効果を明らかにした．

研究業績

論文／査読付国際会議

- ” *Effects of temporal and spectral factors of maskers on speech intelligibility* ”, Applied Acoustics, Vol.73(9), pp.893-899 (2012), Y. Hara, M. Tohyama, and K. Miyoshi (査読付論文)
- ”近接音場における狭帯域包絡線スペクトル分析”, 電子情報通信学会論文誌 A, Vol.J97-A(3), pp.221-223 (2014), 原佳史, 高橋義典, 三好和憲 (査読付論文)
- ” *Method for Estimating Pitch Independently from Power Spectrum Envelope for Speech and Musical Signal* ”, J. Temporal Des. Arch Environ., Vol. 9(1), pp.121-124 (2009), Y. Hara, M. Matsumoto, and K. Miyoshi

国際会議

- ” *Preference of the Transfer Functions for Music Recording in a Coherent Region of a Reverberant Field* ”, The 19th International Congress on Acoustics (ICA2007), CAS-01-007, Madrid Spain, Y. Hara, Y. Takahashi, H. Nomura, M. Tohyama and K. Miyoshi
- ” *Subjective Evaluation for Music Recording Positions in a Coherent Region of a Reverberant Field* ”, AES 2008 124th Convention, 7446, Amsterdam Netherland, Y. Hara, H. Nomura, M. Tohyama and K. Miyoshi

- ” *Perception of Sound Source Distance and Loudness in a Coherent Field of a Reverberant Field* ”, Acoustics '08, 2196, Paris France, Y. Hara, Y. Takahashi, H. Nomura, M. Tohyama and K. Miyoshi
- ” *Estimation of Masking Effects on Speech According to Spectral and Dynamical Characteristics of Maskers* ”, Inter Noise '09, 997, Ottawa Canada, Y. Hara, K. Miyoshi, M. Fujiwara, and Y. Shimizu
- ” *Method for Estimating Pitch Independently from Power Spectrum Envelope for Speech and Musical Signal* ”, The 4th International Symposium on Temporal Design (JTD2009), Kumamoto Japan, Y. Hara, M. Matsumoto, and K. Miyoshi

国内学会

- ”Coherent 領域における音源距離感知覚と伝達関数の最小位相成分”, 2006 年 電子情報通信学会 基礎・境界ソサイエティ大会, A-10-6, 原 佳史, 高橋 義典, 東山 三樹夫, 三好 和憲
- ”Coherent 領域における距離感知覚と伝達系の周波数特性”, 2007 年 電子情報通信学会総合大会, A-10-7, 原 佳史, 高橋 義典, 東山 三樹夫, 三好 和憲
- ”残響音場の Coherent 領域における音楽録音のための主観評価”, 2008 年 電子情報通信学会総合大会, A-10-11, 原 佳史, 野村 博昭, 東山 三樹夫, 三好 和憲
- ”信号の短時間自己相関と包絡線スペクトル分析による音声マスキングに関する検討”, 日本音響学会 2009 年春季研究発表会, 1-3-13, 原 佳史, 東山 三樹夫, 藤原 舞, 清水 寧

謝辞

博士論文研究を遂行するにあたり適切なお指導と議論、そして目標としていた音響工学についての有意義な研究生を送れる環境を与えてくださった工学院大学の三好和憲教授に感謝致します。Speech Privacy に関するマスキング音設計の検討ならびに際立ち音の可視化に関して、多大なお指導ご助言を頂いた東京工業大学の清水寧教授に心より感謝致します。音響工学研究における研究に関して日々御討論いただいた、東山三樹夫教授に感謝致します。授業やゼミを通して数学、統計学に関する適切な御指導をいただいた高橋静昭教授に感謝致します。マスキングシステムの設計並びに評価方法に関して御討論いただいたヤマハ株式会社の山川高史氏に感謝致します。論文の執筆及び実験、有意義な御議論と御指導及び実験に多大な協力をいただいた松本光雄氏、高橋義典氏、後藤理氏、工学院大学計算機構成研究室の皆に感謝致します。音楽理論と音作りに関するこだわり方について御指導をいただいた、河野亮氏に感謝致します。日々、セッションや創作活動を通して研究に対するアイデアとモチベーションを与えていただいた、バンドメンバー、音楽仲間へ感謝致します。学部生の頃から博士論文を執筆するまで、研究生生活をいつも傍で見守ってくれていた愛猫ポッポに感謝致します。最後に、夢であった音響工学についての研究生生活を支えていただいた、家族に感謝致します。

参考文献

- [1] ” *Communication Acoustics* ”, Springer, (2005), J Blauert
- [2] ” *Sound Waves in Rooms* ”, Rev. Mod. Phys, 16, pp. 69-150, (1944), P. Morse, and R. Bolt
- [3] ” 残響音場における位相周波数特性と直接音領域 ”, 電子情報通信学会論文誌, J89-A (4) pp.291-297, (2006), 高橋 義典, 東山三樹夫, 山崎 芳男
- [4] ” *Rendering Spatial Reverberation and Perception on Source Distance According to The Minimum-phase Characteristics* ”, Proc. AES 28th Int. Conf., pp. 203-210, (2006), Y.Takahashi, M.Tohyama and Y.Yamasaki
- [5] ” *Phase Response of Transfer Function and Coherent Field in a Reverberant Room* ”, Electronics and Communications in Japan, Part 3, 90(4), pp.1-8, (2007), Y. Takahashi, M. Tohyama, and Y. Yamasaki
- [6] ” *Temporal Envelope and Fine Structure Cues for Speech Intelligibility* ”, J. Acoust. Soc. Am., 97(1), pp.585-592, R. Drullman
- [7] ” *Speech Recognition with Primarily Temporal Cues* ”, Science, 270, pp.303-304, (1995), R. V. Shannon
- [8] ” *Predicting Speech Intelligibility in Rooms from Modulation Transfer Function. I. General Room Acoustics* ”, Acustica, 46, pp.60-72, (1980), T. Houtgast, H.J.M. Steeneken, and R. Plomp

- [9] " *A Review of the MTF Concept in Room Acoustics and its use for Estimating Speech Intelligibility in Auditoria* ", J. Acoust. Soc. Am., 77(3), pp.1069-1077, (1980), T. Houtgast, and H.J.M. Steeneken
- [10] " *Simulation of Mechanical to Neural Transduction in the Auditory Receptor* ", J. Acoust. Soc. Am., 79(3), pp.702-711, (1986), R. Meddis
- [11] " *Simulation of Auditory-Neural Transduction: Further Studies* ", J. Acoust. Soc. Am., 83(3), pp.1056-1063, (1988), R. Meddis
- [12] " *Virtual Pitch Phase Sensitivity of a Computer Model of the Auditory Periphery* ", J. Acoust. Soc. Am., 89(6), pp.2866-2893, (1991), R. Meddis, and M. J. Hewitt
- [13] " *A Unitary Model of Pitch Perception* ", J. Acoust. Soc. Am., 102(3), PP.1811-1820, (1997), R. Meddis and L. O'Mard
- [14] " *On the Significance of Phase in the Short Term Fourier Spectrum for Speech Intelligibility* ", J. Acoust. Soc. Am., 127, pp.1432-1439, (2010), M. Kazama, S. Gotoh, M. Tohyama, and T. Houtgast
- [15] " *Celebration: 50 Years of Speech Privacy* ", 151st ASA Meeting (2006)
- [16] " *Standard Test Method for Objective Measurement of Speech Privacy in Open Plan Spaces Using Articulation Index* ", Active Standard ASTM E1130
- [17] " *The Role of Perceived Spatial Separation in the Unmasking of Speech* ", J. Acoust Soc Am., 106(6), PP.3578-88, (1999), R.L. Freyman, K.S. Helfer, D.D. McCall, and R.K. Clifton

- [18] " *The Effect of Spatial Separation on Informational and Energetic Masking of Speech* ", J Acoust Soc Am., 112(5 Pt 1), pp.2086-2098, (2002), T.L. Arbogast, C.R. Mason, and G. Kidd Jr
- [19] " *On the Reduction of Masking Effects while Preserving Competing Binaural Audio Streams* ", Signals, Systems and Computers, Conference Record of the 37th Asilomar Conf., 1, (2003), C.D. Schmitz, and N. Iyer
- [20] " *Auditory masking phenomena in the perception of speech* ", pp.79-87, The representation of speech in the peripheral auditory system, R. Carlson and B. Granstroem eds. Elsevier Biomedical Press (1982), M. R. Schroeder, and S. Mehrgardt
- [21] " *Computer Speech* ", Springer, pp.127-130, (1999), M. R. Schroeder
- [22] " *New Method of Measuring Reverberation Time* ", J. Acoust. Soc. Am., 37(3), pp.409-412, (1965), M. R. Schroeder
- [23] " 信号解析と音響学 ", Springer Japan (2006), 東山 三樹夫
- [24] " 音響工学原論 ", コロナ社, (1957), 伊藤 毅
- [25] " *The Influence of a Single Echo on the Audibility of Speech* ", J. Audio Eng. Soc., 20 pp.146-159 (1972), H. Haas
- [26] " 音場評価の三つのパラメータについて ", 日本音響学会講演論文集, 1-5-16 pp.335-336 (1980), 平田 能睦
- [27] " *Concert Hall Acoustics* ", Springer-Verlag Berlin, (1985), Y. Ando,
- [28] " *Fundamentals of Acoustic Signal Processing* ", Academic Press, (1998), M. Tohyama, and T. Koike

- [29] ” *Richtungsverteilung und Zeitfolge der Schallrueckwuerfe in Raeumen* ”, *Acustica*, 3, pp.291-302 (1953), R. Thiele
- [30] ” *Acoustics of the Concert Hall* ”, *IEEE Spectrum*, pp.56-67 (1965), T. J. Schultz
- [31] ” リスニングルームの残響時間と再生音の響きの量 ”, *日本音響学会誌*, 33(7), pp.364-367, (1977), 平田 能睦
- [32] ” *Direct-to-Reverberant Energy Ratio Sensitivity* ”, *J. Acoust. Soc. Am.* 112 (5), pp.2110-2117 (2002), P. Zahorik
- [33] ” *Principles and Applications of Room Acoustics* ”, Applied science publishers. (1982), L. Cremer, and H. Muller
- [34] ” *Statistical Parameters of the Frequency Response Curves in Large Rooms* ”, *J. Audio Eng. Soc.* 35(5) pp.299-306 (1987), M. R. Schroeder
- [35] ” *Assessing auditory distance perception using virtual acoustics* ”, *J. Acoust. Soc. Am.*, 111(4) pp.1832-1846 (2002), P. Zahorik
- [36] ” *Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo* ”, *J. Acoust. Soc. Am.* 62, 1436 (1977), Y. Ando
- [37] ” *Law of Comparative Judgment* ”, *Psychol. Rev.*, 34(4), pp.273-286 (1927), T. T. Thurstone
- [38] ” *Psychophysical Analysis* ”, *Am. J. Psychol.*, 38, pp.368-389, (1927), T. T. Thurstone
- [39] ” 精神測定法 (秋重義治監訳) ”, 培風館, (1959), J.P. Guilford

- [40] ” 心理学的測定法 ”, 東京大学出版会, (1977), 田中良久
- [41] ” 一対比較法における主観的尺度値の信頼性について ”, 日放技学誌, 50(2), pp.176, (1994), 田畑 洋二 et al.
- [42] ” 一対比較法 (ケース V) における主観的尺度値の信頼性について ”, 日放技学誌, 51(4), pp.445-449, (1995), 田畑 洋二 et al.
- [43] ” *Empirical Formula for Creating Error Bars for the Method of Paired Comparison* ”, J. Electron. Imaging, 15(1), 010502.1-010502.3, (2006), E. D. Montag
- [44] ” *The Loudness War: Background, Speculation and Recommendations* ”, Proc. AES 129th Convention, 8175, (2010), E. Vickers
- [45] ” *Algorithms to measure audio programme loudness and true-peak audio level* ” ITU-R BS.1770
- [46] ” *Requirements for loudness and true-peak indicating meters* ”, ITU-R BS.1771
- [47] ” *Loudness normalisation and permitted maximum level of audio signals* ”, EBU R128
- [48] ” デジタルテレビ放送番組におけるラウドネス運用規定 ”, ARIB TR-B32
- [49] ” *Denon Professional Test CD* ”, Track9. II. Anechoically Recorded Samples For Evaluation: No.6 - *Water Music Suite*, Bars 1-11, Track12. II. Anechoically Recorded Samples For Evaluation: *Symphony No.4 In E Minor*, Op.98, Bars 354-362

- [50] "Loudness, Its Definition, Measurement and Calculation", J. Acoust. Soc. Am. 5, pp.82-108 (1933), H. Flecher, and W. A. Munson
- [51] "Acoustics – Method for Calculating Loudness Level", ISO 532, (1975)
- [52] "Loudness constancy with varying sound source distance", Nature Neuroscience 4(1) pp.78-83 (2001), P. Zahorik, and F. L. Wightman
- [53] "Signal Representation Including Waveform Envelope by Clustered Line-Spectrum Modeling", J. Audio Eng. Soc., 51(3), 123-137 (2003), M. Kazama, K. Yoshida, and M. Tohyama
- [54] "Auditory and Visual Sensations", Springer-Verlag, (2009), Y. Ando
- [55] "Evaluation of Speech Transmission Channels by Using Artificial Signals", Acustica, 25, pp.355-367 (1971), T. Houtgast, and H.J.M. Steeneken
- [56] "Sound and Signals", Springer-Verlag, (2011), M. Tohyama
- [57] "From Speech Privacy to Speech Security", Proc. of ICA2007, PL-005. (2007), J. S. Bradley
- [58] "Speech Privacy Between Neighboring Workstations in an Open Office - A Laboratory Study", Acustica, 93, 771-782, (2007), P. Virjonen, J. Keränen, R. Helenius, J. Hakala, O. V. Hongisto
- [59] "執務空間における音環境の最適化技術", 電子情報通信学会誌, 96(8), 643-648 (2013), 清水 寧
- [60] "スピーチプライバシー研究の歴史と近年の動向", 日本音響学会誌, 64(8), 475-480 (2008), 佐藤 洋, 清水 寧

- [61] ”スピーチプライバシーに関する北米の動向と課題について”, 建築音響研究会資料, AA2008-55, (2008), 清水 寧
- [62] ”時間反転残響における音声明瞭度の低下”, 日本音響学会聴覚研究会資料, 35(4), pp.237-242, (2005), 荒井 隆行
- [63] ”*What are the Essential Cues for Understanding Spoken Language?*”, IEICE Trans. on Information and Systems, E87-D(5), pp.1059-1070, (2004), S. Greenberg, and T. Arai
- [64] ”*Energetic and Informational Masking Effects in an Audiovisual Speech Recognition System*”, IEICE Trans. on Audio, Speech, and Language Processing, 17(3), pp.446-458 (2009), J. Barker, and X. Shao
- [65] ”*Using HDC to Evaluate Signal Similarity for Information Masking*”, Applied Acoustics, 70(5), pp.689-694. (2009), S. Gotoh, M. Matsumoto, and Y. Yamasaki
- [66] ”*How much masking is informational masking?*”, J. Acoust. Soc. Am., 88, pp.2607-2610, (1990), R. Lutfi