修 士 論 文

音声波形の振幅包絡線に含まれる個人性の検討

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

朱　治

2015 年 3 月

修 士 論 文

音声波形の振幅包絡線に含まれる個人性の検討

指導教員　鵜木 祐史 准教授

審査委員主査　鵜木 祐史 准教授
審査委員　赤木 正人 教授
審査委員　党 建武 教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

1310029 朱　治

提出年月: 2015 年 2 月

概 要

人の声には，言語情報だけでなく，その人の感情や性別，年齢，健康などの様々な非言語的情報が含まれている．その中でも，話者の特徴（個人性）は非常に重要であり，人はこれを利用して音声を聞くだけで話者を弁別することができる．また，複数の話者が同時に話している状況で特定の話者の音声だけを聴き取る能力（カクテルパーティ効果）にも個人性が重要な役割を果たしている．このように，音声の個人性は音声コミュニケーションにおいて非常に重要な意味を持っている．しかし，人間の聴知覚メカニズムにおいて，どのような特徴が個人性を決定づけるのか未だ完全に明らかになっていない．

音声の個人性が生ずるのは，基本的に音声を生成する発声器官（声帯と声道）が人によって異なることが原因である．更に，発声器官の先天的な形の違いと後天的な動く形の違いが両方とも個人性の知覚に貢献していると考えられる．その考えから，音声の個人性知覚に関する先行研究では，主に音源フィルタ理論に基づいた音声分析合成系を利用し，基本周波数の特徴（声帯音源特性）あるいはスペクトル包絡形状の特徴（声道フィルタ特性）に着目している．更には，各特徴を静的成分と動的成分に分けて検討されている．

一方，風間ら (2009) は，これまでの研究で着目されてきた声帯特性や声道特性ではなく，音声の狭帯域振幅包絡線に含まれる個人性情報に着目し，話者依存の狭帯域振幅包絡線帯域間相関行列（ECM）の相違によって発話内容に依存しない個人性情報の表現が可能であると報告した．しかし，この ECM は帯域間の相関を表しているだけであるため，狭帯域振幅包絡線にあるどのような特徴が個人性に寄与しているのかは分かっていない．また，機械による話者認識性能を評価しているだけであるため，そもそも人が狭帯域振幅包絡線に含まれる個人性情報を話者の違いとして知覚しているのかどうかも分らない．

聴知覚の知見から，振幅包絡線は音声知覚にとって非常に重要な要因と知られている．Drullman ら（1994）は振幅包絡線の変調成分から音声了解度への影響を調査した結果，変調周波数の 4 から 16 までの変調成分は音声了解度に重要であると報告している．更に，Shannon ら（1995）は振幅包絡線情報のみ含まれている雑音駆動音声でも音声の言語情報の知覚が十分に可能であることを報告している．

本研究の目的は，音声波形の振幅包絡線に含まれる個人性情報を聴覚の知覚メカニズムに関連づけて明らかにすることである．そこで，本研究では「話者による違いの大きな物理量が音声の個人性の知覚に寄与している」という仮説をおき，音声波形の振幅包絡線における物理的な個人差の顕著なところを調べ，その物理量と個人性の知覚の関係を調査する．

その第一歩として，複数話者の変調スペクトル間の個人差を分析することで，個人性の情報が含まれていると思われる変調スペクトル帯域を推定した．その結果，20 ERB_N-number 以上の周波数帯域の変調スペクトルの話者間分散が大きいことが分った．また，20 ERB_N-number から 29 ERB_N-number までの変調スペクトルでは，15 Hz 以下の変調周波数帯域の分散がより大きく，30 ERB_N-number 以上の周波数帯域では，全変調周波

数帯域の分散が大きいことが分かった．物理的な差が大きい変調スペクトルを知覚的にも利用していると考えると，これらの変調周波数帯域に個人性情報が含まれていると考えられる．

次に，変調スペクトルの帯域を制限した雑音駆動音声を用いて，その変調成分の変化が個人性知覚に及ぼす影響を調査した．34 帯域と 17 帯域の二種類の雑音駆動音声を利用し，各帯域の振幅包絡を低域通過フィルタにより制御した．XAB 法による個人性知覚実験の結果，34 帯域の場合は変調周波数帯域の上限周波数が低くなると話者弁別率の平均値が低くなるが，分散が大きいため分散分析の結果から有意差は認められなかった．17 帯域の場合は，変調周波数帯域の上限周波数が 8 Hz から低くなるに従って，話者弁別率が有意に下がっており，1 Hz になると話者弁別率が 60 %前後になることが分かった．

これらの結果から，振幅包絡線の変調周波数約 15 Hz 以下の変調帯域に個人性情報が顕著に表れており，音声の個人性知覚に寄与していることが明らかになった．

目 次

第 1 章　序論 **1**
1.1　はじめに 1
1.2　研究背景 1
1.3　本研究の目的 3
1.4　本論文の構成 5

第 2 章　音声の振幅包絡線と音声知覚との関係 **7**
2.1　まえがき 7
2.2　変調周波数成分による音声知覚の影響 7
2.3　振幅包絡線に現れる話者情報 9
2.4　雑音駆動音声 9
2.5　本研究の着目点 10

第 3 章　音声の変調スペクトルに現れる個人差の分析 **11**
3.1　目的 11
3.2　音声データと変調スペクトルの算出法 11
3.3　音声データおよび文章間の分散と話者間の分散 16
3.4　分析結果 16
3.5　考察 21

第 4 章　音声の変調成分の制限が個人性の知覚に与える影響 **22**
4.1　本実験の目的 22
4.2　実験用データベース 22
4.3　実験参加者 24
4.4　刺激音 24
4.5　実験機器 26
4.6　実験方法 27
4.7　実験結果 28
4.8　考察 32

第 5 章　全体考察 **33**

第 6 章　結論 **34**
6.1　本研究で明らかとなったこと 34
6.2　残された課題 34

参考文献 **35**

謝辞 **39**

研究業績 **40**

付録 **43**

図 目 次

1.1 本論文の構成 6

2.1 狭帯域信号の振幅包絡線と時間微細構造（各帯域の中心周波数：$fc_a = 55; fc_b = 603; fc_c = 1963$） 8
2.2 雑音駆動音声の生成法の概要 10

3.1 変調スペクトルの算出法（複数の帯域が点線で省略） 12
3.2 変調スペクトルの例（話者 F101，文章 A01） 15
3.3 文章に関する変調スペクトルの分散. 17
3.4 話者に関する変調スペクトルの分散（全話者）. 18
3.5 話者に関する変調スペクトルの分散（男性）. 19
3.6 話者に関する変調スペクトルの分散（女性）. 20

4.1 実験用データベースの変調スペクトルにおける話者間の分散 23
4.2 刺激音作成のブロックダイアグラム 25
4.3 実験環境 26
4.4 刺激パターン 27
4.5 34 帯域における実験結果 29
4.6 17 帯域における実験結果 31

6.1 話者 F101 の文章 A01 から A06 までの変調スペクトル 44
6.2 話者 F102 の文章 A01 から A06 までの変調スペクトル 45
6.3 話者 F103 の文章 A01 から A06 までの変調スペクトル 46
6.4 話者 F104 の文章 A01 から A06 までの変調スペクトル 47
6.5 話者 F105 の文章 A01 から A06 までの変調スペクトル 48

表 目 次

3.1 各帯域通過フィルタの遮断周波数 13

4.1 等分散性の検定（34 帯域） 28
4.2 分散分析（34 帯域） 28
4.3 等分散性の検定（17 帯域） 30
4.4 分散分析（17 帯域） 30

第1章　序論

1.1　はじめに

人の声には，言語情報だけでなく，その人の感情や性別，年齢，健康などの様々な非言語的情報が含まれている [1][2][3]．その中でも，話者の特徴（個人性）は非常に重要であり，人はこれを利用して音声を聞くだけで話者を弁別することができる．また，複数の話者が同時に話している状況で特定の話者の音声だけを聴き取る能力（カクテルパーティ効果）にも個人性が重要な役割を果たしている [4]．このように，音声の個人性は音声コミュニケーションにおいて非常に重要な意味を持っている．

もし，音声には個人性がなくなると，みんなの声が同じ人のように聞こえられる世界はどんなにつまらないものなってしまうことでしょう．脳機能障害の一種に, 話者の識別ができない phonagnosia と呼ばれる症例がある [5]. この患者は話者の識別はできないが, 音声の言語情報の知覚には問題がない．このような患者の存在は，脳の中で音声の個人性が音声の言語情報を理解するメカニズムとは異なる独立のメカニズムで処理されていることを明示している. しかし，音声知覚の研究に比べて個人性に関する研究は少なく現状では，人間の聴知覚メカニズムにおいて，どのような特徴が個人性を決定づけるのか未だ完全に明らかになっていない．

音声は，声帯で生じた音源波が声道を通過し，空気中に放射されることによって生成される．したがって，音声の個人性情報が声帯音源や声道の音響的特性における話者間の差異に由来すると考えられる．従来研究の多くも発声器官の形状の違いに由来する音響的特徴に着目している．しかし，聴覚的な面から音声の個人性情報は聴覚メカニズムにどのように処理されているのかがほどんと検討されていない．“人はどのように話者を判断しているのか”は音声科学の基本的且つ興味深い課題である．

音声の個人性知覚に寄与している要因を解明できれば，個人性知覚メカニズムの解明に繋がるものである．また，工学的場面では，機器による話者認識のほかに，話者に適応する音声認識システムや個人性を加えた音声を生成できる音声合成法などさまざまな音声処理技術に応用することが可能である．

1.2　研究背景

音声に個人性が生ずるのは，基本的に音声を生成する発声器官（声帯と声道）が人によって異なることが原因である．更に，発声器官の先天的な形の違いだけでなく，後天的

な発話器官の動かし方の違いも個人性の知覚に貢献していると考えられる．その考えから，音声の個人性知覚に関する先行研究では，主に音源フィルタ理論に基づいた音声分析合成系を利用し，基本周波数の特徴（声帯音源特性）あるいはスペクトル包絡形状の特徴（声道フィルタ特性）に着目している．更には，各特徴を静的成分と動的成分に分けて検討されている．

音声の個人性知覚要因となる音響特徴量は，これまでに数多く調べられてきている．伊藤ら（1982）[6] は個人性知覚に影響のあるパラメータが，スペクトル包絡特性，基本周波数，発話時の時間特性（テンポ）の順に大きく，特にスペクトル包絡特性の影響が大きいと報告した．橋本ら（1998）[7] は聴取実験で基本周波数，スペクトル，音素継続時間の三つの音響的特徴の個人性知覚への寄与率を求め，個人性知覚の予測モデルを構築した．その結果，スペクトル包絡，基本周波数は顕著な寄与が認められ，寄与度は音響的特徴の差に依存していることが報告され，伊藤ら [6] と同様の結果を示している．Kasuya ら（1996）[8] は ARX モデルに基づいて，声道の静的特徴と動的特徴が個人性の知覚への寄与を統合的に検討し，動的成分よりも静的成分の寄与が大きかったと報告した．北村ら（1998）[9] は音声のスペクトル遷移パターンの変形が個人性知覚に与える影響を調査した．その結果，スペクトル遷移パターンが話者識別に与える影響は小さい，Kasuya らと同じく音声の動的成分よりも静的成分のほうが話者識別への寄与が大きいという結果が得られた．

声道と声帯のそれぞれの特徴に個別的に着目している研究もある．桑原ら（1986）[10] はホルマント周波数とバンド幅を独立制御できる分析合成システムを構築し，持続５母音を対象に，第３以下のホルマント周波数のシフトが個人性知覚に影響がより大きく，特に F3 が最も重要であることを示した．北村と赤木（1997）[11][12][13] は単母音のスペクトル包絡に着目し，1740 Hz 以上のスペクトル包絡成分の高域に個人性がより多く現れていると報告した．また，スペクトル包絡成分の高域に存在する dip よりも peak が特に重要な意味を持っていることを示唆した．生理学の知見から，Kitamura ら（2005）[14] はスペクトル包絡の高域が発話中にあまり動かない下咽頭腔の形状に由来することを明らかにしている．Amino ら（2006）[15] は，それは鼻腔や鼻咽腔などの調音器官の生理学的特徴に現れる個人差により，鼻音が話者識別に有効である結果を示した．Akagi & Ienaga（1997）[16] は 3 モーラ単語における基本周波数の軌跡が個人性知覚に寄与することを示唆した．

このように，これまでの個人性に関する研究は，主に音声生成の面から周波数成分に含まれる個人性情報に着目している．一方，伊藤らの研究 [6] は，音声を逆再生すると個人弁別が難しくなることも示している．この知見は，音声のスペクトル情報だけでなく，音声波形の時間的変動も個人性知覚に寄与していることを示唆している．しかし，振幅の時間変動，特に振幅包絡の変調成分に含まれる個人性情報に着目した研究はほとんどなかった．近年，風間ら (2009) [17] は，これまでの研究で着目されてきた声帯特性や声道特性ではなく，音声の狭帯域振幅包絡線に含まれる個人性情報に着目し，話者依存の狭帯域振幅包絡線帯域間相関行列（ECM）の相違によって発話内容に依存しない個人性情報の表

現が可能であると報告した．しかし，この ECM は帯域間の相関を表しているだけであるため，狭帯域振幅包絡線にあるどのような特徴が個人性に寄与しているのかは分かっていない．また，機械による話者認識性能を評価しているだけであるため，そもそも人が狭帯域振幅包絡線に含まれる個人性情報を話者の違いとして知覚しているのかどうかも分らない．

聴知覚の知見から，振幅包絡線は音声知覚にとって非常に重要な要因と知られている．ヒトの聴覚末梢系には，音声信号を聴覚フィルタバンクでいくつかの帯域制限した信号に分割する機能（周波数分解機能）がある．さらに内有毛細胞や神経発火のメカニズムによって，帯域分割した信号を半波整流し低域通過フィルタに通すのと同等の処理が行われている [18]．つまり，ヒトの聴覚末梢系では，音声を帯域分割し，各帯域の振幅包絡線を検出するというプロセスで処理を行っている．さらに，聴知覚メカニズムで振幅包絡線情報を処理するときに変調周波数の選択特性（変調フィルタバンク）の存在が Dau ら（1997）[25] の研究から示唆されている．

Drullman ら（1994）[19][20] は振幅包絡線の変調成分から音声了解度への影響を調査した結果，変調周波数の 4 から 16 までの変調成分は音声了解度に重要であると報告した．更に，Shannon ら（1995）[21] は振幅包絡線情報のみ含まれている雑音駆動音声を用いて，音声知覚との関係を調査した．雑音駆動音声ということは，音声信号を複数の帯域に分割し，各帯域の振幅包絡線に基づきその帯域に制限された雑音を変調して合成された音声である．Shannon らの研究によると，その帯域の数が最低 4 つあれば音声の言語情報の聞き取りが可能となる．この雑音駆動音声は人工内耳着用者が聞こえている声を模擬する音声で，人工内耳に関する研究ではよく利用されている．Vongphoe & Zeng（2005）[22] は雑音駆動音声を用いて人工内耳着用者が健聴者より個人性の知覚が難しいことを報告した．Gonzalez & Oliver（2005）[23] は雑音駆動音声の帯域の数と個人性の知覚の関係を調査した結果，帯域の数が多くほど話者の弁別は容易になることを報告した．さらに，Krull & Luo（2012）[24] は，訓練により雑音駆動音声の話者弁別の成績を上げることができると報告した．

これらの研究によると，振幅包絡線に物理的な個人性情報が含まれており，振幅包絡線情報だけでも個人性の知覚が可能であることを示している．しかし，その個人性情報が一体どのような形で現れているのか，またどのようにヒトの個人性知覚に影響しているのかが分かっていない．本研究では，振幅包絡線の周波数成分すなわち変調成分と音声個人性の知覚の関係を調査する．

1.3 本研究の目的

本研究の目的は，音声波形の振幅包絡線に含まれる個人性情報を聴覚の音声知覚メカニズムに関連づけて明らかにすることである．そこで，本研究では「話者による違いの大きな物理量が音声の個人性の知覚に寄与している」という仮説をおき，音声波形の振幅包絡線における物理的な個人差の顕著なところを調べ，その物理量と個人性の知覚の関係を調

査する．

まず，複数話者の変調スペクトルを算出し話者間の分散を求めることにより，音声の振幅包絡線における個人差を分析する．変調スペクトルとは，音声信号を帯域分割し各帯域の振幅包絡線をフーリエ変換により算出された振幅包絡線の周波数スペクトルに相当するものである．ただし，振幅包絡線の周波数は一般的に変調周波数と呼ばれている．そこで個人差の大きな帯域を見つけ出し，音声の個人性の知覚に影響していると予想される物理的な特徴を調査する．

次に，以上の結果に基づき，振幅包絡線を低域通過フィルタによって制限した雑音駆動音声を用いた個人性知覚実験を行い，個人性判断の手がかりを絞り込む．雑音駆動音声とは各帯域の振幅包絡線に基づき，その帯域に制限された雑音を振幅変調して合成されるため，振幅包絡線情報だけが残されている．低域通過フィルタにより振幅包絡の変調成分を制限し，個人性の知覚に影響している変調周波数の帯域を調査する．

最後に物理的な個人差の変調成分と心理的に個人性の知覚に寄与している変調成分を関連付けて，振幅包絡線に含まれる個人性情報を検討する．

1.4　本論文の構成

本論文は，6 章で構成される．

第 1 章

この章では，音声の個人性における研究の背景と問題点を述べる．更に，音声の振幅包絡線情報に関する背景を説明することによって，本研究の目的を明らかにする．

第 2 章

本研究は音声の振幅包絡線情報に着目している．この章では，振幅包絡線または変調スペクトルと音声知覚の関係を説明する．そして，振幅包絡線に個人性が含まれていることを示唆した研究を述べる．これで，本研究の着目点に至る経緯を説明する．

第 3 章

話者による物理的差が大きなところが人の話者認識に利用されていると考えられる．そのため，第 3 章では変調スペクトルにおける話者間の分散を算出することで，振幅包絡線の話者による物理的違いの大きな帯域を見つけ出す．

第 4 章

この章では，第 3 章の結果に基づいた個人性知覚実験を述べる．実験では，雑音駆動音声を利用して低域通過フィルタにより振幅包絡線を制御する．それによって，変調スペクトルの帯域制限と個人性の知覚の関係を検討する．

第 5 章

この章では，本研究における物理的な差と聴取実験で得られた心理的な差の考察を述べる．更に，その結果を従来研究と関連付けて全体考察を述べる．

第 6 章

本研究で得られた結果を要約し，今後の展望を述べる．

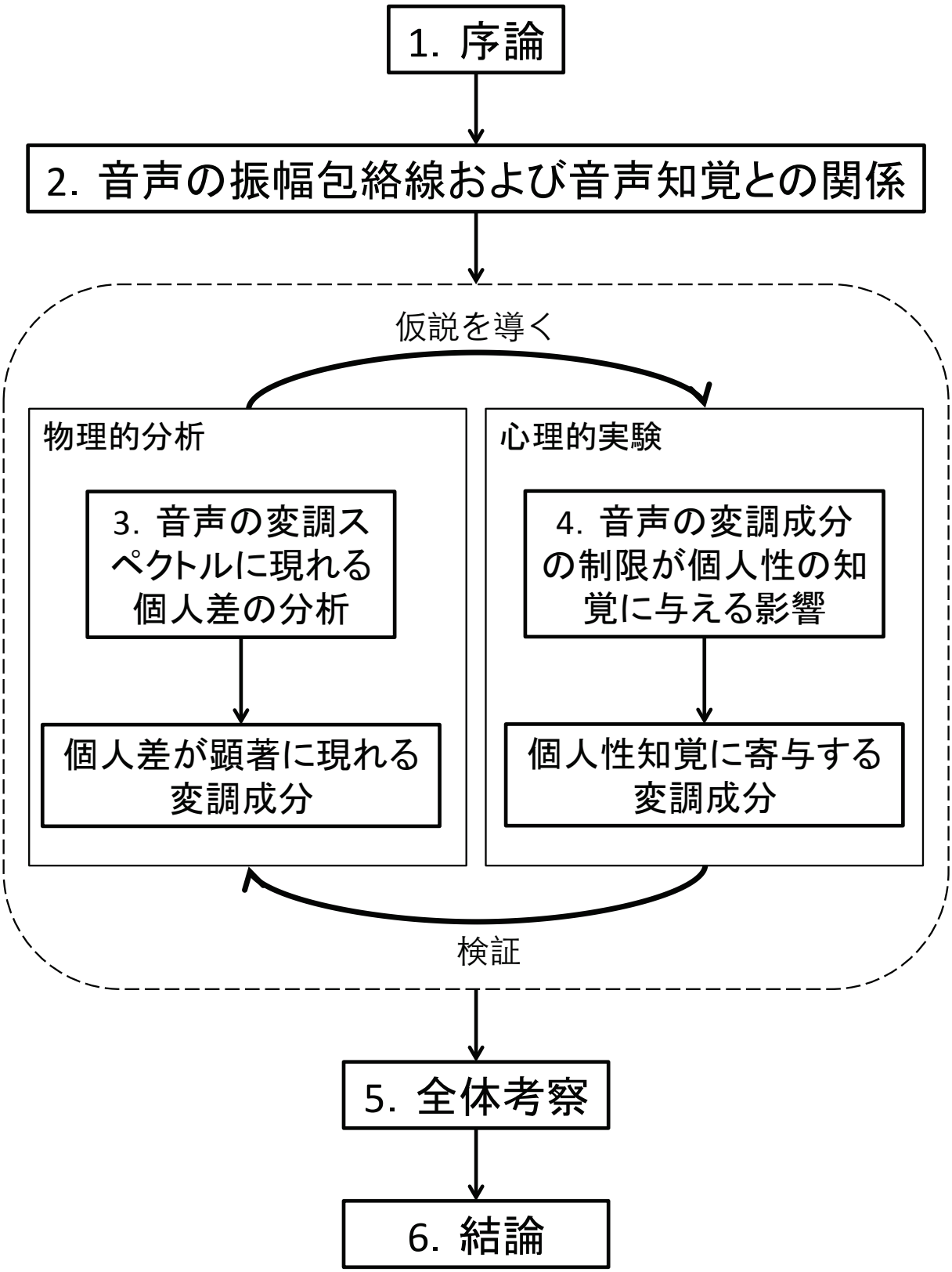
1. 序論
2. 音声の振幅包絡線および音声知覚との関係
仮説を導く
物理的分析
3. 音声の変調スペクトルに現れる個人差の分析
個人差が顕著に現れる変調成分
心理的実験
4. 音声の変調成分の制限が個人性の知覚に与える影響
個人性知覚に寄与する変調成分
検証
5. 全体考察
6. 結論

図 1.1: 本論文の構成

第2章　音声の振幅包絡線と音声知覚との関係

2.1　まえがき

音声知覚には，声道形態を表すスペクトル包絡や音源振動を表す基本周波数に代表される周波数情報が重要な要因であると以前から報告されてきた．しかし，Drullman ら [19] の研究から振幅包絡の変調周波数成分も重要な手掛かりであることが知られた．さらに，ホルマントや基本周波数のような周波数情報が欠落し，振幅包絡線情報だけが残された条件下でも，音声知覚が十分に可能であることが，Shannon ら [21] の雑音駆動音声の研究で明らかになってきた．本研究では，振幅包絡線の変調成分の情報が音声知覚だけでなく，音声に含まれている個人性情報の知覚にも寄与していると考えている．そこで，本章では振幅包絡線と音声知覚の関係を詳しく述べ，振幅包絡線に個人性情報が含まれていることを示唆した研究について説明する．

2.2　変調周波数成分による音声知覚の影響

蝸牛が音声信号を処理する時には，最初に聴覚フィルタバンクにより音声信号を複数の帯域に分割する．各帯域に制限された信号は振幅包絡線と時間微細構造に分けることができる．図 2.1 に，音声信号「あおい」を三つの異なる中心周波数を持つ帯域通過フィルタに通した出力を示す．上から順番に，各帯域の中心周波数は 55, 603, 1963 Hz である．図中の赤い線は波形のピークを縁取った振幅包絡線を示している．この振幅包絡線にさらにフーリエ分析を行うことで得られた振幅包絡線の周波数スペクトルは変調スペクトルと呼ばれている．また，一つ一つの周波数成分は変調周波数成分と呼ばれている．

振幅包絡線または変調スペクトルと音声知覚の関係については Dudley が 1939 年に開発した音声合成システム「VOCODER」[26] から初めて知られた．彼は,「VOCODER」を用いて振幅包絡線の 25 Hz 以上の変調周波数成分をフィルタで除去しても，音声了解度に支障がないことを示し，低変調周波数成分が音声の知覚にとって重要であることを示唆した．さらに，Drullman ら [19] は直接に変調周波数成分から音声了解度への影響を調査した．Drullman らの手法では，まず音声信号を一定のオクターブ距離で複数の帯域に分割した．次に各帯域の振幅包絡線を低域通過フィルタに通してもとの時間微細構造を振幅変調し，音声を再合成した．その再合成した音声を雑音の環境で被験者に呈示し，音声の言

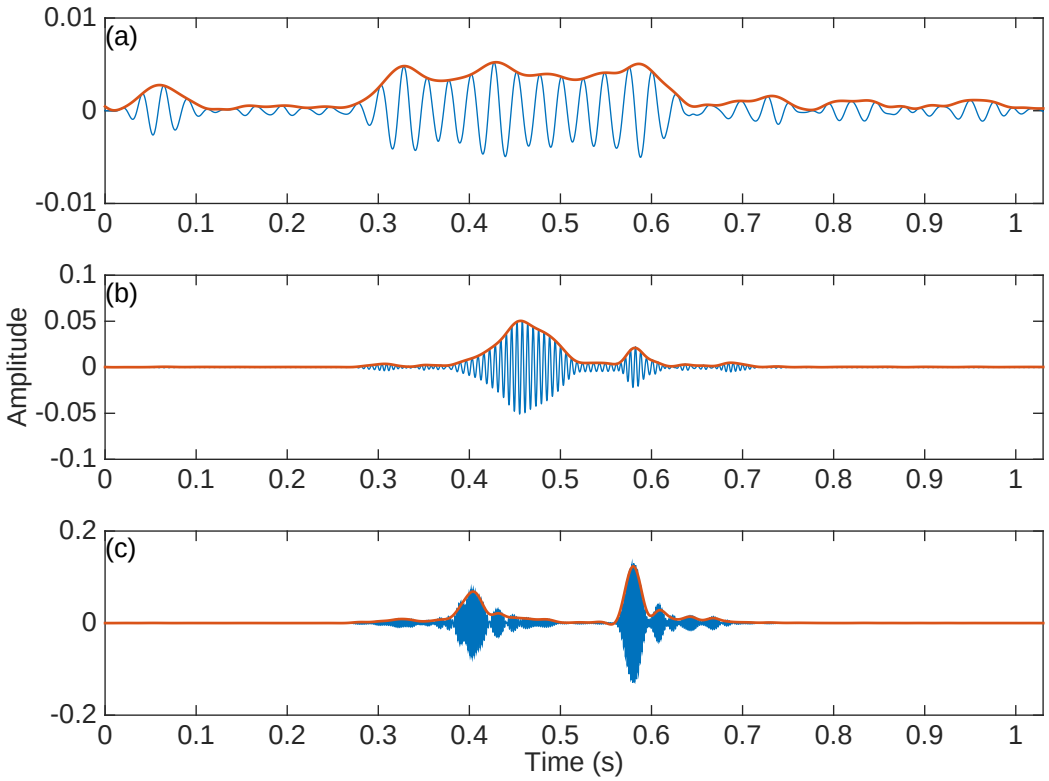

図 2.1: 狭帯域信号の振幅包絡線と時間微細構造（各帯域の中心周波数：$fc_a = 55; fc_b = 603; fc_c = 1963$）

語情報を完全に聞き取ることができる雑音の音圧レベルの閾値いわゆる speech-reception threshold (SRT) を測定した．その結果，低域通過フィルタのカットオフ周波数が 16 Hz 以上では SRT の値にあまり変化がないが，16 Hz 以下になると SRT が急激に上昇することが分かった．つまり，16 Hz 以下の変調周波数成分が音声了解度に重要であることを示した．その後，Drullman ら [20] は振幅包絡線を高域通過フィルタで制御して場合の SRT の値を測定した．高域通過の場合は 4 Hz 以上になると SRT の値が急激に上昇する結果が得られた．

これらの研究は，振幅包絡線の低域の変調周波数成分が音声の言語情報の知覚に重要であることを示している．しかし，音声には言語情報のほかに話者の個人性情報も含まれている．本研究は，振幅包絡線の変調周波数成分が個人性情報の知覚にも重要であると仮定し，その関係を検討する．

2.3 振幅包絡線に現れる話者情報

振幅包絡線に現れる話者情報の存在は風間らの研究から示唆されている [17]．風間らの手法では，まず，文音声信号の 2 kHz 以上の帯域を 1/8 オクターブで 21 個の帯域に分割し，2 kHz 以下の帯域を 1/4 オクターブで 17 個の帯域に分割する．次に，各狭帯域信号を半波整流し，カットオフ周波数 40 Hz の低域通過フィルタを通して狭帯域振幅包絡線を抽出する．最後に，狭帯域振幅包絡線の帯域間相関値を算出し，17×17 と 21×21 の相関行列（ECM）を算出する．この ECM を話者の特徴量として話者認識の実験を行ったところ，2 kHz 以下の帯域の ECM には性別情報が含まれており，2 kHz 以上に話者情報が多く含まれていることがわかった．また，このときの話者認識率は 100 % であった．この結果は，音声信号の ECM が話者認識の特徴として利用できることを示している．しかし，風間らの研究では，機械による話者認識に ECM が有用であることしか検討しておらず，ヒトがこの振幅包絡線の相関を手がかりに話者認識を行っているかは不明である．聴覚の知覚過程では，すべての帯域の振幅包絡線の相関を分析するような処理を想定することは難しく，振幅包絡線そのものに含まれる情報から個人性を知覚していると考えられる．風間らの研究のように挟帯域振幅包絡線の相関ではなく，ヒトの知覚過程に基づいて振幅包絡線そのものに含まれる個人性情報について検討する必要がある．

2.4 雑音駆動音声

本研究は，振幅包絡線に含まれる個人性情報に着目するため，振幅包絡線情報だけが残された雑音駆動音声を用いて刺激音を作る．図 2.2 には雑音駆動音声の生成法の概要を示す．雑音駆動音声を生成するには，まず，原音声信号を帯域通過フィルタバンクで幾つかの帯域に分割する．次に，Hilbert 変換や半波整流と低域通過フィルタの手法で各帯域の信号の振幅包絡線を検出する．その帯域の振幅包絡線をもとに同じ帯域に制限された雑音を振幅変調する．原音声信号の時間微細構造が帯域制限雑音に置き換えられたため，振幅包絡線情報だけが残される．最後に全部の帯域の変調雑音を加算して雑音駆動音声を合成する．

雑音駆動音声の環境では，ホルマントや基本周波数情報などの音声の重要な特徴が欠落しているものの，多くの研究から少ない帯域数だけの状況でも音声の言語情報を聞き取ることができることが報告された [27][28]．日本語音声の 4 モーラ単語を対象として西野ら（2013）[29] が雑音駆動音声の環境で振幅包絡線を低域通過フィルタにより制御し，変調成分の上限周波数と単語の認識率の関係を調査した．約 5 Hz 未満の変調成分を除去した際に平均正答モーラ数が減少する結果から，モーラの時間構造を再現する変調成分を保存さえしていれば，言語情報の取得が可能であることが分かった．

音声の個人性の知覚に関しては，力丸ら（2003）[30] が 1 から 4 帯域の雑音駆動音声を用いて，話者弁別の可能性を検討した結果，話者弁別するのは困難であることが示された．その一方，Gonzalez ら（2005）[23] は 3, 4, 8, 16 帯域の条件を調査し，16 帯域では

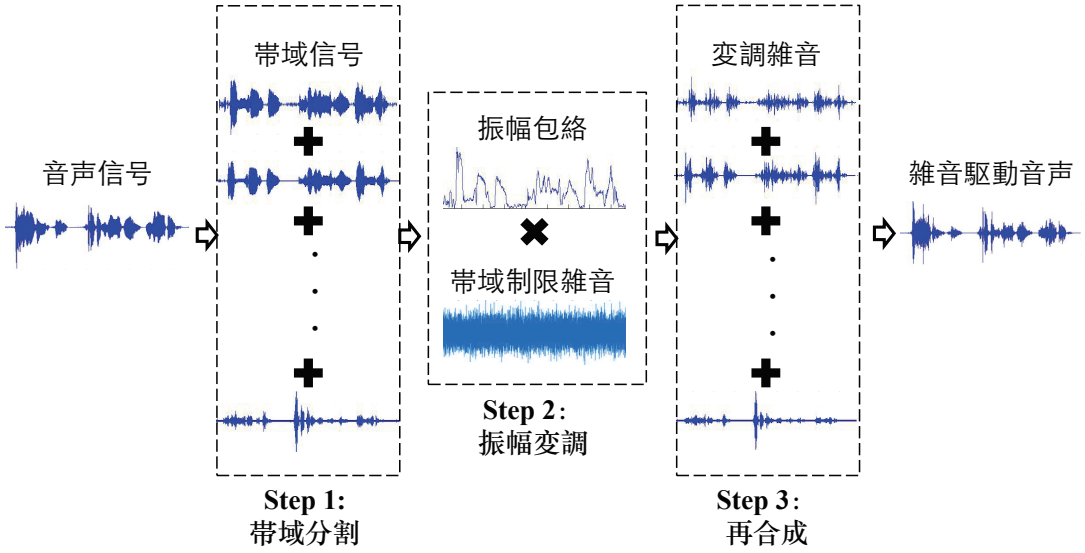

図 2.2: 雑音駆動音声の生成法の概要

高い話者弁別率とことから，振幅包絡線情報だけでも個人性の知覚が十分に可能であることを示唆した．しかし，振幅包絡線に含まれる個人性情報がどのように現れているのかが疑問に残る．

2.5 本研究の着目点

これまでは，音声の振幅包絡線の変調成分は音声の言語情報の知覚には重要な手掛かりであり，特に低域の変調周波数成分が音声了解度に強く影響していることが分かった．また，振幅包絡線情報が機器による話者認識システムにも有用であることから，振幅包絡線には物理的な個人性情報が含まれていることが示唆された．さらに，振幅包絡線情報だけが残された雑音駆動音声の環境では音声の言語情報また個人性情報の聞き取ることができることが分かった．従って，振幅包絡線に含まれている個人性情報が個人性の知覚に寄与していることを示唆している．しかし，振幅包絡線にどのような形式で個人性が表現されているかまた個人性知覚とどのような関係があるかが解明されていない．

本研究では，振幅包絡線の変調成分に着目している．まず，振幅包絡線の変調スペクトルにおける話者間の分散を算出することで，話者による物理的な違いが大きなところを調査する．次に，各帯域に低域通過フィルタにより振幅包絡線の変調成分を制御した雑音駆動音声を利用して，音声の個人性知覚実験を行う．その制御により，振幅包絡線の変調成分の上限周波数を低くしていくと，個人性情報の取得は段階的に難しくなると予測し，個人性の知覚にとって重要な変調周波数帯域を検討する．

第3章　音声の変調スペクトルに現れる個人差の分析

3.1　目的

本研究では,「話者による違いの大きな物理量が音声の個人性の知覚に寄与されている」という仮説をおき，音声波形の振幅包絡線における物理的な個人差の顕著なところを調べ，その物理量と個人性の知覚の関係を調査する．振幅包絡線に含まれる変動成分の分布を示す変調スペクトルに着目し，話者によって変動の大きな変調成分が個人性知覚に大きな影響を与えるかどうかを明らかにすることが目的である．その第一歩として，複数話者の変調スペクトル間の個人差を分析することで，個人性の情報が含まれていると思われる変調スペクトル帯域を推定する．

3.2　音声データと変調スペクトルの算出法

変調スペクトルとは，振幅包絡線の周波数スペクトルに相当するもので，振幅包絡線の時間的変動の特性を表現することが可能である．図 3.1 に変調スペクトルの算出法の概要を示す.

まず，人間の聴覚メカニズムをできるだけ忠実に模擬するために，聴覚フィルタバンクに基づいた ERB_N-number [31] を利用して，周波数帯域分割を行った．ERB_N-number と周波数との関係は下記の式で定義される．

$$\mathrm{ERB}_N\mathrm{-number} = 21.4 \log_{10}\left(\frac{4.37f}{1000} + 1\right) \tag{3.1}$$

ここで，f は周波数である．ERB_N-number はヒトの聴覚フィルタの周波数帯域幅を等価矩形帯域幅で近似し，その幅を 1 として周波数軸を変形したものである．そのため，この尺度を利用して帯域分割することで，基底膜の周波数分解機能をより忠実に近似することができる．本研究では，音声信号を 2 ERB_N-number から 35 ERB_N-number まで，33 個の帯域に分割した．各帯域通過フィルタの遮断周波数は表 3.1 に示す.

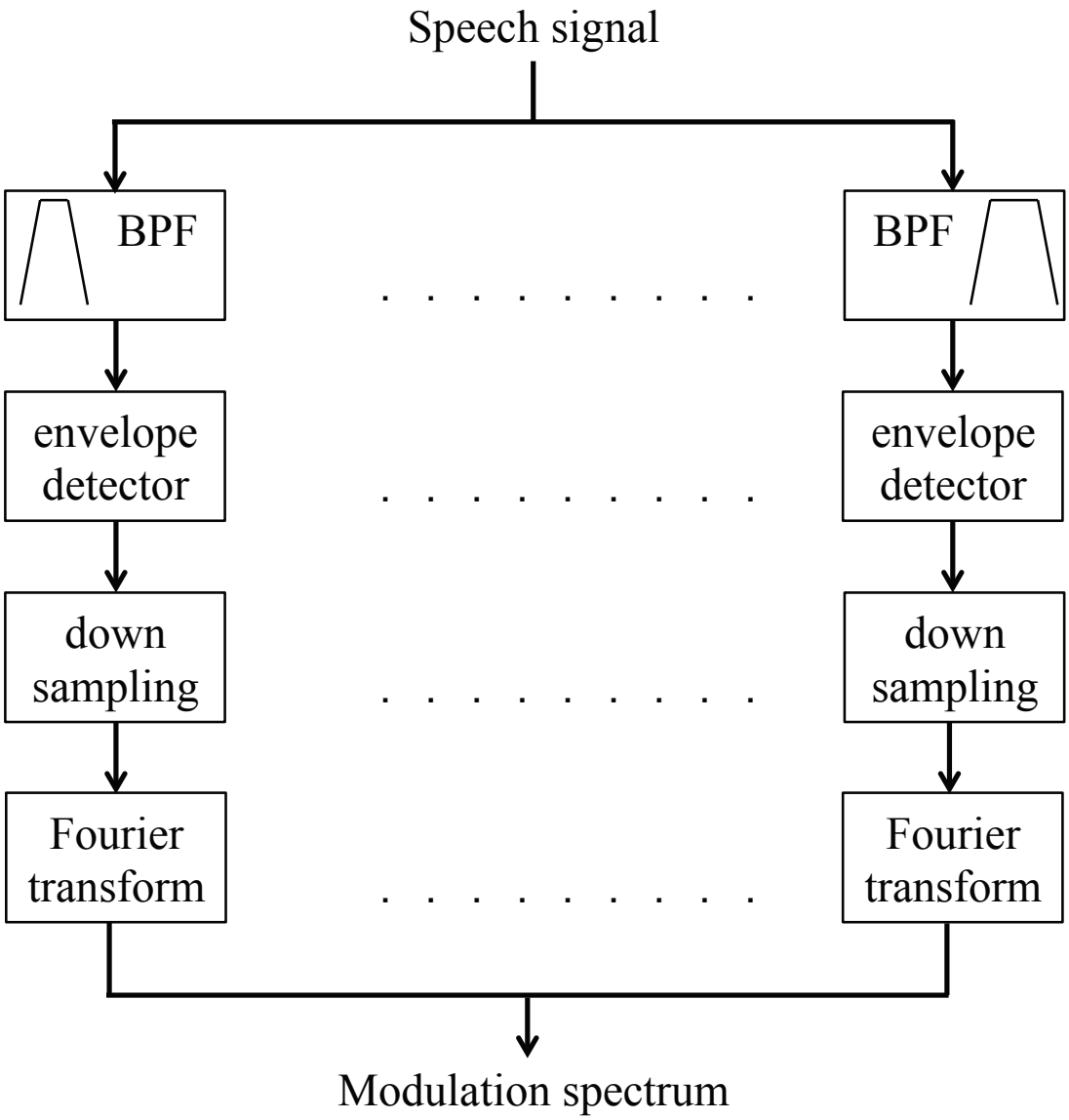

図 3.1: 変調スペクトルの算出法（複数の帯域が点線で省略）

表 3.1: 各帯域通過フィルタの遮断周波数

帯域番号	ERB_N-number	Frequency [Hz]
1	2 〜 3	55 〜 87
2	3 〜 4	87 〜 123
3	4 〜 5	123 〜 163
4	5 〜 6	163 〜 208
5	6 〜 7	208 〜 257
6	7 〜 8	257 〜 312
7	8 〜 9	312 〜 374
8	9 〜 10	374 〜 442
9	10 〜 11	442 〜 519
10	11 〜 12	519 〜 603
11	12 〜 13	603 〜 698
12	13 〜 14	698 〜 803
13	14 〜 15	803 〜 921
14	15 〜 16	921 〜 1051
15	16 〜 17	1051 〜 1196
16	17 〜 18	1196 〜 1358
17	18 〜 19	1358 〜 1539
18	19 〜 20	1539 〜 1739
19	20 〜 21	1739 〜 1963
20	21 〜 22	1963 〜 2212
21	22 〜 23	2212 〜 2489
22	23 〜 24	2489 〜 2798
23	24 〜 25	2798 〜 3142
24	25 〜 26	3142 〜 3525
25	26 〜 27	3525 〜 3951
26	27 〜 28	3951 〜 4426
27	28 〜 29	4426 〜 4955
28	29 〜 30	4955 〜 5544
29	30 〜 31	5544 〜 6200
30	31 〜 32	6200 〜 6930
31	32 〜 33	6930 〜 7743
32	33 〜 34	7743 〜 8649
33	34 〜 35	8649 〜 9657

原音声信号を帯域通過フィルタバンクで分割した各狭帯域信号を $s_k(n)$ とする．k は帯域の番号である．次に，下記の式により，Hilbert 変換を利用して各帯域の振幅包絡を算出した．

$$e_k(n) = \sqrt{s_k^2(n) + \mathcal{H}^2\{s_k(n)\}} \tag{3.2}$$

ここで，$\mathcal{H}[\cdot]$ は Hilbert 変換である．本研究では，50 Hz 以下の変調成分に注目するために，$e_k(n)$ をダウンサンプリングして取り扱う（サンプリング周波数 100 Hz）．ダウンサンプリングした信号を $e_k(m)$ で示す．最後に各帯域の振幅包絡線を離散 Fourier 変換し，dB 尺度に変換して，図 3.1 に示したように，変調スペクトルを求めた．

$$E_k(f_m) = 20\log_{10}|\mathcal{F}[e_k(m)]| \tag{3.3}$$

ここで，$\mathcal{F}[\cdot]$ は Fourier 変換，f_m は変調周波数である．図 3.2 には，ATR 音声データベース [32]C セットの文音声データ（話者 F101，文章 A01）の変調スペクトルを一例として示す．

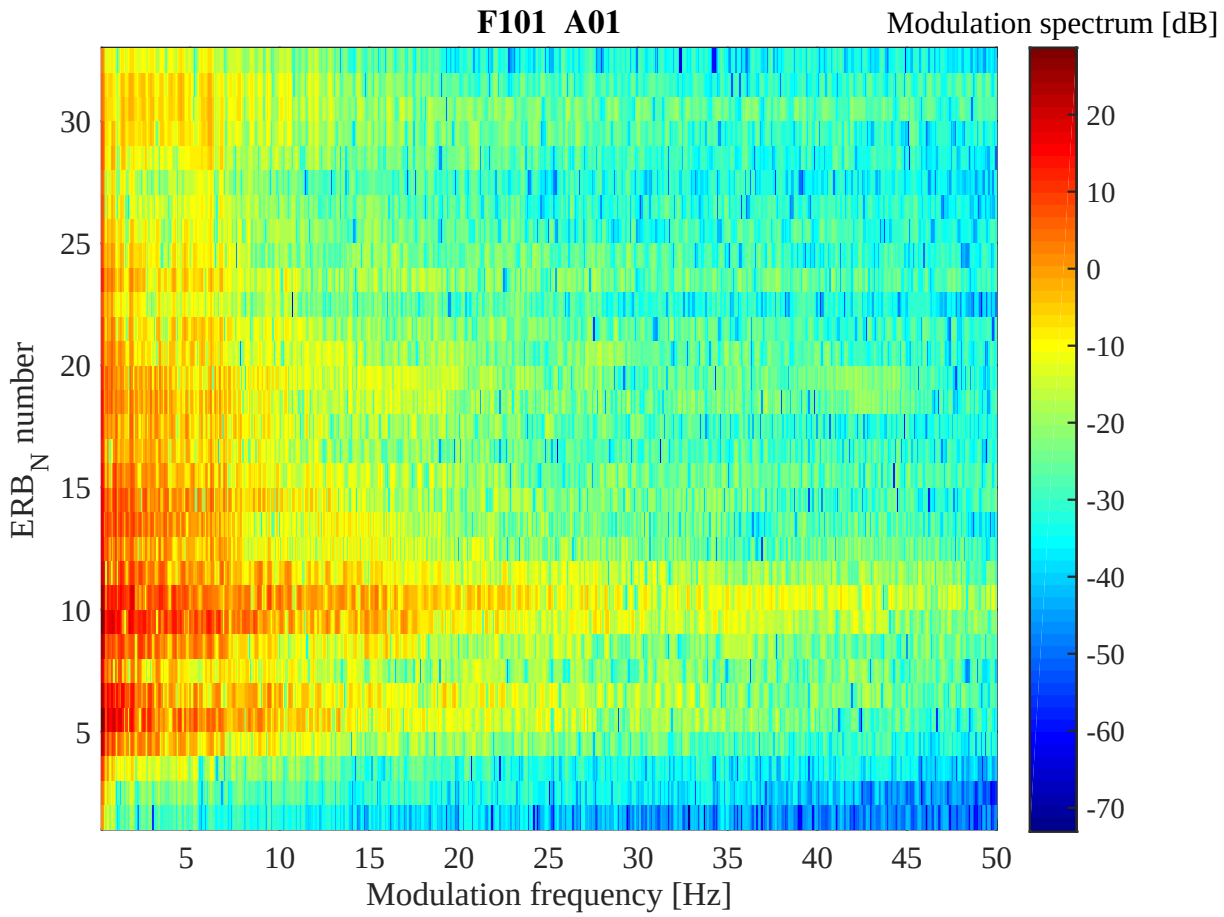

図 3.2: 変調スペクトルの例（話者 F101，文章 A01）

3.3 音声データおよび文章間の分散と話者間の分散

音声データは ATR 音声データベース C セットにある，男女それぞれ 10 名，一人 10 文章の音声データを利用した．一文章の長さが 3〜7 秒であり，サンプル周波数は 20 kHz である．分散を求めるために，最も長い文章を基準に，これより短い文章の最後にゼロ埋め処理を施すことで，すべての音声データを同じ長さにした．

これからは話者 p により発声された文章 q の変調スペクトルを $E_{pq}(k, f_m)$ と表す．ここで，k は帯域番号であり，f_m は変調周波数を表す．

まず，同じ話者であれば，発話内容による変調スペクトルの変化が小さいことを確認するために，各話者の 10 文章の変調スペクトルの文章間分散を次式により算出する．

$$\sigma_p{}^2(k, f_m) = \frac{1}{10}\sum_{q=1}^{10}\{E_{pq}(k, f_m) - \mu_p(k, f_m)\}^2 \tag{3.4}$$

ここで，μ_p は話者 p の 10 文章の音声の平均変調スペクトルであり，

$$\mu_p(k, f_m) = \frac{1}{10}\sum_{q=1}^{10} E_{pq}(k, f_m) \tag{3.5}$$

で与えられる．

次に，各話者の 10 文章の変調スペクトルを平均したものを，その人の特徴となる変調スペクトルとし，$E_p(k, f_m)$ で表す．その特徴変調スペクトルの話者間の分散は次式により算出する．

$$\sigma^2(k, f_m) = \frac{1}{N}\sum_{p=1}^{N}\{E_p(k, f_m) - \mu(k, f_m)\}^2 \tag{3.6}$$

ここで，N は話者間分散を求めるときに利用した話者数を表す．μ は変調スペクトルの平均であり，

$$\mu(k, f_m) = \frac{1}{N}\sum_{p=1}^{10} E_p(k, f_m) \tag{3.7}$$

で与えられる．

3.4 分析結果

変調スペクトル自体を比較すると（付録参照），直感的にではあるが，一人の話者の 10 文章の変調スペクトルの間の違いは小さく同じような形状になっていることが分かった．さらに，違う話者の変調スペクトルの形状は，異なる印象を受けた．

そこで，同じ話者であれば，発話内容による変調スペクトルの変化が小さいことを数学的に確認するために，各話者の 10 文章の変調スペクトルの文章間分散を算出した．その例として話者 F101 の結果を図 3.3 に示す．この結果から，文章間の変調スペクトルの分

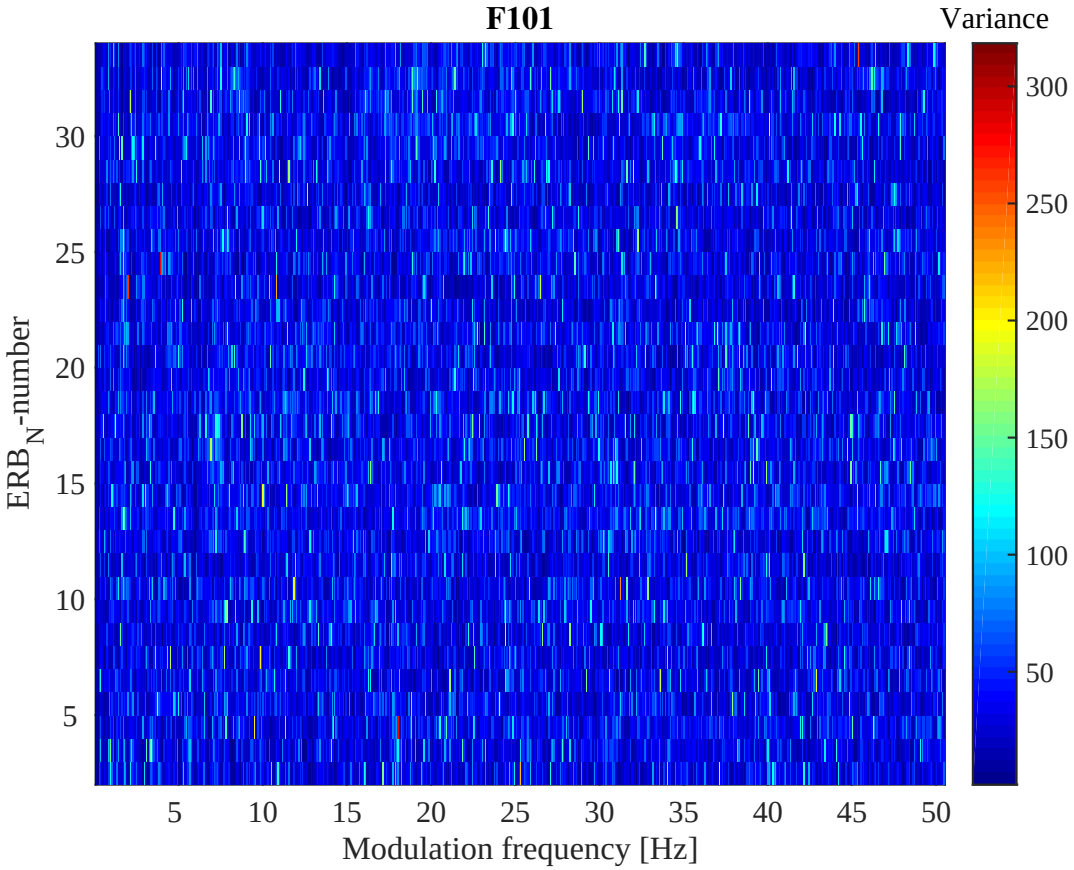

図 3.3: 文章に関する変調スペクトルの分散.

散は，全体的に小さいことが分かった．他のすべての話者の変調スペクトルの分散も似た傾向が得られた．したがって，話者固定で長文章の変調スペクトルは，発話内容によって大きな影響を受けず，文章に依存しないといえる．

次に，各話者の 10 文章の変調スペクトルを平均したものを，その人の特徴となる変調スペクトルとし，話者間の分散を求めた．図 3.4 は男女を含む 20 名分の話者間分散である．この結果から，5 ERB$_N$-number（163.1 Hz）以下の周波数帯域に非常に大きな分散があることが分かった．これは，男性と女性の声帯音源波が存在する周波数帯域が違うことによって大きなばらつきが生じたためと考えられる．しかし，今回の分析では男女の違いではなく，それ以外の話者による違いを見ることが目的である．そこで，男性と女性の話者を分けて性別間で変調スペクトルの違いを分析した．

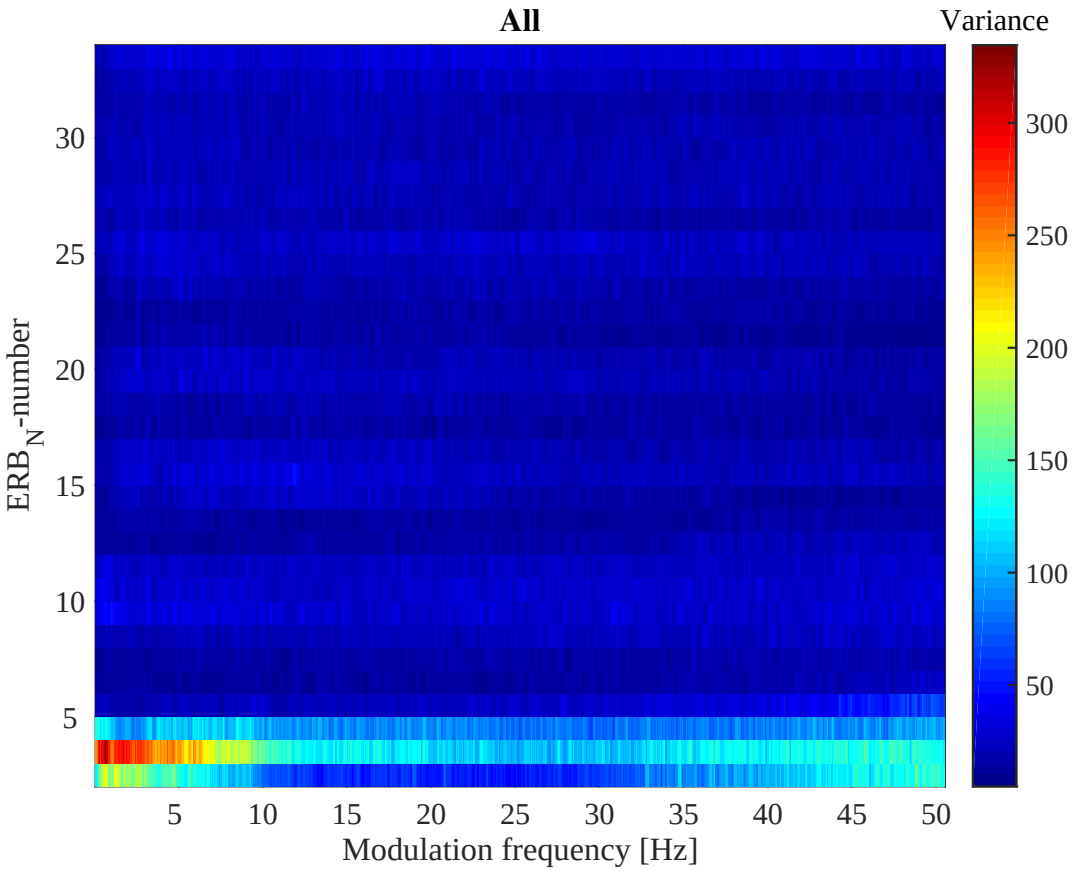

図 3.4: 話者に関する変調スペクトルの分散（全話者）.

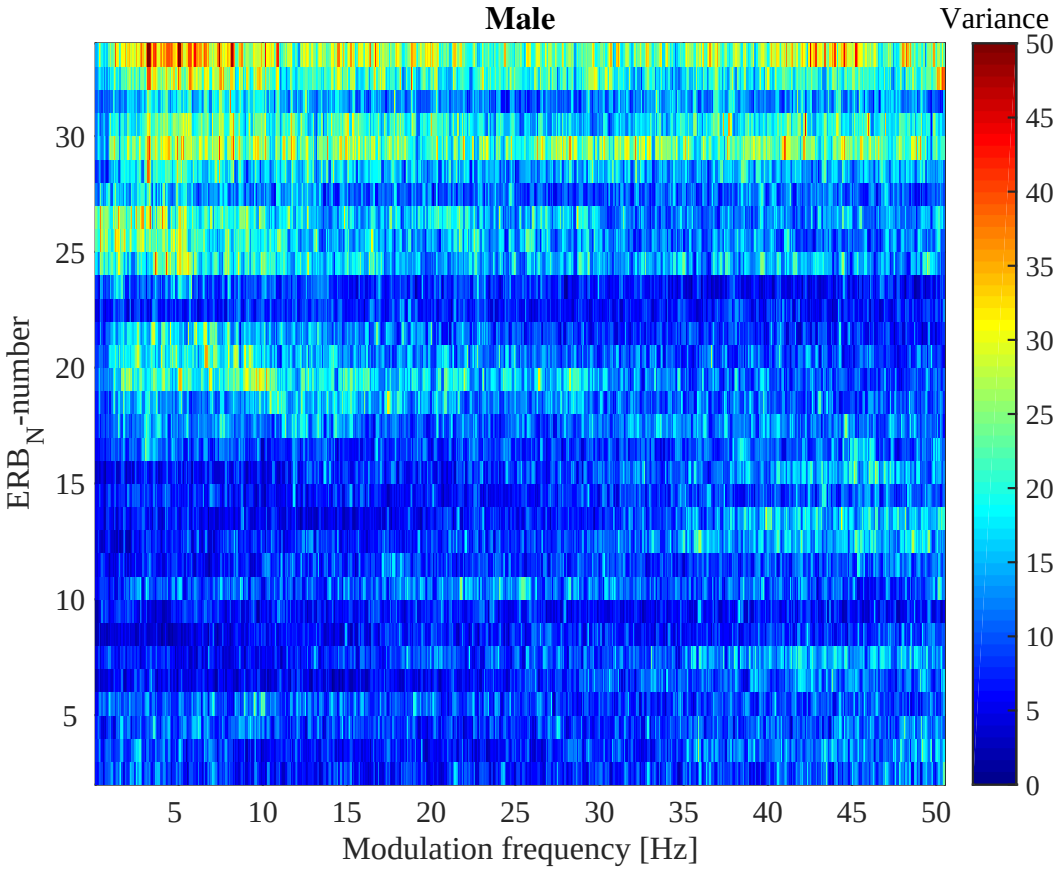

図 3.5: 話者に関する変調スペクトルの分散（男性）.

男性 10 名，女性 10 名それぞれの話者間分散を図 3.5 と図 3.6 に示す．これらの結果から，20 ERB_N-number を境界として，それより下の周波数帯域よりも，それより上の周波数帯域での変調スペクトルに関する分散が大きいことがわかった．次に，変調周波数軸上での違いを見ると，20 ERB_N-number から 29 ERB_N-number までの周波数帯域では，15 Hz 以下の変調周波数帯域の分散が大きく，30 ERB_N-number 以上の周波数帯域では，全変調周波数帯域の分散が大きいことが分かった．物理的な差が大きい変調スペクトルを知覚的にも利用していると考えると，これらの変調周波数帯域に個人性情報が含まれていると考えられる．

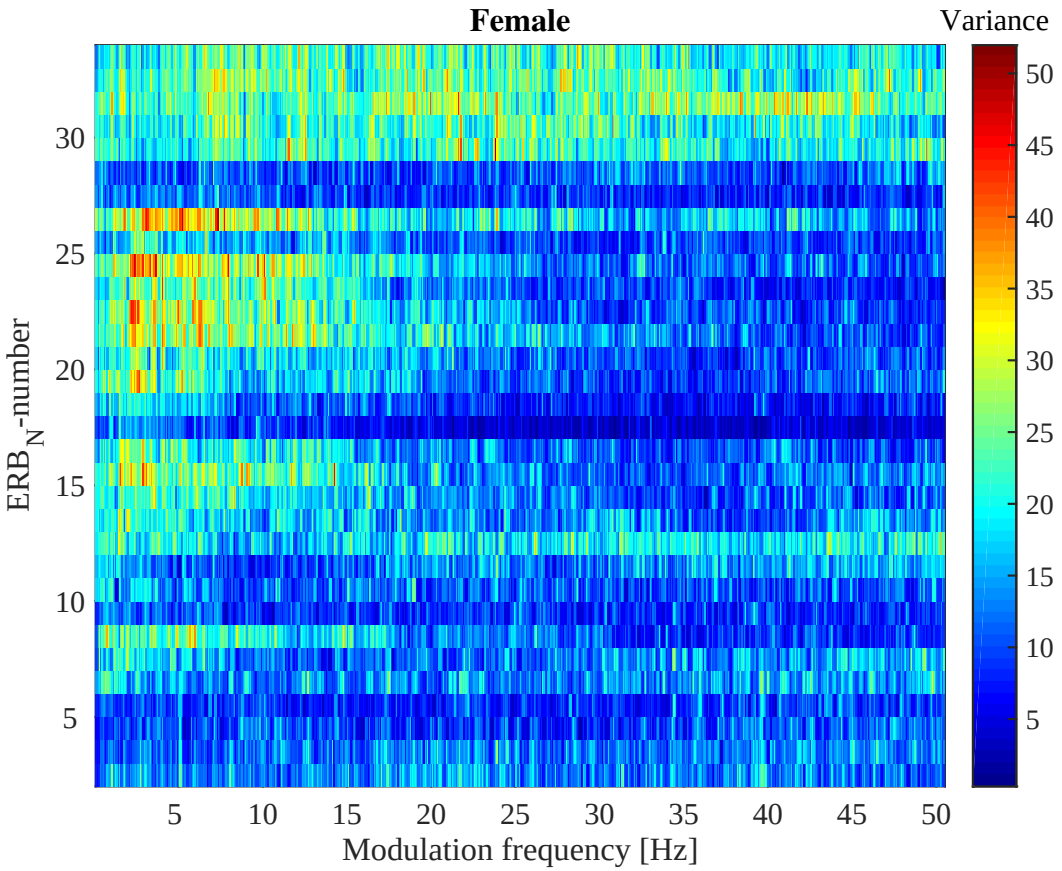

図 3.6: 話者に関する変調スペクトルの分散（女性）.

3.5 考察

変調スペクトルに現れる個人差が主に低域の変調周波数帯域に存在することから，改めて低域変調周波数成分の重要性を示唆した．本研究の変調スペクトルの話者間分散の結果において，男女の差が主に 5 ERB_N-number（163.1 Hz）以下の周波数帯域に現れており，男女それぞれの個人差が同じく 20 ERB_N-number（1739 kHz）以上の帯域に存在する．そこで，振幅包絡線を手掛かりとした機器による話者認識の従来研究において，風間らの研究 [17] では，狭帯域振幅包絡情報の 2 kHz 以下の帯域の ECM には主に性別情報が含まれており，2 kHz 以上に話者情報が沢山含まれている結果が示された．周波数帯域における話者の特徴になる帯域が本研究の結果とほぼ一致している．

さらに，20 ERB_N-number から 29 ERB_N-number までの周波数帯域では，15 Hz 以下の変調周波数帯域の分散が大きく，個人差が顕著に表れている結果になっている．そこで，Falk ら（2010）[33] は変調スペクトルを話者の特徴として Gaussian mixture model (GMM) に基づいた話者認識の研究では，変調周波数 3〜15 Hz の間の変調成分が話者認識に有用であることが分かった．変調周波数軸上の特徴もほぼ一致している．

従って，本研究の結果がそれぞれの研究を支持するものとなっており，それらの結果の裏つけになる物理的な要因かを考えられる．

第4章　音声の変調成分の制限が個人性の知覚に与える影響

4.1　本実験の目的

音声の変調スペクトルにおける個人差の分析で，話者による差異が顕著な帯域が分かった．その個人差が人の個人性知覚に対して，どのような影響があるのかを解明することが本実験の目的である．3章で明らかにした個人差の大きな振幅包絡線の変調成分が低域通過フィルタにより除去されると個人性の知覚がむずかしくなると仮定し，XAB法による話者弁別実験でそれを調査する．

4.2　実験用データベース

本実験で利用する音声データは音声の個人性の類似性を考慮して選んだ．川元と北村（2013）[34]は本研究で使用した話者セットを含むATR音声データベースセットCの男性話者20名による個人性の類似度評価を行った．各話者ペアの類似度を多次元尺度構成法を用いて分析し，20名の話者を知覚空間上に布置した．その結果，20名の話者には，類似性の高いペアと類似性の低いペアが混在していることを示した。本実験では，幅広い話者性について検討するため，川元と北村の結果に基づき話者間距離が一定以上離れた10名の話者（M211, M318, M409, M508, M517, M519, M601, M603, M710, M718）を選択した．

第3章において，ATRデータベースにある男性10名，女性10名の音声データにおける変調スペクトルの個人差を分析し，20 ERB_N-number以上の周波数帯域にある，約15 Hz以下の変調スペクトルに特徴的な個人差が現れることを示した．しかし，本実験で用いた話者セットは第3章とは異なるものである．そこで，本実験で用いる話者セットにおいても同様の特徴が現れるかを確かめるため，上述した10名の話者の変調スペクトルを第3章と同じ方法で分析し，個人差の現れる傾向を確かめた．

図4.1に，変調スペクトルにおける10名の話者の話者間分散を示す．その結果，20から29 ERB_N-numberの周波数帯域で変調スペクトルの分散が大きくなることが分かった．また，変調周波数が低くなるほど，話者間の分散が大きくなり，個人差が主に低変調周波数領域に現れていることも分かった．この傾向は，第3章の結果とほぼ一致している．物理的な違いが大きい変調スペクトルを知覚的にも利用していると考えると，これらの変調

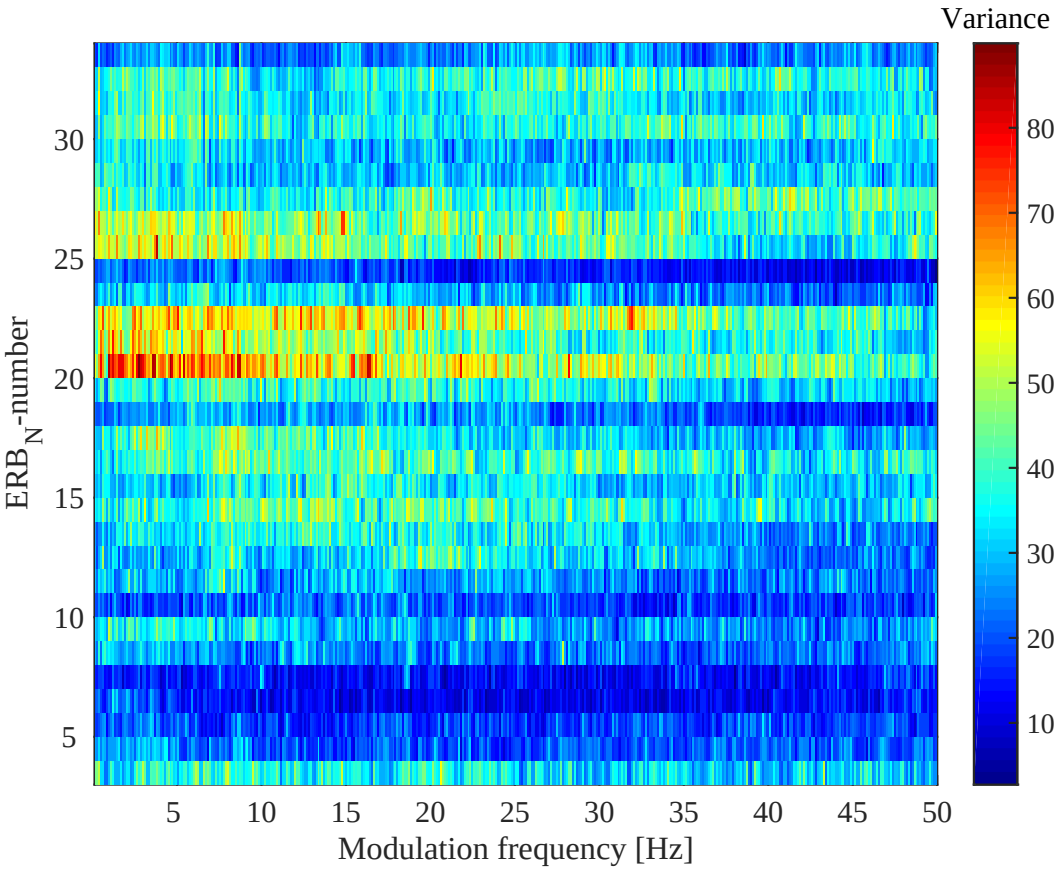

図 4.1: 実験用データベースの変調スペクトルにおける話者間の分散

周波数帯域に個人性情報が含まれていると考えられる．そこで，雑音駆動音声を利用した個人性知覚実験により，これらの個人差が個人性知覚に与える影響を調査する．

4.3 実験参加者

北陸先端科学技術大学院大学の大学院生 6 名（男性：5 名，女性：1 名）が実験に参加した．実験参加者全員の両耳に対して聴力検査を行い，正常な聴力を有することを確認した．

4.4 刺激音

本実験では，音声の振幅包絡の個人性情報に着目するため，振幅包絡線情報だけを残し，雑音キャリアを乗じて音声合成する雑音駆動音声を利用して刺激音を作成した．図 4.2 には刺激音作成のブロックダイアグラムを示す．

まず，変調スペクトルの分析と同じように，ERB_N-number を利用して，周波数帯域分割を行う．本実験では，1 ERB_N-number から 35 ERB_N-number まで，1 ERB_N-number ずつで 34 帯域に分割した刺激と，2 ERB_N-number ずつで 17 帯域に分割した刺激の二種類を用いた．

次に，次式を利用して各帯域信号 $s_k(n)$ の振幅包絡線 $e_k(n)$ を抽出した．

$$e_k(n) = \mathrm{LPF}\left\{\sqrt{s_k(n)^2 + \mathcal{H}[s_k(n)]^2}\right\} \tag{4.1}$$

ただし，k はその帯域の番号，$\mathcal{H}[\cdot]$ と $\mathrm{LPF}\{\cdot\}$ はそれぞれ Hilbert 変換と低域通過フィルタである．

実験では，低域通過フィルタのカットオフ周波数を変化させることで，振幅包絡線の変調成分の上限周波数を変化させた雑音駆動音声の刺激を作成した．2 節で行った変調スペクトルの個人差の分析結果から，変調スペクトルの低域に大きな個人性が含まれていることが分かっている．そこで，低域の変調スペクトルの影響をより細かく調査するために，本実験では，低域通過フィルタのカットオフ周波数を，1, 2, 4, 8, 16, 32, 64 Hz の 7 つとした．最後に，各帯域の振幅包絡線と同じ周波数帯域で帯域制限された白色雑音を掛け合わせて，帯域ごとに変調雑音を求め，最終的に，全帯域の変調雑音を加算することにより，刺激音を作成した．

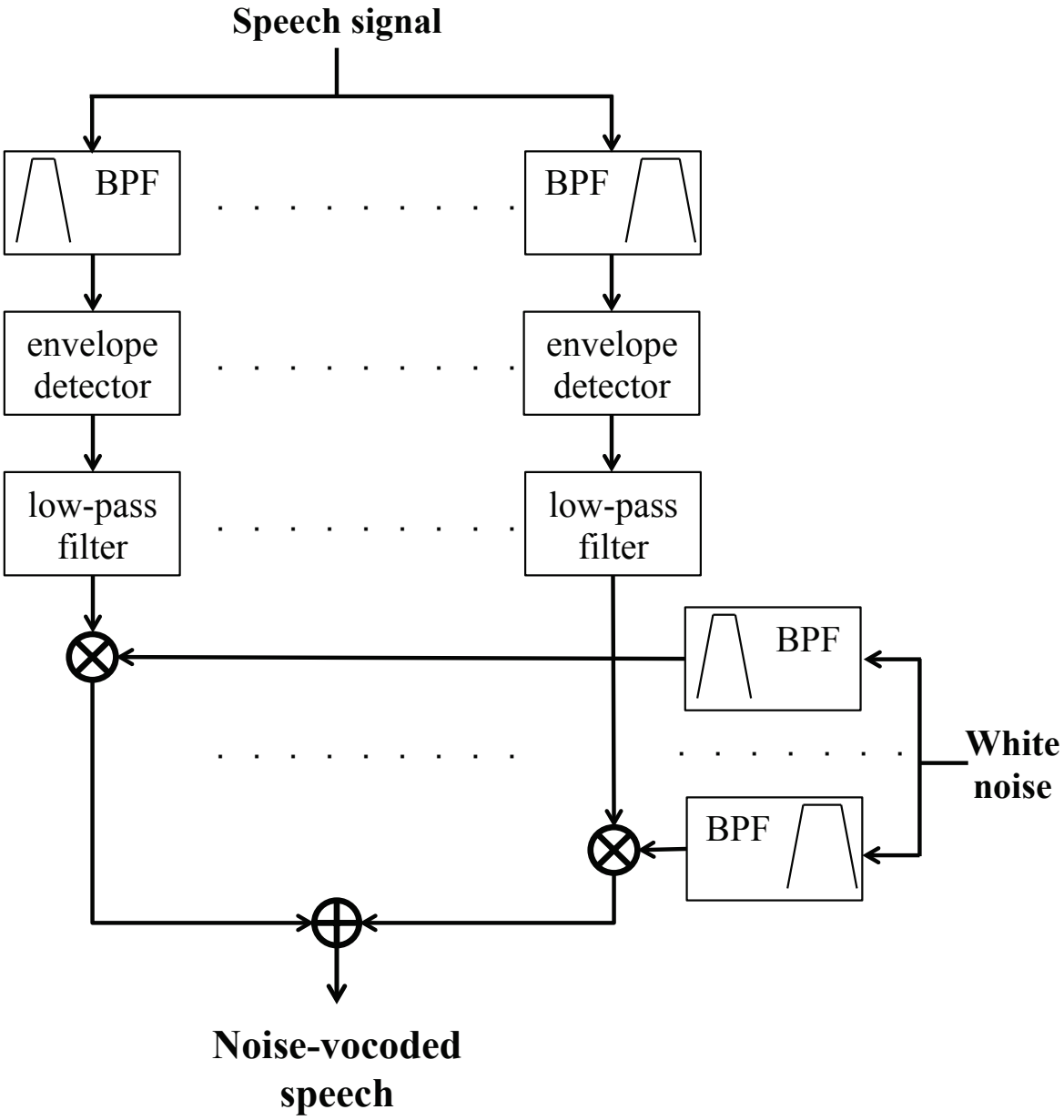

図 4.2: 刺激音作成のブロックダイアグラム

4.5 実験機器

実験は，防音室（暗騒音 26.7 dB）にて行った．刺激の呈示には，MAC（Windows 7），オーディオインターフェース（Fireface UCX），ヘットホン（SENNHEISER HDA 200）を使用した．ヘッドホンからの出力レベルは，B&K NEXUS, B&K type 2231 モジュール型精密騒音計を利用して，実験前に毎回聞きやすいレベル（約 65 dB 前後）に校正された．コンピュータは防音室の外に置き，防音室内にはモニターを設置する．被験者はモニターの画面によりマウスで操作する．

図 4.3: 実験環境

4.6 実験方法

前述した話者間距離が一定以上に離れた ATR データベースの男性話者 10 名の文音声データ（サンプリング周波数：20 kHz，長さ：約 5 s 前後）を利用した．実験は，XAB 法により行った．刺激音 X, A, B の内容を以下に示す．

X: 原音声

A: X と同じ話者，違う文章の雑音駆動音声

B: X と違う話者，A と同じ文章の雑音駆動音声

以上の刺激音を 0.5 s の無音区間を挟んで呈示し（図 4.4），X の話者が A と B の話者のどちらと同じであるかを強制判断させた．順序効果を打ち消すために，XBA の順についても実験を行った．X, A, B の三つの刺激音の組を 1 刺激とする．1 刺激につき XAB, XBA を各一回呈示する．実験条件は，２種類の帯域分割方法と 7 種類の低域通過フィルタのカットオフ周波数で計 14 種類ある．1 つの条件につき，10 回の違う話者ペアによる刺激を呈示した．そのため，1 回の実験では 280 刺激を呈示した．各刺激は 1 回だけ呈示し，聞き直しはできない．

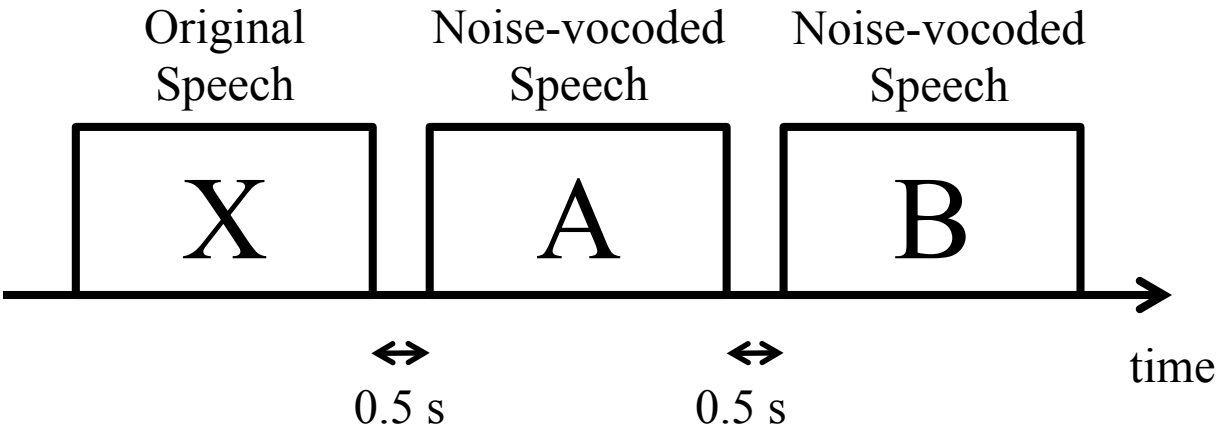

図 4.4: 刺激パターン

表 4.1: 等分散性の検定（34 帯域）

Lenvene 統計量	自由度 1	自由度 2	有意確率
1.696	6	35	.151

表 4.2: 分散分析（34 帯域）

	平方和	自由度	平均平方	F 値	有意確率
グループ間	.185	6	.031	2.117	.076
グループ内	.508	35	.015		
合計	.693	41			

4.7 実験結果

図 4.5 に，34 帯域の条件での話者弁別率の平均と標準偏差を示す．低域通過フィルタのカットオフ周波数が低い場合，話者弁別率の平均値が低くなることが分かった．しかし，分散が大きく，特に 1 Hz の場合は話者弁別に被験者の個人差が大きかった．

統計解析ソフトウェア IBM SPSS Statistic（SPSS）により分散分析を行った．まず，実験結果の等分散性の検定の結果を表 4.1 に示す．

有意確率が $p = 0.151 >$ 有意水準 0.05 なので，仮説「各グループの母分散は等しい」は棄てられない．従って，等分散性が認められた．

分散分析の結果（$F(6, 35) = 2.117, p = 0.076 > 0.05$）から，各カットオフ周波数の条件間に有意差があるとは言えないことが分かった．

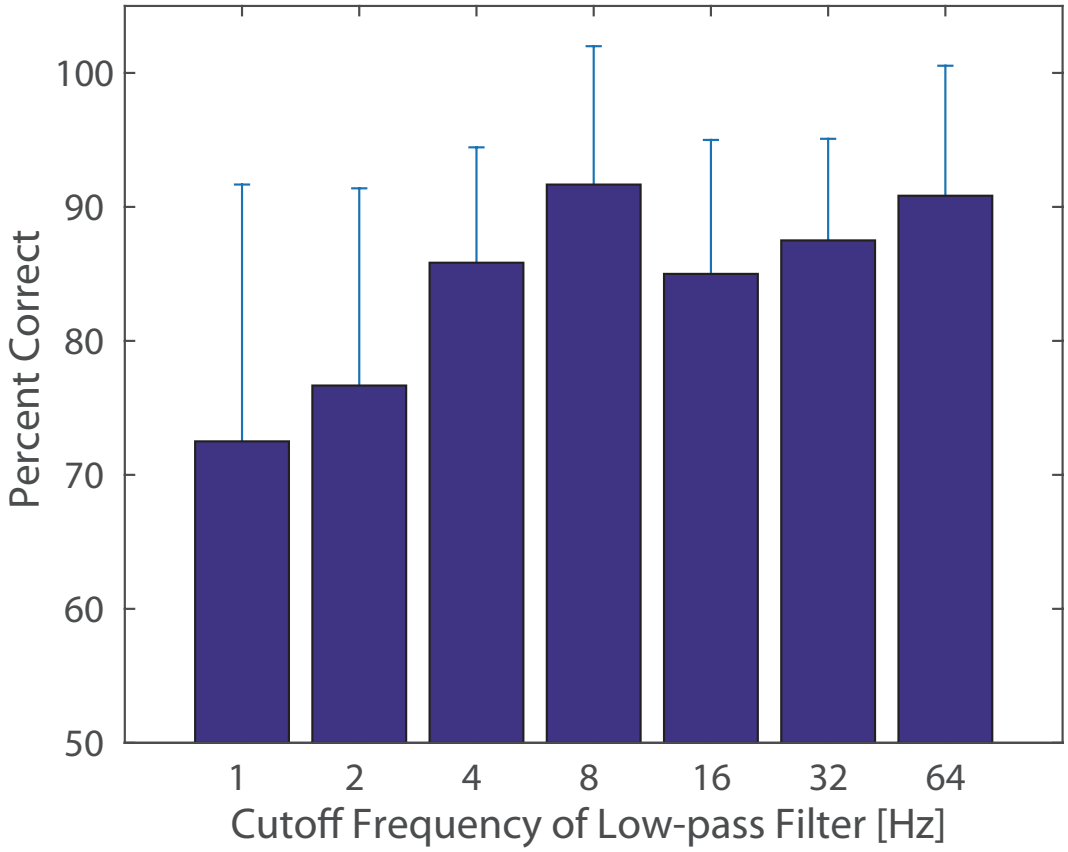

図 4.5: 34 帯域における実験結果

表 4.3: 等分散性の検定（17 帯域）

Lenvene 統計量	自由度 1	自由度 2	有意確率
.663	6	35	.679

表 4.4: 分散分析（17 帯域）

	平方和	自由度	平均平方	F 値	有意確率
グループ間	.379	6	.063	4.591	.002
グループ内	.482	35	.014		
合計	.861	41			

図 4.6 は 17 帯域の条件での話者弁別率の平均と標準偏差である．低域通過フィルタのカットオフ周波数が 8 Hz より低くなるにつれて，話者弁別率が徐々に下がっており，1 Hz になると弁別率が 60 %前後になることが分かった．この結果についても SPSS により分散分析を行った．

まず，表 4.1 には等分散性の検定の結果を示す．有意確率が $p = 0.679 >$ 有意水準 0.05 なので，仮説「各グループの母分散は等しい」は棄てられなく，等分散性が認められた．

分散分析を行ったところ（$F(6, 35) = 4.591, p = 0.002 < 0.05$），カットオフ周波数による効果が有意であることが確認できた．更に，Tukey 法による多重比較を行った結果，1 Hz の条件と 16, 32, 64 Hz の条件の間に有意差が認められた．

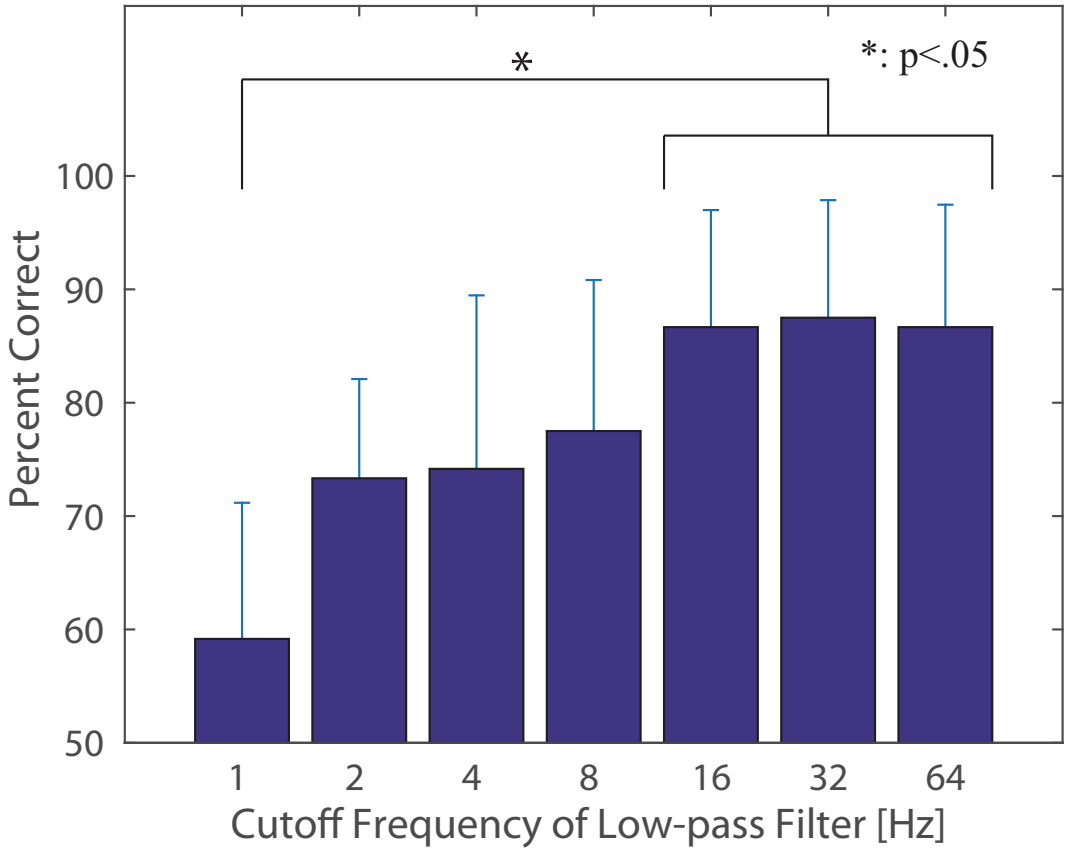

図 4.6: 17 帯域における実験結果

4.8 考察

34 帯域の場合は，変調成分が大きく削られても，ある程度の話者弁別が可能であった．周波数帯域を 1 ERB_N-number ずつ分割した場合は，17 帯域の場合よりも，周波数成分の情報がより多く残されている．そのため，時間的に平均された周波数成分のスペクトルキューが個人性の知覚に強く寄与し，振幅包絡線に含まれる個人性の情報の影響が現れにくかったと考えられる．また，被験者により弁別率にばらつきがあったことから，スペクトル情報に含まれる個人性と振幅包絡線に含まれる個人性のどちらに注目して個人性を知覚するのかが，被験者によって異なる可能性もある．

一方，17 帯域の場合は 34 帯域のものよりも，周波数帯域の分割が粗いため，変調成分に含まれる個人性情報の影響が相対的に大きくなった．そのため，カットオフ周波数による影響が強く現れた．さらに，16 Hz 以下になると話者弁別率が徐々に低くなるため，約 16 Hz 以下の変調周波数帯域に個人性情報が含まれていることが示唆された．変調スペクトルにおける個人差の分析では，20 ERB_N-number 以上の周波数帯域にある，約 15 Hz 以下の変調スペクトルに特徴的な個人差が現れることを示した．聴取実験の結果と比べると変調周波数帯域についてはほぼ一致してるように見える．従って，変調周波数 15 Hz 以下の変調成分に顕著に現れた個人差が話者の違いとして音声の個人性の知覚に利用されていることを示している．

第5章　全体考察

第 3 章では，振幅包絡線の変調スペクトルにおける話者間の分散を算出することにより，物理的に話者による違いを調査した結果，主に 15 Hz 以下の変調周波数帯域に個人差が顕著表れていることが分かった．さらに，第 4 章では，雑音駆動音声の環境で振幅包絡線に低域通過フィルタで制御することにより変調成分の上限周波数の変化が音声の個人性の知覚への影響を調査した．その結果，変調成分の上限周波数 16 Hz 以下になると音声の個人性の知覚が段階的にむずかしくなり 1 Hz の条件ではほぼ話者を認識できるないような状態になっていることが明らかになった．それら結果を関連付けると振幅包絡線の 15 Hz 以下の変調成分に物理的な個人差が顕著に含まれており，人の音声の個人性知覚に寄与されていることが明らかになった．本研究における「話者による違いの大きな物理量が音声の個人性の知覚に寄与されている」という仮説にも支持していると考えられる．

変調スペクトルにおける個人差の分析では 15 Hz 以下の変調周波数帯域に表れる個人差が主に 20 ERB_N-number から 29 ERB_N-number までの周波数帯域に存在する結果が得られた．この分散が大きい帯域は北村ら（1995）[35] の単母音のスペクトル包絡の話者間分散の結果とほぼ同じ帯域である．しかし，北村らが着目したところは単母音の平均的スペクトルであり，本稿で着目したところは狭帯域包絡線の変調スペクトルである．つまり，単純な周波数帯域が個人性に関係があるのではなく，その周波数帯域における時間的な変動の違いに個人性の含まれている可能性が示唆される．

本研究で，着目した振幅包絡線の変調成分における個人性情報を音声生成の面から追求すると音声のスペクトル包絡の時間的変動から生ずることと考えられる．動的な個人性情報に関して，基本周波数の時間的変動に着目した研究がしばしばある [16][36]．音声のスペクトル包絡の時間的変動に関連すること研究においては [37][8]，基本周波数と静的なスペクトル包絡情報も混在する環境で実験を行ったため，時間的変動の特性を抽出することできなかった．本研究は雑音駆動音声を利用したため，時間的変動の特性以外のものができるだけ除去されたため，時間的変動の特性だけに注目することができた．

第6章　結論

6.1　本研究で明らかとなったこと

本研究では，まず，音声の振幅包絡線を Fourier 変換により算出した変調スペクトルにある個人差を調査した．その結果，20 ERB_N-number 以上の周波数帯域の変調スペクトルの話者間分散が大きいことが分った．また，20 ERB_N-number から 29 ERB_N-number までの変調スペクトルでは，15 Hz 以下の変調周波数帯域の分散がより大きく，30 ERB_N-number 以上の周波数帯域では，全変調周波数帯域の分散が大きいことが分かった．物理的な差が大きい変調スペクトルを知覚的にも利用していると考えると，これらの変調周波数帯域に個人性情報が含まれていると考えられる．

次に，変調スペクトルの帯域を制限した雑音駆動音声を用いて，その変調成分の変化が個人性知覚に及ぼす影響を調査した．34 帯域と 17 帯域の二種類の雑音駆動音声を利用し，各帯域の振幅包絡を低域通過フィルタにより制御した．XAB 法による個人性知覚実験の結果，34 帯域の場合は変調周波数帯域の上限周波数が低くなると話者弁別率の平均値が低くなるが，分散が大きいため分散分析の結果から有意差が認められなかった．17 帯域の場合は，変調周波数帯域の上限周波数が 16 Hz から低くなるに従って，話者弁別率が有意に下がっており，1 Hz になると話者弁別率が 60 %前後になることが分かった．

これらの結果から，振幅包絡線の変調周波数約 15 Hz 以下の変調帯域に個人性情報が顕著に現れており，音声の個人性知覚に寄与されていることが明らかになった．

6.2　残された課題

- **変調スペクトル分析方法の改善**
  本研究では，音声信号の全サンプル点にフーリエ変換を行ったため，短時間フーリエ変換のように変調スペクトルの時間的な変動を調査することができない．さらに，フーリエ変換の点数が音声の長さにより変わっている．その一方，窓処理をすると得られた変調スペクトルは 4 次元（時間，周波数，変調周波数，変調スペクトル）のデータになるため，分析方法の改良が必要である．

- **変調スペクトルにおける個人性情報が含まれる周波数帯域の検討**
  本研究では，雑音駆動音声の刺激音を作成するときに全部の周波数帯域に同じ低域通過フィルタをかけたため，変調周波数軸上の影響だけを調査した．変調スペクト

ルに含まれる個人差の分析では，20 ERB_N-number 以上の周波数帯域の変調スペクトルの話者間分散が大きいことが分った．その結果と個人性知覚の関係調査するために，新たな実験方法また刺激音の作成方法が必要である．

- **個人性知覚のメカニズムへの発展**
  本研究で得られた結果は，個人性知覚の物理的な要因の一端にしか過ぎない．ヒトの聴知覚メカニズムにおける個人性知覚メカニズムの解明に繋がるものである．近年，時間的に変動の特徴すなわち Temporal Cue は聴覚メカニズムや音声知覚の領域によく注目されている課題となっている．さらなる研究で，個人性の知覚メカニズムを解明する必要があると考えられる．

参考文献

[1] 粕谷 英樹，楊 長盛, “音源から見た声質,” 日本音響学会誌，Vol. 51, No. 11，pp. 869–875，1995.

[2] 粕谷 英樹, “声質の伝える情報とその関連量,” 日本音響学会誌，Vol. 68, No. 10，pp. 520–526，2012.

[3] 森 大毅, 前川 喜久雄, 粕谷 英樹, “音声は何を伝えているか,” コロナ社, pp. 131–191, 2014.

[4] 古井 貞熙, “声の個人性の話,” 日本音響学会誌，Vol. 51, No. 11，pp. 876–881，1995.

[5] L. Garrido, F. Eisner, C. McGettigan, L. Stewart, D. Sauter, J.R. Hanley, S.R. Schweinberger, J.D. Warren and B.Duchaine, “Developmental phonagnosia: A selective deficit of vocal identity recognition,” Neuropsychologia, Vol. 47, No. 123–131, 2009.

[6] 伊藤 憲三，斉藤 収三, “音声の音響的特徴パラメータが個人性の知覚に及ぼす影響,” 電子通信学会論文誌，Vol. J65–A，pp. 101–108，1982.

[7] 橋本 誠, 北川 敏, 樋口 宜男, “音声の個人性知覚に影響を及ぼす音響的特徴の定量的分析,” 日本音響学会誌, Vol. 54, No. 3, pp. 169–178, 1998.

[8] H. Kasuya，W. Zhu，M. Matsuda，and C. Yang, “Voice quality conversion based on an ARX speech analysis-synthesis method and its application to the study of speaker individuality,” J. Acoust. Sco. Am., Vol. 100, No. 4, pp. 2600, 1996.

[9] 北村 達也, 赤木 正人, 北澤 茂良, “スペクトル遷移パターンが個人性知覚に与える影響 について,” 日本音響学会聴覚研究会資料, H-98-97, pp. 1–8, 1998.

[10] 桑原 尚夫，大串 健吾，“ホルマント周波数・バンド幅の独立制御と個人性判断,” 電子通信学会論文誌 A，Vol. 69, No. 4，pp. 509–517, 1986.

[11] 北村 達也，赤木 正人，“単母音の話者識別に寄与するスペクトル包絡成分,” 日本音響学会誌，Vol. 53, No. 3，pp. 185–191，1997.

[12] T. Kitamura, M. Akagi, “Speaker individualities in speech spectral envelopes,” J. Acoust. Soc. Jpn.(E), Vol. 16, No. 5, pp. 283–289, 1995.

[13] T. Kitamura and T. Saitou, “Effects of acoustic modification on perception of speaker characteristics for sustained vowels,” Acoustical Science and Technology, Vol. 28, No. 6, pp. 434–437, 2007.

[14] T. Kitamura, K. Honda and H. Takemoto, “Individual variation of the hypopharyngeal cavities and its acoustic effects,” Acoustical Science and Technology, Vol. 26, No. 1, pp. 16–26, 2005.

[15] K. Amino, T. Sugawara, T. Arai, “Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties,” Acoustical Science and Technology, Vol. 27, No. 4, pp. 233–235, 2006.

[16] M. Akagi and T. Ienaga, “Speaker individuality in fundamental frequency contours and its control,” J. Acoust. Soc. Jpn. (E), Vol. 18, No. 2, pp. 73–80, 1997.

[17] 風間 道子，東山 三樹夫，山崎 芳男，“狭帯域音声波形包絡線の帯域間相関行列に現れる話者情報,” 電子通信学会論文誌，Vol. J92–A, No. 4，pp. 205–215，2009.

[18] T. Dau and D. Puschel, “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure,” J. Acoust. Sco. Am., Vol. 99, No. 6, pp. 3615–3622, 1996.

[19] R. Drullman, J. M. Festen, and R. Plomp, “Effect of temporal envelope smearing on speech reception,” J. Acoust. Sco. Am., Vol. 95, No. 2, pp. 1053–1064, 1994.

[20] R. Drullman, J. M. Festen, and R. Plomp, “Effect of reducing slow temporal modulations on speech reception,” J. Acoust. Sco. Am., Vol. 95, No. 5, pp. 2670–2680, 1994.

[21] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, “Speech recognition with primarily temporal cues,” Science, Vol. 270, pp. 303–304, 1995.

[22] M. Vongphoe, and F. G. Zeng, “Speaker recognition with temporal cues in acoustic and electric hearing,” J. Acoust. Sco. Am., Vol. 118, No. 2, pp. 1155–1061, 2005.

[23] J. Gonzalez, and J. C. Oliver, “Gender and speaker identification as a function of the number of channels in spectrally reduced speech,” J. Acoust. Sco. Am., Vol. 118, No. 1, pp. 461–470, 2005.

[24] V. Krull, and Xin Luo, “Talker–identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition,” J. Acoust. Sco. Am., Vol. 131, No. 4, pp. 3069–3078, 2012.

[25] T. Dau, and B. Kollmeier, “Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,” J. Acoust. Sco. Am., Vol. 102, No. 5, pp. 2906–2919, 1997.

[26] H. W. Dudley, “The vocoder,” Bell Labs Rec., Vol. 18, pp. 122-126, 1939.

[27] R. V. Shannon, F. G. Zeng, and J. Wygonski, “Speech recognition with altered spectral distribution of envelope cues,” J. Acoust. Sco. Am., Vol. 104, No. 4, pp. 2467–2476, 1998.

[28] P. C. Loizou, M. Dorman, and Z. Tu, “On the number of channels needed to understand speech,” J. Acoust. Sco. Am., Vol. 106, No. 4, pp. 2097–2103, 1999.

[29] 西野 恭生, 宮内 良太, 鵜木 祐史，“音声の各周波数帯域の振幅包絡に含まれる言語情報,” 日本音響学会聴覚研究会資料，Vol. 43, No. 7, pp. 547–552，2013.

[30] 力丸 裕, 片山 貴史, “劣化雑音音声の知覚はどこまで可能か？話者弁別,” 日本音響学会聴覚研究会資料，Vol. 33, No. 1, pp. 25–27，2003.

[31] B. C. J. Moore, “An introduction to the psychology of hearing, sixth edition,” BRILL, Sixth Edition, pp. 74–80, 2013.

[32] 匂坂 芳典, 浦谷 則好, “ATR 音声・言語データベース,” 日本音響学会誌, Vol 48, No. 12, pp. 878–882, 1992.

[33] T. H. Falk and W. Chan, “Modulation Spectral Features for Robust Far-Field Speaker Identification,” IEEE Trans. Audio, Speech Lang. Process., Vol. 18, No. 1, pp. 90–100, 2010.

[34] 川本 広樹, 北村 達也, “ATR 音声データベースセット C の文音声の個人性類似度,” 電子情報通信学会技術研究報告，音声, Vol. 112, No. 450, pp. 33–34, 2013.

[35] 北村 達也, 高木 直子, 赤木, 正人, “個人性情報を含む周波数帯域について,” 電子情報通信学会技術研究報告, 音声, Vol. 95, No. 140, pp. 1-6, 1995.

[36] 大野 宏, 赤木, 正人, “文音声中の基本周波数パターンに含まれる個人性の検討,” 電子情報通信学会技術研究報告, 音声, Vol. 97, No. 586, pp. 89–96, 1998.

[37] 出水田 剛志, 赤木 正人, “聴取印象に着目した音声の個人性知覚に関する基礎研究,” 日本音響学会聴覚研究会資料, Vol. 41, No. 7, pp. 551–554, 2011.

謝辞

本研究を行うに際して, 終始御指導ならびに御助言頂いた鵜木祐史准教授, 赤木正人教授に心から御礼申し上げます.

本論文を執筆するにあたり, 有益なる御助言, 適切なる御指摘を頂きました北陸先端科学技術大学院大学情報科学研究科 党建武教授, 田中宏和准教授に心より感謝致します.

北陸先端科学技術大学院大学情報科学研究科宮内良太助教には, 本研究を進めるにあたり数多くの助言を賜り, 原稿執筆に関して添削をして頂き，心から感謝致します.

研究室会議において数多くの御意見と多面に渡るご協力を頂いた赤木鵜木研究室の皆さんに感謝致します.

貴重な個人性類似度データをいただいた甲南大学知能情報学部北村達也教授に深く感謝致します.

実験のために貴重な時間を割いて頂いた多くの実験参加者の方々に，深く感謝申し上げます.

最後に，これまでの学生生活を経済的にも精神的にも支え頂き，温かく見守ってくれた両親に心から最大の感謝を申し上げます.

研究業績

本研究に関する研究業績

国際会議

- Zhi Zhu, Ryota Miyauchi, and Masashi Unoki, “Analysis of Speaker Individual Differences on Modulation Spectrum,” Proc. 2015 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, Kuala Lumpur, Malaysia, February 2015. (Accepted)

研究会

- 朱 治, 宮内 良太, 鵜木 祐史, “音声の変調スペクトルに現れる個人差の分析,” 日本音響学会聴覚研究会資料, Vol. 44, No. 7, pp. 457–460, 和歌山, Octorber 2014.

口頭発表

- 朱 治, 宮内 良太, 鵜木 祐史, “変調スペクトルの帯域を制限した雑音駆動音声の個人性知覚に関する研究 ,” 日本音響学会 2015 年春季研究発表会, 2–Q–6, 東京, March 2015.

そのほかの研究業績

論文

- Zhi ZHU, Katsuhiko YAMAMOTO, Masashi UNOKI, and Naofumi AOKI, “Study on scramble method for speech signal by using random bit shift of quantization,” Journal of Signal Processing, Vol. 18, No. 6, pp. 303–307, 2014.

- Katsuhiko YAMAMOTO, Zhi ZHU, Masashi UNOKI, and Naofumi AOKI, “Semi-Scramble Method for Speech Signals Based on Phonemic Restoration,” Journal of Signal Processing, Vol. 18, No. 4, pp. 205–208, 2014.

国際会議

- Zhi ZHU, Katsuhiko YAMAMOTO, Masashi UNOKI, and Naofumi AOKI, “Study on scramble method for speech signal by using random bit shift of quantization,” Proc. 2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 1PM1-2-2, pp. 109–102, Hawaii, USA, March 2014.

- Katsuhiko YAMAMOTO, Zhi ZHU, Masashi UNOKI, and Naofumi AOKI, “Semi-Scramble Method for Speech Signals Based on Phonemic Restoration,” Proc. 2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 1PM2-2-1, pp. 201–204, Hawaii, USA, March 2014.

研究会

- 朱治, 山本克彦, 鵜木祐史, 青木直史, “量子化ビットのランダムシフトを利用した音声スクランブル法,” 電子情報通信学会技術研究報告, マルチメディア情報ハイディング・エンリッチメント研究会, Vol. 113, No. 480, pp. 57–62, 石川, March 2014.

- 山本克彦, 朱治, 鵜木祐史, 青木直史, “音韻修復現象に着目した音声半開示スクランブル法,” 電子情報通信学会技術研究報告, マルチメディア情報ハイディング・エンリッチメント研究会, Vol. 113, No. 290, pp. 59–64, 広島, November 2013.

口頭発表

- 朱治, 山本克彦, 鵜木祐史, 青木直史, “量子化ビットのランダムシフトによる音声スクランブル法の検討,” 平成 25 年度電気関係学会北陸支部連合大会, G–17, pp. 21, 石川, September 2013.

- 山本克彦, 朱治, 鵜木祐史, 青木直史, “音韻修復現象に着目した音声半開示スクランブル法,” 平成 25 年度電気関係学会北陸支部連合大会, G–18, pp. 22, 石川, September 2013.

付録

ここでは，本論文の第 3 章第 4 節に掲載しない変調スペクトルの図の一部を下記に示す．ATR 音声データベース [32]C セットにある女性話者 F101 から F105 まで 5 人分，各人の文章 A01 から A06 まで 6 つ文章の変調スペクトルの図を示している．まったく違う文章の変調スペクトルでも話者が同じであればその形状も類似していることが確認できる．また，話者により形状が異なっていることも確認できる．A01 から A06 までの文章の内容は以下に示す．

A01 あらゆる現実をすべて自分の方へねじ曲げたのだ.

A02 一週間ばかりニューヨークを取材した

A03 テレビゲームやパソコンでゲームをして遊ぶ

A04 物価の変動を考慮して給付水準を決める必要がある

A05 救急車が十分に動けず救助作業が遅れている

A06 言論の自由は一歩譲れば百歩も千歩も攻め込まれる

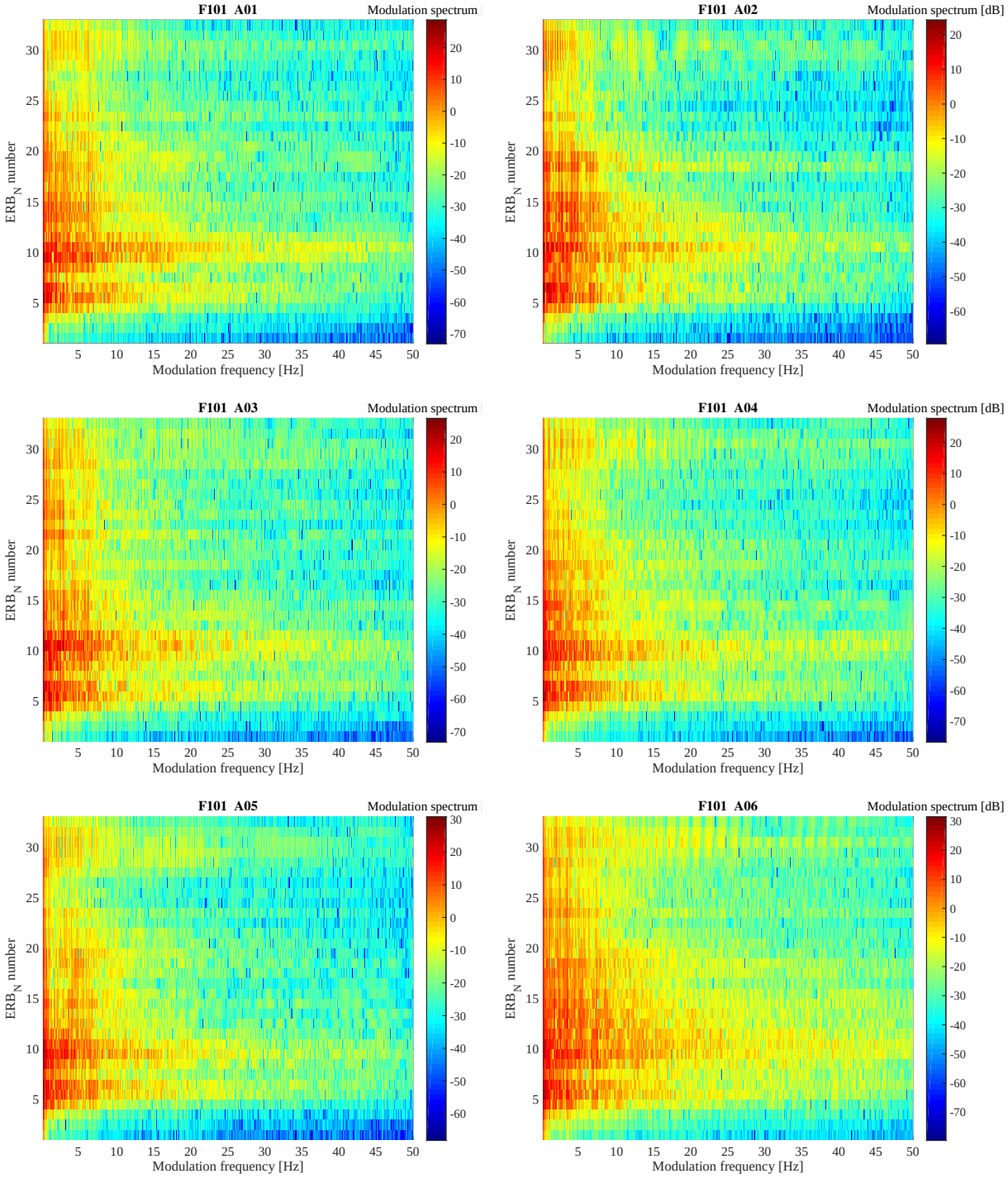

図 6.1: 話者 F101 の文章 A01 から A06 までの変調スペクトル

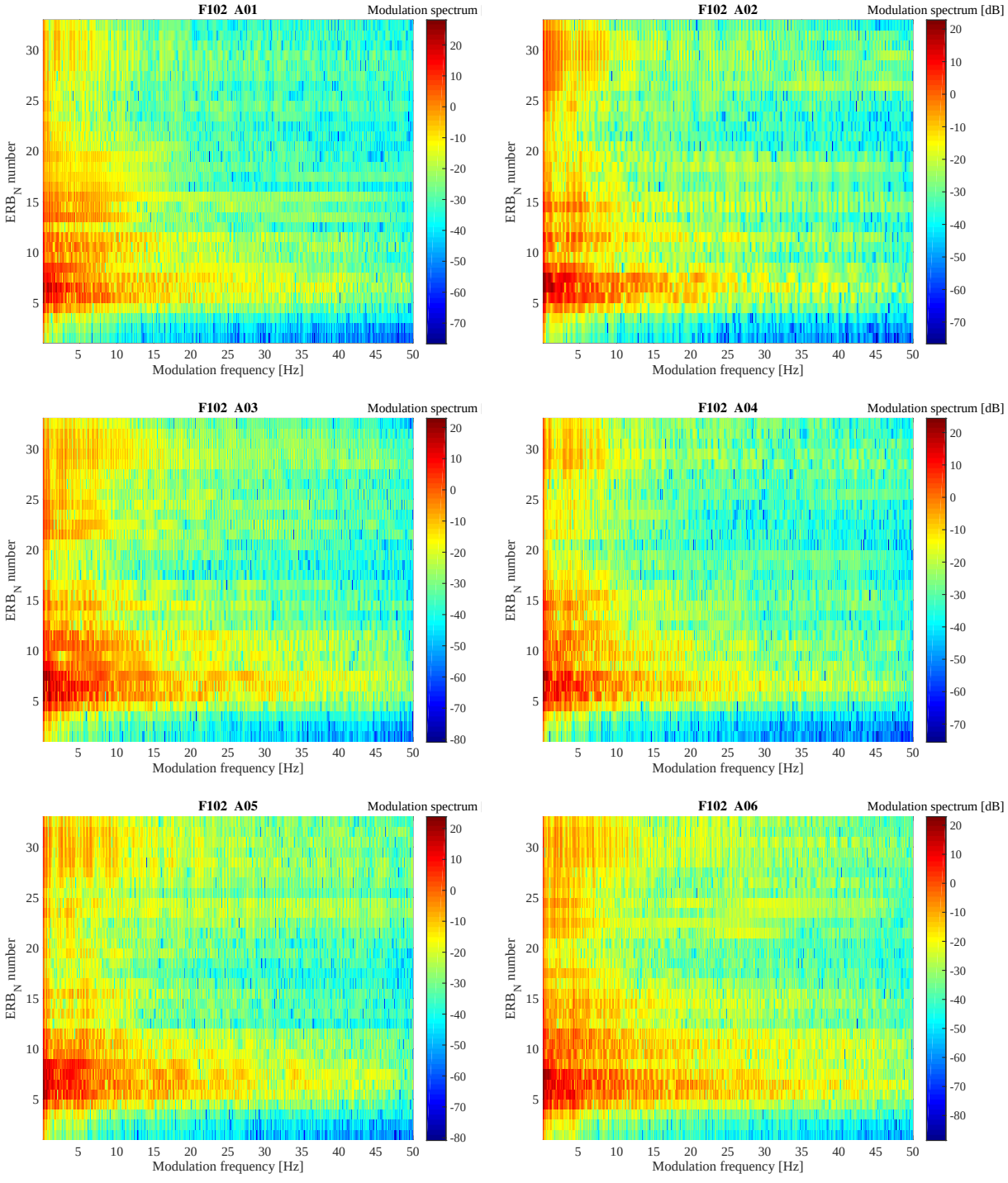

図 6.2: 話者 F102 の文章 A01 から A06 までの変調スペクトル

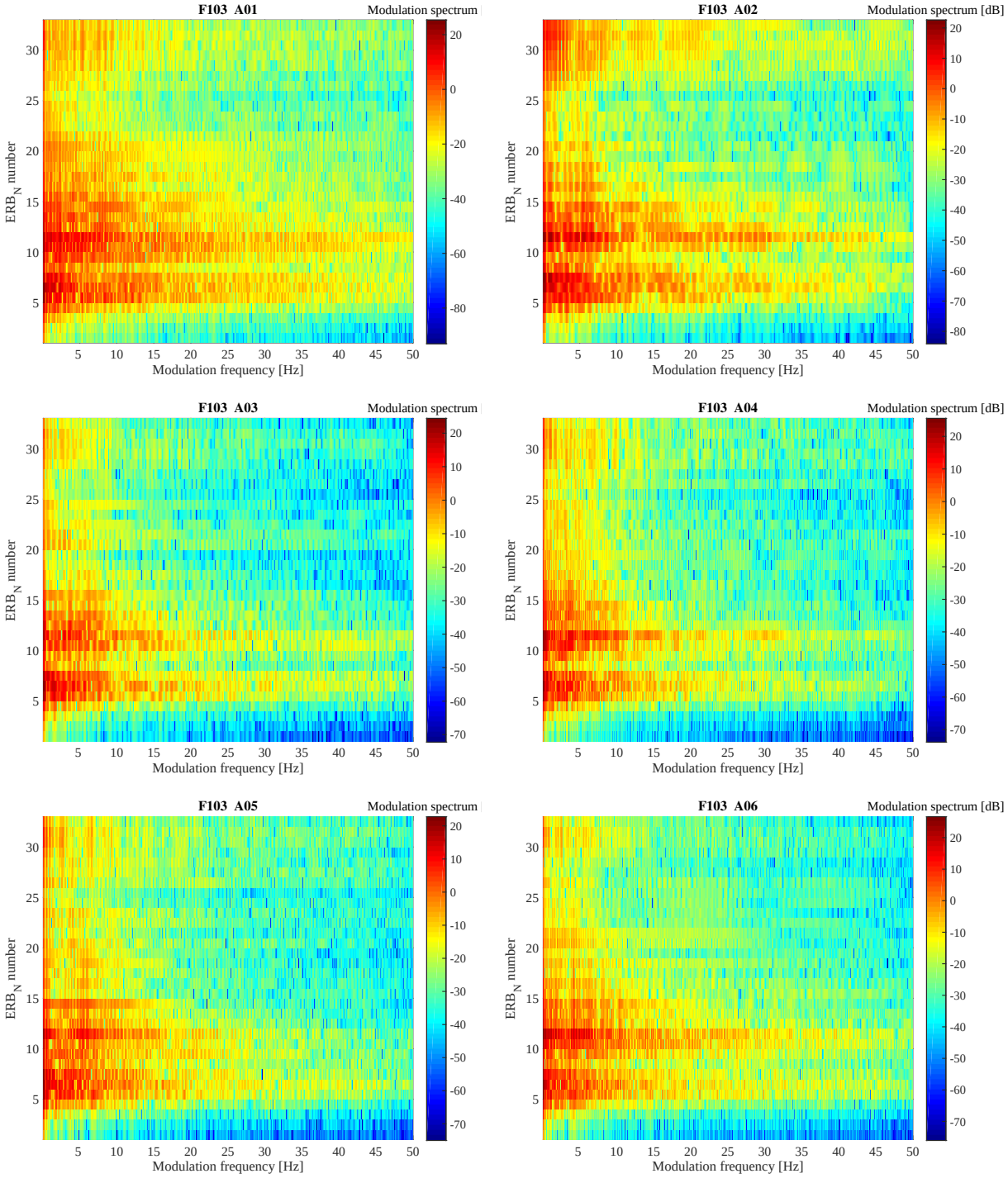

図 6.3: 話者 F103 の文章 A01 から A06 までの変調スペクトル

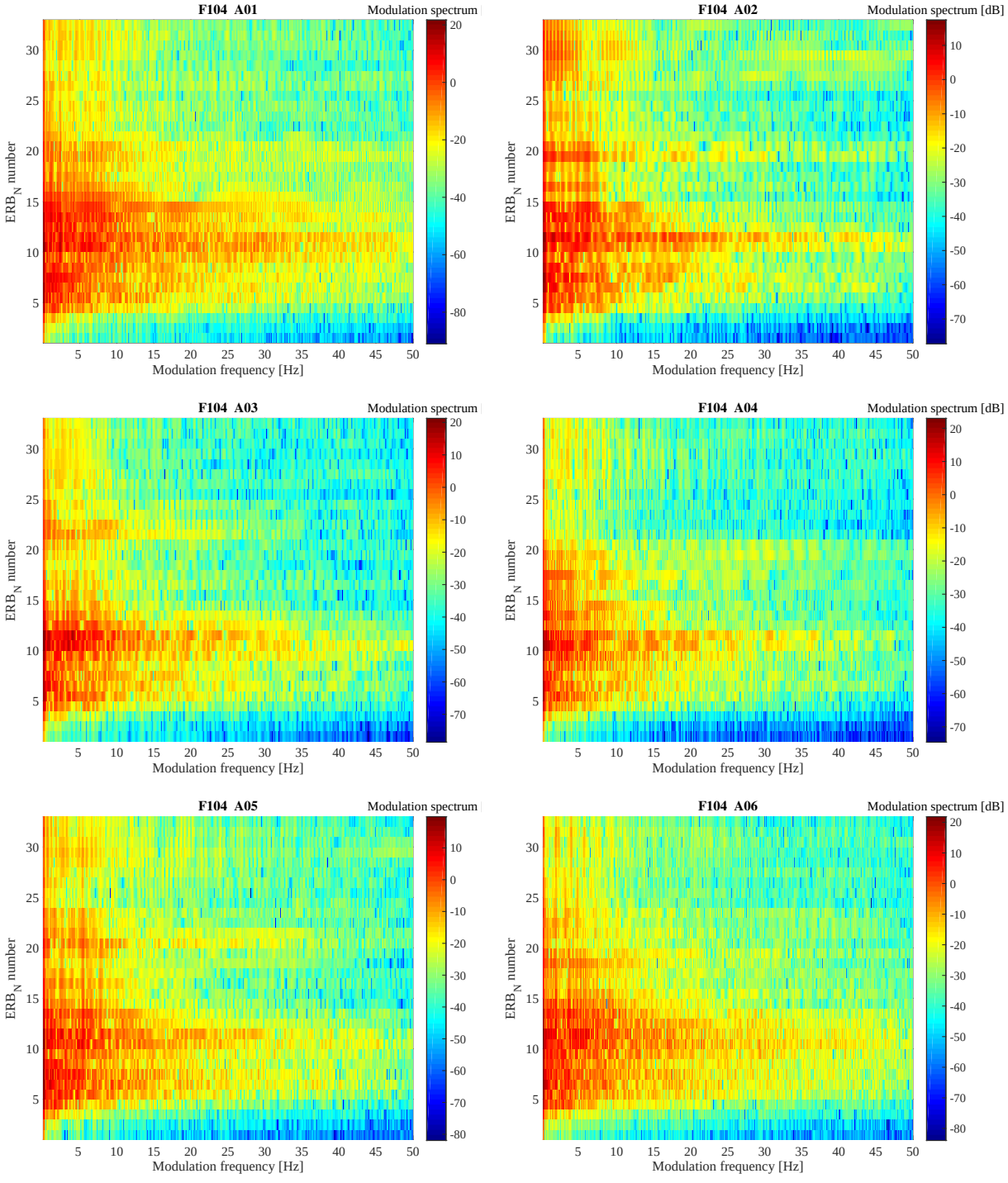

図 6.4: 話者 F104 の文章 A01 から A06 までの変調スペクトル

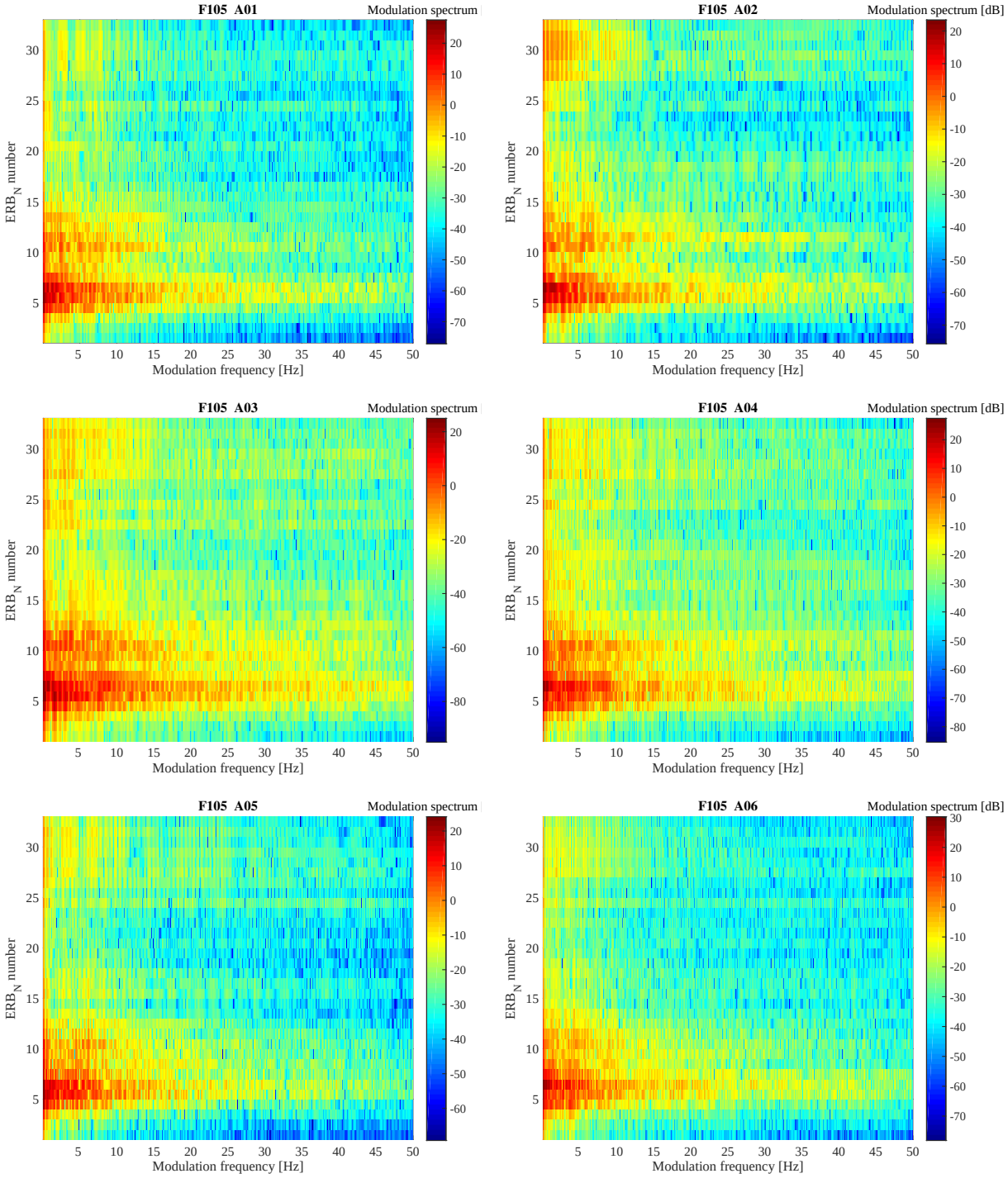

図 6.5: 話者 F105 の文章 A01 から A06 までの変調スペクトル