

TELECOM FRONTIER

Support Center for Advanced Telecommunications
Technology Research

SCAT

TECHNICAL JOURNAL

NO. **60**

2008 SUMMER

劣化合成音声を用いた模擬難聴の基礎検討

4

入野 俊夫 和歌山大学システム工学部教授

難聴者の聞こえを健聴者が体験できるような模擬難聴を実現することを将来の目標として研究を進めてきた。まず、難聴者の通常音声の知覚特性と健聴者の劣化音声知覚特性の対比を行った。もし両者の知覚特性が何らかの尺度上で一致した場合、劣化音声を作成した時に用いた音声分析合成系のパラメータ値として難聴者の聞こえの一面を定量化できる。ここでは、4名の難聴者の異聴表と、多数の健聴者が様々な劣化音声聞いた時の379個の異聴表の類似度を計算し、難聴に対応したパラメータ値を推定し、妥当な結果を得た。この実験では、単語音声を用いたが、データの収集に労力と時間が非常にかかる問題点があった。そこで、過去に電話品質評価等で多量に収集されてきた単音節の異聴表も推定に利用可能とすることを考えた。このためには、単語と単音節の知覚特性の相違を定量化する必要がある。ここでは、調音結合や統計的な音節遷移情報の影響について調べ、その傾向を定量化した。

多周波 FSK 信号エネルギー検波方式の移動体通信への適用に関する研究

14

岩波 保則 名古屋工業大学大学院工学研究科教授

ミリ波帯や準ミリ波帯などはマイクロ波帯に比べ広帯域幅の通信路を確保しやすいが、信号のコヒーレンスや周波数安定度が必ずしも良くない。また、このように高い周波数帯を用いる移動体の端末では、電力効率の良い飽和形の増幅器の使用が好ましく、定包絡線の信号形式が適している。このような状況下で利用可能な無線変復調方式の一つに、多周波 FSK(Frequency Shift Keying) 信号の非同期(non-coherent) エネルギー検波方式がある。FSK 信号のエネルギー検波方式は、白色ガウス雑音(AWGN, Additive White Gaussian Noise)の加わる無線(有線)通信路において、位相の不連続性や位相ジッタの存在する FSK 信号に対し、最小のビット誤り率(BER, Bit Error Rate)を与える非同期検波方式(キャリア位相同期を必要としない検波方式)である。エネルギー検波方式は、受信側でキャリア位相同期の必要がないので、受信キャリアの位相ジッタ(位相雑音)に対して強く、キャリア位相再生回路を必要とする同期検波方式に比べ、比較的簡易な受信機構成で実現できる検波方式として知られている。また、多周波数の FSK 信号を用いることにより、帯域幅は増加するが、高ビット速度の通信が可能となる。エネルギー検波を用いれば、受信 FSK 信号は位相連続である必要はなく、送信 FSK 信号は位相連続、不連続のどちらでも使用することができる。また受信キャリア信号の周波数オフセットの影響に対しても強い。このようなロバストな特徴は、移動体通信路などのマルチパス遅延波の存在する周波数選択性フェージング(周波数特性を有するフェージング)環境において、ミリ波や準ミリ波帯を用いた通信、例えば ITS(Intelligent Transportation System)通信などに適していると考えられる。本稿では、周波数選択性のマルチパスフェージング通信路に対する多周波 M-FSK のエネルギー検波送受信方式の有効性につき述べる。

劣化合成音声を用いた模擬難聴の基礎検討

Simulation of Hearing Impaired Listeners Through Perception of Degraded Speech



入野 俊夫 (Toshio IRINO, Dr. Eng.)

和歌山大学システム工学部教授

(Faculty of Systems Engineering, Professor, Wakayama University)

IEEE 米国音響学会(ASA) 米国聴覚研究学会 (ARO)
国際音声コミュニケーション学会 (ISCA) 電子情報通
信学会 日本音響学会 会員

受賞：2000年 日本音響学会 第40回佐藤論文賞 1989
年 日本音響学会 栗屋潔学術奨励賞

著書：「新編感覚知覚心理学ハンドブック Part 2」(第
III部聴覚「聴覚初期過程の機能モデル」pp.349-358)(共
著) 誠信書房 ISBN 978-4-414-30504-3 615p 2007年

"The Dynamics of Speech Production and Perception,"
("Vowel normalisation: Time-domain processing of the
internal dynamics of speech", pp.153-170)(共著), NATO
Science Series, Series A: Life Sciences, IOS press,
Amsterdam, ISBN 1-58603-666-1, 369p. 2006年 他

研究専門分野：聴覚計算理論 聴覚心理 音声信号処理

あらまし 難聴者の聞こえを健聴者が体験できるような模擬難聴を実現することを将来の目標として研究を進めてきた。まず、難聴者の通常音声の知覚特性と健聴者の劣化音声知覚特性の対比を行った。もし両者の知覚特性が何らかの尺度上で一致した場合、劣化音声を合成した時に用いた音声分析合成系のパラメータ値として難聴者の聞こえの一面を定量化できる。ここでは、4名の難聴者の異聴表と、多数の健聴者が様々な劣化音声を聞いた時の379個の異聴表の類似度を計算し、難聴に対応したパラメータ値を推定し、妥当な結果を得た。この実験では、単語音声を用いたが、データの収集に労力と時間が非常にかかる問題点があった。そこで、過去に電話品質評価等で多量に収集されてきた単音節の異聴表も推定に利用可能とすることを考えた。このためには、単語と単音節の知覚特性の相違を定量化する必要がある。ここでは、調音結合や統計的な音節遷移情報の影響について調べ、その傾向を定量化した。

1. はじめに

高度高齢化社会を迎えようとしている現在、「耳が遠くなる」人の数も増加することは確実であろう。騒音をなるべく避けて平静な環境で暮らせば、難聴になる確率も低くなると考えられている。しかし、自動車・電車・テレビ・音楽プレーヤー等は現代の日常生活で欠かせないものとなっており、騒音源と認識して避けて生きることはなかなか難しいであろう。一度難聴になってしまうと、再生医療の方法がまだ確立されていないため根本的な治療は難しい。そこで、補聴器を利用するしかなくなるが、すべての症状に対応できるものは開発できておらず、音声を満足に聞き取れるように戻るとは限らない。一方、高齢者に限らず、難聴に代表される音声コミュニケーション失調の問題は、健聴者が思っている以上に深刻で、言語聴覚士という臨床資格まで設けられている。これらの状況の改善に、工学の立場からも積極的に貢献する必要がある。特に情報通信技術分野では、「高齢者や障害者に優しい技術」の基礎を固め、サービスとして提供することは重要な課題となっている。

1.1 模擬難聴

いままで、視覚の緑内障や筋力の衰えなどの「模擬体験」をできる器具が開発され、若い健常者でも老人の苦労がわかる機会が増えてきた。ところが、聴覚器官の衰えを体験できる「模擬難聴」はまだ開発されておらず、難聴者の苦労を感じとれる機会もあまりない。もし「模擬難聴」が実現できると、(1)高齢化社会到来で急務にもなっている究極の補聴器の開発をきわめて効率化できること、(2)臨床現場で患者の聞こえの状態を正確に把握できるようになって治療法も画期的に改善できること、(3)聴覚障害への誤解や偏見を生んでしまう深刻な社会問題を根本から解決できること等、社会的なインパクトはきわめて大きい。

1.2 研究の流れ

この模擬難聴は歴史的に数多くの研究があるにもかかわらず未だに実現できていない。難聴と一言言っても様々な要因があるためである。この難問に、最新の音声分析合成技術と聴覚モデルに基づく信号処理を

用いて新たな観点から挑戦した。ここでは、難聴者の通常音声の知覚特性と健聴者の劣化音声の知覚特性の対比を行った⁽¹⁾⁻⁽⁴⁾。もし両者の知覚特性が何らかの尺度上で一致した場合、音声を劣化させた時に用いた音声分析合成系のパラメータ値として難聴者の聞こえの一面を定量化できる。第2章では、このアプローチと検討結果について紹介する。

ここでは、単語音声を用いて実験を行ってきた。これは、音声コミュニケーションの基本単位として、単音節（カナ1文字分）ごとの知覚特性を測定するより現実に近いと考えたからである。しかしながら、実験は時間と労力のかかる作業であり、十分なデータ量を集めるのは容易ではない。このような劣化音声の知覚に関しては、1960年代に電話品質を評価するための単音節明瞭度試験がすでに行われており⁽⁵⁾、得られた多量のデータが存在する。劣化条件は異なるが、このデータも対比に利用できるようにすることは、推定精度向上のためにも重要である。そのためには、単語と単音節の知覚特性の相違を定量化する必要がある。このための検討も行ったので、第3章で紹介する。

2. 健聴者と難聴者の音声知覚の対比

難聴は、外耳・中耳の障害に主に起因する伝音性難聴と、聴覚末梢系の内耳やそれ以降の部位の障害に起因する感音性難聴の2つに大きく分類される。前者は比較的単純であるが、後者の場合障害部位によって症状が異なる上に、部位の特定が難しい場合もある。そもそも、内耳より後の聴覚系の情報処理機構自体があまり解明されておらず、現在の重要な研究課題となっている。

2.1 研究の方略

ここでは情報処理機構の解明にこだわらず、図1に示す方略を取って難聴のパラメータ化を目指す。まず、図1左上示すように、難聴者が外界の音声を知覚する場合に、何らかの情報欠落あるいは劣化が起きていると考える。この劣化の状況を何らかの形で定量化/パラメータ化できれば良い。そこで、図1右図のように、外界の音声を実験者がパラメータ制御できる音声分析合成法を用いて劣化させた音声を作成し、それを健聴

者に聞かせる。ここで、多数の音声データを用いて、難聴者が通常音声を聴取して書き取った結果と、健聴者が劣化音声を聴取して書き取った結果を比較する。もし、書き取った単語内の音素の正解や間違いが一致したとすると、難聴者内の劣化特性の一面を実験者が制御した劣化パラメータとして定量化できることになる。

2.2 異聴表

実際の比較では、たとえば図1左下の4つの図（難聴者4名分）で示される「異聴表」を用いる。異聴表とは、提示した単語音声内の音素に対して、被験者が回答した音素を2次元のマス目に1つずつ加算し、全音素分累積させた表である。この図では濃度で度数を表現している。もし、提示音素と回答音素が完全に一致した（完全正解した）とすると、対角成分だけ数値が存在し（濃い色となり）、他は零（白色）となる。逆に提示と回答で音素が異なると非対角成分に値が出て、この割合は間違いの増加とともに増加する。

図1左下の難聴者の異聴表では、非対角成分が多く、音素誤りの数が多いことがわかる。これに対して、健聴者6名分を合計した異聴表を、ある実験の劣化条件ごとに示したのが図1右下の3×3分割図である。#1～#9まで番号の小さい方がおおむね劣化の度合いが大きい。実際、#1では対角成分がほとんど存在せず、非対角成分のみである。これに対して、劣化条件が緩和され原音声に近くなる（条件番号が増える）に従い対角成分の度数が多くなっていることがわかる。

2.3 聴取実験

上記の異聴表を得るために、単音節でなく単語音声を用いた。これは、刺激音をなるべく日常会話に近づけるためである。ただし、単語の辞書的な知識の影響を最小限にするため、「親密度別単語了解度試験音表(FW03)」⁽⁶⁾の日本語4モーラ単語群の中から親密度が最も低い単語群（親密度2.5～1.0）を用いた。これは、国語辞典には掲載されているものの、あまり耳馴染みがなく意味を知らないことが多い単語群に相当する。また、ここでのモーラとは、日本語のカナ1文字分（音節としては数えない長音「ー」、促音‘Q’「ッ」、撥音‘N’「ン」を含む）に相当する。つまり4モーラ単語は

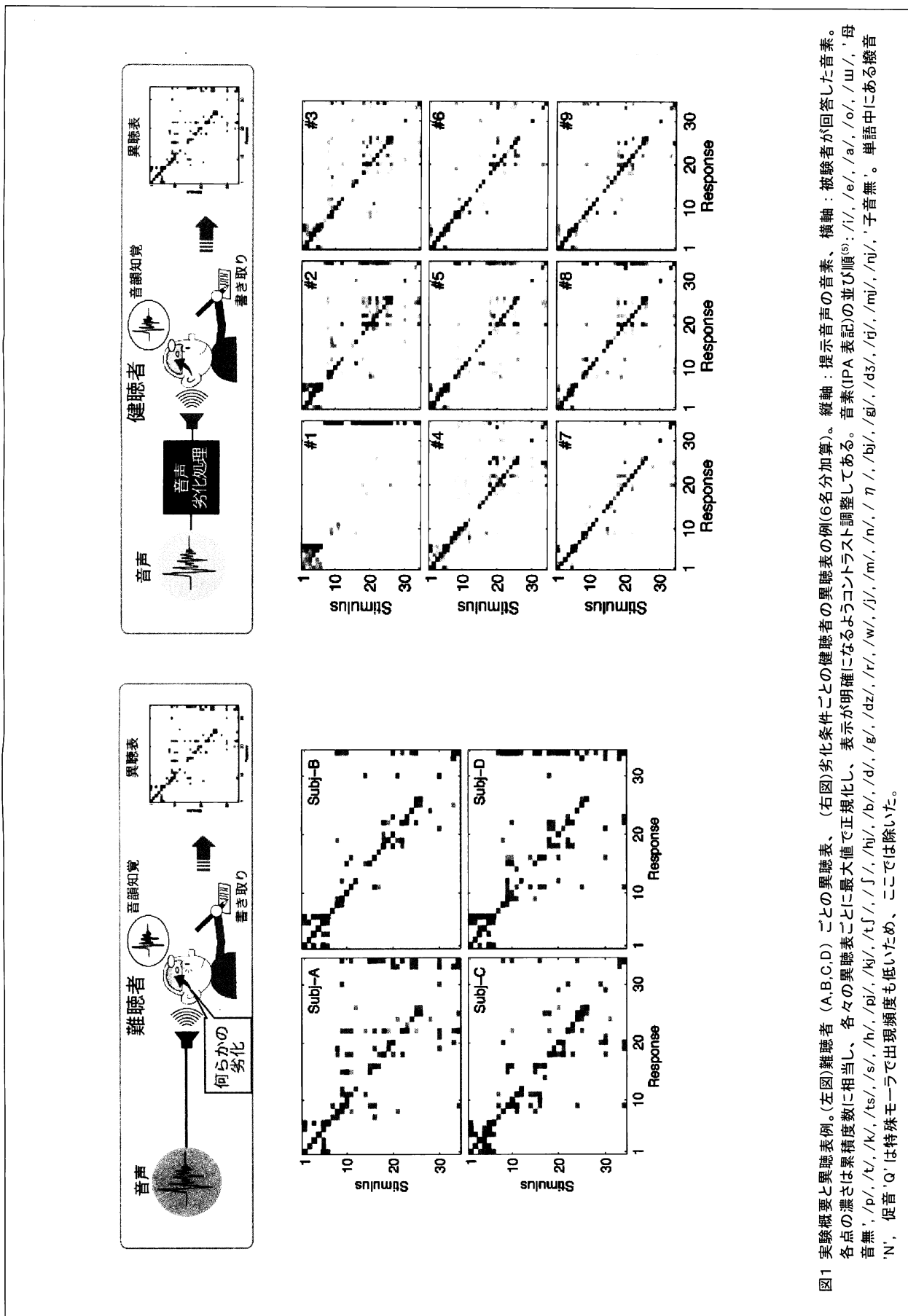


図1 実験概要と異聴表例。(左図)難聴者 (A,B,C,D) ごとの異聴表、(右図)劣化条件ごとの健聴者の異聴表の例(6名分加算)。縦軸：提示音声の音素、横軸：被験者が回答した音素。各点の濃さは累積度数に相当し、各々の異聴表ごとに最大値で正規化し、表示が明確になるようコントラスト調整してある。音素(IPA表記)の並び順⁽⁶⁾：/i/, /e/, /a/, /o/, /u/, '母音無', /p/, /t/, /k/, /ts/, /s/, /h/, /pj/, /kj/, /tʃ/, /ʃ/, /tʃ/, /kʃ/, /b/, /d/, /g/, /dz/, /r/, /w/, /j/, /m/, /n/, /ŋ/, /bj/, /gi/, /dʒ/, /dʒ/, /mj/, /nj/, '子音無'。単語中にある撥音 'N', 促音 'Q' は特殊モーラで出現頻度も低いいため、ここでは除いた。

カナ4文字で書ける。この低親密度の1000単語(50単語x20セット)について、男女各1名分の発話音声データを原音声として実験に用いた。表1に、難聴者と健聴者の実験条件を示す。

2.3.1 難聴者による聴取実験

上記のデータベース中の1セット50単語の女性話者の原音声を、難聴者(4名A,B,C,D)に提示して書き取り実験を行った⁽⁷⁾。このとき、難聴者ごとに数条件の音圧レベルで提示した。ここでは、このうち最も成績が良かった条件の場合の異聴表(図1左下)を用いることにする。

2.3.2 健聴者による聴取実験

健聴者には、STRAIGHT^{(8),(9)}のパラメータを操作して作成した劣化音声を実験刺激として提示した。劣化合成音声作成の詳細な説明は、文献(3),(10)を参照されたい。劣化条件は、周波数軸劣化(スペクトログラムを周波数軸方向に平滑化、Warped-DCT係数で制御)、時間軸劣化(スペクトログラムを時間軸方向に平滑化、変調上限周波数(Hz)で制御)、音声駆動音源劣化の3通り

がある。この組み合わせを様々に変えて、6実験を実施した(表1)。Warped-DCT係数も変調上限周波数も数値が小さいほど、スペクトル変形が大きくなり劣化度合いが増大して聞き取りにくくなる。音声駆動音源劣化は数値が低いほどささやき声に近くなる。しかし、他のパラメータほど知覚に影響がないことがわかった。また、被験者にヘッドホンで音声を提示するが、この時の音圧も重要なパラメータで、音圧が低いほど聞き取りにくくなる。70dB SPLは通常の会話程度で、40dB SPLではやや聞き取りにくくなる。

全劣化条件数は49で、原音声も分析合成系の品質確認のために利用した。なお、劣化条件あたり男女1話者ずつ各1リスト(50個)の音声試料を、リストの重複が最小限にして選び、可能な限り記憶で回答できないようにした。被験者は20歳代の健聴者の男女のべ35名である。得られた異聴表は、音声話者1名分ごとに379個である。図1右下図は、実施したうちの1つの実験において、劣化条件ごとに6名の被験者の回答を集計した異聴表である。

2.4 パラメータ推定

難聴者1名ずつの異聴表と、健聴者の379個の異聴表(女性話者分)間で類似したものを探す。もし極めて似ているものが見つければ、難聴者の聞こえの劣化は、劣化合成音声を作成した時のパラメータとして定量化できるであろう。まず、各異聴表(34x34の行列)を、母音と子音の間の異聴成分(すべて0)を取り除いた上で、1次元ベクトル(820成分)として表現しなおし、正規化した。難聴者と健聴者の異聴表間の類似度を、ベクトルどうしの内積値として定義した。両者が完全に一致する場合に1となり、相違が大きくなるほど値が小さくなる。ただし、難聴者と健聴者で聴取する単語音声セットが必ずしも一致しないため、異聴表が完全に一致することはまずありえない。

表2に、難聴者(A~D)ごとに類似度が最も高くなった劣化条件を示す。類似度2位以降もおおよそ同様な傾向であった。これを見ると、音声の提示音圧が共通で40dB SPLであることがわかる。類似度が0.97以上の難聴者B,Cでは、周波数軸方向劣化が無く、時間軸方向劣化も軽度の変調上限周波数15Hzである。類

表1 実験条件と設定した劣化パラメータ値

	難聴者	健聴者
被験者数	4	のべ35(6実験)
全刺激条件数	原音声のみ1	劣化合成音 49 原音声 5
周波数方向劣化 Warped-DCT次数	-	0.1,3,4,5,7, 10,15,∞
時間軸方向劣化 変調上限周波数	-	0.1,3,5,7, 10,15,20,∞
音源波形劣化 合成音源制御値	-	delsp=0.5,20 Cornf=0,1500,4000
提示音圧(dBSPL)	70,80,80,115	40,70,78
FW03音声話者	女性1名(fhi)	男性1名(mis) 女性1名(fhi)

表2 類似度の最も高い条件のパラメータ値

難聴者	A	B	C	D
周波数方向劣化 Warped-DCT次数	15	∞	∞	7
時間軸方向劣化 変調上限周波数	7	15	15	15
音圧(dBSPL)	40	40	40	40
類似度	0.88	0.98	0.97	0.86

似度が0.9未満の難聴者A,Dでは、周波数軸方向劣化あるいは時間軸方向劣化の度合いが中程度の7となっている。これらのことより、音圧が重要な要因であることと、スペクトル的な平滑化も同時に必要であることがわかる。

さらなる精度向上のためには、実験条件と被験者数を増やすことが必要である。しかし、実験には労力と時間がかかるため、十分なデータ量を集めるのは決して容易ではない。そこで、別の目的で過去実施された研究において得られている単音節の異聴表データを推定に利用できるようにすることも重要である。次章では、そのために単語音声と単音節の関係を明確化する実験について述べる。

3. 単語音声と単音節の知覚特性

単語音声を聞き取る場合、最後の音節がたとえ聞き取りにくかったとしても、予測して答えることができる。これは、単語中の音節には、並び方の規則性があり、ランダムではないためである。このような予測は、言語の辞書的な知識により可能になると一般的には考えられている。そこで、今回の実験ではこの要因をなるべく排除するべく親密度の最も低い単語セット⁽⁶⁾を用いた。しかしながらこの場合でも、単語音声の中の個々の音素の正解率は、音節の出現頻度と単音節正解率から計算される値よりも高くなり、しかも劣化の度合いに関わらずこの傾向があることがわかった⁽³⁾。この原因を特定し、モーラや音素の正解率(明瞭度)と単語の正解率の間に一定の法則性が見つけられれば、両

者の異聴表を適切に重み付けして利用できるようになると考えられる。

3.1 作業仮説と実験設計

音素や音節は言語学的なラベルであり離散量である。これに対して、通常発声される単語や文は、口の中の舌の動きによって生成されるため、離散的な量として表すことはできない。たとえば「アイ」を発声した場合、2つの母音/a/と/i/の間では連続的にスペクトルが遷移する。これが調音結合と呼ばれ、前後の音素の情報の両方を共有している中間状態が存在することになる。従って、次の音素を「確実に」聞く前に、予測できるだけの情報を得られる可能性がある。

そこで、追加実験1として、この調音結合の影響を取り除くことを考えた。前章で用いたFW03の単語の音節順番を保持したまま、各音節部分を単独発声音節(=単音節=1文字分)に置き換えて調音結合をなくした「有意味単音節系列音」を作成した。この系列音から、前章の自然発話単語の実験で用いた劣化条件(表3)と同一にして劣化音声を作成し、聴取実験を行った。

今回のFW03データベースは、4モーラの単語で構成されていて、その単語としての親密度別に分類されている。しかし、その単語はそもそも既知の漢字2文字で書ける場合が多いため、ランダムに並べた場合に比べて次のモーラの予測可能性は高くなる。(例「あいきゃく」=「相客」。)これは上記の音響的な調音結合とは異なる情報であるため、(日本語)単語に特有な音節の統計的遷移情報として扱う。

そこで、追加実験2として、単語セット中にある音節をランダムに並べ替えたリストに従って、単音節を4モーラ分並べて「無意味単音節系列音」を作成した。これも表3の劣化条件に従って劣化音声を作成して、聴取実験を行った。

上記の2種類の単音節系列音実験の基本的な実験条件等は、対比を行う対象の自然発話単語の場合(前章)と同一である。有意味/無意味の単音節系列音の具体的な作成法や被験者等の実験依存の部分の詳細情報は文献(10),(11)にゆずり、ここでは実験結果のみを図2、図3にまとめて示す。

表3 実験内の条件番号と劣化パラメータ値

条件番号	1	2	3	4	5	6	7	8	9	10	11
周波数方向劣化 Warped-DCT次数	0	7	7	7	15	15	15	∞	∞	∞	原 音 声
時間軸方向劣化 変調上限周波数	0	7	15	∞	7	15	∞	7	15	∞	

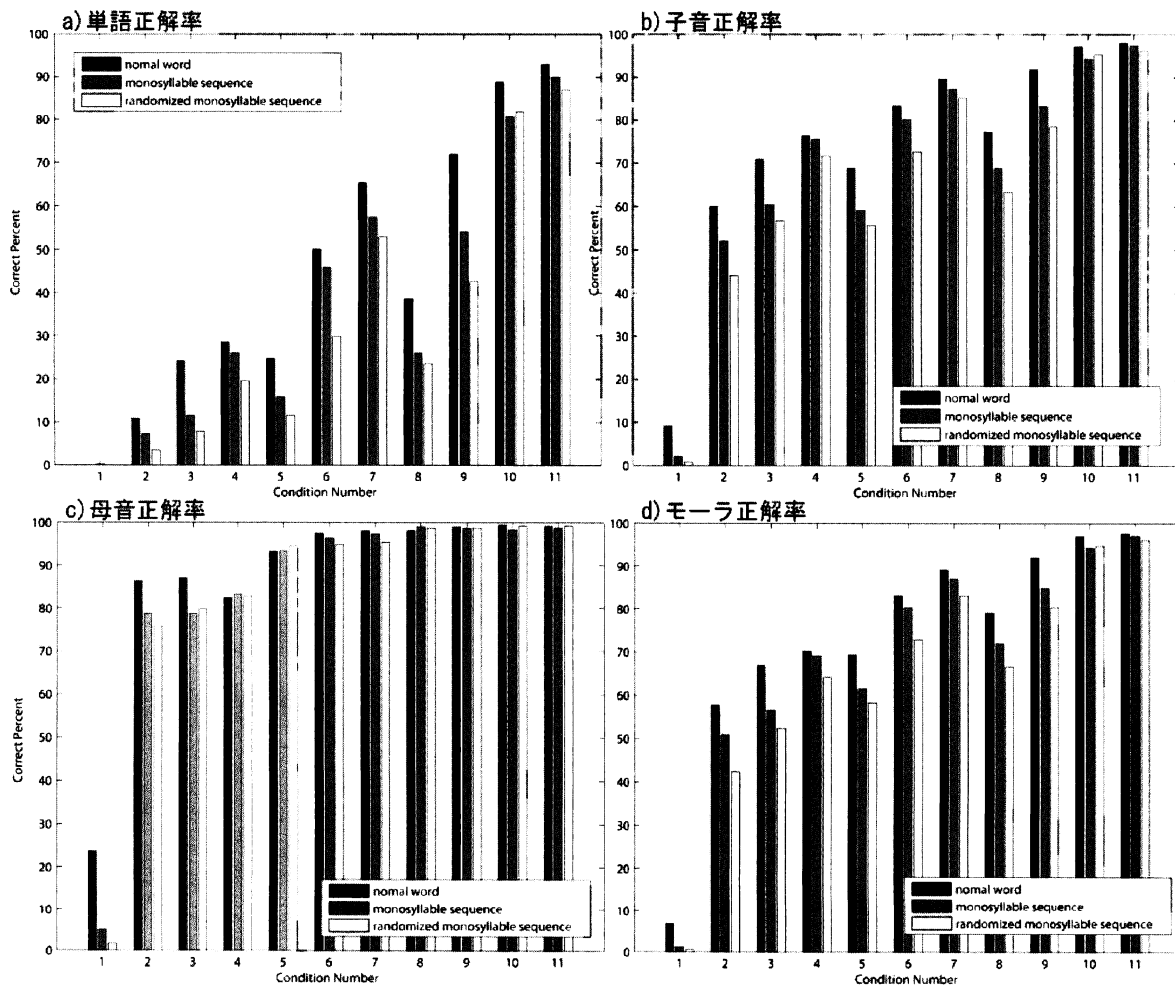


図2 (a)単語・(b)子音・(c)母音・(d)モーラ正解率。横軸は表3に示した劣化条件番号。縦軸は正解率(%)。棒グラフは、
 各々自然発話単語(1)-(3)(黒)・有意義単音節系列音(灰色)・無意味単音節系列音(白)の場合の実験結果を示す。

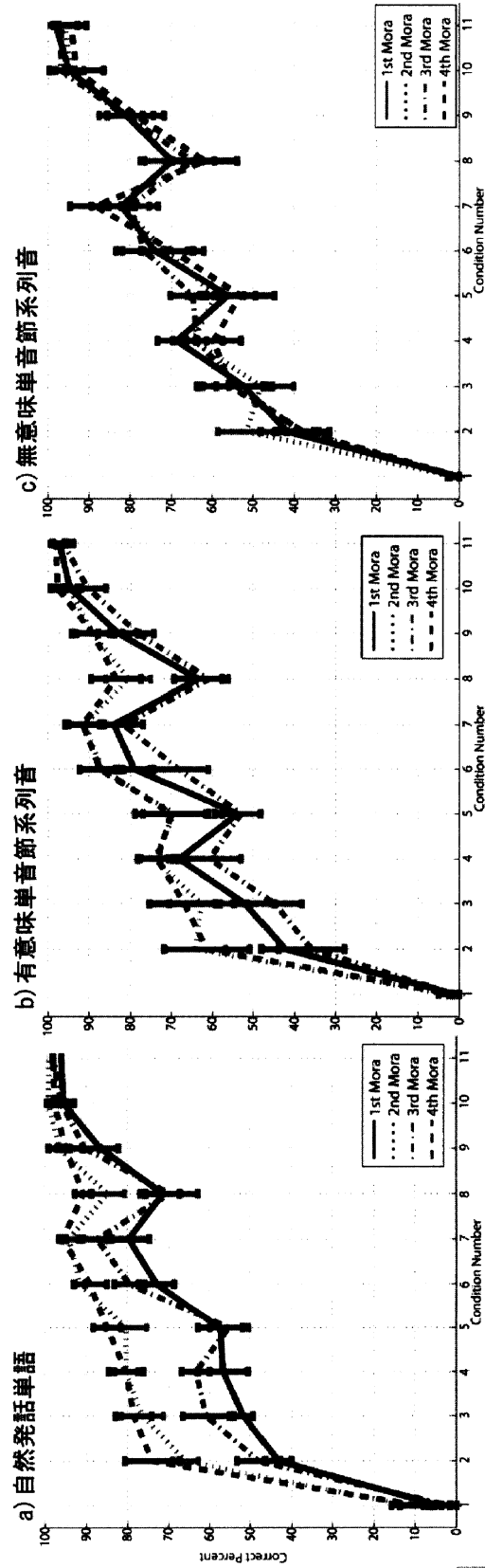


図3 モーラ順番ごとの正解率。横軸は表3に示した劣化条件番号。縦軸は正解率(%)。実線：1モーラ目・点線：2モーラ目・鎖線：3モーラ目・破線：4モーラ目。縦棒は被験者間の標準偏差を示す。

3.2 実験結果

まず、調音結合の影響を調べるために、自然発話単語(図2黒棒、図3(a))と有意義単音節系列音(図2灰色棒、図3(b))の実験結果を比較する。図2より自然発話に比べて有意義単音節系列音の場合、全体的に正解率が低いことがわかる。図2(c)の母音に関しては、条件4~11で低下は少ないが、これは天井効果で差が見えにくくなっているためと考えられる。また、図3(a)の自然発話単語のモーラ順番の正解率の実験結果より、1・3モーラ目の正解率よりも2・4モーラ目の正解率の方が高いことがわかる。同様な傾向が、図3(b)の有意義単音節系列音の場合においても見られる。しかし、その差は自然発話単語の場合に比べて小さくなっている。このことから、調音結合による影響はあるものの、さらに別の要因でこの差が出ていることがわかる。

次に、音節の統計的な遷移情報の影響を調べるため、有意義単音節系列音(図2灰色棒、図3(b))と無意味単音節系列音(図2白棒、図3(c))の結果を比較する。図2より単語・子音・モーラの正解率は、両者の条件で異なっているが、母音の正解率はほとんど差がみられない。図3(c)においては、モーラ順番によって正解率が異なるという一貫した傾向がないことがわかる。このことから、今回使用した低親密度の単語の場合でも、音節の並び方の統計的な情報が、知覚系の単語認識において用いられていることがわかる。

3.3 単語と音素やモーラの正解率の関係

単語正解率 P_w を、子音正解率 P_c と母音正解率 P_v 、あるいはモーラ正解率 P_m から以下の式で推定することを考える。

$$P_w = P_c^M P_v^N, \quad P_w = P_m^K \quad (1)$$

もし、音素どうしの情報が独立であれば、 $M \cdot N$ は4モーラの単語中に現われる子音・母音・モーラの平均出現頻度に一致するはずである。FW03データベース中の単語から求めた子音と母音の4モーラ単語内での平均出現頻度は、表4の2行目に示すように3.5と3.6である。またモーラ数はもちろん4である。

図2に示した通り、各実験において、単語正解率の他に子音・母音・モーラの正解率も求まっている。この正解率を式1に代入し、実験した11劣化条件について、最小自乗法を用いて $M \cdot N \cdot K$ の最適値とそれぞれの場合の推定誤差の平均値を算出した。

この結果を実験ごとに表4の3~5行目に示す。まず4列目と6列目を見ると、どの実験条件においても平均誤差は2%程度と十分小さくなることがわかる。2列目の子音正解率に対するべき乗数 M の値は全ての実験において平均出現頻度の値とほぼ同じである。これに対し、3列目の母音正解率に対するべき乗数 N の値は、実験ごとに異なる。ここで、自然発話単語・有意義単音節系列音・無意味単音節系列音の順番に次第にべき乗数 N の値が大きくなっていることがわかる。これは、音

表4 子音・母音・モーラの単語内の平均出現頻度、および認識率から推定したべき乗値と平均誤差(%)

条件	M (子音)	N (母音)	平均 誤差(%)	K (モーラ)	平均 誤差(%)
平均出現頻度	3.5	3.6	—	4	—
自然発話単語	3.6	1.6	2.4	3.8	2.5
有意義単音節系列音	3.4	2.1	2.5	3.9	2.1
無意味単音節系列音	3.4	2.8	2.2	3.8	2.1

素間の独立性の度合いを反映した結果であることがわかる。ただし、無意味単音節系列音の場合でも平均出現頻度と一致していない。これは、無意味単語であっても、日本語の子音-母音音節（1モーラ）ごとに並び方を変えただけであり、音節構造はそのまま保持されているためであると考えられる。

表4の5列目のモーラ正解率に対するべき乗数 K の値に関しては、用いた単語の種類に関わらずほぼ一定で4(モーラ)より若干低い程度であった。これは、モーラ単位で計ると子音・母音の区別がなくなるため、 $M \cdot N$ の場合のような差異が明確に出ないことを示している。

日本語の音素正解率と単語正解率の関係に関して、べき乗数の形で求めた。この結果は、データ全体に対する傾向であり、個別の音素の異聴に関しては語っていない。しかし、単語と単音節の異聴表を統合する場合の制約条件としては使えるのではないかと考えている。

4. まとめと将来展望

本報告では、模擬難聴実現への1ステップとして実施した研究を紹介した。まず、音声合成系のパラメータを系統的に操作し合成した劣化音声に対する健聴者の知覚特性と、通常音声に対する難聴者の知覚特性の対比を行い、難聴に対応したパラメータ値を推定した。精度はまだ高くないと考えられるが、数値的には妥当な結果が得られた。また、収集に労力と時間がかかる単語の異聴表だけではなく、既存の単音節異聴表も推定に用いるために、調音結合や統計的な音節遷移情報の影響について、傾向を定量化した。

最近、パソコン上で実時間動作するSTRAIGHTシステム⁽¹²⁾を開発した。これを用いて、実環境において模擬難聴を試験的に体験することも可能となっている。あとは、劣化パラメータ設定を適切に行うだけで良い。今後、推定したパラメータがどの程度難聴者の状態を模擬できているのかを検討するつもりである。また、模擬の対象になる難聴者の数を増やすのであれば、健聴者の劣化音声知覚のデータをそれに従って大幅に増やす必要がある。このデータ収集の労力を減らすためにも、既存の単音節明瞭度データとの統合や、推定法を改善して精度を向上させることも重要な研究課題である。

すべての難聴者に対応できる模擬難聴実現のためには、地道なデータ収集と革新的な手法開発の両者が必要であろう。今後も積極的に研究を推進するつもりである。

参考文献

- (1) Jin Jiang, Hideki Banno, Hideki Kawahara, and Toshio Irino, "Intelligibility of degraded speech from seared STRAIGHT spectrum", *Proc. ICSLP 2004*, 4, pp.530-533, (2004).
- (2) 佐藤諭、入野俊夫、坂野秀樹、河原英紀、“スペクトル時間変化を制限して合成した劣化音声の知覚”、日本音響学会春季研究発表会講演論文集、pp.251-252、(2005)。
- (3) Toshio Irino, Satoru Satou, Shunsuke Nomura, Hideki Bano, Hideki Kawahara, "Speech intelligibility derived from time-frequency and source smearing", *Proc. Interspeech 2005*, pp.1737-1740, Lisbon, Portugal, Sep.4-8, (2005).
- (4) 中家諒、綿貫敬介、坂本真一、入野俊夫、“健聴者の劣化音声知覚と難聴者の通常発声知覚の対比”、日本音響学会春季講演論文集、pp.483-484、(2006)。
- (5) 齊藤収三、“日本語音韻異聴表”、電気通信研究所研究実用化報告別冊 **6** 号、(1961)。
- (6) 坂本修一、鈴木陽一、天野成照、小澤賢司、近藤公久、曾根敏夫、“親密度と音韻バランスを考慮した単語了解度試験用単語リストの構築”、日本音響学会誌 **Vol.54**、pp.842-849、(1998)。
- (7) 大多和史子、原田美加、坂本真一、西野安男、大屋茂喜(リオン)、“親密度で統制した単語リストを用いた補聴効果の評価に関する一考察”、日本音響学会聴覚研究会資料、H-2004-110、**Vol.34(9)**、(2004)。
- (8) Hideki Kawahara, Ikuyo Masuda-Katsuse and Alain de Cheveigne: "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-

frequency-based F0extraction: Possible role of a repetitive structure in sounds”, *Speech Communication*, **27**, pp.187-207, (1999).

- (9) 河原英紀、“Vocoder のもう一つの可能性を探る - 音声分析変換合成システム STRAIGHT の背景と展開 - “、日本音響学会誌、**Vol.63**, **No.8**, pp.442-449、(2007)。
- (10) 森本隆司、入野俊夫、河原英紀、“劣化处理した単音節系列音の知覚に関する検討”、日本音響学会：聴覚研究会資料 H-2007-135, **37 (10)**, pp.775-780, 熊本大学、熊本、2007年12月6日～7日。
- (11) 森本隆司、入野俊夫、河原英紀、“劣化处理した単音節系列の知覚に関する検討”、日本音響学会：秋季研究発表会講演論文集、pp.595-596、山梨大学、甲府、2007年9月19日～21日。
- (12) Hideki Banno, Hiroaki Hata, Masanori Morise, Toru Takahashi, Toshio Irino, and Hideki Kawahara, “Implementation of realtime STRAIGHT speech manipulation system: Report on its first implementation,” *Acoust. Sci. & Tech.* (ed. by the Acoustical Society of Japan), **28(3)**, pp. 140-146, (May, 2007).

この研究は、平成16年度SCAT研究助成の対象として採用され、平成17年度～19年度に実施されたものです。