「ことばの鎖」に基づく音声明瞭度評価手法の確立

研究代表者　田　村　俊　介　九州大学　医学研究院　特任助教
共同研究者　水　町　光　徳　九州工業大学　大学院工学研究院　教授

1 背景と目的

音声明瞭度とは，雑音下や残響下など様々な環境下で評価される音声の聴き取りやすさのことであり，情報・通信の分野においては，音声信号処理によって音声明瞭度を予測する指標を開発する研究が進められている。例えば，室内音場における音声の聴き取りやすさを予測する speech transmission index (STI)[1]がよく知られている。他にも，様々な指標の開発が進められており，それぞれの指標は特定の状況下においては有効性が示されている。しかしながら，汎用性の高い音声明瞭度指標が提案されているとは言い難い状況である。

我々の研究では,音声コミュニケーションの概念モデルである「ことばの鎖」[2]に基づいて，音声明瞭度を評価する方法を再考し，新たな予測指標を提案することを長期的な目標としている。「ことばの鎖」では，発話者から聴取者に至るまでの音声コミュニケーションの流れが，音響学的・生理学的・言語学的レベルに分けて考えられている（図 1)。まず，発話者は脳内で言語情報を形成し（言語学的レベル），その情報を基に調音器官の制御を行うことで（生理学的レベル），音声信号を発する（音響学的レベル)。一方で，聴取者は発話者が発した音声信号を聴取し，その信号を聴覚器官での処理を通して神経信号に変換することで（生理学的レベル），最終的に脳内で言語情報を抽出する（言語学的レベル)。本研究調査期間内での目標は，聴取者の視点から見た音響学的・生理学的・言語学的レベルの各段階に注目しながら音声明瞭度を評価する枠組みを作ることである。具体的には，まず音響的レベルに注目して，ロンバード音声（雑音環境下で発声される明瞭度の高い音声）の音響的分析を行うことで，明瞭度の高い音声信号が持つ音響的特徴を明らかにする。さらに，ロンバード音声の音響的特徴を様々に操作した分析合成音声を作成し，その分析合成音声を用いた脳機能計測及び主観評価実験を行うことで，音響学的なレベルで見られるロンバード音声の特徴が生理学的レベルでどのように表現され，さらに，生理学的レベルでの表現がどのように言語学的なレベルで音声明瞭度に反映されるのかを調べる。

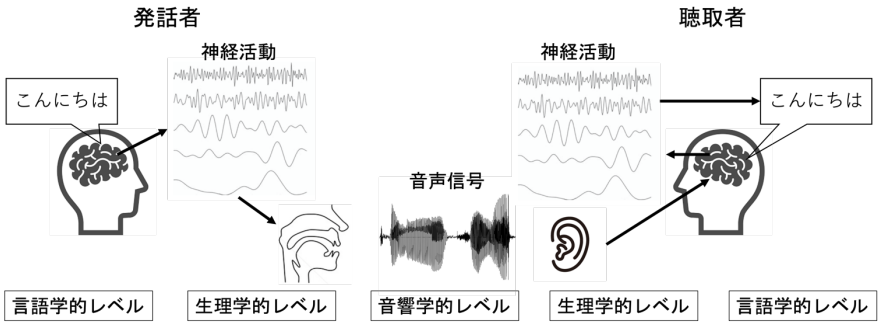

図 1. ことばの鎖のイメージ図

本研究調査で行うロンバード音声の生成や知覚に関する基礎的な研究の知見は，情報通信分野では，ロンバード音声の特性を反映した音声強調技術としてすでに実用化もされている（HOYA サービス株式会社 VoiceText シリーズ)。そのため，ロンバード音声の持つ音響的特徴とその明瞭度への影響度については，先行研究でも一定程度の知見が蓄積されてきている。しかしながら,「ことばの鎖」で考慮されている生理学的レベルを考慮した研究は申請者の知る限り存在しない。つまり，音声コミュニケーション研究として，人間科学的なエビデンスが充分ではない状態である。一方で，神経科学分野の研究では，音声信号の持つ音響的特徴と神経活動の関係性を調べる研究が進められており，脳波や脳磁図といった脳機能計測手法を用いて，聴取者の音声の聴き取り能力を他覚的に評価する方法なども検討されている[3]。しかしながら，雑音環境下

公益財団法人電気通信普及財団
研究調査助成報告書　第38号　2023年度

での音声を知覚する際の脳活動を調べる上で，ロンバード音声が用いられている研究は申請者の知る限り存在せず，主に静寂下で発声された音声に背景雑音を被せるという生態学的妥当性の低い刺激設定での検討が行われている。つまり，神経科学分野での研究では，本研究調査で注目している「ことばの鎖」の観点から見ると，音響学的レベルが上手く考慮されない状態で研究が進められている。そのため，本研究での検討は，同分野の研究手法にも大きな示唆を与えると考えられる。

2 研究方法と結果

2-1 ロンバード音声の分析

日本語母語話者 15 名（女性 7 名，男性 8 名，平均年齢 20 歳，18 歳から 28 歳まで）から 4 種類の雑音下で発声したロンバード音声及び静寂下で発声した音声を録音した。

背景雑音として，広帯域雑音（non-AM）と 2，4，8Hz の変調周波数（変調の深さは 0dB）を持つ正弦波振幅変調広帯域雑音（AM 2 Hz，AM 4 Hz，AM 8 Hz）を使用した。これらの刺激を USB オーディオインターフェース（RME，Fireface UC）で増幅し，ヘッドフォン（Sennheiser，HD599）を介して発話者に呈示した。音圧は，広帯域ノイズの音圧レベルが 80dB になるように調整し，正弦波振幅変調広帯域雑音は同じ RMS（二乗平均平方根）値を持つように設定をした。

静寂下（Quiet）条件を含む 5 つの背景条件は，同日に別々のブロックで実施し，その実施順序は参加者ごとにランダムで決定した。参加者は，各背景条件において，音素バランス文[4]を 50 文音読した。発話音声は，ヘッドセットマイク（SHURE，BETA54）と USB オーディオインターフェース（RME，Fireface UC）を用いて，サンプリングレート 44100Hz で録音した。マイクは話者の口元から約 5cm の位置にセットされた。

発話音声の分析は 1 文ごとに行い，まず初めに，録音した音声の波形を目視で確認することで発話区間を切り出した。なお，発話者が提示された文章を流暢に読み上げることができなかった場合や，被験者の声以外の音が発話区間に混入していた場合は，分析に使用しないこととした。切り出された音声信号の振幅は，全ての話者の全ての発話音声で RMS が同じになるように正規化された。音声の分析では，音声の主要な時間的特徴である時間微細構造と振幅包絡（図 2A）に注目して分析を行った。解析の流れは，図 2B に示す通りである。具体的には，まず，音声信号を，55～9657Hz の周波数帯域を 33 個に分割する矩形の帯域通過フィルタ群を用いて帯域分割した（矩形のフィルタは，ゼロ位相の 6 次バターワース無限インパルス応答フィルタを用いて作成し，使用されたバンドパスフィルタの境界周波数は，ヒトの聴覚フィルタ形状を推定するために行われた行動実験の結果[5]に基づいて決められたものに設定した）。

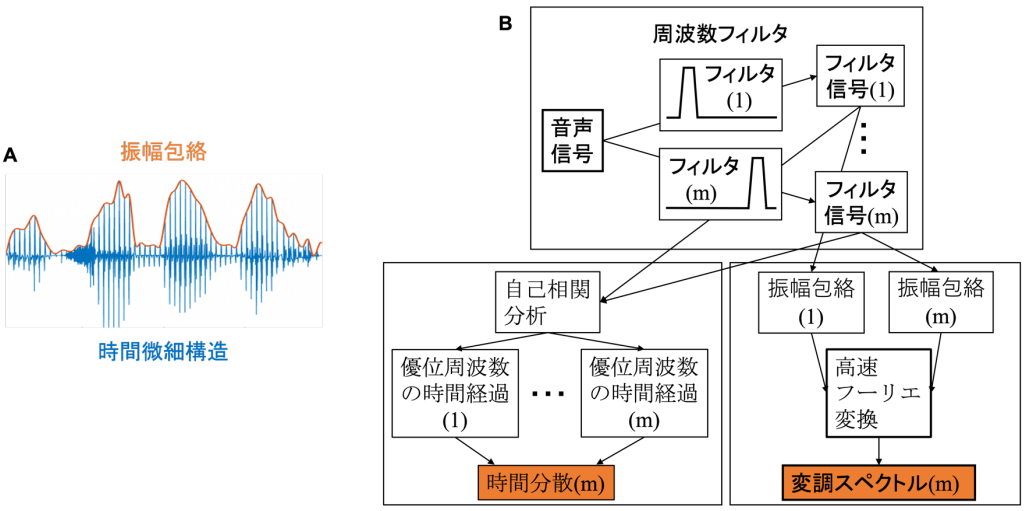

図 2. 音声分析の流れ

時間微細構造の評価は，各周波数帯域において，優位周波数が時間的にどの程度変化するか分析することで行った。帯域制限された信号における優位周波数の時間変化は，52ms の時間窓（42ms の時間的重なりを持つ）を用いて信号を切り出した後に，自己相関分析を用いて推定された。推定された優位周波数の時間分散

（標準偏差）を周波数帯域ごとに求め，優位周波数がどの程度ばらつきを持っているかを評価した。振幅包絡については，各周波数帯域において音声信号の変調スペクトルを計算することで評価した。変調スペクトルは，帯域制限信号の振幅包絡をヒルベルト変換で抽出し，その後に振幅包絡に高速フーリエ変換を適用することで求めた。参加者ごと，試行ごとに得られた優位周波数の時間分散，変調スペクトルは，参加者ごと，背景条件ごとに分けて文章間で平均化した上で，統計解析に掛けられた。

図 3A に各周波数フィルタにおける優位周波数の時間分散の結果を示す。雑音下条件（non-AM、AM 2Hz、AM 4Hz、AM 8Hz）では、静寂下条件（Quiet）と比較して，どの周波数フィルタにおいても優位周波数の時間変動が大きいことが観察された。統計解析として Cluster-based permutation ANOVA を行うことで，4-21 番目の周波数フィルタ（163-2489 Hz）において，背景条件の主効果があることが明らかになった（$p = 0.001$）。図 3B に，背景条件の主効果が見られた周波数フィルタから抽出した，支配周波数の時間変動を示す。多重比較の結果，雑音環境下では，静寂条件と比較して優位周波数の時間変動が有意に大きかった（Quiet vs. non-AM：$p < 0.001$; Quiet vs. AM 2 Hz：$p < 0.001$; Quiet vs. AM 4 Hz：$p < 0.001$; Quiet vs. AM 8 Hz：$p < 0.001$）。4 種類の背景条件間で有意な差は見られなかった（non-AM vs AM 2 Hz：$p=0.537$； non-AM vs AM 4 Hz：$p=0.287$； non-AM vs AM 8 Hz：$p=0.105$； AM 2 Hz vs AM 4 Hz：$p=1.000$； AM 2 Hz vs AM 8 Hz：$p=1.000$； AM 4 Hz vs AM 8 Hz：$p=1.000$）。

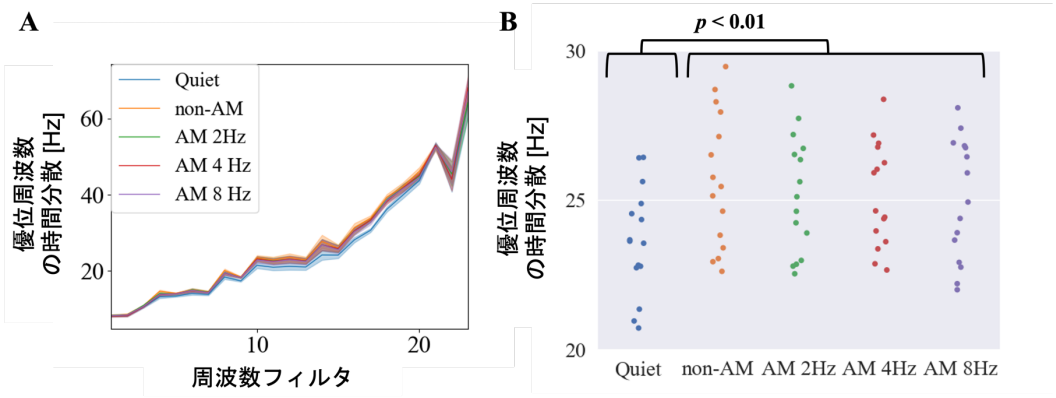

図 3. 時間微細構造の解析結果

図 4A に，各周波数フィルタにおける変調スペクトルを示す。すべての背景条件で，1-15 の周波数数フィルタ（55-1196Hz）において，10Hz 未満の大きな変調スペクトルパワーが観察された。Cluster-based permutation ANOVA では，背景条件の主効果を持つクラスターが 2 つ示された（右下パネル）。クラスターの 1 つ目のクラスターは，主に 1-5 番目の周波数フィルタ（55-257Hz）において，変調周波数全体を通して観察された（$p = 0.005$）。2 つ目のクラスターは，9-25 番目の周波数フィルタ（442-3951Hz）を中心に，変調周波数全体を通して観察された（$p=0.001$）。図 4B は，2 つのクラスターから抽出した，背景条件ごとの変調スペクトルを示したものである。多重比較の結果，2 つ目のクラスターでは，雑音下でのロンバード音声は静寂下での発話音声に比べて振幅変調が強かった（Quiet vs non-AM：$p = 0.003$; Quiet vs AM 2 Hz：$p < 0.001$; Quiet vs AM 4 Hz：$p < 0.001$; Quiet vs AM 8 Hz：$p = 0.002$ ）。一方で，4 つの背景雑音条件間では有意な差は認められなかった（non-AM vs. AM 2 Hz: $p = 1.000$, non-AM vs. AM 4 Hz: $p = 1.000$, non-AM vs. AM 8 Hz: $p = 1.000$; AM 2 Hz vs. AM 4 Hz: $p = 1.000$; AM 2Hz vs. AM 8 Hz: $p = 1.000$; AM 4 Hz vs. AM 8 Hz: $p = 1.000$）。

2-2 主観評価実験

ロンバード音声の分析に続いて，録音した音声やその音声から作成した分析合成音声の聞き取りやすさを評価する主観評価実験を行なった。実験には，24 名の日本語母語話者（女性 14 名，男性 10 名，平均年齢 35 歳，20 歳から 54 歳まで）が参加した。

音声刺激として，**2-1 ロンバード音声の分析**の課題で録音した 1 名の参加者のロンバード音声と静寂下での発話音声を使用した。この参加者では，すべての雑音下（non-AM，AM2Hz，AM4Hz，AM8Hz）において，静寂下（Quiet）と比べて，振幅包絡（変調スペクトル）と時間微細構造（優位周波数の時間変化）の違いが明確に観察された。すべての背景条件において，50 文から同じ 4 文の発話音声が選ばれた。選択した発話音声の

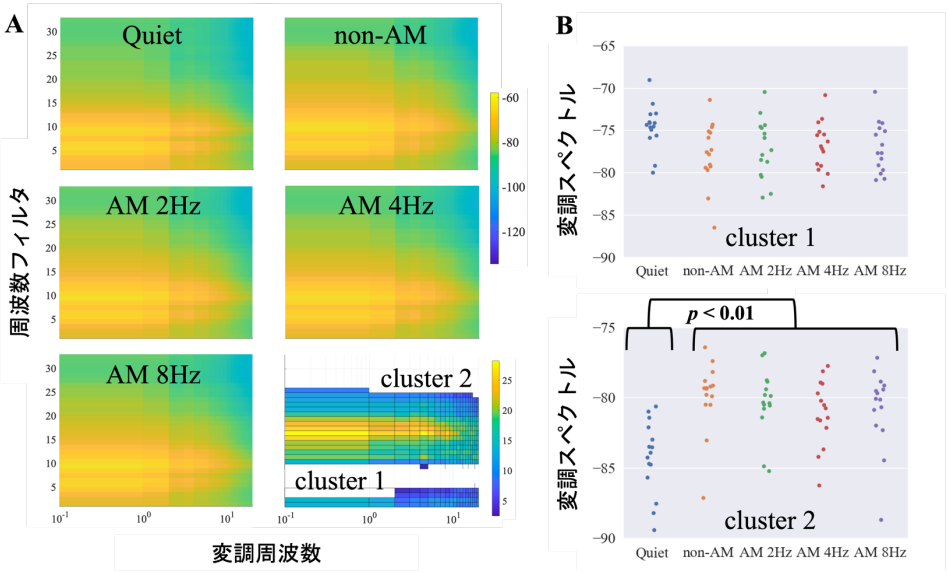

図 4.振幅包絡の解析結果

振幅は，使用するすべての音声で同じパワーになるように正規化した。

背景雑音には，ロンバード音声の録音時に使われた 4 種類の雑音を用いた。各背景騒音条件では，その背景雑音下で録音した文音声と静寂下で発話された文音声を音声刺激として用い，それぞれの文音声から，SN 比 +6，0，-6dB のスピーチインノイズ（SiN）刺激を作成した。SN 比は RMS 値に基づき，背景雑音刺激の RMS 値を増減させることで調整した。これらの刺激は，ヘッドフォン（ゼンハイザー，HD599）を通して実験参加者に呈示された。ブリュエル・ケアー社の騒音計（2260 型），1/2 インチコンデンサーマイク（4192 型），人工耳（4153 型）を用いて，SNR が 0dB のときに背景騒音の音圧レベルが 80dBA になるように音圧を調整した。

また，ロンバード音声と静寂下での発話音声から作成した純音駆動音声を使用した主観評価実験も行なった。純音駆動音声は、以下の手順で合成された。まず，ロンバード音声の分析にも用いた周波数フィルタ群を用いて帯域分割した。次に，ヒルベルト変換を用いて，各周波数フィルタにおける信号の振幅包絡を抽出した後，カットオフ周波数 10Hz のゼロ位相ローパスフィルタ（2 次バターワース無限インパルス応答フィルタ）を振幅包絡信号に適用した。最後に，各周波数フィルタから抽出した振幅包絡を，各周波数フィルタの中心周波数に相当する周波数を持つ純音に掛け合わせた。作成された純音駆動音声の振幅は，元の音声と同じ RMS 値を持つように調整された。オリジナルの音声と同様，各背景雑音条件で，その背景雑音下で録音した文音声と静寂下で発話された文音声から作成した純音駆動音声それぞれで，SNR が+6，0，-6dB の SiN 刺激を作成した。

実験の条件は全てで 8 つ（2 つの刺激条件（オリジナル音声と純音駆動音声）×4 つの背景雑音条件）で，それらは別々のブロックで実施をし，その実施順序は参加者間でカウンターバランスを取った。各背景雑音条件には，50 文から選ばれた 4 文のうち 1 文のみが割り当てられた。各条件への文の割り当ては，24 人の参加者間でカウンターバランスを取った。各実験条件における 6 つの SiN 刺激における音声の聴き取りやすさは，サーストンの一対比較法を用いて評価された。音声の聴き取りやすさは，6 つの SiN 刺激から作成された 15 個の刺激対で比較された。呈示順序の効果を考慮し，これら 15 組の比較は 2 回行われ，2 つの SiN 刺激の呈示順は 2 回の試行で逆転された。各試行では，15 個の刺激対からランダムに選んだ 1 個の刺激対を被験者に呈示し，どちらの文音声が背景雑音と分離してより聴き取りやすいかを回答するように教示した。

図 5 は，6 つの SiN 刺激における相対的な音声の聴き取りやすさを，実験条件ごとに示したものである。その結果，同じ SN 比では，ロンバード音声から作成された SiN 刺激は，静寂下での発話音声から作成されたものに比べて，オリジナル音声条件と純音駆動音声条件の両方で聞き取りやすいことが確認された。また，ロンバード音声と静寂下での発話音声の間の聴き取りやすさの差は，オリジナル音声の方が雑音駆動音声よりもやや大きい傾向にあることがわかった。主観評価データに，Bradley-Terry model を用いた検定を適用し

たところ，AM 4Hz の雑音下では，オリジナル音声条件において，静寂下での発話音声に比べてロンバード音声の方が有意に聴き取りやすい傾向があることが示されたが（z = −1.92，p = 0.05），純音駆動音声条件ではロンバード音声と静寂下での発話音声の間で聴き取りやすさに違いが見られなかった（z = −1.32，p = 0.16）。その他の雑音条件下では，オリジナル音声と TVS 音声のいずれの条件においても，ロンバード音声の方が静寂下での発話音声よりも聴き取りやすいことが分かった（non-AM & オリジナル音声：z = −6.08，p < 0.001；non-AM & 純音駆動音声：z = −4.63，p < 0.001；AM 2 Hz & オリジナル音声：z = −4.66，p < 0.001；AM 2 Hz & 純音駆動音声：z = −2.01，p = 0.04；AM 8 Hz & オリジナル音声：z = −6.98，p < 0.001；AM 8 Hz & 純音駆動音声：z = −2.61，p = 0.009）。

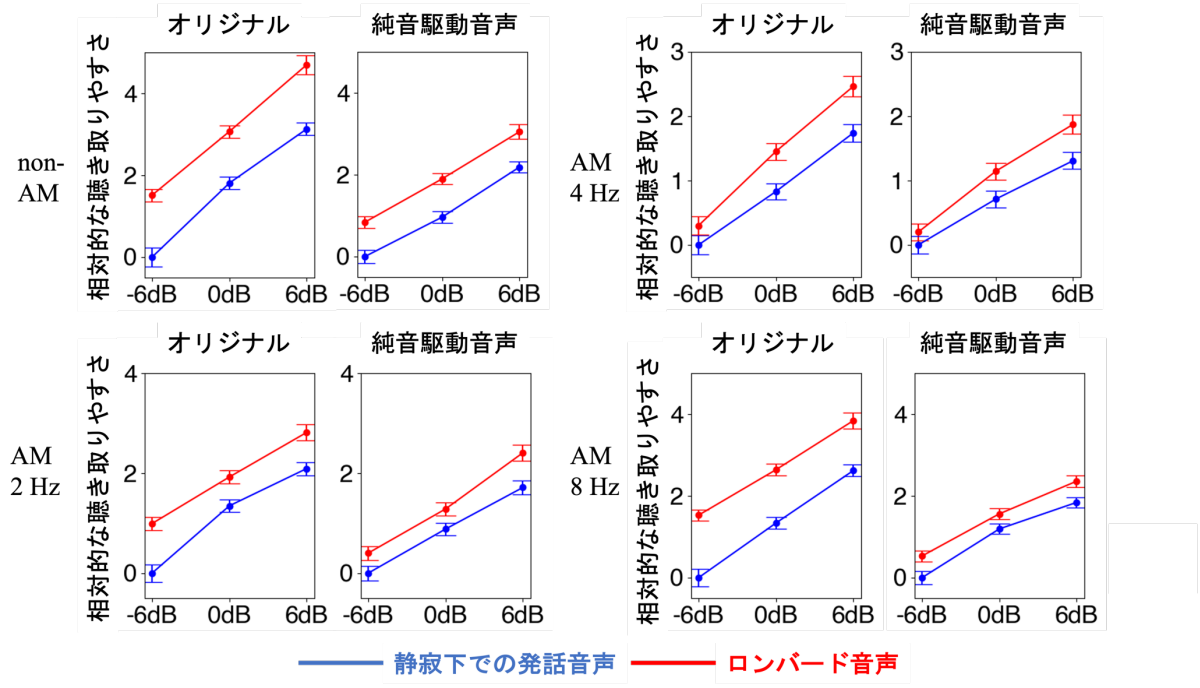

図 5. 主観評価実験の結果

2-3 脳機能計測実験

ロンバード音声やその音声から作成した分析合成音声の聞き取りやすさを評価する主観評価実験を行なっている際の脳活動を脳磁図で計測する実験を行なった。実験には，18 名の日本語母語話者（女性 10 名，男性 8 名，平均年齢 40 歳，21 歳から 56 歳まで）が参加した。音声刺激として，**2-1 ロンバード音声の分析**の課題で録音した 1 名の参加者のロンバード音声（non-AM 条件飲み）と静寂下での発話音声を使用した。non-AM 条件と Quiet 条件において，50 文の発話の中から同じ 1 文の発話音声を選び，それらが同じ RMS 値を持つように正規化した。また，対照刺激として，ロンバード音声と静寂下での発話音声から作成した純音駆動音声も使用した。

ロンバード音声の録音時に使われた non-AM を背景雑音，non-AM 条件で録音した文音声と Quiet 条件で発話された文音声を音声刺激として用いた。それぞれの文音声から，SN 比 0 dB のスピーチインノイズ（SiN）刺激を作成した。SN 比は RMS に基づき，背景雑音刺激の RMS を増減させることで調整した。これらの刺激は，ヘッドフォン（ゼンハイザー，HD599）を通して実験参加者に呈示された。ブリュエル・ケアー社の騒音計（2260 型），1/2 インチコンデンサーマイク（4192 型），人工耳（4153 型）を用いて，背景騒音の音圧レベルが 80dBA になるように音圧を調整した。オリジナルの音声と同様に，non-AM 条件及び Quiet 条件での発話音声のそれぞれから作成した純音駆動音声でも SN 比 0 dB の SiN 刺激を作成した。

作成した 4 つの SiN 刺激における音声の聴き取りやすさは，サーストンの一対比較法を用いて評価した。音声の聴き取りやすさは，4 つの SiN 刺激から作成された 6 個の刺激対で比較された。これら 6 組の比較は，それぞれ 28 回行われ，呈示順序の効果を考慮して，2 つの SiN 刺激の呈示順は半分の試行で逆転された。つ

まり，課題の中で 4 つの SiN 刺激はそれぞれ 84 回呈示された。各試行では，6 個の刺激対からランダムに選んだ 1 個の刺激対を被験者に呈示し，どちらの文音声が背景雑音と分離してより聴き取りやすいかを回答するように教示した。

主観評価課題時の脳活動は 306 チャンネル の全頭型脳磁計(Elekta, Neuromag)を用いて計測した(サンプリング周波数 1000 Hz)。 また，脳磁図の計測前に，3D 磁気センサーを用いて，頭部に設置した 4 つのコイルと頭部形状を記録するとともに，脳磁図ヘルメット内における 4 つのコイルの位置情報を取得した。これらのデータを用いて，各参加者で撮像した MRI 構造画像データの座標位置と MEG データの座標位置を合わせた。脳磁図データの解析には MNE 及び freesurfer を用いた。データの前処理では，Maxfilter，帯域通過フィルタ（1-100 Hz），ノッチフィルタ（60 Hz）をデータに適用した。続いて，各 SiN 刺激で刺激呈示の 1 秒前から 4 秒後までの波形を抽出して，その後に加算平均波形を求めた。信号源解析では，まず，各参加者の MRI 構造画像に基づいて脳内に信号源（全脳で 20484 点）を設定し，各信号源で特定の電流が生じた場合に脳磁図センサーでどのような磁場が得られるかを予測する順モデルを計算した。続いて，最小ノルム法を用いて加算平均データから信号源の活動分布を求める逆推定を行った後，刺激呈示前 100 ms の脳活動を用いて dSPM 法によるデータの正規化を行った。

音声知覚時の神経活動について，音声の振幅包絡に同期した周期的な神経活動 (4-8 Hz の θ 帯域神経振動) が生じることが,数多くの研究で示されている[6,7 など]。そこで，本研究では，各信号源で推定された脳活動の時系列データに高速フーリエ変換を適用し，音声の振幅包絡に同期した δ，θ 帯域神経振動が生じているかを確かめるとともに，4 つの SiN 刺激の間でそれらの神経振動にどのような違いが見られるのかを調べた。まずは，SiN 刺激の作成に用いた 4 つの音声刺激の分析を行い，それらの振幅包絡が持つ変調周波数成分を調べた。具体的な手順は，**2-1 ロンバード音声の分析**に述べた変調スペクトルの計算と同様であるが，今回の分析では全周波数フィルタで得られた変調スペクトルを平均化した結果を図 6 に示す。 分析結果から,non-AM 条件下で発話された音声は 2 Hz 以下，4 Hz，5-6Hz，7-8 Hz 付近に顕著なスペクトルパワーが見られることがわかった。2 Hz 以下，4 Hz 以下に関しては，Quiet 条件での発話音声でも特徴的なスペクトルパワーのピークが見られるが，特に後者に関しては non-AM 条件に比べてパワーが小さいことがわかった。5-6 Hz，7-8 Hz 付近のスペクトルパワーに関しては Quiet 条件ではピークが確かめられず，ロンバード音声のみが持つ変調周波数成分であった。 そこで，本研究では，δ，θ 帯域神経活動の解析を行う上で，上述の変調周波数と一致した成分に注目して解析を行なった。

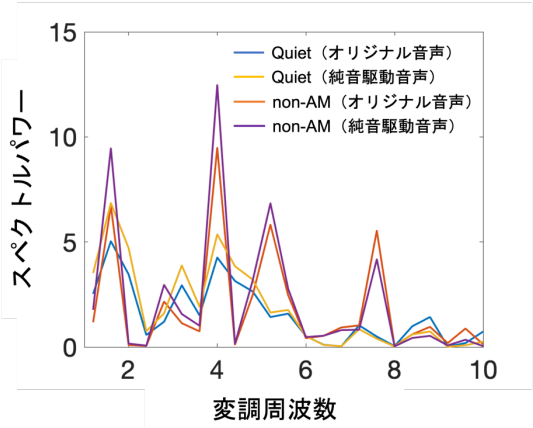

図 6. 脳機能計測実験で用いた刺激が持つ振幅包絡特徴

脳磁図データで 2 Hz 以下，4 Hz，5-6Hz，7-8 Hz の周波数を持つ神経活動の信号源を調べたところ，どの周波数帯においても，両半球の横側頭溝，上側頭回，上側頭溝付近で強いパワーが見られた（参考として，図 7A に non-AM 条件での発話音声から作成された SiN 刺激に対する 5.333...Hz の神経活動マップを示す)。そこで，これらの脳領域における神経活動のスペクトルを調べたところ，横側頭溝における 5-6 Hz 付近の周波数帯の神経活動において，non-AM 条件で発話されたロンバード音声によって誘発されるパワーが，Quiet 条件で発話された音声によって誘発されるパワーよりも明確に大きくなることが分かった（図 7B)。その他の領域や周波数帯では，non-AM 条件と Quiet 条件の差はそれほど明確ではなかった。オリジナル音声と純音駆動音声の違いについては，どの領域どの周波数帯でも確かめられなかった。

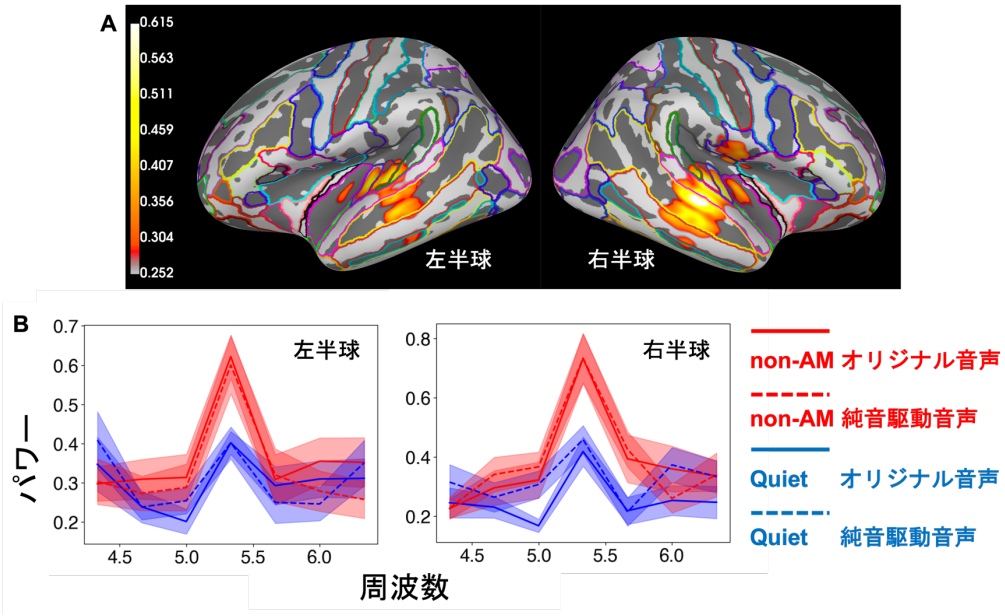

図 7. 脳磁図の解析結果

3 考察

ロンバード音声の分析では，静寂下で発話された音声と比較して，442-3951Hz の周波数帯域において強い変調スペクトルパワーを持つことが分かった。この結果は，ロンバード音声の変調スペクトルを分析した先行研究[8]とほぼ一致していた。また、ロンバード音声と静寂下での発話音声の変調スペクトルに有意差が認められた周波数帯域は，先行研究でロンバード音声と静寂下での発話音声のスペクトル形状に特徴的な差があると報告された周波数帯と大凡一致していた[9, 10 など]。したがって、ロンバード音声の変調スペクトルパワーは，ロンバード音声が持つ特徴的なスペクトル形状に大きく依存すると考えられる。次に，時間微細構造における周波数変調の度合いを調べるために，各周波数フィルタにおけるの優位周波数の時間変化を評価した。その結果、ロンバード音声では，163～2489Hz の周波数帯域において，静寂下での発話音声に比べて優位周波数の時間変化が大きくなっていることが分かった。各周波数フィルタの優位周波数は，音声の基本周波数とその高調波の影響を大きく受けていると考えられる。したがって，ロンバード音声における周波数変調の増加は，ピッチの時間変化を反映していると考えられる。もしそうだとすれば，今回の結果は，特に時間変調雑音やバブル雑音などの雑音下において，ピッチの時間変化が音声コミュニケーションに重要な役割を果たすことを示唆する知見と一致する[11, 12 など]。

次に，ロンバード音声が持つ振幅変調と周波数変調の特徴が，その高い聴き取りやさを生み出す上でどのような貢献をしているのかを調べるための主観評価実験を実施した。具体的には，ロンバード音声と静寂下での発話音声の聴き取りやすさの違いを，オリジナル音声と時間微細構造を削除した合成音声である純音駆動音声の間で比較することで，ロンバード音声の聴き取りやすさにおける振幅包絡と時間微細構造の貢献度を分離して評価した。その結果，non-AM，AM 2 Hz，AM 8 Hz の雑音下では，オリジナル音声条件でも純音駆動音声条件でも，静寂下での発話音声と比較して，ロンバード音声の方がより背景雑音から分離して知覚されることがわかった。この結果は，音声が時間微細構造を持つか否かにかかわらず，ロンバード音声が静寂下での発話音声よりも聴き取りやすいことを示しており，ロンバード音声の持つ振幅包絡の特徴がその聴き取りやすさに大きく寄与していることが示唆される。ただし，AM 4 Hz の雑音条件下では，オリジナル音声でも純音駆動音声でも，ロンバード音声と静寂下での発話音声の間に有意な聴き取りやすさの違いは見られなかった。音声の振幅包絡が持つ変調周波数（4 Hz）のピークに近い振幅変調を持つ背景雑音では，ロンバード音声が持つ振幅包絡の特徴がその聴き取りやすさに反映されにくいということが考えられる。オリジナル音声条件と純音駆動音声条件の結果の違いは明確ではなかったが，オリジナル音声条件では，純音駆動音

声条件と比較して，ロンバード音声と静寂下での発話音声の聴き取りやすさの違いがやや大きい傾向があった。このことから，ロンバード音声における時間微細構造の特徴が，ロンバード音声の聴き取りやすさを補足的に支えていることが示唆される。

脳機能計測実験では，ロンバード音声が持つ振幅包絡や時間微細構造の特徴がδ，θ帯域神経活動にどのように反映されるかについて検討を行なった。主観評価実験と同様，この実験でも純音駆動音声を用いることで，振幅包絡と時間微細構造のどちらがδ，θ帯域神経活動と関係をしているのかについて検討を行なった。その結果，non-AM条件で録音されたロンバード音声は，純音駆動処理されるか否かに関わらず，静寂下での発話音声に比べて，聴覚関連領域において強いパワーを持つθ帯域神経活動を惹起することが分かった。この結果は，ロンバード音声の持つ強い振幅包絡成分は，それと同じリズムで生じる脳内のθ帯域神経活動に反映されることを示している。一方で，ロンバード音声における時間微細構造の特徴はδ，θ帯域神経活動には反映されないことが分かった。主観評価実験でも，時間微細構造の特徴が音声の聴き取りやすさに及ぼす影響は明確ではなかったため，神経活動にもうまく反映されていなかったのかもしれない。あるいは，時間微細構造に関連する神経活動を捉えるためには，δ，θといった遅い周波数の神経活動ではなく，30 Hz以上の周波数を持つ神経活動であるγ帯域神経活動に注目した解析が必要となるかもしれない。研究代表者の最新の研究で，音声聴取時に時間微細構造の主要な構成要素である基本周波数に同期してγ帯域神経活動が生じることや，γ帯域神経活動の緩やかなパワーの変化が振幅包絡に同期して起こることなどを報告している。γ帯域神経活動は振幅包絡と時間微細構造の処理を同時に行なっていると考えられるため，今後はγ帯域神経活動の分析を行うことでδ，θ帯域神経活動とは異なる結果が得られることが予想される。

4 課題と今後の展望

本研究調査では，「ことばの鎖」における音響学的，生理学的，言語学的レベルに注目して，ロンバード音声が持つ明瞭度について検討をした。具体的には，ロンバード音声が持つ時間的特徴に注目して，その特徴が神経活動にどのように活かされるのか，音声の聴き取りにどのように貢献するのかを評価する枠組みを作ることができた。しかしながら，本研究調査にはいくつかの課題が存在する。本研究調査では，主に音声信号の持つ時間特徴に注目した検討を行ったが，今後はロンバード音声の更なる分析を進め，ロンバード音声において他にどのような音響的特徴が強調されているのか，さらには，その特徴がどのように生理学的に脳内表現され，明瞭度につながるかを検討していく必要があると思われる。音声刺激についても，主観評価課題と脳機能計測実験ともに特定の文音声を用いた課題設定であったため，日常の音声コミュニケーション場面を想定した実験系ではなかった。今後は連続的な音声を使って知覚課題をデザインすることで，その結果をより一般化して解釈することができると思われる。連続音声を知覚している際の神経活動評価については，近年の脳波・脳磁図解析技術の発展によって，様々な音響特徴と神経活動の相関関係が調べられるようになってきているため[14,15 など]，現在はその実験系の構築に取り組んでいるところである。背景雑音の種類についても，本研究調査で用いる白色雑音やバブルノイズだけでなく，より日常的な生活環境下での背景雑音なども用いることも考えている。将来的に，様々な音環境を想定したデータを蓄積し，本研究調査で作った研究の枠組みで音声明瞭度に関する研究を進めていくことで，最終的にはその本質的理解や汎用性の高い音声明瞭度指標の開発につながると考えている。

本研究調査やその後の研究で得られた音声データについては，最終的に脳機能計測や主観評価実験での結果をラベル付けしてデータベース化し，他の研究機関でも利用できる形で公開することを考えている。また，最終的に汎用性の高い音声明瞭度指標の開発が進めば，室内音場における音声の聴き取りやすさを予測するために開発されたSTI[1]に比べて，より幅広い音環境のデザインに役立つと考えられる。さらに，本研究で得られた知見及びデータは，電車のホームや災害時の拡声音声などの場面を想定した音声強調技術への応用についても，人間科学的エビデンスの強い形での開発が可能になると考えられる。

【参考文献】

[1] Houtgast, T., & Steeneken, H. J. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, *77*(3), 1069-1077.

[2] Danes, P. B. & Pinson, E. N., The Speech Chain: The Physics and Biology of Spoken Language (Worth Publishers, New York, 1993).
[3] Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *Journal of the Association for Research in Otolaryngology*, *19*, 181-191.
[4] Kurematsu, A., Takeda, K., Sagisaka, Y., Katagiri, S., Kuwabara, H., & Shikano, K. (1990). ATR Japanese speech database as a tool of speech recognition and synthesis. *Speech communication*, *9*(4), 357-363.
[5] Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1-2), 103-138.
[6] Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378-1387.
[7] Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, *39*(29), 5750-5759.
[8] Bosker, H. R., & Cooke, M. (2018). Talkers produce more pronounced amplitude modulations when speaking in noise. *The Journal of the Acoustical Society of America*, *143*(2), EL121-EL126.
[9] Cooke, M., Mayo, C., & Villegas, J. (2014). The contribution of durational and spectral changes to the Lombard speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *135*(2), 874-883.
[10] Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. *Computer Speech & Language*, *28*(2), 629-647.
[11] Shen, J., & Souza, P. E. (2017). Do older listeners with hearing loss benefit from dynamic pitch for speech recognition in noise? *American Journal of Audiology*, *26*(3S), 462-466.
[12] Wu, M. (2019). Effect of F0 contour on perception of Mandarin Chinese speech against masking. *PloS One*, *14*(1), e0209976.
[13] Tamura, S., & Hirano, Y. (2023). Cortical representation of speech temporal information through high gamma-band activity and its temporal modulation. *Cerebral Cortex*, bhad158.
[14] Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, *18*, 25-31.
[15] Gillis, M., Van Canneyt, J., Francart, T., & Vanthornhout, J. (2022). Neural tracking as a diagnostic tool to assess the auditory pathway. *Hearing Research*, 108607.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Cortical representation of speech temporal information through high gamma-band activity and its temporal modulation.	Cerebral Cortex	2023 年 5 月