

ガンマチャープ聴覚フィルタに基づく模擬難聴システム WHIS の新実装

入野俊夫[†]

[†] 和歌山大学 システム工学部 〒640-8510 和歌山市栄谷 930

E-mail: tirino@wakayama-u.ac.jp

あらまし 模擬難聴システム WHIS の新実装を行ったので報告する。まず、ガンマチャープ聴覚フィルタバンク (GCFB) に、絶対閾値と難聴特性を導入した。それを分析系の基盤として、難聴処理を新規に開発した。この結果見通しの良い実装となった。さらに、信号処理的にどの程度忠実に難聴が模擬できているかの評価を行った。模擬難聴音を健聴条件の GCFB で分析した場合と、難聴条件の GCFB で分析した場合との比較を、蝸牛の入出力関数や音声に対する出力のスペクトル距離に関して行った。この結果、旧実装の WHIS と同等の出力が得られること、Cambridge 版模擬難聴システム (CamHLS) に比べて音声の歪みが小さいことがわかった。

キーワード 聴覚モデル、オーディオグラム、内外有毛細胞、機能低下、入出力関数

A new implementation of hearing impairment simulator WHIS based on the gammachirp auditory filterbank

Toshio IRINO[†]

[†] Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama, 640-8510 Japan

E-mail: tirino@wakayama-u.ac.jp

Abstract A new implementation of Wakayama-university (Wadai) Hearing Impairment Simulator, WHIS, is reported. The output level at absolute threshold and hearing levels of persons with hearing loss were introduced into the gammachirp auditory filterbank (GCFB) and a new algorithm for hearing loss simulation was developed. Comparisons were performed between the output of GCFB with hearing loss (HL) and the output of GCFB with normal hearing (NH) after WHIS processing in terms of the input-output function of sinusoids and the distance between speech spectrograms. As a result, it was found that the outputs of the new WHIS were almost compatible with those of the conventional WHIS and that the spectral distortion of WHIS processed sound was smaller than that of Cambridge version of hearing loss simulator (CamHLS). (See Annex: Extended abstract)

Key words auditory model, audiogram, inner/outer hair cell, dysfunction, input/output function

1. はじめに

日本は世界に先駆けて超高齢社会に突入した。高齢化率増加に伴い老人性難聴者の割合も増加が見込まれる。この状況における音声コミュニケーション支援は喫緊の課題であろう。そこで我々はこの問題解決へのアプローチとして、模擬難聴システム WHIS を提案し、音声知覚特性の解明・発声訓練・言語聴覚士養成課程教育プログラム開発に取り組んできた [1]。

従来から模擬難聴は様々な開発されてきた (紙幅の都合で経緯は略す)。そこでは主に音声了解度/明瞭度実験の道具立てとして利用されてきた。しかし、他の聴覚心理実験では使われず、さらに信号処理的にどの程度忠実に難聴が模擬できているかの評価は十分されてこなかった。そのため、どの程度の難聴や信号処理であれば、良く模擬できるのかの限界もわからなかった。

本報告では、模擬難聴システム WHIS の新実装を、ガンマチャープ聴覚フィルタバンク [2] の改良を通して行った。さらに、旧実装の WHIS や Cambridge 版模擬難聴 (以下 CamHLS) [3,4] と蝸牛の入出力関数や音声分析出力のスペクトル距離を用いて比較評価したので報告する。

2. GCFB の改良

模擬難聴 WHIS の新実装のために、動的圧縮型ガンマチャープ聴覚フィルタバンク (dcGC-FB) [2] の改良を行った。具体的には、1) 高速化のためのフレーム処理の導入、2) 絶対閾値の導入、3) 難聴特性の導入を行った。ここでは、圧縮型ガンマチャープ聴覚フィルタの定義から始めて、順番に説明する。

2.1 圧縮型ガンマチャープフィルタ

聴覚フィルタの特性を反映させた、圧縮型ガンマチャープ

フィルタの周波数特性は以下で定式化されている [5]。

$$|G_{CC}(f)| = \{a_T \cdot |G_T(f)| \cdot \exp(c_1 \theta_1)\} \cdot \exp(c_2 \theta_2) \quad (1)$$

$$= |G_{CP}(f)| \cdot \exp(c_2 \theta_2) \quad (2)$$

$$= |G_{CP}(f)| \cdot H_{HPAF}(f) \quad (3)$$

$$\theta_1 = \arctan\left(\frac{f - f_{r1}}{b_1 \text{ERBN}(f_{r1})}\right) \quad (4)$$

$$\theta_2 = \arctan\left(\frac{f - f_{r2}}{b_2 \text{ERBN}(f_{r2})}\right) \quad (5)$$

ここで、 b_1, c_1, b_2, c_2 はフィルタ係数。ERBN(f) は標準的な健聴者の等価矩形帯域幅。 f_{r1}, f_{r2} はそれぞれのフィルタの漸近周波数を表す。圧縮型ガンマチャープフィルタの周波数特性 $|G_{CC}(f)|$ は、受動的ガンマチャープ (pGC) の周波数特性 $|G_{CP}(f)|$ と高域通過非対称フィルタ (HP-AF) の周波数特性 $H_{HPAF}(f) (= \exp(c_2 \theta_2))$ の積で表せる。pGC のピーク周波数を f_{p1} とし、pGC フィルタ出力から得られる推定音圧 P_{gcp} (dB) を用いて、 f_{r2} の値を以下の式で表す。

$$f_{r2} = f_{rat}(P_{gcp}) \cdot f_{p1} \quad (6)$$

$$f_{rat}(P_{gcp}) = f_{rat}^{(0)} + f_{rat}^{(1)} \cdot P_{gcp} \quad (7)$$

ここで、 $f_{rat}^{(0)}, f_{rat}^{(1)}$ は係数で、他の係数 b_1, c_1, b_2, c_2 といっしょに、ノッチ雑音マスキング法によって被験者データより推定された値になっている。

2.2 動的圧縮型ガンマチャープフィルタバンク

上記の圧縮型ガンマチャープフィルタの定式化を用いて、動的圧縮型ガンマチャープフィルタバンク (dcGC-FB) は構成されている [2]。図 1 に 1ch 分のブロック図を示す (以下、GCFB_{v211})。信号経路 (Signal path) を見ると、pGC と HP-AF が縦続接続されている。前者の pGC は、帯域幅の広い受動的な基底膜振動を模擬している。難聴特性を導入する際でも、pGC は健聴者と難聴者で共通であると仮定する。難聴者の聴覚フィルタは健聴者よりもかなり広く推定される場合が報告されているが [6]、生理学的な von Békésy の高音圧での観測結果 [7] とは整合性のある仮定と考える。後者の HP-AF は外有毛細胞 (Outer Hair Cell, OHC) による増幅処理と周波数分解能向上 (=帯域幅減少) を反映させている。中段に配置されたレベル推定回路によって、信号経路の 2 つのフィルタの相対周波数比を制御する。なお、この推定回路は、2 音抑圧の影響も考慮に入れたため、ノッチ雑音マスキング法において係数推定に用いたもの (2.1 項) よりやや複雑になっている。レベル推定回路の出力音圧 (P_c) によりサンプル点ごとに時々刻々フィルタ係数に変化する。したがって、推定音圧にしたがって帯域幅と利得が変化することとなる。この処理のため、原音声長の数十倍の計算時間がかかった。

このままでは、模擬難聴システムの GUI におけるインタラクティブな処理が時間がかかりすぎて使えない。そこで、高速化のためのフレームベース処理を導入することとした。励起パターン (Excitation Pattern) だけを見ればよく、時間微細構造 (TFS) を無視できる場合には近似度がある程度高ければ十分活用できる。模擬難聴はこの場合の一例で、他のアプリケーション (たとえば音声了解度客観指標 GEDI) においても有効に活用できる。実装した 0.5 ms シフトでフレーム出力する設定 (サンプリング周波数は 2000 Hz) とした場合、ほぼ原音声と同じ時間 (real time) で処理が完了することを確認した。詳細な処理過程に関しては 2.3.4 節で述べる。

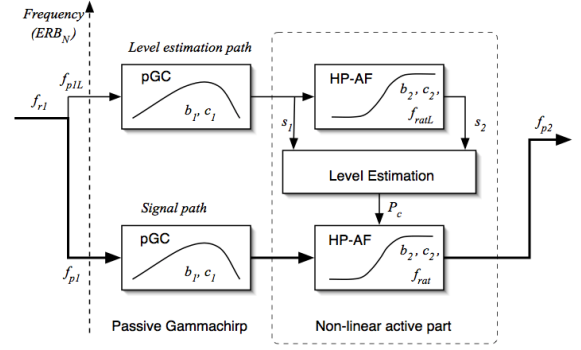


図 1: 従来法 (GCFB_{v211}) : サンプル毎にフィルタ計算する dcGC-FB の 1ch 分の処理のブロックダイアグラム [2]

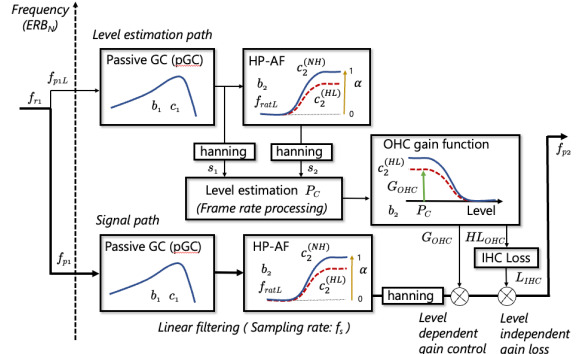


図 2: 提案法 (GCFB_{v231}) : フレーム処理・絶対閾値・難聴者特性を導入した dcGC-FB の 1ch 分の処理

2.3 OHC の機能低下の導入と入出力特性

ここで、聴力損失をこの dcGC-FB に導入することを考える。

2.3.1 聴力損失特性の考え方

高齢難聴者の場合、オーディオグラムにおいて高域で聴力損失 (Hearing Loss, HL) が大きくなる。この聴力レベル低下は OHC による増幅作用の機能低下と、内毛細胞 (Inner Hair Cell, IHC) の振動から神経発火への変換の機能低下の両方の影響による。オーディオグラムでは、それぞれの劣化の度合いを判明できず、全体的な閾値上昇だけを示すことになる。Moore と Glassberg [8] は、全体の聴力損失 (HL_{total}) が OHC による損失 (HL_{OHC}) と IHC による損失 (HL_{IHC}) の dB 上での加算として表せると考え、ラウドネスモデルを作成した。すなわち、

$$HL_{Total} = HL_{OHC} + HL_{IHC} \quad (8)$$

である。dcGC-FB でも同じ式で難聴特性を導入する。

2.3.2 OHC の機能低下の導入

難聴者における OHC の機能低下を反映させるために、HP-AF の利得のダイナミックレンジを決める式 2 の c_2 に OHC の健全度を導入することを考える。健全度を $\alpha \{0 \leq \alpha \leq 1\}$ と置いて、難聴者の $c_2^{(HL)}$ と健聴者の $c_2^{(NH)}$ との関係を以下で定式化する。

$$c_2^{(HL)} = \alpha \cdot c_2^{(NH)} \quad (9)$$

$\alpha = 1$ の場合、OHC の機能低下が無い健聴者の特性で、 $\alpha = 0$ の場合は OHC が完全に機能不全になっている場合に相当する。

難聴特性を導入して、式 3 の $|G_{CC}(f)|$ を書き直す。

$$|G_{CC}^{(HL)}(f)| = |G_{CP}(f)| \cdot H_{HPAF}^{(HL)}(f) \quad (10)$$

対応する時間応答は以下になる。

$$g_{cc}^{(HL)}(t) = g_{cp}(t) * h_{HPAF}^{(HL)}(t) \quad (11)$$

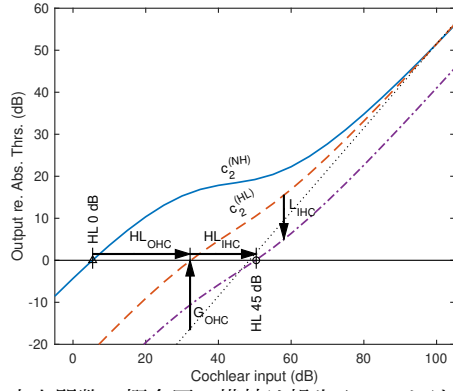


図 3: 入出力関数の概念図。横軸は蝸牛入口における音圧レベル (dB)。縦軸は、絶対閾値を 0 dB となるように正規化した蝸牛の出力レベル。HL 0 dB は、健聴者の平均聴力レベル (HL) に対応するレベル¹。ここでは例題として聴力レベル HL 45 dB の難聴者の入出力関数を示す。青実線：健聴者の入出力関数で $c_2^{(NH)}$ を使用。黒点線：入力と出力が 1:1 の関係になる場合。赤破線：式 9 において健全度 $\alpha = 0.5$ とした場合の $c_2^{(HL)}$ を用いた場合の入出力関数で HL_{OHC} 分だけ閾値が上昇する。紫一点鎖線： HL_{IHC} の分が加わり、ちょうど HL 45 dB のところを入出力関数が通過する。

このままでは線形フィルタで、入力音圧に応じた利得の変動が表現できない。フレーム処理を導入するため、時間的に変動する帯域幅の近似はあきらめ、利得だけを近似することとする。時間変動する利得関数として $G^{(HL)}$ を、この式に追加する。フレーム時刻 τ における推定音圧を $P_c(\tau)$ として、以下のように変形する。

$$g_{cc}^{(HL)}(t) = \{g_{cp}(t) * h_{HPAF}^{(HL)}(t)\} \cdot G^{(HL)}(P_c(\tau)) \quad (12)$$

ここで、 $G_{HPAF}^{(HL)}(P_c(\tau))$ の利得は式 5 の θ_2 において、pGC のピーク周波数 f_{p1} と HPAF の中心周波数 f_{r2} の差分から計算される量で近似できるので、以下のように表すことができる。

$$\begin{aligned} G^{(HL)}(P_c(\tau)) &= H_{HPAF}^{(HL)}(f_{p1}, P_c(\tau)) \quad (13) \\ &= \exp \left\{ c_2^{(HL)} \cdot \arctan \left(\frac{f_{p1} - f_{r2}}{b_2 \text{ERBN}(f_{r2})} \right) \right\} \\ f_{r2} &= f_{rat}(P_c(\tau)) \cdot f_{p1} \quad (14) \end{aligned}$$

ここで、 $f_{rat}(P_c(\tau))$ は、式 7 と同じ形式で pGC と HPAF の間の周波数位置の関係式で、フレーム毎に推定された $P_c(\tau)$ から算出される。なお、実装においてフレームレートやサンプリング周波数の各処理間の相違は再サンプリングで対応する。

OHC の利得は、低音圧で大きく高音圧で小さくなることが知られている。これは次節で述べる蝸牛の入出力関数の圧縮特性として見る事ができる。HP-AF はその特性を良く近似できる。利得 0 dB となる音圧 (例えば 100 dB) を P_{Norm} とすると、OHC の利得 G_{OHC} は以下の式で近似できる。健聴者でも難聴者でも P_{Norm} は共通のため、同じ式を使える。

$$G_{OHC}(P_c) = \frac{H_{HPAF}(f_{p1}, P_c(\tau))}{H_{HPAF}(f_{p1}, P_{Norm})} \quad (15)$$

2.3.3 蝸牛の入出力特性

図 3 に蝸牛の入出力関数の模式図を示す。ここで、OHC と IHC の機能低下による影響を説明する。横軸は蝸牛入口におけ

る音圧レベル (dB) で、縦軸は絶対閾値を 0 dB の基準とした場合の蝸牛の出力レベルを示す。健聴者で $c_2^{(NH)}$ を使用した場合、青実線に示されるような入出力関数となる。蝸牛入力レベルに換算した絶対閾値の HL 0 dB を通り、傾きが緩やかになって圧縮特性を示したあと、音圧が高くなるにつれて傾きが増加して 1:1 の黒点線に収束する。

これに対して式 9 において、健全度 α を 0.5 とした $c_2^{(HL)}$ を用い、OHC の機能低下だけを導入した入出力特性を赤破線で示す。OHC の利得 G_{OHC} (式 15) は、健聴者における OHC 利得より小さく、ダイナミックレンジも狭くなる。この結果、入出力関数の傾きは全体的に増加し、圧縮特性が小さくなる。これを入力側のレベルで換算すると、健聴者の絶対閾値よりも HL_{OHC} の分だけ、閾値が上昇することとなる。さらに、この絶対閾値 HL 45 dB の難聴者の場合、IHC の機能低下によって出力が低下する。この利得の減少分を L_{IHC} とすると、赤破線の入出力関数は下方に並行移動して紫一点鎖線の特性となる。この IHC による出力レベル減少度を、入力換算で見たものが HL_{IHC} である。

オーディオメータによる測定により絶対閾値がわかって、 HL_{OHC} と HL_{IHC} の大きさの割合はわからない。モデルにおいてはこの割合にかかわらず、入出力関数が出力レベル 0 dB を横切る場所 (すなわち絶対閾値) が移動しないようにする必要がある。このモデルでは、事前に利用者が設定した健全度 α を元に HL_{OHC} を計算し、そこから絶対閾値までの差分として HL_{IHC} を算出するようにした。

2.3.4 GCFB における難聴特性の導入

前節の入出力関数を踏まえたうえで、dcGC-FB における実装を考える。図 2 に、難聴者特性を考慮し、フレーム処理も導入した 1ch 分のブロック図を示す (以下 GCFB_{v231})。まず、上段のレベル推定経路を見る。pGC と HP-AF が縦続接続されている。ここで、式で表される HP-AF の $f_{(rat)}$ は、レベル推定用の定数 $f_{(ratL)}$ に固定されているため線形フィルタとなる。フィルタ出力は、入力信号と同じサンプリング周波数のままで時間応答波形となる。難聴者の広がったフィルタ形状を反映させるため、OHC の健全度 α を用いて式 9 から決まる難聴者の $c_2^{(HL)}$ を HP-AF において用いる。ここまでは線形フィルタにすぎないので、OHC の増幅作用による音圧依存利得 G_{OHC} (式 15) を掛ける必要がある。

そのため、pGC と HP-AF の時間応答波形出力からそれぞれ hanning 窓でフレームを切り出し、その RMS レベルから利得を制御するための音圧レベル P_c を推定する。推定式は [2] で示されたものと同じである。この P_c を用いて式から式 15 の利得 G_{OHC} を求める。

下段の信号経路は、上段のレベル推定経路のフィルタと中心周波数以外全く同じもので、線形フィルタとなっている。これを hanning 窓でフレーム処理を行い、式 15 の $G_{OHC}(P_c)$ を掛ける。その後、式 8 から算出された IHC の機能低下分の損失 L_{IHC} (図 3 参照) が掛けられ、出力される。

2.4 計算結果: 入出力関数

上記のモデルを用いて、オーディオグラムの周波数 (125 Hz ~ 8000 Hz) における入出力特性を、健聴者の場合と高齢難聴者 (80 歳男性の平均値 [9]) の場合について計算した。オーディオグラムだけでは OHC の健全度 α の情報はまったく無いため、 α の値を 0, 0.5, 1 の 3 段階に設定した。なお、 $\alpha = 0$ と初期設定した場合、オーディオグラムにおける絶対閾値の制約から、 α の値が 0 より大きく自動設定直されている場合もある。入出力関数の計算においては、入力音 200 ms の正弦波を用いて、蝸牛入口の

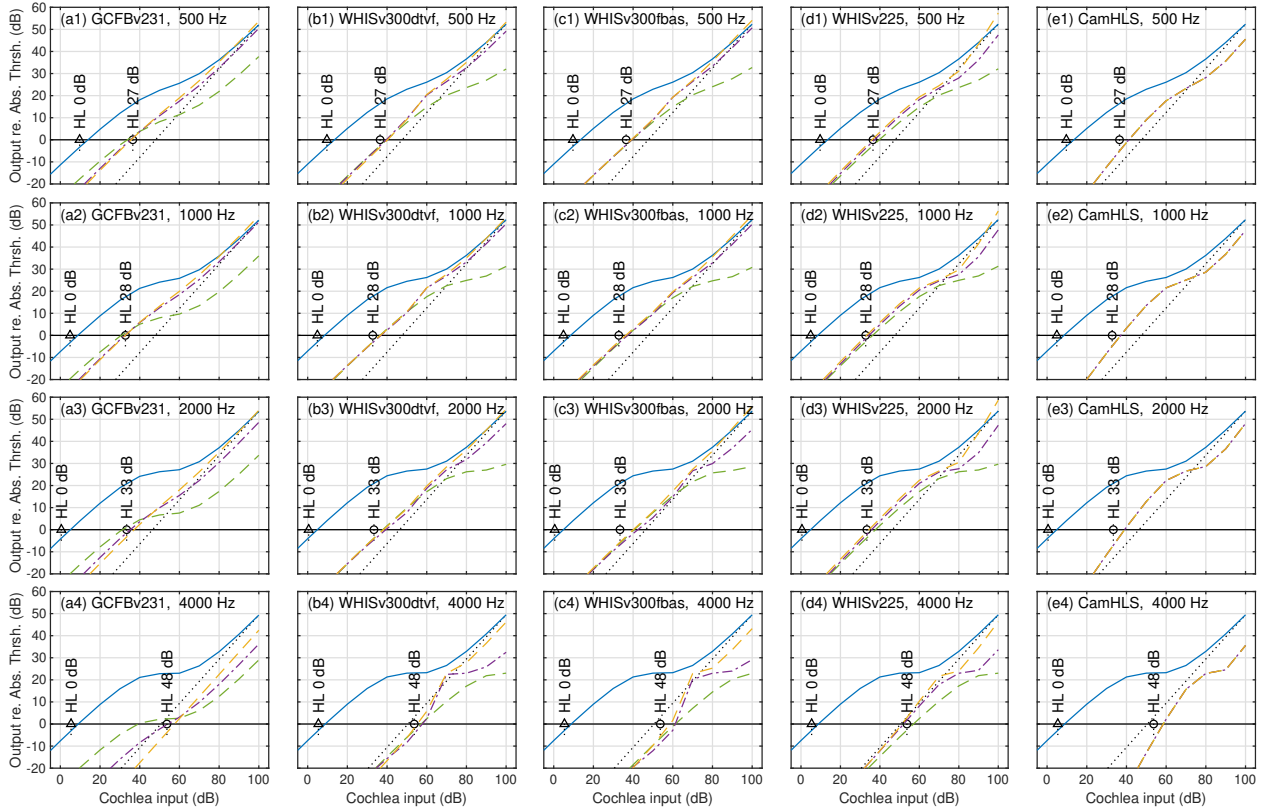


図 4: 信号音周波数 500, 1000, 2000, 4000 Hz における入出力関数。横軸：蝸牛入口におけるレベル。縦軸：絶対閾値を基準 (0 dB) とした場合の出力レベル。HL 0 dB は健聴者の聴力レベル。各図の右側にある HL は、高齢難聴の 80 歳男性の平均聴力レベル [9]。青実線：健聴者特性。緑破線：高齢難聴者、健全度 $\alpha = 1$ 、紫一点鎖線：同、健全度 $\alpha = 0.5$ 、橙一点鎖線：同、健全度 $\alpha = 0$ 。(a1)~(a4) GCFB_{v231} による分析。その他は模擬難聴処理音を GCFB_{v231} の NH 条件で分析した場合の出力。

音圧レベルで -30 dB から 110 dB まで 10 dB 毎に設定した。出力は、フィルタバンク出力の励起パターン (excitation pattern) の最大値とした。図 4 の (a1) ~ (a4) に 500 ~ 4000 Hz の入出力関数を示す。健聴者 NH の入出力関数は、各周波数で絶対閾値 HL 0 dB の値に対して数 dB 以内のところで出力レベル 0 dB の水平線を横切っている。オージオメータでは 5 dB ステップで閾値を測定することを考えると、それと同等の誤差範囲内にはいっていることがわかる。さらに、難聴者の入出力特性は、健全度 α の値が 0, 0.5, 1 と異なっても、概ねの絶対閾値の近くを通っていることがわかる。

このことから、健聴者、難聴者を問わず絶対閾値を適切に導入でき、入出力関数を概ね良く近似できることがわかった。

3. WHIS の新実装

前節の GCFB_{v231} を基盤として、模擬難聴システム WHIS の新実装 (以下 WHIS_{v300}) を行う。従来手法 (以下 WHIS_{v225}) においては、圧縮特性のキャンセルを行ったのちに IHC の損失分を勘案していた。ここでは、発想を転換して、GCFB_{v231} の入出力関数に着目して構成しなおす。

3.1 分析系

図 5 に提案する WHIS_{v300} の 1 ch 分の分析系のブロック図を示す。左側半分のフィルタ部分と音圧推定部分は、図 2 の GCFB_{v231} と全く同じである。実際に GCFB_{v231} の処理をそのまま利用している。違いは中段右側にあった OHC の利得関数

が入出力関数 (IO function) に置き換わったところである。ここでは、OHC の機能低下による NH の場合からの利得の減少を、入力レベルに換算した L_{OHC} を求めている。これは、 $c_2^{(NH)}$ と $c_2^{(HL)}$ の場合の入出力関数の差分となる。この L_{OHC} は推定音圧 P_c に依存する。入力レベルを P_{in} (dB) として、出力レベルを P_{out} (dB) とすると、式 15 の G_{OHC} を用いて、入出力関数 F_{IO} は以下の式で表すことができる。

$$P_{out} = F_{IO}(P_{in}) = G_{OHC}(P_{in}) + P_{in}. \quad (16)$$

模擬難聴のレベル調整は入力換算で計算しないといけないため、この逆関数 $P_{in} = F_{IO}^{-1}(P_{out})$ を準備しておく。

ここで、NH と HL の入出力関数が、それぞれ $F_{IO}^{(NH)}$ と $F_{IO}^{(HL)}$ とすると、ある出力レベル P_{out} における入力換算の損失 L_{OHC} は以下の式で与えられる。

$$L_{OHC} = F_{IO}^{(HL)-1}(P_{out}) - F_{IO}^{(NH)-1}(P_{out}). \quad (17)$$

この時の P_{out} は、HL の時のレベル推定回路出力の音圧 $P_c(\tau)$ で求められる。

$$P_{out} = F_{IO}^{(HL)}(P_c(\tau)) \quad (18)$$

このようにして、フレームごとのレベル $P_c(\tau)$ に依存する OHC 機能低下分の損失 $L_{OHC}(P_c(\tau))$ が求まる。なお、 $P_{out} = 0$ (dB) の時、すなわち絶対閾値の出力レベルの場合、図 3 で述べ

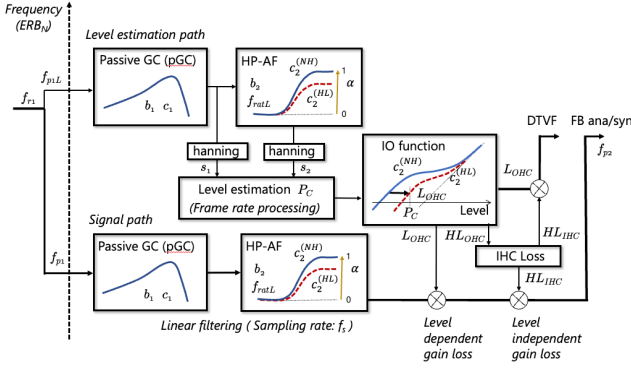


図 5: 提案法の模擬難聴 (WHIS_{v300}): 分析系の 1 ch 分のブロック図。詳細は本文と図 2、図 3 参照

た HL_{OHC} となる。これが決まると、 HL_{IHC} は式 8 より HL 時の絶対閾値 HL_{total} から求まる。さらに、そのチャンネルにおけるフレーム毎の減衰量 $L_{total}(P_c(\tau))$ は、 $L_{OHC}(P_c(\tau))$ と、一定の HL_{IHC} を加え合わせたものとなる。

$$HL_{OHC} = L_{OHC} |_{P_{out}=0} \quad (19)$$

$$HL_{IHC} = HL_{total} - HL_{OHC} \quad (20)$$

$$L_{total}(P_c(\tau)) = L_{OHC}(P_c(\tau)) + HL_{IHC} \quad (21)$$

このように、式 3 や式 15 の GCFB における HPAF の関数を用いて、入出力関数とその逆関数から、見通し良く減衰量を決めることができるようになった。

3.2 合成系

前節の分析系により、フレーム毎の減衰量 $L_{total}(P_c(\tau))$ と、線形フィルタ出力にこの減衰量を掛けたものが出力される。この 2 種類の出力を用いて、2 種類の音合成信号処理を行うことができる。

3.2.1 フレーム毎の減衰量から時変フィルタで合成

一つ目は WHIS_{v225} で用いていた直接時変フィルタ (Direct time-varying filter, DTVF) の方法である (詳細は [1] 参照)。入力信号を、hanning 関数の平方根の窓でフレーム分割し、シフト量を半分にする。一方 $L_{total}(P_c(\tau))$ から算出した、フィルタバンクチャンネルに渡る減衰パターン (Gain reduction pattern) を計算しておく。この減衰パターンのパワーを線形周波数軸に変換した後、最小位相減衰フィルタを算出する。信号のフレームごとにこの最小位相減衰フィルタを掛けたのち、平方根 hanning 関数をかけて、overlap-add する。入力信号を 1 ch のまま直接フィルタするため、フィルタバンク処理上の位相ずれの影響を小さくできて、比較的出力音の歪みは小さい。

しかし、当初の目的の一つであった時間特性劣化の導入は、この枠組みでは難しい。一度合成音にした後、処理時間の短い 1/3 オクターブフィルタバンク等で分割し、包絡線を低域通過した後、合成する実装を行ったが、一体感はない。これを以下では WHIS_{v300}^{DTV} と呼ぶ。

3.2.2 フィルタバンク分析合成

もう一つは、フィルタバンクで各周波数ごとに分割した波形を、チャンネル間の位相を補正した後、overlap-add して合成音とする手法である [2]。この場合は、図 5 の線形フィルタ出力の振幅を $L_{total}(P_c(\tau))$ で利得減衰させて、位相はそのまま用いて合成することになる。チャンネル間の位相補正をある程度行っても、上記の WHIS_{v300}^{DTV} に比べわずかに歪みが大きいように感じる。

しかし、この手法の利点は、フィルタバンク出力ごとに、振幅包絡に対して低域通過フィルタをかけて時間応答特性の劣化

を再現できることにある。これを以下では WHIS_{v300}^{FBAS} と呼ぶ。

4. 模擬難聴の評価

上記の WHIS により模擬難聴音を合成して、それを NH の GCFB_{v231} で分析する。この結果を HL の GCFB_{v231} の出力と比較することにより、難聴をどの程度良く模擬できるかを評価できるであろう。これは、いままでの模擬難聴では十分に行われてこなかった。ここでは、80 歳男性の平均聴力レベル [9] を用いて、末梢系の機能低下について比較検討を行った。なお、以下の $\alpha = 0$ の条件はプリセットした時の値で、WHIS で計算する時はオーディオグラムの閾値から最小の $\{\alpha \mid 0 \leq \alpha < 0.5\}$ が自動的に選ばれている。

4.1 入出力関数

ここでは、2.4 節と同じ条件で正弦波を用いた。図 4 に、模擬難聴出力を健聴者特性 (NH) の GCFB_{v231} で分析した結果の 500 Hz~4000 Hz の入出力関数を示す。(b1)~(b4) が WHIS_{v300}^{DTV}, (c1)~(c4) が WHIS_{v300}^{FBAS}, (d1)~(d4) が従来法の WHIS_{v225}, (e1)~(e4) が CamHLS の結果である。

図 4(a1)~(a4) の GCFB_{v231} と (b1)~(b4) の WHIS_{v300}^{DTV} を対比すると、OHC 健全度 $\alpha = 1$ の場合 (緑波線) の特性は、大きくずれていることがわかる。これは以下のように説明できる。GCFB_{v231} の出力レベルの調整は、図 3 で示される G_{OHC} と L_{IHC} の縦軸方向の利得で行われる。これに対し模擬難聴においては、蝸牛への入力音圧レベル HL_{OHC} と HL_{IHC} の横軸方向の損失で行われる。健全度 $\alpha = 1$ の場合は圧縮特性が健全なため、縦方向と横方向の利得調整で大きな違いが生じる。これに対し、 $\alpha \leq 0.5$ においては、両者の特性は $\alpha = 1$ の場合よりは近くなっている。老人性難聴者の場合、OHC だけが完全に健全で IHC の損失だけということも多くないであろう。このことから、 $\alpha \leq 0.5$ 程度に相当する OHC 損失をいれた近似である程度模擬できるであろう。一方、OHC が健全で IHC の機能低下がある Synaptopathy の場合は、模擬するのが難しいということになる。さらに、耳栓やグラフィックイコライザーといった線形の挿入損失は、OHC の健全度の制御がないため模擬難聴としては不十分であることを示している。

図 4(b1)~(b4) の WHIS_{v300}^{DTV} と (c1)~(c4) の WHIS_{v300}^{FBAS} の入出力特性はほぼ一致している。さらに、(d1)~(d4) の WHIS_{v225} の場合は、500 Hz と 1000 Hz の入力レベル 80 dB 以上で $\alpha = 0.5$ の条件でやや違いが見られるが、他の条件でほぼ一致している。このことは、WHIS_{v225} による今まで行ってきた実験結果と、WHIS_{v300} による今後の実験結果の間で整合性が取れることを示唆している。

最後に (e1)~(e4) の CamHLS においては、OHC の健全度の設定がなく、オーディオグラムから自動的にリクルートメントの特性が決まり、それによって利得が制御される。そのため、難聴者の入出力関数は 1 つだけである。この 80 歳条件の場合、WHIS の $\alpha = 0.5$ に近い特性となっていることがわかる。逆に言うと CamHLS で実装されているリクルートメント現象の模擬を、WHIS は OHC の制御により原因側で実現できていることがわかる。

4.2 聴覚スペクトル距離

模擬難聴は非線形システムであるため、正弦波で評価するだけでは不十分である。ここでは、信号処理の対象となる音声の出力表現における対比を行う。まず、基準として、GCFB_{v231} を上記と同様に 80 歳男性 [9] のオーディオグラムの設定で、音声分析して聴覚スペクトログラム ($S(\tau, ch)_{G-HL}$) (τ フレーム

番号で 0.5 ms ごとに計算。ch はフィルタチャンネル) を算出する。この時、入力音声レベルを RMS レベルで 50 dB SPL と 80 dB SPL として、OHC の健全度 α を 1, 0.5, 0 の 3 段階として、計 6 処理条件で分析を行った。音声データとしては、日本語単語データベース FW07 の親密度の高い 1 セットの中で、男女 2 名が発話した 20 単語をランダムに選んだ。

模擬難聴は、上記処理条件と全く同じにして、音出力を GCFB_{v231} の健聴者 (NH) 特性で分析して聴覚スペクトログラム ($S(\tau, ch)_{G-NH}^{WHIS}$) を算出した。これら 2 つの聴覚スペクトログラムの正規化距離 d_S (dB) を以下で定義した。(22)

$$d_S = 20 \log_{10} \left\{ \frac{\text{rms}(S(\tau, ch)_{G-NH}^{WHIS} - S(\tau, ch)_{G-HL})}{\text{rms}(S(\tau, ch)_{G-HL})} \right\}$$

なお、模擬難聴の合成音は原音声と開始時刻がずれる場合があるため、ある程度の範囲内でフレームシフトをして距離計算し、最小の d_S を選んだ。この値が小さいほどスペクトログラム間の類似度が高いことになる。

4.2.1 結果

図 6 に棒グラフで結果を示す。聴覚スペクトル距離はそのままでは解釈しにくいので、音声信号にピンク雑音を対比のため SNR 3 dB(*), 0 dB(o), -3 dB(Δ) で重畳した場合の距離もプロットした。これは目安で、処理が全く違うため、同じ d_S で聴取特性 (たとえば了解度) が同じとなるわけではないことに注意は必要である。

分散分析： 模擬難聴条件、音圧、健全度 α の 3 要因の分散分析を行うと、主効果も交互作用も全て有意 ($p < 0.0001$) であった。このことは、条件ごとに違いが大きいことを意味する。模擬難聴の良否は個別条件ごとに行う必要がある。

音圧条件での違い： まず上段の 50 dB と下段の 80 dB を全体的に比較すると、50 dB の場合の方が平均的に誤差が小さいことがわかる。これは、図 4 で示された、高音圧条件における入出力関数の大きな相違に起因すると考えられる。

模擬難聴の種類による違い： 一般的に WHIS に比べて、CamHLS のスペクトル距離 d_S がどの条件でも大きいことがわかる。このことは、CamHLS において帯域幅を広げるためにスペクトル平滑化 [3] を行っているためと考えられる。フィルタバンクの 1 ch 分の情報が他の ch に入りこんだ後で合成すると、信号処理的に歪みが生じる。この歪みの成分が d_S の増加につながっていると考えられる。

一方、WHIS_{v300}^{DTVF} と WHIS_{v300}^{FBAS} を比較すると、80 dB の α が 0.5 の時以外ほぼ同じであることがわかる。これらに対し、WHIS_{v225} は 50 dB の α が 0 と 0.5 の時に d_S が小さい以外、同等かやや低いことがわかる。ほぼ互換性は保たれていると考えられる。

健全度 α による違い： 健全度 α が 1 の時は、図 4 からわかるように高音圧条件で誤差が大きくなるため、 d_S も大きくなる。 $\alpha = 0, 0.5$ では $\alpha = 1$ の場合にくらべて全般に d_S が小さい。80 dB 条件では、SNR=3 dB の雑音重畳条件に対し、WHIS は $\alpha = 0.5$ の時は同程度、 $\alpha = 0$ の場合は d_S がやや大きくなっている。

まとめ： もともと難聴者にとって聞き取りにくい 50 dB SPL の音声は、模擬する機会も多くないであろう。実際の場面で話しかける「大きめの音声」である 80 dB の音声がどのように聞こえるか模擬することを考えると、WHIS_{v300}^{DTVF} が一番精度が高いと考えられる。実際に模擬難聴音の歪みは少ない感じがする。

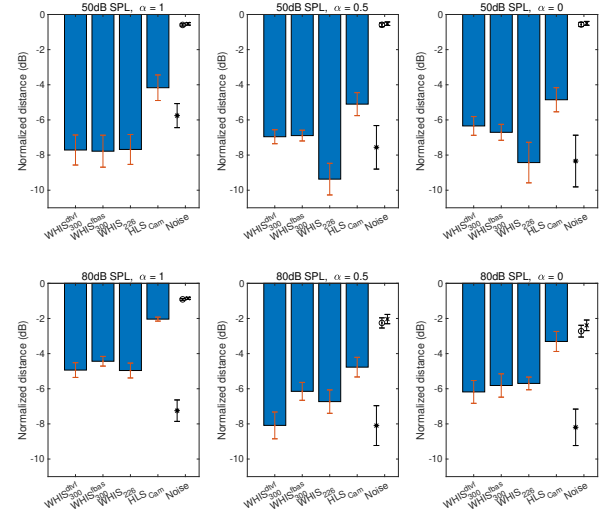


図 6: 聴覚スペクトル距離 d_S 。上段：入力音声レベル 50 dB SPL、下段 80 dB SPL。OHC 健全度 α : 左 1, 中 0.5, 右 0。棒グラフは WHIS 条件の d_S 。エラーバーは標準偏差を示す。対比のため GCFB(HL) に雑音を重畳した信号を入力したときの d_S を SNR 3 dB(*), 0 dB(o), -3 dB(Δ) で示した。

5. まとめ

開発した WHIS_{v300} は実験刺激作成用の Batch 処理プログラムだけではなく、WHIS_{v225} で供用していた GUI をそのまま継承しインタラクティブな模擬難聴体験ができる。また、OHC の健全度の制御により、リクルートメント現象を含む幅広い特性に対応できる。さらには CamHLS にくらべて処理に伴う歪みも小さいこともわかった。これらのことから音声了解度/明瞭度以外の聴取実験にも応用でき、使い勝手の良いシステムになっていると考える。ただし、オーディオグラムからは HL_{OHC} と HL_{IHC} の割合は決められないため効果的な測定手段が必要で、今後の課題である。なお WHIS_{v300} [10] とその処理に不可欠な GCFB_{v231} [11] は、GitHub において公開予定である。

謝辞 CamHLS の MATLAB コードは、Dr. Michael A. Stone から提供を受けた。ここに感謝する。この研究は、科研費 16H01734, 18K10708, 21H03468, 21K19794 の支援を受けた。

文 献

- [1] T. Irino and R.D. Patterson, “-,” Acoust. Sci. and Technol., vol.41, no.1, pp.99–107, 2020.
- [2] T. Irino and R.D. Patterson, “-,” IEEE trans. ASLP, vol.14, no.6, pp.2222–2232, 2006.
- [3] T. Baer and B.C. Moore, “-,” JASA, vol.94, no.3, pp.1229–1241, 1993.
- [4] M.A. Stone and B.C. Moore, “-,” Ear and Hearing, vol.20, no.3, pp.182–192, 1999.
- [5] T. Irino and R.D. Patterson, “-,” JASA, vol.109, no.5, pp.2008–2022, 2001.
- [6] B.C.J. Moore, An introduction to the psychology of hearing, 6th edition, Brill, 2013.
- [7] G. von Békésy and W.T. Peake, Experiments in hearing, Acoustical Society of America, 1990.
- [8] B.C. Moore, B.R. Glasberg, and T. Baer, “-,” J. Audio Eng. Soc., vol.45, no.4, pp.224–240, 1997.
- [9] 立木孝他, “-,” Audiology Jpn, vol.45, no.3, pp.241–250, 2002.
- [10] <https://github.com/AMLAB-Wakayama/WHIS/WHISv300>
- [11] <https://github.com/AMLAB-Wakayama/gammachirp-filterbank/GCFBv231>

Annex: Extended abstract

A new implementation of hearing impairment simulator WHIS based on the gammachirp filterbank
Toshio IRINO

A new implementation of Wakayama-university (Wadai) Hearing Impairment Simulator, WHIS, is reported. For this purpose, the gammachirp auditory filterbank (GCFB) was renovated to represent absolute thresholds for persons with normal hearing (NH) and with hearing loss (HL) (Fig.2, GCFB_{v231}). Frame-base processing was also introduced to improve processing speed to be almost real time while the conventional GCFB (Fig.1, GCFB_{v211}) took more than a few ten times of real time. Moreover, the outer haircell gain (G_{out}) in Eq. 15 and the inner haircell gain loss (L_{out}) was included into GCFB_{v231}. The OHC health $\{\alpha|0 \leq \alpha \leq 1\}$ was introduced into the parameter c_2 as in Eq. 9 to represent the hearing loss caused by the outer haircell dysfunction. The hearing loss by OHC and IHC was formulated as shown in Eq. 8. Figure. 3 shows schematic cochlear input/output (IO) functions where the output level is relatively defined from the absolute threshold. In contrast with the NH case (blue solid line), the OHC health α of 0.5 reduces the degree of compression (red dashed curve). The IHC loss L_{IHC} is further introduced to make the IO function across the point of HL 45 dB (purple dashed and dotted curve).

Figure 5 shows a new implementation of the analysis part of WHIS (WHIS_{v300}) based on GCFB_{v231} in Fig. 2. The main filters are the same and there is difference in the used of the IO function. Two synthesis methods were developed: direct time-varying filter (DTVf) used in the previous WHIS (WHIS_{v225}) [1] and filterbank analysis/synthesis (FBAS). DTVf is advantageous in the quality of simulated sound while FBAS is advantageous when simulating degradation of temporal resolution in the auditory pathway.

Comparisons were performed between the output of GCFB with hearing loss (HL) condition (GCFB^(HL)) and the output of GCFB with normal hearing (NH) condition followed by WHIS (WHIS + GCFB^(NH)). Figure 4 shows the IO functions of NH (blue solid line) and HL of the average 80 years-old Japanese male [9] when the OHC health α is 0 (dark red dashed and dotted line), 0.5 (purple dashed and dotted line), and 1 (green dashed line). Panels (a1)sin(a4) show the results of GCFB^(HL). Other panels show the results of WHIS + GCFB^(NH) when using WHIS_{v300}^{DTVf}, WHIS_{v300}^{FBAS}, WHIS_{v225}, and Cambridge version of hearing loss simulator (CamHLS) [3,4]. The IO functions of GCFB^(HL) are similar to those of WHIS + GCFB^(NH) except when $\alpha = 1$, i.e., healthy OHC. Figure 6 shows spectral distance d_S (Eq. 22) calculated between GCFB^(HL) and WHIS + GCFB^(NH) when SPL level (L_{eq}) is 50 dB (upper panels) and 80 dB (lower panels) and when $\alpha = 1, 0.5$, and 0. d_S values of WHIS_{v300}^{DTVf}, WHIS_{v300}^{FBAS}, and WHIS_{v225} are almost the same and smaller than those of CamHLS. This might be because CamHLS performs spectral smearing which increases the distortion.

WHIS_{v300} would be useful because it is possible to use the same GUI as in WHIS_{v225} for interactive experience of hearing loss as well as to use batch processing for producing experimental stimuli. The software of WHIS_{v300} [10] and GCFB_{v231} [11] will be provided in GitHub.