

リアルタイム音声処理を用いたデジタル聴覚プロテクターの開発

奈良工業高等専門学校 電子制御工学科 ○栗生 小百合 ◎西田 茂生

要旨

聴覚過敏により学習環境で不利益を被るのは児童が多い。この問題を解決するため、周囲のざわつきを低減し、先生の音声を聞き取りやすくするデジタル聴覚プロテクターを開発する。リアルタイム音声処理のために、マイコンを用いてノイズゲート及び各種フィルタ処理を行った。その結果、ノイズを低減し、設定した必要な音を抜き出せることが確認できた。

1. 緒言

聴覚過敏とは、日常で耳にする程度の環境音を不快に感じ、時には耳を覆いたくなるほど頭に響く症状である。不快に感じる音は人によって様々で、食器の接触音などの甲高い音、キーボードの打鍵音などの小さい音、飲食店など日常生活における騒がしい環境の音などがあげられる。聴覚過敏者は特に自閉症や発達障害の子供に多く、これらの症状のために授業に集中できない場合がある。聴覚過敏は客観的な評価が難しいため、治療法などの確立は遅れている。¹⁾ 現在その処方薬として、神経・精神安定剤が処方されているが、聴覚過敏の症状自体を治す薬はない。また音刺激を低減するために耳栓やイヤーマフなどの聴覚プロテクターが使用されているが、必要な音も減衰し聞き取りにくくなる問題点がある。

現在、飛行機や工事現場の騒音を削減するノイズキャンセリングを搭載したヘッドホンが発売されているが、この技術はすべての環境音を打ち消したい場合には有効であるが、騒音中での会話相手の音声などのみを聴取することはできない。

本研究では、上記のような問題点の解決策を提案する。イヤーマフにより外部の環境音を低減した上で、必要な音声のみを聴取することのできる聴覚プロテクターを開発する。そのためにマイコンによるデジタル信号処理により人の音声の周波数以上の音声を低減し、必要な音声以外の音声をカットする。また聴覚過敏により学習環境で不利益を被るのは児童が多いことから、開発する聴覚プロテクターは児童が授業中に使用することを想定する。

2. 聴覚プロテクターの概要

2.1. 本体の概要

本研究で開発する聴覚プロテクターはイヤーマフ内部にスピーカーとマイコン、外部にマイクを取り付けることで、マイクから外部の環境音を取得し、マイコンにより音声処理を行い、必要な音のみをスピーカーから出力する。マイクは無指向性マイクと単一指向性マイクの両方を取り付け、使用する状況により使い分けすることができるようにする。また、聴覚プロテクター本体には外部の環境音を遮断できるオーバーヘッド型のイヤーマフを採用する。オーバーヘッド型のイヤーマフはマイクや音声処理を行うマイコンを内蔵することが可能である。さらに、児童でも正しい取り付けが簡単に行えるという利点がある。

2.2. 音声処理の概要

マイクに入力された環境音に対して、周波数スペクトル、音量を解析し以下の処理を行うことで、会話相手の声や先生の声などの必要とされる音声を抽出する。

(1)ノイズゲート

ノイズゲートの特性を Fig.1 に示す。図のようにノイズゲートとは入力信号の振幅が閾値以下の場合、出力信号振幅を 0 にする処理である。²⁾ この処理では必要な音より振幅が小さい音声であるひそひそ声やエアコンの音、スピーカーからのノイズなどのカットを目的としており、予備実験の結果、閾値を-25 dBとした。

(2)低域通過フィルタ

低域通過フィルタには、FIR フィルタの一種であるウィンドウ処理法を使用した。理想的な FIR 型低域通過フィルタ(LPF)のフィルタ係数は次式となる。³⁾

$$b(m) = 2f_e \text{sinc}(2\pi f_e m) \quad (1)$$

ここで f_e はエッジ周波数、 m はフィルタ係数の個数、 $b(m)$ は m 目の乗算器にセットされるフィルタ係数である。また、 $\text{sinc}(x)$ はシンク関数と呼ばれ、 $-\infty$ から $+\infty$ に渡り定義される。そのため理想的

なフィルタ係数の個数 m は無限となる。しかしコンピュータ上ではこうしたフィルタは実現できない。そこで窓関数を用いて、フィルタを打ち切る。これがウィンドウ処理法である。ウィンドウ処理法はフィルタ係数の個数 m は多くなるが、遷移域は小さくなるという特徴を持つ。今回カットする高周波数の音声は小音でも響くことから、遷移域を小さくすることが可能であるウィンドウ処理法を採用した。

また本研究では人の音声域である 100Hz から 4000Hz の音を必要な音声としている。しかし食器の接触音やセミの鳴き声などは、その周波数のピークが 3000Hz から 4000Hz であり、人の音声の周波数と被っている。したがって本研究では人の音声の音質は少し変化するが、より多くの雑音を遮断することを目的とするため、3000Hz 以下の音声を通過するフィルタを作成する。

(3)帯域遮断フィルタ

上記の処理のみでは 100Hz から 3000Hz の音声は-25dB 以上の場合、必要な音のみを取得することができないため、授業中の児童の音声の全てをカットすることができない。そこで上記の LPF に加え、IIR フィルタを用いた帯域遮断フィルタ(BEF)を適用することで、児童の音声の基音と倍音を減衰させ、児童と先生の音声の混在する音声から先生の音声のみを抽出する。

Figure2 に IIR フィルタのブロック線図を示す。IIR フィルタはフィードバック経路を内部に持つため、インパルス応答が無限に続く。また FIR フィルタと異なり、IIR フィルタは演算が単純で高速なためリアルタイム音声処理に適している。

IIR フィルタはフィルタ係数を変化させることでフィルタの特性を変化させることができる。今回使用する BEF は特定の帯域だけを非常に低いレベルに減衰させるフィルタ回路であり、そのフィルタ係数は式(2)~(6)となる。ここで f_{c1} 、 f_{c2} は遮断周波数と呼ばれ、通過域と阻止域の境、 Q は遮断周波数における増幅率であり、 $f_{c1}-f_{c2}$ より求められる。²⁾

$$b(1) = \frac{8\pi^2 f_{c1}^2 - 2}{1 + \frac{2\pi f_{c1}}{Q} + 4\pi^2 f_{c1}^2} \quad (2)$$

$$b(1) = \frac{8\pi^2 f_{c1}^2 - 2}{1 + \frac{2\pi f_{c1}}{Q} + 4\pi^2 f_{c1}^2} \quad (3)$$

$$b(2) = \frac{4\pi^2 f_{c1}^2 + 1}{1 + \frac{2\pi f_{c1}}{Q} + 4\pi^2 f_{c1}^2} \quad (4)$$

$$a(1) = \frac{8f_{c1}^2 - 2}{1 + \frac{2\pi f_{c1}}{Q} + 4\pi^2 f_{c1}^2} \quad (5)$$

$$a(2) = \frac{1 - \frac{2\pi f_{c1}}{Q} + 4\pi^2 f_{c1}^2}{1 + \frac{2\pi f_{c1}}{Q} + 4\pi^2 f_{c1}^2} \quad (6)$$

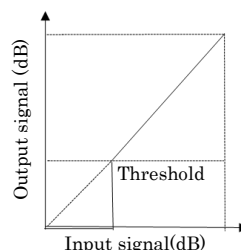


Fig.1 ノイズゲート

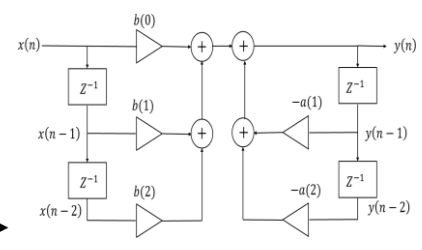


Fig.2 IIR フィルタ

3. 実験方法

3.1. 実験装置

本研究では、イヤーマフ本体として遮音性が高い KOSS 社製ヘッドホン QZ99 を使用した。ヘッドホンを使用することで、イヤーマフにスピーカーや信号入力部を取り付ける必要がなくなった。また、信号処理部には Raspberry Pi3、マイクには SANWA 社製 USB マイク MM-MCU02BK を使用した。



Fig.3 実験装置

3.2. 実験方法

以下の音声信号に対し、上記の処理をマイコンで実行することで、人の音声などの必要とする音声以外の音声が減衰しているか検証を行った。

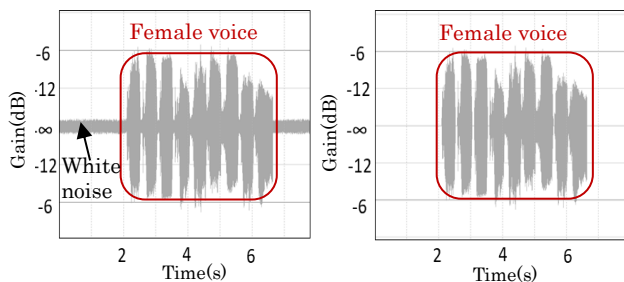
- (1)-30dB のホワイトノイズと女声の合成音声
- (2)ホワイトノイズ
- (3) LPF 適用後の 10 歳女児と 20 歳男性の「あいうえお」の音声

(1)の音声に対しノイズゲートを、(2)の音声に対し LPF を、(3)の音声には BEF のプログラムを実行し、適切に処理が行えているか検証を行った。ホワイトノイズとは全ての周波数域で同じ強度となるノイズである。ホワイトノイズは、Wavagane、女性の音声は UTAU を使用し作成した。また、周波数帯域の検証には Audacity、音量変化の検証には SoundEngine を用いた。

4 評価結果

4.1. ノイズゲート

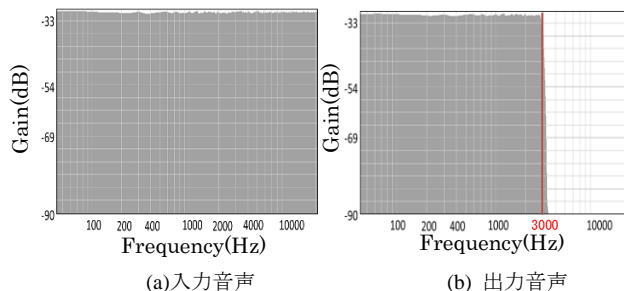
-30dB のホワイトノイズと女声の合成音声にノイズゲート処理を実行した結果を Fig.4 に示す。Fig.4 のように、ホワイトノイズと女声の合成音声から、ホワイトノイズのみのカットが可能であることが確認できた。



(a)入力音声 (b)出力音声
Fig.4 ノイズゲートによる音量変化の検証結果

4.2. 低域通過フィルタ

ホワイトノイズに LPF を行う前後の音声の周波数特性を Fig.5 に示す。図のように、入力音声に対して出力音声では、設定した 3000Hz 以上の音声を大幅にカットできていることがわかる。

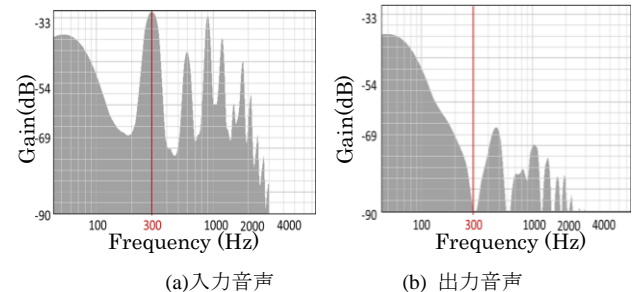


(a)入力音声 (b)出力音声
Fig.5 LPF による周波数スペクトル変化の検証結果

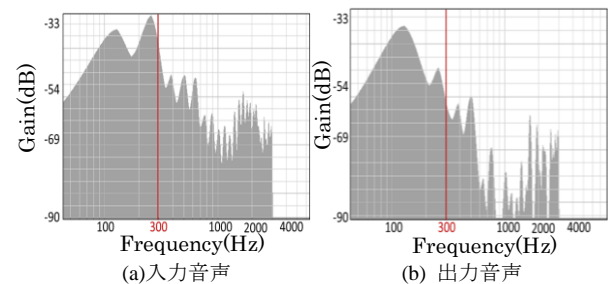
4.3. 帯域遮断フィルタ

3000Hz 以上の周波数をカットした 10 歳女児の音声と 20 歳男声それぞれに、同一の BEF を適用した。本研究ではカットオフ周波数を児童の音声の基音である 300Hz から 8 倍音までのそれぞれ $\pm 100\text{Hz}$ とした。

Figure6, Fig.7 に児童の音声と大人の音声それぞれに BEF 処理を行う前後の周波数スペクトルの変化を示す。児童、大人の音声共に設定した周波数が減衰することを確認できた。また児童の音声、大人の音声共に音声データ全体のゲインは減衰するが、児童の音声は大人の音声より大幅に減衰しており、ほぼ聞こえなくなった。それに対し大人の音声は、音は少しこもるが、発音もはっきりと聞き取ることができた。



(a)入力音声 (b)出力音声
Fig.6 児童の音声の周波数スペクトル変化の検証結果



(a)入力音声 (b)出力音声
Fig.7 大人の音声の周波数スペクトル変化の検証結果

5. 考察

作成したノイズゲートのプログラムにおいて、設定したゲイン未満の音声をカットすることができた。ノイズゲートは必要な音声以下のノイズのカットに適切であると考えられる。

LPF については、設定した周波数以上の音声をカットすることができたが、必要な音も減衰してしまう点、人の音声がかもってしまう問題がある。高周波域の音声をカットした上で、人の音声はこもらない手法の開発が必要である。

BEF については、児童の音声と大人の音声の合成音から、大人の音声のみを抜き出すことができたが、LPF 同様、大人の音声も減衰してしまう点、子供の音声に大きな抑揚がある場合、児童の音声をカットしきれない問題がある。また実用化のためには、複数の児童の音声が増減する周波数帯を探す必要がある。

さらにマイコンを用いることによりリアルタイム音声入出力が可能になったが、現在入力と出力に 0.5 秒の時間差がある。DMA 転送方式を使用することでリアルタイム出力により近づくと考えられる。

6. 結言

マイコンを用いることにより、リアルタイム音声入出力が可能になった。さらに、ノイズゲートによって、閾値以下の音声をカットし、必要な音声以外のノイズをカットすることができた。また、LPF により 3000Hz 以上の周波数域をカットし、BEF により大人と児童の合成音声から、児童の音声のみを減衰できることを確認した。

参考文献

- 1) デービッド・M・バグリー, ゲルハルト・アンダーソン, 聴覚過敏 仕組みと診断そして治療法, 海文堂出版(2012), p.322
- 2) 小坂直敏, サウンドエフェクトのプログラミング C による音の加工と音源合成, オーム社(2012), p.105