聴感特性を考慮した音声分離手法の聞き取りやすさ評価

藤田 佑樹[†1]　中村 美惠子[†1]

概要：本研究はインターレースオーディオによるアナウンス音声分離を提案している．音声を櫛形フィルターに通す本手法は音声分離可能であるが周波数情報の欠落した音声になる．周波数情報が欠落しても聞き取り可能なように人の聴覚特性を考慮した設計を行うため，フィルターの設計とフィルター音声のノイズ耐性について調べた．結果として低周波数成分の音声のノイズ耐性が高く高周波成分の音声のノイズ耐性が低い傾向にあると言えた．そのためフィルタ設計の際は低い周波数成分が多く含まれるよう設計が必要である．また周波数成分の偏りにかかわらず数字の書き取りの正解率がほぼ 100%になるには-9 dB から-12 dB 程度の SN 比が必要であった．この SN 比は一般的なイヤホンやヘッドホンで得られる SN 比よりも十分に低いものであるため，システムで生成される音声が十分に聞き取りやすい音であったと言える．

1. はじめに

1.1 背景

ショッピングモールやイベント会場などではガイド音声を来場者に向けて発信している．そのガイド音声は来場者全員に向けた連絡以外にも，来場者の一部に向けた様々な音声が流れている．図 1 のように複数の音声が同時に再生されれば，来場者は聞き取りづらくなる．このような公共空間では同じ時間と空間を共有するため，音声が被らない工夫が必要となる．そこで筆者らはあえて音を混合させ発信し，来場者の手元で分離できるインターレースオーディオシステムを提案している．

1.2 先行研究

混合音声の分離は，音源分離の研究がなされてきた．例えば，マイクロホンアレイを用いて音源の空間的方向・位置を求める手法[1]や，収録後の音声を分離するブラインド信号分離や，独立成分分析に基づいて音を分離する手法[2,3]が存在する．これらの手法は，分離のために高コストの計算が求められる

図 1　アナウンス音声の問題と解決

1.3 目的

本稿では提案手法であるインターレースオーディオシステムの実装と共に，聴覚実験を通してシステムにより生成される音声のノイズ耐性を調べ，実装後のシステムで音声の聞き取りやすさの検証した結果を報告する．

2. インターレースオーディオによる多重音声配信手法

提案手法であるインターレースオーディオによる多重音声配信の流れを図 2 に示す．インターレースオーディオは伝達関数 H_1, H_2 を畳み込むことで混合可能な音声を作成する前段処理から始まる．この伝達関数が周波数空間上で重

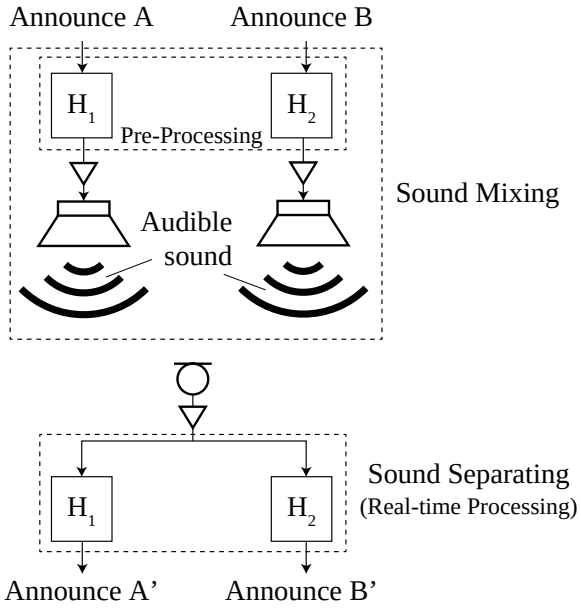

図 2　インターレースオーディオによる多重音声配信手法

†1 東京藝術大学　芸術情報センター

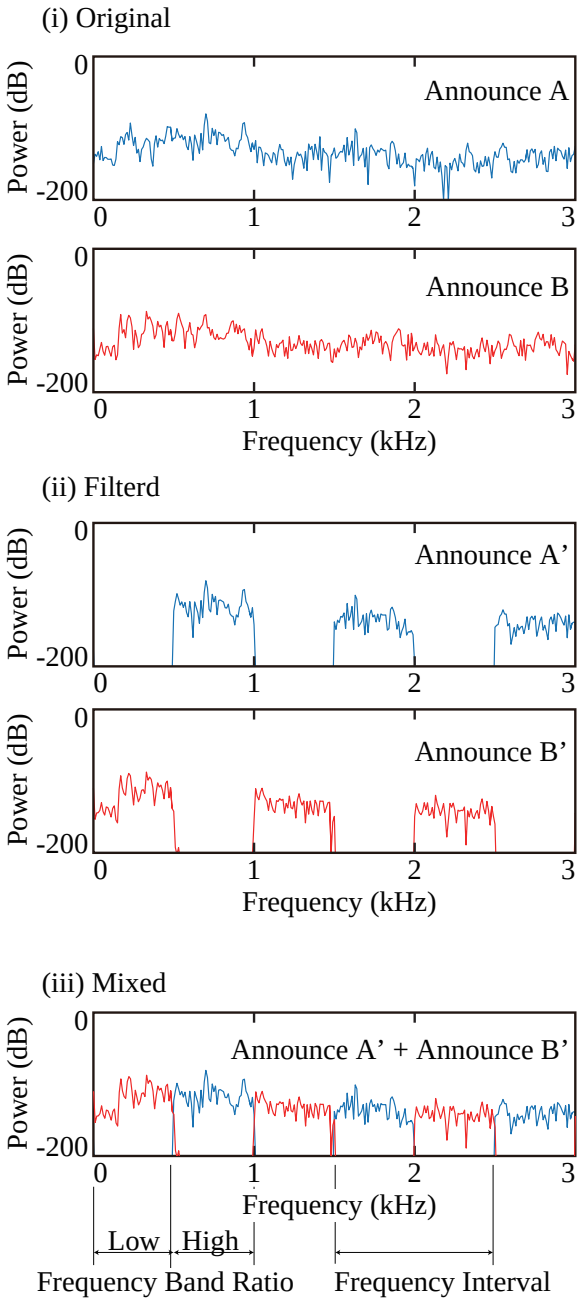

図 3　アナウンス音声とフィルターを適用した音声のスペクトル

複しない櫛形フィルターである．フィルター通過後の音声は，聴衆の耳に届く時に音声が混合される．このときに周波数空間上で音声情報が交錯していることからインターレースオーディオと呼んでいる．インターレースオーディオはマイクで取得後に，再び伝達関数 H_1, H_2 を畳み込むことで音が分離可能である．

図 3 に示すのがシステムで生成される音声のスペクトルである．櫛形フィルターを通過した音声は多くの情報が欠落する．周波数空間上では音声の倍音列に抜けが発生し，声道特性に対応するスペクトル包絡が変わる．また櫛形フィルターはエコーを発生させる．そのため音声情報の欠落による聞きにくさの問題が生じるが，先行研究の実験によって，人間の聴感に合わせて比較的聞き取りやすいフィルターを設計することが可能であることを示した．具体的には図 3 に示すように，Frequency Band Ratio と Frequency Interval を決めることで設計できる．結果として Frequency Interval が 700 Hz か 1000 Hz で，Frequency Band Ratio の Low : High が 1:1 のフィルターが聞きやすいという評価を得ている．

3. 櫛形フィルターを通過する音声のノイズ耐性の聴覚実験による評価

3.1 評価音声の設計

提案手法を利用した場合，利用者は図 4 の(ii) を聞きたいが，実際にはイヤホンで遮音しきれない図 4 (iii) が聞こえるため聞き取りが阻害される. 図 4 (iii) をノイズととらえ，提案手法のノイズ耐性を聴覚実験により調べた．

以前のフィルター設計の調査から Frequency Interval が 700 Hz か 1000 Hz で，Frequency Band Ratio の Low : High が 1:1 であるフィルターが適しているとわかっているため，同様のフィルターを用いた．

本実験で使用した音声は大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 音声資源コンソーシアムの筑波大 多言語音声コーパス (UT-ML) を使用した．内容は日本語の数字を読み上げたものである．これを Frequency Band Ratio の Low と High で別々の数字を読み上げノイズを作成し．聞き取りたい音声として Frequency Band Ratio の Low または High を加え，SN 比を調整した音声用意した．例えば，Frequency Band Ratio の Low「8354」と High「0342」が読み上げられる音声のうち，聞き取りたい音声が Low「8354」の場合は，「8354」が大きく，「0342」が小さく聞こえる音声となる．

実験では Frequency Interval が 700 Hz と 1000 H の 2 つのフィルターを用意し，聞き取りたい音声が Low と High の 2 つの場合で計 4 パターンに対して評価を行なった．各パターンにつき 40 種類の音声を用意し，実験参加者には計 160 問を聞き取ってもらいその数字の書き取りを行い，数字ひとつずつで正答率を評価した．

3.2 SN 比が悪い -6 dB の場合の聞き取り

まず実験参加者らに対し SN 比が-6 dB の音声を評価してもらった．実験参加者は 20 代の女性の 7 名である．図 4 は数字書き取りの正解率である．

特に High 側の音声の評価が悪く，正答率は約 51%と半分の数字が聞き取れていない．アナウンスのような長い音声の半分が聞き取れていないことなり，実用に耐えられないといえる．

3.3 SN 比を-9 dB から-12 dB まで下げた場合

3 章 2 節の Low 側と同等の正答率になる SN 比を探るさ

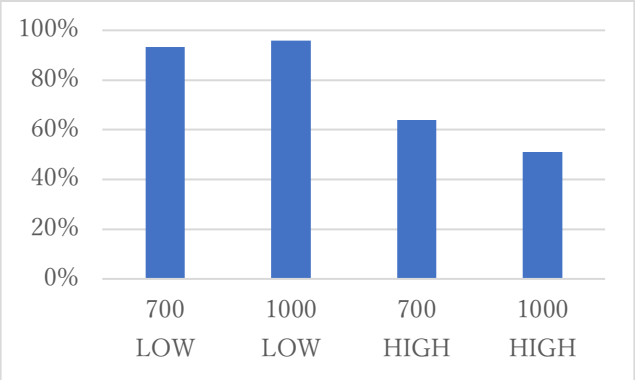

図 4　SN 比 -6 dB

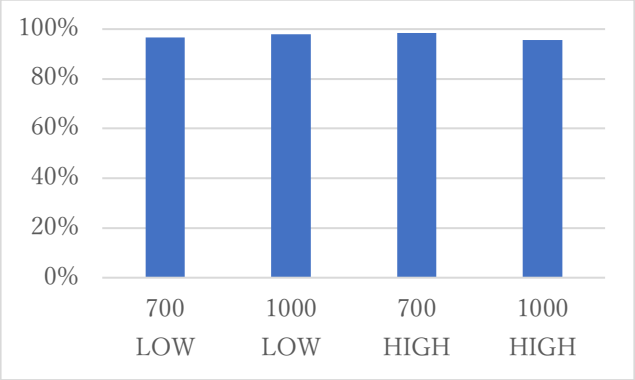

図 5　SN 比 -9 dB

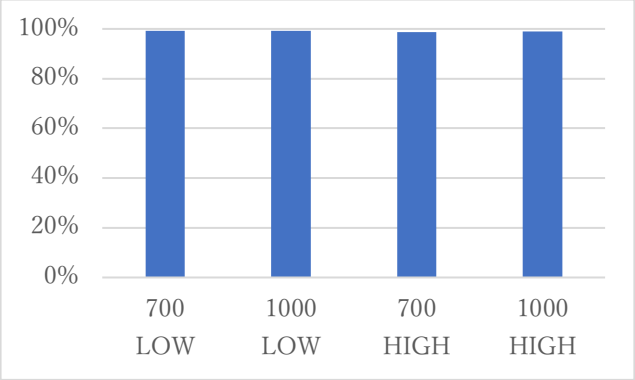

図 6　SN 比 -12 dB

めに SN 比を変えた音声で評価を行なった．実験参加者は 20 代の女性の 5 名である．図 5 は SN 比が -9dB，図 6 は SN 比が -12dB の場合の正答率である．

-9 dB の SN 比でも一番低い正答率でも 96%であり 3 章 2 節の Low 側と同等以上の正答率を得ている．-12 dB の SN 比では聞き取れない数字がほぼなかった．実際に-12 dB の SN 比の音を聞くとほぼノイズは聞こえていなかったためである．

以上のことから低周波数成分の音声はノイズ耐性が高く高周波成分の音声はノイズ耐性が低い傾向にあると言える．これは聴感上，聞き取りには比較的低い周波数成分を頼りにしているためと考えられる．提案手法においてはノイズ耐性の低い高周波数成分に合わせてシステムを設計すれば良いと言える．図 1 のようにイヤホンを使用する提案手法では，SN 比がイヤホンの音響遮音特性で決定できる．イヤホンで実現できる遮音特性は最大-20 dB 程度が得られる[4]ことから，提案手法では十分な SN 比が得られ，十分に聞き取り可能であると言える．

4. まとめ

本研究は周波数フィルタによるアナウンス音声分離手法において周波数情報の欠落した音声の聞き取り可能なように人の聴覚特性を考慮した設計を行うために，周波数フィルターの設計とフィルター音声のノイズ耐性について調べた．

フィルター音声の聞き取りやすさは周波数フィルターの周波数に依存し，エコーのかかった音声が聞き取りづらいという評価が得られていたため，エコーのかかりにくい，聞き取りやすい評価が得られた周波数フィルターのノイズ耐性を調べた．ノイズ耐性を調べるために，数字の書き取りを行なった．低周波数成分の音声はノイズ耐性が高く高周波成分の音声はノイズ耐性が低い傾向にあると言える．これは聴感上，聞き取りには比較的低い周波数成分を頼りにしているためと考えられる．周波数成分の偏りにかかわらず数字の書き取りの正解率がほぼ 100%になるには-9 dB から-12 dB 程度の SN 比が必要であった．この SN 比は一般的なイヤホンやヘッドホンで得られる SN 比よりも十分に低いものであるため，システムで生成される音声が十分に聞き取りやすい音であったと言える．

今後はシステムを実装したデバイスを利用することで，イベント会場での社会実装が課題となる．

本研究は東京藝術大学 芸術情報センターの倫理審査を受け，実験を行なった．

謝辞　本研究は公益財団法人中山隼雄科学技術文化財団の平成 29 年度助成研究 B を受けている．

参考文献

[1] 大須寿郎，山崎芳男，金田豊，"音響システムとディジタル処理"，電子情報通信学会，1995.
[2] A. Cichocki and S. Amari, "Adaptive Blind Signal and Image Processing", John Wiley & Sons, 2002.
[3] Aapo Hyvarinen, Juha Karhunen, and Erkki Oja, "Independent Component Analysis", John Wiley & Sons, 2001.
[4] 青山 裕樹，大谷 真，平原 達也，"聴覚実験用イヤホンの諸特性"，電子情報通信学会技術研究報告. SP，音声 107(165), 25-30, 2007-07-19