第８回　学生のための研究発表会

講演論文集

2009 年 11 月 29 日

熊本大学工学部

総合研究棟 204･208 号室

（社）日本音響学会　九州支部

2009 年度 日本音響学会九州支部
第８回　学生のための研究発表会　プログラム

10:00 開会挨拶

10:10－11:10 修士２年の部（１）　座長：積山薫（熊本大）

（休憩 11:10－11:20）

11:20－12:20 修士２年の部（２）　座長：尾本章（九州大）

（昼食休憩 12:20－13:20）

13:20－14:20 修士１年の部（ポスターセッション）　座長：川井敬二（熊本大）

（休憩 14:20－14:30）

14:30－16:00 学部学生の部　座長：水町光徳（九工大）

（休憩 16:00－16:10）

16:10－17:10 修士２年の部（３）　座長：山内勝也（長崎大）

音声画像ベースの構音障害診断における障害音声の定量的評価法の検討*

米倉達郎 富田翔 坂田聡 上田裕市 (熊本大院 自然科学研)

1 まえがき

構音障害は発音が正しくできない症状を指し，音声器官における形態上の異状により引き起こされる器質性構音障害，異状はないにもかかわらず正しい構音が行えない機能性構音障害，音声器官の麻痺による発音上の障害である運動性構音障害に大別される．本稿では，器質性構音障害のうち代表的な口蓋裂を対象の構音障害とする．この口蓋裂とは，先天性異常の一つであり，軟口蓋あるいは硬口蓋またはその両方が閉鎖しない状態のことを指す．このような口蓋裂特有のことばの異常は，鼻にぬけるような声の開鼻声と異常構音による音の歪みが主である．

音声障害の診断・訓練への応用として先行研究であるPCベースのリアルタイム音声画像化システム [1] の機能が期待される．これは，音素の置換・省略・歪みとして現れる構音障害による音声現象に関する視覚的・客観的評価を行うことを意味する．現在臨床やリハビリの現場では，音素の置換・省略・歪み等の音声現象について聴覚判断によって診断を行っているため，評価が主観的にならざる得ない．したがって，客観的な構音診断を行うために，発語中の音素(の有無を含めた)位置の検出とその音響的歪みを定量化する必要がある．

本研究では，ＤＰマッチング手法に基づく自動セグメンテーションを用いて音素の置換・省略・歪みの定量化を行うことを目的とする．音素の置換･省略ではセグメンテーション精度の検討や検出される音素位置の視覚化について述べる．音素の歪みでは，ＤＰマッチングにおける累積距離に関して健常音声群での距離基準を定め，音響歪みの定量化について述べる．

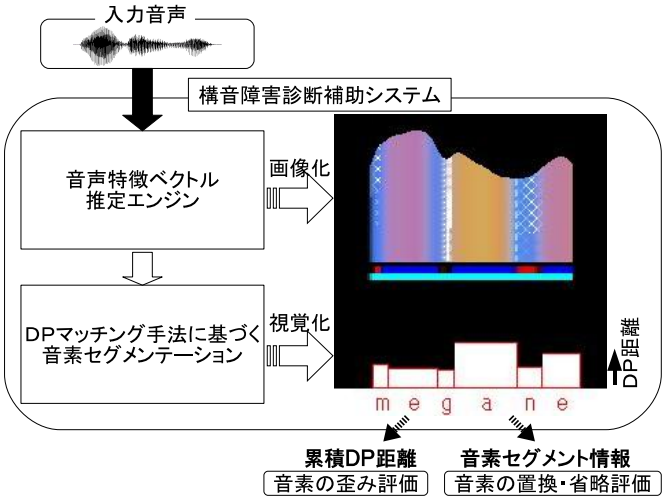


Fig. 1 Illustration of the proposed diagnosis system, where speech features are visualized directly and phoneme segments based on DP-matching are displayed graphically.

2 構音障害診断補助システム

Fig.1 に本研究で構音障害診断システムの概略図を示す．図中の音声特徴ベクトル推定エンジンと音素セグメンテーション処理について次に述べる．

2.1 音声特徴ベクトル推定エンジン [2]

Fig.2 に構音障害診断補助システムの音素セグメント情報までの流れを示す．音声画像表現に用いている複合パラメータに関する処理は，先に提案された音声特徴ベクトル推定エンジン [2] で行う（Fig.2 左)．この推定エンジンはフレーム周期 (10ms) 毎にフォルマント周波数やメルLPCケプストラム係数などの音声特徴ベクトル (複合パラメータ) を出力する．同時に，それらのパラメータを入力とするニューラルネットワークにより，母音性，破裂性，摩擦性などの音素特徴ベクトルを得る．また，これらの音声，音素特徴ベクトルと日本語音素 28 音素の標準パタンと

*Investigation of a quantitative evaluation method of dysarthric speech for a diagnosis system based on speech visualization. by YONEKURA Tatsurô, TOMITA Kakeru, SAKATA Tadashi, and UEDA Yuichi(Graduate School of Science and Technology,Kumamoto University)

の距離をフレームごとに並べた音素距離行列を作成する．

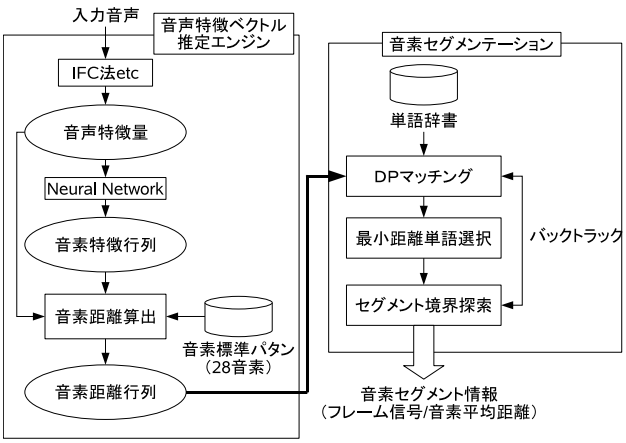


Fig. 2　Block diagram of phoneme segmentation, where phoneme distances are estimated in an engine estimating speech feature vector and using them phoneme segmental information is obtained by the DP-matching.

2.2 DP マッチングによる音素セグメンテーション [3]

音素セグメンテーション（Fig.2 右）においては，入力音声と単語辞書間で音素系列の音素距離の比較を行う．単語辞書には単語ごとに音素記号系列で表記されており，本研究での発話診断では，発話意図している単語が既知であるため，検査語毎に，癖や発話障害等（音素の置換・省略) が予想される単語については，単語ひとつにつき複数個の音素記号系列を単語辞書に登録する．継続時間長を適合させるためにフラグ付 DP マッチング [4] を用いる．これにより入力単語音声と比較する候補音素列の最小累積ＤＰ距離を決定する．本研究では，この累積ＤＰ距離を発話における歪みの評価基準として検討する．この距離を単語辞書中のすべての候補音素列について求め，距離が最も小さくなる候補音素列を選択する．その後，DP マッチング結果のバックトラックにより，音素セグメンテーション境界を探索し，単語中の音素セグメンテーションを行う．同時に，各音素セグメントでの音素距離平均値を算出する．

これらの音素情報 (音素系列～セグメント長/音素距離) から発話評価の定量化を目指すと共に，これらを音声画像表現と並列可視化して診断補助として提供する．

3 実験

実験試料は Table.1 の規定の構音障害検査語群 50 単語音声（2～5 モーラ）の健常者 7 名 (成人男性 5 名，成人女性 2 名) と口蓋裂患者 2 名 (口蓋化構音 22 単語 (男性)，咽頭破裂音 8 単語 (女性)) を用いる．

Table 1　The inspection word list.

パンダ	ポケット	バス	ぶどう
まめ	めがね	みかん	たいこ
とけい	テレビ	でんわ	ないてる
ねこ	にんぎょう	カニ	コップ
ケーキ	くち	キリン	ガム
ごはん	ぎゅうにゅう	さかな	そら
せみ	すいか	つみき	ぞう
ズボン	しんぶん	ちょうちょ	ちいさい
じゃんけん	ジュース	じてんしゃ	ふうせん
ひこうき	はっぱ	はさみ	らっぱ
ロボット	れいぞうこ	りんご	やきゅう
ようふく	あし	あひる	えんぴつ
うさぎ	いぬ		

3.1 ＤＰ累積距離による歪み評価実験

3.1.1 実験手順

全ての実験試料の累積ＤＰ距離を単語毎に算出し，健常音声と口蓋裂音声で比較を行う．話者毎の単語音声長の違いを規格化するために，各単語について累積ＤＰ距離を総フレーム数で除した値を評価値として用いる．

一方，口蓋裂音声については聴取実験を行い，聴覚的な音素の歪みの評価した．被験者は健聴男子大学生 4 名で，各単語音声をランダムな順序で聴取し，音素の歪みを三段階 (1:歪みがない，2:歪みが少しある，3:歪みが多い) で評価した．この聴覚的評価とＤＰ距離評価値の比較により，音素歪みの評価基準としての妥当性を考察する．

3.1.2 実験結果

Fig.3 は，健常音声 (7 名の最大値，最小値，標準偏差) と口蓋裂音声の評価値を表した図である．50 単語は健常音声の平均値について昇順に並べており，この健常音声の結果を評価基準とする．多くの口蓋裂音声は健常音声より値が大きく，一方で両者が同等の値に

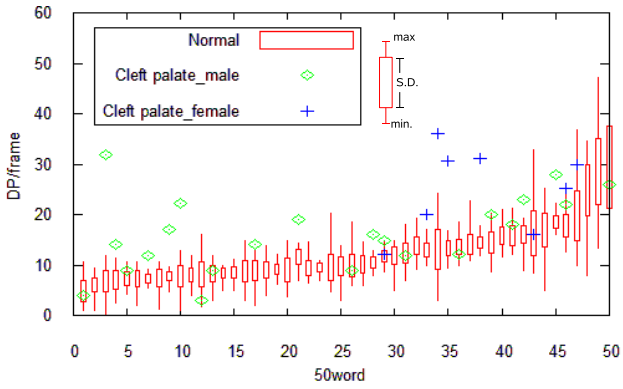


Fig. 3 Result of the DP-evaluated value for inspection words. "Normal" as criterion is represented by maximum, minimum and standard deviation.

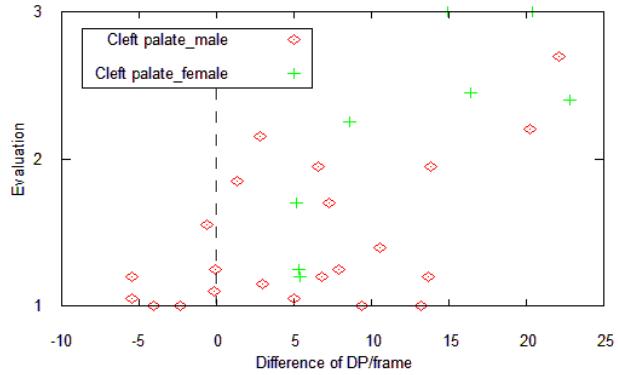


Fig. 4 Result of the DP-evaluated value and auditory evaluation.

なるものも存在した．このことが聴取評価とどう対応するかを次に検討した．

Fig.4 は，口蓋裂音声と評価基準 (Normal) の平均値間の差と聴取評価との相関図である．差が 0 以下の場合，口蓋裂音声は評価基準と同等のＤＰ距離である．

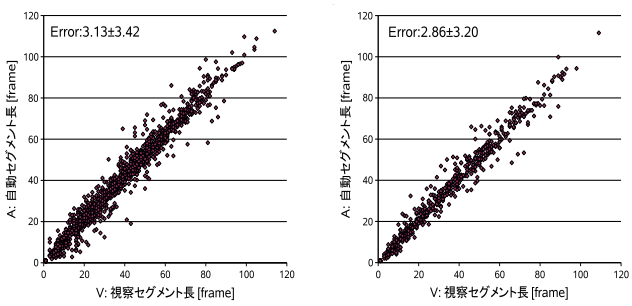
聴覚評価値が大きいほど聴覚的に歪みがあることを表す．この図から聴覚的に歪みがあるほど評価基準との差が大きくという相関がみられ，聴覚的印象の傾向との一致が確認できる．これにより，音素の歪み評価として累積ＤＰ距離を用いることが有効であると考えられる．

3.2 音素ＤＰ距離での検討

3.2.1 セグメンテーション実験

累積ＤＰ距離について検討を行ったが，次に音素毎のＤＰ距離について行う．しかしながら，この音素ＤＰ距離を算出する際に音素セグメンテーションを行うため，まずはその精度を調べる必要がある．

全健常者の音声試料について，2.2 節の自動セグメンテーション処理を適用して，全構成音素の境界をフレーム単位 (10ms) で求めた．また，波形の視察による音素セグメンテーション (V) を行い，両者の測定結果をフレーム単位で比較した．



(a) Male voice (5 male)　(b) Female voice (2 female)

Fig. 5 Accuracies in phoneme segmentation based on DP-matching.

Fig.5 は男女声群についてのセグメンテーション結果で，視察による結果 (V) と自動セグメンテーションによる結果 (A) の相関図である．図中の Error は視察 (V) と自動 (A) の結果の絶対誤差の平均と標準偏差である．結果として，ＤＰ処理に基づく自動セグメンテーション手法は視察による結果と平均で 3 フレーム程度 (30ms) のエラーの推定精度が得られた．推定例の傾向として，無音部に続く破裂部での誤りがみられ，これが Error の変動の主因となっている．また，この Error が音素ＤＰ距離との相関がないことを確認した．

3.2.2 音素ＤＰ距離

Fig.6 は健常音声の音素ＤＰ距離（7 名の全音素の最大値，最小値，標準偏差）を音素毎に表した図である．音素は平均値について昇順に並べている．50 単語音声の音素のため音素数はすべて異なる．音素によってＤＰ距離にばらつきがあり，/f/や/k/などの音素でのＤＰ距離が大きい結果となった．これら

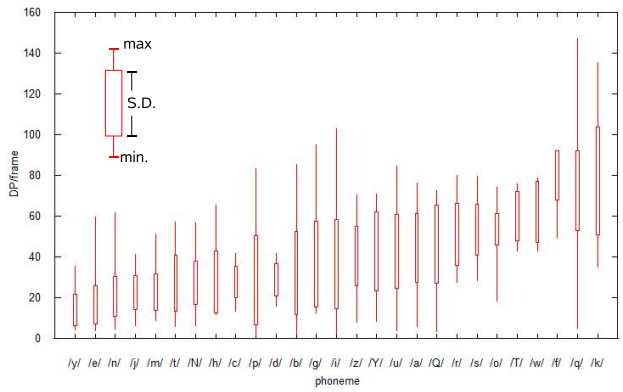
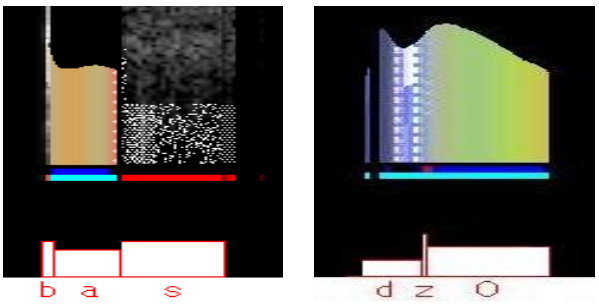


Fig. 6　Result of DP-matching distances of phoneme for normal speakers, where maximum, minimum and standard deviation(SD) are marked respectively.

の音素が含まれる単語は累積ＤＰ距離が大きいと考えられる．また，各話者で音素毎の平均ＤＰ距離を算出し，それが話者によって変動があるか検討した結果，/k/と/q/(無音)以外の音素で標準偏差が 10 以下となり，話者の個人差はほとんどないと考えられる．

3.3 音素ＤＰによる置換・省略評価

Fig.7 に口蓋化構音男性の/basu/，咽頭破裂音女性の/zO/の音声画像と音素セグメント表示例を示す．(a) では/u/に対して母音の無声化がおきており，(b) では/z/が/dz/に置換しているのが確認できる．このように音素の置換や省略がセグメンテーションの結果から確認でき，今後は聴覚評価との関連を検討していく必要がある．



(a) /basu/　　(b) /zO/

Fig. 7　Example of speech visualization and phoneme segmentation.(a)/basu/ and (b)/zo/ uttered by a cleft palate male (a) and a female (b), respectively.

4 まとめ

本稿では, 構音声画像ベースの構音障害診断補助システムとして，ＤＰによる累積距離や音素セグメント情報による障害音声の定量的評価法を検討した．

音素の歪み評価では，ＤＰマッチング距離の評価値を用いて，健常音声の単語ごとの評価基準を定め，口蓋裂音声との比較を行った．評価基準との差異がみられ，それが「音素の歪み」の聴覚的印象の傾向と一致することを確認した．さらに多くの音声を用いて評価基準の妥当性を検討する必要がある．音素の置換・省略評価では，セグメンテーション精度の評価を行い，音素セグメント情報と聴覚評価の比較を行う必要がある．

今後はこれらの評価法を構音障害音声の診断評価システムに用いることを検討していく必要がある．

謝辞

本研究の一部は，平成 21 年科研費（基盤研究 (C)，No.21500476）の補助を受けた．また，本研究テーマに関する助言と提案をいただいた鹿児島大学大学院医歯学総合研究科平原講師・五味助教に感謝する．

参考文献

[1] 上田裕市 他，”リアルタイム音声画像化処理に基づく発話訓練システムの構築”，信学技報 WIT2007-104，pp.79-84，2008.
[2] 上田裕市 他，” 音声応用システムのためのリアルタイム音声特徴推定エンジンの構築 ”，信学技報 SP2008-67，pp.61-66，2008.
[3] 富田翔 他，”音声画像ベースの発話評価のための自動音素セグメンテーションの検討”，電気関係学会九州支部連合大会予稿集 (CD-ROM)，03-2P-07，2008.
[4] 池田直光 他，”複合パラメータを用いた音声認識に対する声道長比正規化の効果”，信学論 (D-Ⅱ)，vol.J87-D-Ⅱ，no.7，pp.1416-1427，JUL.2004.

周波数領域両耳聴モデルにおける音源分離性能の検討
-音源分離に関するパラメータの分離性能への影響- *

冨田美奈子 苣木禎史 宇佐川毅 (熊本大)

1 はじめに

これまで様々な音源分離手法が提案されているが, 多くの音源分離手法における解決すべき問題として, 音質の劣化が挙げられる。雑音推定の誤り等によって音源分離が適切に行われなかった場合, 信号歪みやミュージカルノイズによる分離音の音質劣化が生じる。分離音は人によって聴取されるため, 聴取者にとって不快となる音質の劣化は改善すべきである。特にミュージカルノイズへの対策として, スペクトル減算法において, 人間の聴覚特性に基づいた減算係数調整を行う手法 [1] や, ミュージカルノイズ発生量の測量尺度として対数カートシス比を用いた自動最適化法 [2] などが提案されている。また先行研究においても周波数領域両耳聴モデル (FDBM) [3] におけるミュージカルノイズ抑制を試みているが, 十分な性能は得られていない [4]。本稿では, FDBM における音源分離性能向上という目的のため, 具体的な手法の検討に先立って FDBM の音源分離に関するパラメータが分離性能にどのように影響しているかを検討する。なお分離性能の評価手法として電話帯域音声の客観的音声品質評価手法である Perceptual evaluation of speech quality (PESQ) [5] を用いるが, 先行研究において FDBM の評価へ適用可能であることを確認している [6]。

2 周波数領域両耳聴モデルによる音源分離

本章では Fig. 1 に示す FDBM のアルゴリズムについて述べる。

2.1 両耳間位相差 (IPD) 及び両耳間レベル差 (ILD) の算出

両耳入力信号 $l(n)$, $r(n)$ をフーリエ変換することで得られるスペクトル $L(k)$, $R(k)$

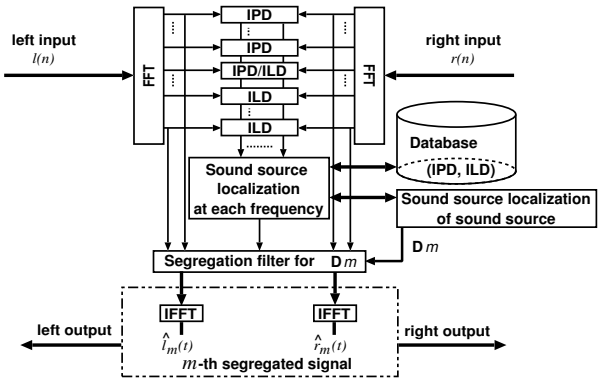


Fig. 1 Block diagram of frequency domain binaural model.

を用い, 両耳入力信号間のクロススペクトル $C_{lr}(k)$ を求める。

$$C_{lr}(k) = L^*(k)R(k) \tag{1}$$

ここで k は周波数に対するインデックスであり, $L^*(k)$ はスペクトル $L(k)$ の複素共役を示す。またパワースペクトルを $C_{ll}(k)$ とすると, 両耳間位相差 (IPD) $\theta_{lr}(k)$ と両耳間レベル差 (ILD) $\xi_{lr}(k)$ はそれぞれ式 (2), (3) から求められる。

$$\theta_{lr}(k) = \tan^{-1}\left[\frac{\mathrm{Im}[C_{lr}(k)]}{\mathrm{Re}[C_{lr}(k)]}\right] \tag{2}$$

$$\xi_{lr}(k) = 20\log\left|\frac{C_{lr}(k)}{C_{ll}(k)}\right| \tag{3}$$

2.2 周波数帯域毎の到来方向推定

あらかじめ左右耳の頭部伝達関数 (HRTF) により作成したデータベースの IPD $\theta_{map}(k,\phi,\psi)$ 及び ILD $\xi_{map}(k,\phi,\psi)$ と, 入力信号から求められた IPD $\theta_{lr}(k)$ 及び ILD $\xi_{lr}(k)$ の比較によって推定された方位角, 仰角の組合せ, $(\phi,\ \psi)$ を, それぞれ $D_\theta(k,\phi,\psi)$, $D_\xi(k,\phi,\psi)$ として得る。なお,

* A study on sound source segregation performance of frequency domain binaural model - Effect from segregation parameters - by TOMITA, Minako, CHISAKI, Yoshifumi, and USAGAWA, Tsuyoshi (Kumamoto University)

本稿では MIT より提供されている HRTF データベースを使用する [7]。次式により方向推定情報 $D_m(k,\phi,\psi)$ を得る。

$$D_m(k,\phi,\psi) = \beta(k)\cdot D_\theta(k,\phi,\psi) + (1-\beta(k))\cdot D_\xi(k,\phi,\psi) \quad (4)$$

ここで $\beta(k)$ は周波数毎の加重係数である。低周波数領域での回折現象の容易さによるレベル差の減少や, 高周波数領域における位相回転に伴う多義性から, 低域では IPD, 高域では ILD が強調されるよう定義されている。ILD を用いる帯域においても高域では ILD の遷移が非線型であり方向推定の不確定性が高まる。

2.3 音源分離フィルタ

音源の分離は, 音源方向推定情報 $D_m(k,\phi,\psi)$ を基に行う。目的音源方向から到来したと推定されたスペクトル成分を抽出し, 逆 FFT 処理することで分離信号を得る。分離信号 $\hat{l}_m(t)$ および $\hat{r}_m(t)$ は入力信号スペクトル $L(k)$, $R(k)$, 信号分離フィルタ $H_m(k)$ から

$$\hat{l}_m(n) = \mathrm{IFFT}[H_m(k)\cdot L(k)] \quad (5)$$
$$\hat{r}_m(n) = \mathrm{IFFT}[H_m(k)\cdot R(k)] \quad (6)$$

として得られる。なお, 観測信号スペクトル $L(k)$ および $R(k)$ に対して同様のフィルタを用いて分離処理を行っているため, 分離信号の IPD および ILD は観測信号のそれと同一あることから, 分離信号も空間情報を保持していると考えることができる。

3 シミュレーション

3.1 シミュレーション条件

シミュレーションで用いる音源分離パラメータの値を Table 1 に示す。複数の値がある項目にて下線を付したものが, 従来より使用しているパラメータである。この値の結果を基準に音源分離性能を検討する。方向角は正面を 0° , 右回りを正の方向として, ターゲット方向を $(\phi, \psi) = (0°, 0°)$ とし, 雑音源を $(-30°, 0°)$, $(60°, 0°)$ のいずれかに配置する。ターゲット音源は男声スピーチを用い,

Table 1 Values of segregation parameters.

FFT tap [samples]	256, <u>512</u>, 1024
Frame shift [samples]	32, 64, <u>128</u>, 256

雑音源には女声スピーチもしくはピンクノイズを用いた。音声試料は日本音響学会研究用連続音声データベース [8] に収録されている試料を使用しており, サンプリング周波数は 16 kHz, 量子化ビット数は 16 ビットである。FFT 処理ではフレーム毎にハニング窓を適用する。入力信号の SNR はドライソースの段階で設定するものとし, $-20 \sim 20$ dB の範囲で 10 dB 刻みで変化させる。

3.2 性能評価

(1) 指向特性

本稿では $(0°, 0°)$ をターゲット方向としているため, 仰角 0° の水平面上で方向角 $-90° \sim 90°$ の 10° 間隔に音源を配置し, それぞれの方向での入力信号に対する出力信号のゲインを算出することで指向特性を求めた。ゲイン G は入力信号を $s(t)$, 出力信号を $\hat{s}(t)$ としたとき次式で定義される。

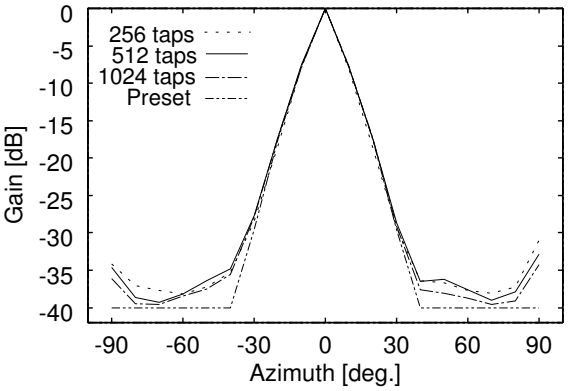
$$G = 10\log\frac{\overline{\hat{s}(t)^2}}{\overline{s(t)^2}} \quad (7)$$

(2) PESQ

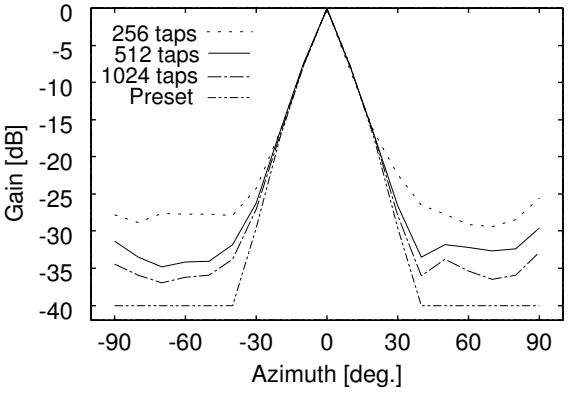
PESQ [5] は電話帯域音声の客観的音声品質評価手法である。PESQ 評価値は $-0.5 \sim 4.5$ の範囲で与えられ, 参照信号と評価対象が同質である場合に 4.5 となる。本稿では理想的な FDBM 出力であるターゲット信号を参照信号として評価を行う。PESQ は雑音抑圧アルゴリズムの影響について十分に検証されていないため, 本稿では入力信号と出力信号の評価値の相対比較に用いることとし, それにより FDBM の分離性能を評価する。

3.3 シミュレーション結果

Fig. 2 に音源に女声スピーチを用いた場合とピンクノイズを用いた場合のフレームシフト 128 サンプルでの指向特性を示す。Preset は FDBM で定義している重みであり, 方向推定が適切に行われた場合はこの値に近付



(a) Female speech



(b) Pink noise

Fig. 2　Directivity patterns of female speech and pink noise when the frame shift was 128 samples.

く。方向角が ±30° 以上の範囲は, IPD, ILD が非線型となるため方向推定誤差が生じやすいことから, Preset の値より減衰量が小さいと考えられる。Fig. 2 より ±30° の範囲ではいずれもほぼ同様の特性だが, ±30° 以上の範囲では FFT タップ長が長いほど特性が Preset に近付いていることがわかる。Fig. 2 (b) は音源にピンクノイズを用いた場合だが, ±30° 以上の減衰量が Fig. 2 (a) よりも減少しているのは, 高域での方向推定の不確定性のためであると考えられる。しかし, タップ長が長いほど減衰量が大きい傾向は女声スピーチと同様である。

Fig. 3 に PESQ による評価結果を示す。雑音源に (60°, 0°) に配置した女声スピーチを用いた入力 SNR = 0 dB での左耳の信号の結果である。図中の二点鎖線は入力信号の評価値を示し, 一定値である。Fig. 3 より, FFT タップ長によらずフレームシフトが短いほど

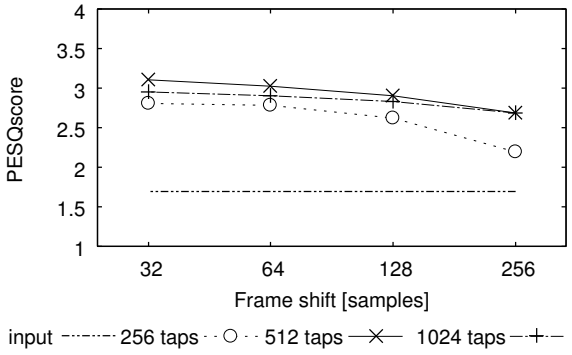


Fig. 3　PESQ scores at the left channel signal, when the target signal was male speech at (0°, 0°) and the interference signal was female speech at (60°, 0°).

改善量が大きいことがわかる。オーバーラップが増えることで出力信号の歪みが改善されるためであると考えられる。また, 同じフレームシフトで比較すると, わずかな差であるが, FFT タップ長が 512 の場合の改善量が大きい。ここで Fig. 3 におけるフレームシフト 128 サンプルのときの各 FFT タップ長での出力信号のスペクトログラムを Fig. 4 に示す。各信号は左耳の信号である。1.5 ~ 1.8 s の区間を比べると, Fig. 4 (b) ではほとんど雑音成分がないが Fig. 4 (d) では残留雑音成分が顕著である。また, Fig. 4 (a) と比較すると Fig. 4 (d) ではターゲット成分の損失が少ないが, Fig. 4 (b) では高周波数帯域の成分の損失が確認できる。以上の結果から, タップ長 256 の場合, タップ長 512 に比べて改善量が小さい原因として, 周波数分解能が低いため, 全体的に信号歪みが生じていることがあげられる。またタップ長 1024 の場合には周波数分解能の高さから方向推定誤差が生じやすくなり, 雑音抑圧性能が低下したと考えられる。雑音源の配置や種類, 入力 SNR の値を変化させた場合にも, タップ長 512 の改善量がもっとも大きいという傾向が見られた。また, フレームシフトについては上述のように短くすることで音質が改善しているが, わずかな差であり処理時間の増加を考慮すると, 本稿で検討した範囲では従来より使用しているタップ長 512 が適切な選択であると考えられる。

4　おわりに

本稿では FDBM の音源分離に関するパラメータによる, 分離性能への影響を検討した。指向特性より, タップ長を長くすることで定義した特性へと近づくことが確認されたが, PESQ の評価では雑音源を配置した場合には従来より用いているタップ長 512 の改善量がもっとも大きいという結果であった。

参考文献

[1] N. Virag, “Single channel speech enhancement based on making properties of the human auditory system,” *IEEE Trans. Speech Audio Process.*, **7**, 126-137 (1999).

[2] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, K. Kondo, “Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics,” *Proc. IWAENC 2008* (2008).

[3] H. Nakashima, Y. Chisaki, T. Usagawa and M. Ebata, “Frequency domain binaural model based on interaural phase and level differences,” *Acoust. Sci. & Tech.*, **24**, 172-178 (2003).

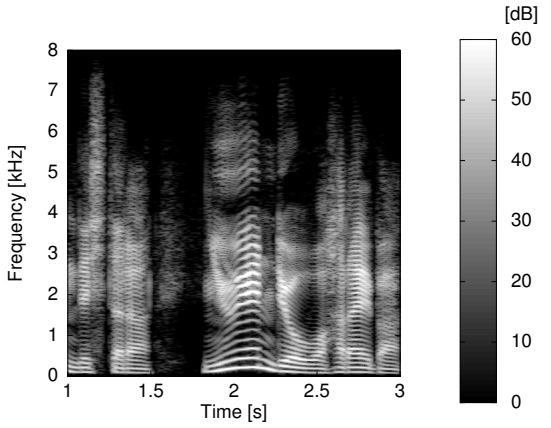
[4] 冨田美奈子, 河野亮詞, 苣木禎史, 宇佐川毅, “周波数領域両耳聴モデルによる音源分離におけるミュージカルノイズ抑制の試み - 時間-周波数フィルタリングを用いた分離フィルタの事後処理 -,” 音講論集, pp. 795-798 (2009.3).

[5] ITU-T Recommendation P.862, “Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs” (2001).

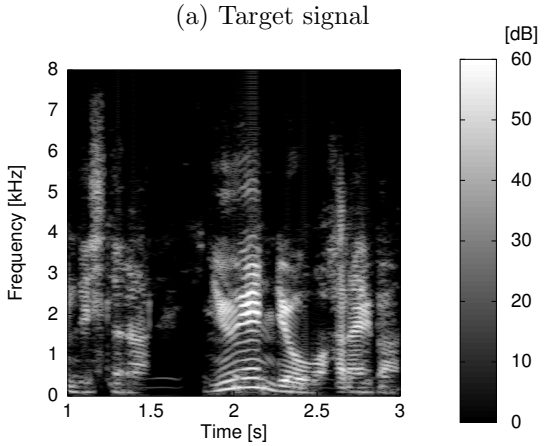
[6] M. Tomita, S. Saon, Y. Chisaki, T. Usagawa, “Quantitative evaluation of segregated signal with frequency domain binaural model,” *Acoust. Sci. & Tech.* (in press).

[7] Bill Gardner and Keith Martin, “HRTF measurements of a KEMAR dummy head microphone,” *MIT Media Lab Perceptual Computing Technical Report*, #280 (1994).

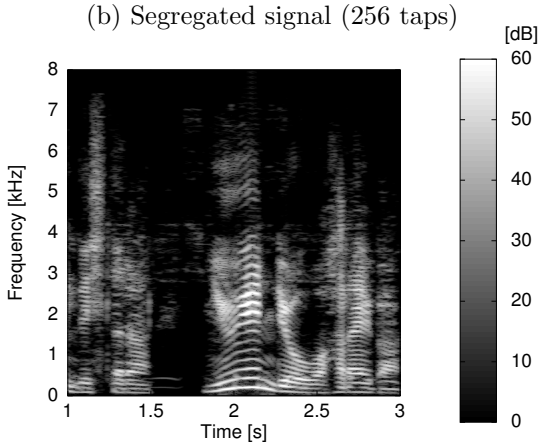
[8] 小林哲則, 板橋秀一, 速水悟, 竹澤寿幸, “日本音響学会研究用連続音声データベース,” 音響学会誌, **48**, 888-893 (1992).



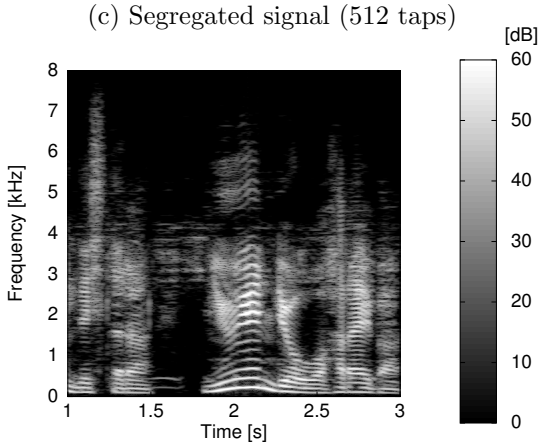
(a) Target signal



(b) Segregated signal (256 taps)



(c) Segregated signal (512 taps)



(d) Segregated signal (1024 taps)

Fig. 4　Spectrograms of the target and segregated signals at the left channel.

発声障害音声復元のための劣化音声特徴量推定とその補正手法に関する研究*

横田 豊和　坂田 聡 上田 裕市 (熊本大院 自然科学研)

1　はじめに

音声は，情報を伝える手段として人間が持っている基本的なものの中でも，重要性が高い．しかし，発声障害者は，発声器官をうまく制御できないために自発話音声 (劣化音声) に明瞭度低下等の声質劣化が生じ，正常な音声コミュニケーションを行うことができない場合がある．しかしながら，そのような劣化した音声でも発話者の意図する音韻情報が残存している場合には，それらを用いることで劣化音声を復元可能であると考えられる．先に，復元方式として，歪みや変動性で特徴づけられる音声パラメータ群を忠実に推定して目的に合わせた補正・変換処理の後，フォルマント合成を行う方式を提案した [1]．さらに，その応用として，構音訓練現場での効率的な訓練のために，本方式を自発話音声に適用して，正常構音方向に変形した目標音声を生成する手法を提案した [2]．

本稿では，これまで提案した手法を統合した発声障害音声復元のための分析・補正手法について述べる．

2　障害音声の分析

劣化音声の音声試料として，口蓋裂患者音声 (M:1 名・F:1 名・男児:1 名, 計 3 名)，顎変形患者音声 (F:1 名)，及び嗄声 (M:3 名 (それぞれ粗ぞう性, 無力性, 努力性)・F:1 名 (気息性), 計 4 名) の各 5 母音を用いる (M,F は，それぞれ成人男性および成人女性を示す)．

2.1　IFC 法によるフォルマント推定

音声分析におけるフォルマント推定には逆フィルタ制御法 (IFC 法)[3] を用いる．IFC 法は，スペクトルマッチングに基づく誤差最小化規準ではなく，逆フィルタ出力波形の零交差数分布の荷重平均に基づく推定であるため，スペクトル形状に依存せず，高精度で安定したフォルマント周波数を推定することができる．ピッチ (F_0) は，フォルマント推定値が零点に対応する可変逆フィルタ群に原信号を通すことで声帯音源微分波 (DGLW) を得て，その自己相関関数のピークピッキングにより推定する．

2.2　障害音声における分析の頑健性

IFC 法は健常音声の分析には有用であることが示されているが，本研究では分析対象が障害音声であるため，その有用性を検証する必要がある．Fig. 1 に口蓋裂音声 (M) 連続母音/aiueo/ の音声波形，10ms 周期の FFT スペクトル遷移，及び IFC 法で抽出したフォルマント周波数軌跡 (F_1〜F_4) を示す．F_1〜F_4 がスペクトルピークを正しくトラッキングしていることから，IFC 法は障害音声のフォルマントトラッキングにも有用であると言える．

2.3　母音発声における音声劣化の特徴

口蓋裂音声 (M), 顎変形音声 (F), 嗄声 (M : 粗ぞう性) についてのピッチ・フォルマント分析結果をそれぞれ Fig. 2(a)〜(c) に示す．(d) には比較のため健常話者音声の分析結果を示す．各図は上から音声波形, ソナグラム上のフォルマント軌跡, ピッチ軌跡である．(a) の母音/i/の F_3 や母音/o/の F_2，また，(c) の

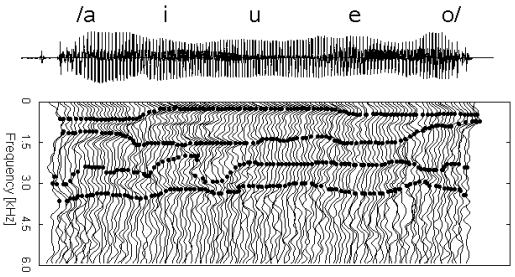


Fig. 1　Formant tracking based on the IFC method.

* A study of estimation of speech features and those correction method for restoring the degraded speech. by YOKOTA,Toyokazu, SAKATA,Tadashi, and UEDA,Yûichi(Graduate School of Science and Techunology, Kumamoto University)

母音/i/の F_0 でそれぞれ大きな変動が確認できる．

次に，F_1, F_2 の $F_1 - F_2$ 分布を Fig. 3 に示す．図中の直線は各母音の平均値を結んだ”日本語母音五角形”を表している．(a)〜(c) では，音韻性の決定に重要な母音五角形の形状が，健常話者 (d) と比較して歪んでいることが確認できる．健常話者音声の母音五角形を母音バランス基準とした場合，形状が歪んでいることは，いずれかの母音の構音異常による歪みがあると考えることができる．他の音声試料でも同様の特徴が見られたことから，音声劣化の特徴として：

1 フォルマントやピッチ軌跡の大きな変動
2 母音バランスの歪み (音韻性劣化)

を挙げることができる．

3 音声特徴補正・再合成手法

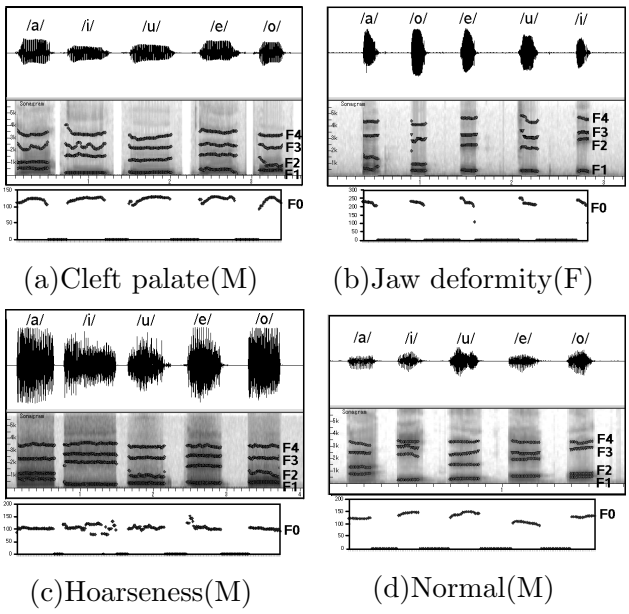
2.3 節で述べたように，障害音声の音声特徴量 (F_0, F_1〜F_4) は，その軌跡に大きな変動 (上記 1) が見られ，そのまま再合成に用いると音質が低下する．そのため，変動補正によって変動の軽減を検討する．この補正で除去不可能な傾斜に対して，傾斜補正により更なる変動の軽減を検討する．また，音韻性劣化と関係がある母音バランス歪み (上記 2) については，フォルマント偏り補正によって改善する．補正したパラメータの再合成にはフォルマント分析合成方式を用いる．

3.1 フォルマント分析合成方式

本方式は音源と声道の特性をそれぞれ独立に制御する方式で，音源として F_0 より生成した声帯音源微分波 ($OQ = 0.8, SQ = 2.3$) を，声道特性近似として F_1〜F_4 それぞれから生成したカスケードタイプの単共振フィルタ群を用いる (Fig.4)．

3.2 変動補正 (1)

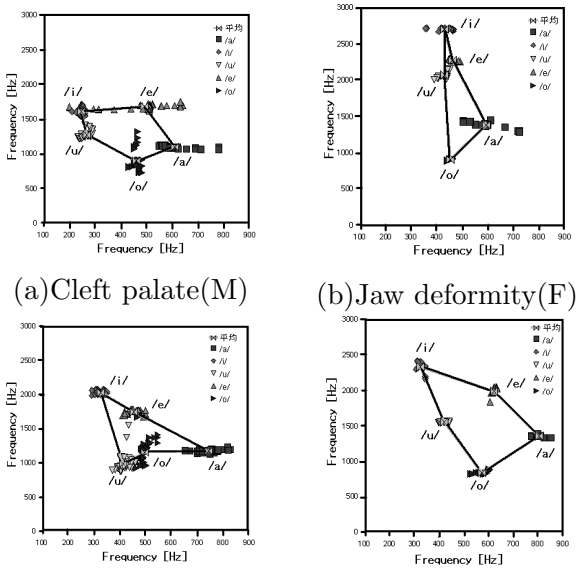
Fig. 2(a) の口蓋裂音声のフォルマント軌跡に対してハミング窓 (窓長 20) を用いて移動平均を行い，それらをパラメータとして再合成した結果を Fig. 5(a) に示す．この窓長は，ハミング窓の周波数特性と音声パラメータ軌跡のゆらぎの周波数成分を考慮して決定した．母音/i/ の F_3 において，補正前の大き



(a)Cleft palate(M) (b)Jaw deformity(F)

(c)Hoarseness(M) (d)Normal(M)

Fig. 2 Speech features extracted from the degraded Japanese five vowels.(M:male,F:female)



(a)Cleft palate(M) (b)Jaw deformity(F)

(c)Hoarseness(M) (d)Normal(M)

Fig. 3 F_1-F_2 scatter plots and averages of five vowels calculated by use of the extracted formant frequencies in Fig. 2.

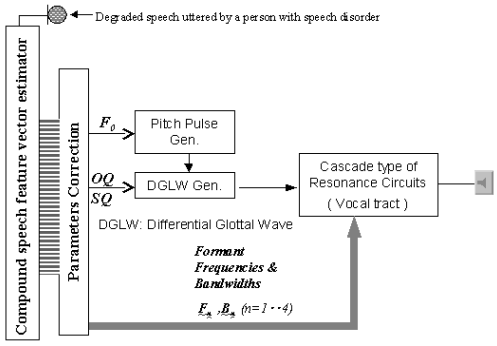


Fig. 4 System flow of formant-based synthesizer used in our study.

Table 1　Criteria of articulatory distortion(Normal:median(25%～75%)) and Mahalanobis' generalized distances of the degraded vowels.(mean(SD))

	/a/	/i/	/u/	/e/	/o/
Normal(M:male)	2.25(1.44～3.79)	2.24(1.15～3.83)	2.14(1.28～3.81)	1.92(0.79～4.13)	2.50(1.30～3.70)
Cleft palate(M)	5.7(0.8)	40.1(3.4)	6.5(2.0)	4.3(0.7)	11.6(2.0)
Hoarseness(M:Rough)	3.2(0.3)	9.2(1.5)	8.1(1.2)	4.6(1.7)	28.5(4.0)
Hoarseness(M:Asthenic)	37.7(4.2)	11.7(1.1)	14.7(2.3)	66.8(2.3)	54.4(8.8)
Hoarseness(M:Strained)	2.6(1.3)	30.6(2.1)	7.1(2.8)	12.3(4.2)	5.7(2.7)
Normal(F:female)	2.44(1.27～3.94)	2.32(1.65～3.54)	2.47(1.23～3.69)	2.35(1.11～3.94)	2.14(1.19～3.87)
Cleft palate(F)	2.6(0.6)	9.6(1.2)	7.6(0.9)	1.9(1.4)	77.5(3.5)
Jaw deformity(F)	12.0(4.4)	6.2(0.8)	10.2(2.8)	4.4(2.3)	9.5(2.4)
Hoarseness(F:Breathy)	2.5(0.3)	8.9(0.7)	0.6(0.3)	11.1(3.1)	23.4(2.7)
Cleft palate(male child)	32.5(1.5)	19.2(1.8)	19.1(1.8)	1.2(0.1)	55.4(3.0)

な変動 (Fig. 2(a)) を軽減できていることが確認できる．しかし，母音/o/ の F_2 では，移動平均で除去できない傾斜が残るために，再合成音声の聴覚的な歪みのすべてが改善できるわけではない．

3.3　傾斜補正 (1)

異常傾斜 (Fig. 5(a)) を除去するために，パラメータ軌跡の傾斜を健常者音声から算出した変動閾値 (標準偏差) 内におさえる．傾斜が閾値を超えている場合，パラメータ列の標準偏差が変動閾値と等しくなるよう，各要素毎に算出した値を乗算する．Fig. 5(a) の音声の F_1～F_4 に本補正手法を適用して再合成した結果を Fig. 5(b) に示す．母音/o/の F_2 に注目すると，異常傾斜が改善されていることが確認できる．その結果，再合成音声の聴覚的な歪みを改善することができた．

3.4　フォルマント偏り補正 (2)

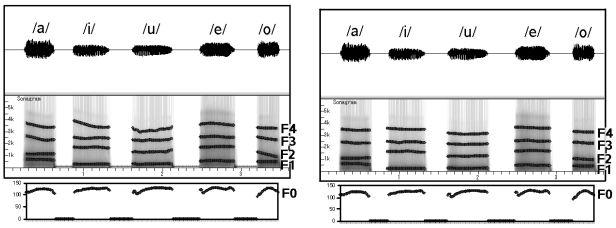
3.4.1 構音歪み判定

母音の構音歪み判定基準として，多数話者の 5 母音試料 (成人男女各 100 名の単母音) の F_1～F_3 から作成した健常構音データベース (DB) を用いた各母音群との距離ベースの母音バランス基準を用いる．健常者 DB の全話者のマハラノビス距離を算出し，男女別に母音毎の中央値と第 1, 第 3 四分位数 (25%～75%) を Table. 1(Normal) に示す．また，各障害音声試料の F_1～F_3 について算出した同様の値も Table. 1 に示す．

マハラノビス距離が自由度 3 のカイ二乗分布に従うことを利用して，検定により構音歪みを判定する．有意水準を 1%として判定を行った結果を Table. 2 に示す．表中の 0 は構音歪みがあることを示し，1 は正常であることを示す．Table. 2 において歪みがあると判定された母音と，Table. 1 においてマハラノビス距離が基準と比較して大きくなっている母音が対応していることが確認できる．

Table 2　Determination of articulatory distortion using Mahalanobis' generalized distances shown in Table.1(1:normal, 0:distortion)

	/a/	/i/	/u/	/e/	/o/
Cleft palate(M)	1	0	1	1	0
Hoarseness(M : Rough)	1	1	1	1	0
Hoarseness(M : Asthenic)	0	0	0	0	0
Hoarseness(M : Strained)	1	0	1	0	1
Cleft palate(F)	1	1	1	1	0
Jaw deformity(F)	0	1	1	1	1
Hoarseness(F : Breathy)	1	1	1	1	0
Cleft palate(male child)	0	0	0	1	0

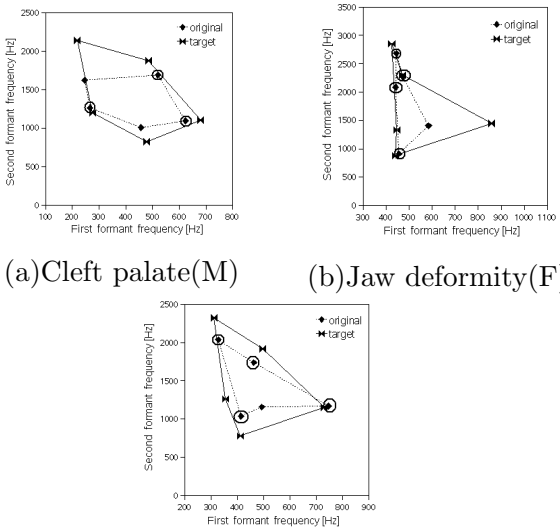


(a)Fluctuation correction　(b)Grade correction

Fig. 5 Results of speech features' correction.(M:Cleft palate)

3.4.2 フォルマント補正のための目標話者選択

異常と判定した母音についてフォルマント変換を行い，母音バランスを改善する．劣化音声の正常化のためのフォルマント変換時の目標フォルマント周波数の決定においては，原話者のフォルマント空間，すなわち声道長などの生理学的特徴を考慮する必要がある．そのため，検定により正常であると判定された母音を基準とし，DB の全話者について式 (1) に示すマハラノビス距離を母音毎に算出して，基準母音についての総和が最も小さい



(a)Cleft palate(M)　(b)Jaw deformity(F)

(c)Hoarseness(M)

Fig. 6　Formant frequencies of normal speakers as a target for correcting formant bias.

$$D_M^2 = (\boldsymbol{F}' - \boldsymbol{F})^T \Sigma^{-1} (\boldsymbol{F}' - \boldsymbol{F}) \quad (1)$$

$$where \quad \boldsymbol{F} = (F_1, F_2, F_3)^T : Original$$

$$\boldsymbol{F}' = (F_1', F_2', F_3')^T : Target$$

$$\Sigma : Covariance\ matrix$$

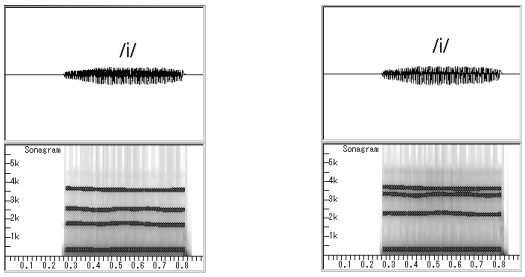
話者を目標話者として選択する．Fig. 2(a)～(c) の各音声について，原話者と選択された目標話者の母音五角形を Fig. 6(a)～(c) に示す．○印は基準として用いた母音である．

3.4.3 フォルマント変換

目標話者の F_1～F_4 と原話者のそれらとの差分を原パラメータ列に加算して，パラメータ列の平均値を目標話者のフォルマント周波数と等しくする．Fig. 2(a) の音声について，Fig. 6(a) に示す目標話者にフォルマント変換を行った結果を Fig. 7 に示す．フォルマントの変動は変わらずに，平均値のみ変化していることが確認できる．この補正により，音韻性を改善することができた．

4　まとめ

IFC 法の有用性を確認し，様々な症状の劣化音声を分析することにより，劣化音声にはフォルマントやピッチなどの音声特徴の軌跡や母音バランスに異常 (音韻性異常) があることを確認した．軌跡の異常に対しては，変動補正・傾斜補正により軽減でき，再合成音声の聴覚的な歪みを改善できた．また，母音バランスの異常に対しては，母音バランス基準に基づいた統計的な構音歪み判定を行い，マハラノビス距離を用いて選択した目標話者へ声質を考慮したフォルマント偏り補正を行うことで，音韻性を改善できた．



(a)No correction　(b)Correction

Fig. 7　Results of the formant bias corrections.

今後の課題は，フォルマント変換再合成音声の聴覚評価，及び合成方式としてスペクトルモーフィング [4] を用いた劣化音声復元を行うことである．

謝辞　本研究の一部は，平成 20 年度電気通信普及財団研究助成及び平成 21 年度科研費 (基盤研究 (C);No.21500476) の補助を受けた．研究遂行にあたり御助言いただいた平原講師・五味助教 (鹿児島大学大学院医歯学総合研究科) に感謝の意を表する．

参考文献

[1] 横田豊和 他，”発声障害音声復元のための劣化音声特徴量の検討”，電気関係学会九州支部連合大予稿集 (CD-ROM)，03-2P-04，2008

[2] 横田豊和 他，”構音訓練における目標音声提示のための劣化音声復元に関する検討”，電気関係学会九州支部連合大予稿集 (CD-ROM)，03-2P-03，2009

[3] Akira Watanabe，”Formant Estimation Method Using Inverse-Filter”，IEEE TRANS. on Speech and Audio PROC.，Vol.9，pp.314-326，2001

[4] 上田 裕市, 廣田 美菜子, ”スペクトルモーフィング手法に基づく母音合成と話者変換への応用”，電子情報通信学会論文誌，Vol. J90-D，No.5，2007

声門流の境界層解析と音源-フィルタ相互作用を考慮した音声生成モデル

大毛勝統, 鏑木時彦 (九州大院)

1 はじめに

発声時の声門流の挙動を詳細に記述するためには, 高レイノルズ数流れである声門流に対して, 速度の急勾配を伴う境界層の影響や, 境界層の剥離などの解析を行なう必要がある。本研究では, 声門流を非粘性の主流と粘性を考慮した境界層によって近似し, その相互作用を考慮した解析を行う。また, 声門流のモデル化に加え, 音源-声道フィルタ間の相互作用を考慮するため, Sondhi ら [1] による音響管モデルを用いる。ここでは声道・声門・声門下をひと連なりの音響管とみなし, 声門間音圧差や, 声帯にかかる平均音圧を求め, 音源機構への声道の音響作用を考慮した上で, 音声の生成を行なう。

2 境界層解析と音源-フィルタ相互作用を考慮した音声生成モデル

肺からの呼気圧によって生じる声門流は, その時間変動の形で音波となり音声を形作る。よって, この声門流を精緻に解析することは, 音声の生成過程を理解する上で重要な意味を持つ。声門流の生成には呼気流の流体力学的な挙動と, 声帯の動力学的な運動とが関与する。空気の粘性によって, 声帯壁面付近に粘性流れである境界層が発達し, 流れが通過する実効的な声門面積を減少させる。また, 流れの剥離が声帯壁面のどこで生じるかによって, 声門流量の推定に大きな違いが生じることから, 任意の声門形状において適応的に剥離位置を推定することは, 声門流のモデル化において必要不可欠である。

一方で, 声道・声門下の音響的なフィードバックが音源機構に与える影響も無視できない。Titze[2] は, 声道・声門下の音響特性が, 声帯の自励振動を抑制したり, 逆に促進させる働きを持つことを示した。

これらの点を踏まえ, 本論では Fig.1 のような音声生成モデルを構築した。以下, 声門流の粘性の影響や, 流れの剥離位置を考慮した境界層解析法 (2.1 節) と, 声道フィルタの音源機構に対する音響的な寄与を考慮するための声門間圧力差・声門間平均音圧の考え方

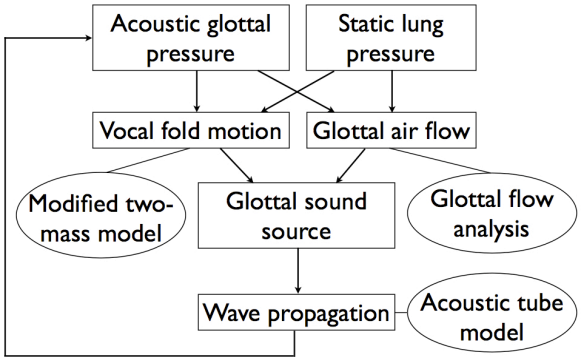


Fig. 1 音声生成モデル概要

(2.2 節) について説明する。また 2.3 節では, 声帯閉鎖前後での, 声門流全体が粘性流であるポアズイユ流れになった場合の解析方法を説明する。

2.1 粘性-非粘性の相互作用を考慮した境界層解析

音声の音源を成す声門流の解析においては, 流れのレイノルズ数や境界層厚さを考慮する必要がある。ここでは境界層解析の方法と, 主流と境界層の間に存在する相互作用について説明する。実際の声門幅から境界層の厚さ $\delta(x)$ を排除した実効的な声門幅より, 非粘性主流部での実効的な速度は,

$$v(x) = \frac{u_g}{(h(x) \quad 2\delta(x))l_g} \tag{1}$$

となる。ここで u_g は声門流量, $h(x)$ は声門の開口幅, l_g は声門の長さである。x 軸は声門の対称軸に取る。δ や剥離位置を与える境界層解析は, Kármán-Pohlhausen による運動量積分方程式に基づいている。

$$\frac{d}{dx}\{v(x)^2 \ (x)\} + \delta(x)v(x)\frac{d}{dx}v(x) = \frac{\tau(x)}{\rho} \tag{2}$$

ここで, (x) は運動量厚さ, $\tau(x)$ は壁面での粘性応力, ρ は密度である。$v(x)$ は境界層外縁での主流速度であることから, 主流（非粘性）と境界層（粘性）とは相互に依存関係に

Speech Production Model considering Boundary-layer Analysis and Source-filter Interaction by DAIMO, Katsunori and KABURAGI, Tokihiko (Kyushu University)

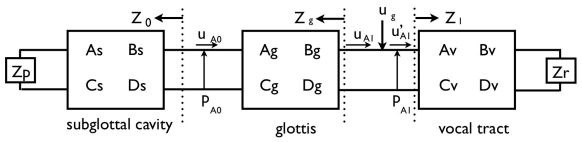


Fig. 2　発話系の音響管モデル

ある。この境界層方程式は, 次のように書き換えられる。[3]

$$v(x)\frac{\delta(x)}{\nu}\frac{d}{dx}\frac{\delta(x)}{H(x)}+\left(1+\frac{2}{H(x)}\right)F_1(H(x)) = H(x)F_2(H(x)) \quad (3)$$

$$\frac{\delta(x)^2}{\nu}\frac{dv(x)}{dx}=F_1(H(x)) \quad (4)$$

ここで $F_1(H(x))$ = 2.4{1 exp(0.43(2.59 $H(x)$))}, $F_2(H(x))$ = $4/H(x)^2$ $1/H(x), H(x) = \delta(x)/$ (x) である。流れの剥離位置は壁面での粘性応力が零になる地点, つまり $F_2 = 0, H = 4$ になる位置を探索することで求められる。[3]

しかしながら, 声門閉鎖の前後では, 流れのレイノルズ数が低下し, 境界層近似が成り立たなくなるため, このような境界層解析は適用できない。よってこの場合においては, 声門流全体を粘性流であるポアズイユ流れと捉える。この場合の解析法は 2.3 節で詳述する。

2.2　音源-フィルタ相互作用の考慮

2.2.1　声門間音圧差・声門間平均音圧の推定

本研究では音源-フィルタ相互作用を考慮する上で, 音響的な声門間音圧差と平均音圧を導入する。

$$\Delta p_A = p_{A0} \quad p_{A1} \quad (5)$$

$$\overline{p_A}=\frac{p_{A0}+p_{A1}}{2} \quad (6)$$

ここで, p_{A0}, p_{A1} はそれぞれ時間領域における声門下部・上部での音響的な圧力を表す。Δp_A は声門流量の推定 (2.2.2 節) に用い, $\overline{p_A}$ は声帯の駆動力算出 (2.2.3 節) に用いる。

これを考えるために, Sondhi ら [1] による音響管モデルを用いて, 声道・声門・声門下をひと連なりの音響管とみなす (Fig.2)。声門をひとつの均一音響管で近似することによって, 周波数領域において次式が成り立つ。

$$\begin{pmatrix} P_{A1} \\ U_{A1} \end{pmatrix} = \begin{pmatrix} A_g & B_g \\ C_g & D_g \end{pmatrix} \begin{pmatrix} P_{A0} \\ U_{A0} \end{pmatrix} \quad (7)$$

ここで, 声門の伝搬行列を表す要素は $A_g = \cosh(\sigma l_g/c), B_g = \ (\sigma c/S_g)\gamma\sinh(\sigma l_g/c), C_g = \ (S_g/\sigma c)(\sinh(\sigma l_g/c))/\gamma, \ D_g = \cosh(\sigma l_g/c)$ と表せる。これらの関係を用いると, 周波数領域での声門間音圧差・声門間平均音圧は以下の形で与えられる。[4]

$$\frac{\Delta P_A}{U_g}=\frac{P_{A0} \quad P_{A1}}{U_g}=\frac{\{B_g \quad (A_g \quad 1)Z_0\}Z_1}{Z_D} \quad (8)$$

$$\frac{\overline{P_A}}{U_g}=\frac{1}{2}\frac{P_{A0}+P_{A1}}{U_g}= \quad \frac{\{B_g \quad (A_g+1)Z_0\}Z_1}{2Z_D} \quad (9)$$

ここで, $Z_D = (D_g \quad C_gZ_0)Z_1 \quad (B_g \quad A_gZ_0)$, Z_0, Z_1 はそれぞれ声門下・声道の入力インピーダンスであり, 以下のように与えられる。

$$Z_0=\frac{P_{A0}}{U_{A0}}= \quad \frac{A_sZ_p+B_s}{C_sZ_p+D_s} \quad (10)$$

$$Z_1=\frac{P_{A1}}{U'_{A1}}=\frac{D_vZ_r \quad B_v}{A_v \quad C_vZ_r} \quad (11)$$

ただし, Z_p は肺の終端インピーダンス, Z_r は口唇での放射インピーダンスを表す。

2.2.2　声門流量の推定

次に, 声門間音圧差を用いた声門流量の推定法について説明する。非粘性流れである主流において, 流れの剥離位置での圧力が大気圧に等しいとすると, ベルヌーイの定理より声門入口と流れの剥離位置の間の圧力差 Δp は,

$$\Delta p=\frac{1}{2}\rho\left(\frac{u_g(t)}{S_s}\right)^2 \quad (12)$$

と表せる。ここで, S_s は剥離位置における声門面積である。また, (8) 式の $\Delta P_A/U_g$ の逆フーリエ変換を $z_\Delta(t)$ とし, これを用いて先ほどの声門間圧力差 Δp を表すと,

$$\Delta p = p_{F0} + z_\Delta(t) * u_g(t) \quad (13)$$

となる。ただし, p_{F0} は静的な声門下圧を表す。これら (12) 式と (13) 式を連立させて離散時間表現すると,

$$p_{F0}+\sum_{k=0}^{K \ 1} z_\Delta(k)u_g(n \quad k)=\frac{\rho u_g(n)^2}{2S_s^2} \quad (14)$$

となる。(14) 式を u_g について解くことで,

$$u_g(n)=\frac{z_\Delta(0)S_s^2+S_s\sqrt{(z_\Delta(0)S_s)^2+2\rho P}}{\rho} \quad (15)$$

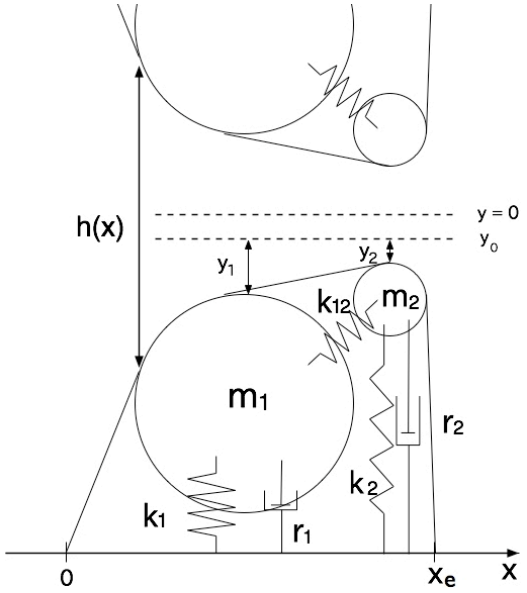


Fig. 3　声帯の２質量モデル

と u_g を推定できる。ただし, $P = p_{F0} + \sum_{k=1}^{K\ 1} z_\Delta(k)u_g(n \quad k)$ である。K は z_Δ の長さとする。

2.2.3　声帯の駆動力

続いて, 声門間平均音圧を用いた声帯の駆動力の求め方について説明する。本論では, 声帯の時間的な動きを表現するために, バネ-質量で表された声帯の機械モデル（Fig.3）を用いる。境界層解析を安定に行なうためには, 声門形状に不連続が生じないようにする必要があるため, 声門形状を円弧とそれを繋ぐ線分で近似した Pelorson の２質量モデル [5] を採用した。各質量の支配運動方程式は, 以下のように与えられる。

$$m_1\frac{d^2y_1}{dt^2} + r_1\frac{dy_1}{dt} + k_1y_1 + k_{12}(y_1 \quad y_2) = f_1 \tag{16}$$

$$m_2\frac{d^2y_2}{dt^2} + r_2\frac{dy_2}{dt} + k_2y_2 + k_{12}(y_2 \quad y_1) = f_2 \tag{17}$$

ここで, m_1 は声門下側の質量, m_2 は声門上側の質量を指し, y_1, y_2 は各質量の平衡位置からの変位を表す。Pelorson[5] に従い, 声門上側の質量には駆動力は働かないとした。声門下側の質量に働く駆動力は以下に説明する手順で求める。

声帯表面に働く圧力分布 $p(x)$ は, 流体的圧力分布に音響的な圧力を含め,

$$p(x) = p_{F0} \quad \frac{1}{2}\rho v(x)^2 + z_M(t) * u_g(t) \tag{18}$$

とする。ここで, $z_M(t)$ は (8) 式の $\overline{P_A}/U_g$ の逆フーリエ変換である。よって, 声門下側の質量に働く駆動力 f_1 は,

$$f_1 = l_g\int_{x_0}^{x_s} p(x)dx \tag{19}$$

となる。ここで, x_s は剥離位置の x 座標である。また, 声門が完全に閉鎖している場合の駆動力は,

$$f_1 = \lambda l_g\{p_{F0} + z_0(t) * u_g(t)\} \tag{20}$$

とする。ここで, z_0 は声門から声門下部を見た入力インピーダンスの逆フーリエ変換, λ は,圧力が働く実効的な声帯の長さを表し, ここでは $\lambda = 0.25$ とした。

2.3　低レイノルズ数の場合の解析法

2.1 で述べたように, 声門閉鎖前後では流れのレイノルズ数が低下し, 境界層近似が成り立たなくなる。よってこの場合は, 声門流全体を粘性流れであるポアズイユ流と捉える。このときの声門流量の推定法, 声帯駆動力の算出法は以下の通りである。流れの声門間圧力差 Δp は,

$$\Delta p = \frac{12\mu}{l_g}u_g\int_{x_0}^{x_e}\frac{1}{h(x)^3}dx \tag{21}$$

で表される [4]。 x_e は声門出口, x_0 はよどみ点を表す。これに音響成分を含めて離散時間で表現すると,

$$p_{F0}+\sum_{k=0}^{K\ 1} z_\Delta(k)u_g(n \quad k) = \frac{12\mu}{l_g}u_g\int_{x_0}^{x_e}\frac{1}{h(x)^3}dx \tag{22}$$

となることより, 声門流量は,

$$u_g(n) = \frac{p_{F0} + \sum_{k=1}^{K\ 1} z_\Delta(k)u_g(n \quad k)}{\frac{12\mu}{l_g}\int_{x_0}^{x_e}\frac{1}{h(x)^3}dx \quad z_\Delta(0)} \tag{23}$$

と推定される。また声帯表面の圧力分布 $p(x)$ は,

$$p(x) = p_{F0} \quad \frac{12\mu}{l_g}u_g\int_{x_0}^{x}\frac{1}{h(\chi)^3}d\chi + z_M(t) * u_g(t) \tag{24}$$

と表すことができる。よって, 声門下側の質量に働く駆動力は,

$$f_1 = l_g\int_{x_0}^{x_e} p(x)dx \tag{25}$$

となる。ここで, x_e は声門出口の x 座標である。

3　シミュレーション手順と結果

声帯質量の位置, 静的な声門下圧を与えると声門面積が決まり, 声道や声門下の断面積が分かっている場合, 発話系全体の音響特性が決まる。流れのレイノルズ数がある閾値よりも高い場合は境界層解析を適用し, (15) 式より流量を推定し, (19) 式で声帯駆動力を求める。レイノルズ数が閾値よりも低い場合は, 流れをポアズイユ流れとみなし, (23) 式で流量, (25) 式で声帯駆動力を求める。声門が閉鎖している場合は, 声門流量は 0 とし, 声帯駆動力は (20) 式より求める。声帯モデル [5] に従って声帯の運動方程式を 4 次精度の Runge-Kutta 法で解き, 次のステップの声帯変位を求めることによって, シミュレーションを進めて行く。

Fig.4 は母音/a/の声道断面積を用いた際の定常部のシミュレーション結果である。サンプリング周波数は 20 kHz とした。 y_1, y_2 は実線が声門下部の質量, 点線が声門上部の質量の変位を表す。u_g は声門流量, p_Δ は声門間音圧差, p_M は声門間平均音圧差, $speech$ は口唇からの放射音圧を表している。音声の基本周波数は 128 [Hz] となった。声門開大期では, 声門間音圧差は負に減少している。これによって声門体積流の生成を阻害するような働きが起こり, 結果として波形の立ち上がりが遅くなる。一方で声門が閉じようとする際は, 声門間音圧差が正まで増加する。これにより波形が膨らみ, 全体としてみると若干右に傾いたような波形になる。声門間平均音圧は声帯が開く際は正の値になり, 声帯を押し開く働きをする。一方で, 閉じる際には負圧となり, 声帯を引き寄せる働きをする。これは, 声門間平均音圧が, 声帯の自励振動を維持させるような働きをしていることを意味する。

4　まとめ

本論では, 1 次元声門流の境界層解析とともに, 音源-フィルタ間の相互作用を考慮した音声生成のシミュレーションモデルを構築した。音源-フィルタの相互作用を考慮するために, 発話系全体の音響管モデルを考え, 声門間音圧差・声門間平均音圧を導入した。その結果, 音源-フィルタ相互作用によって引き起こされる体積流量の傾きが, 声門間音圧差の働きによって説明でき, また声門間平均音圧が, 声帯の自励振動を促す働きをしていることがわかった。今後の検討事項としては, 音源-声道フィルタ相互作用の発声効率への影響、声帯の緊張パラメータと音声の基本周波数との間にある非線形性、等を中心に検討していく予定である。

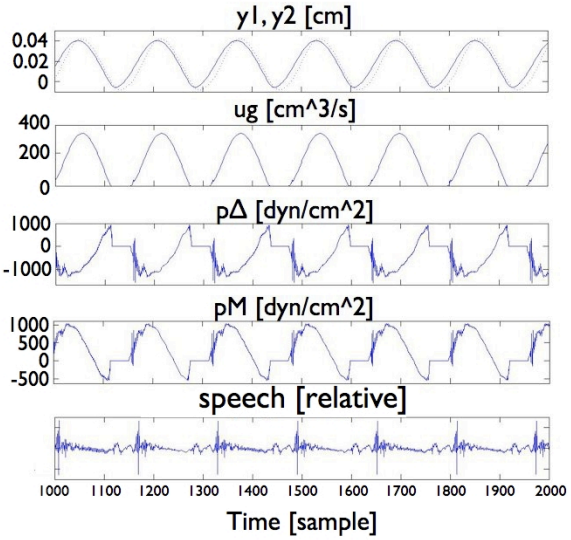


Fig. 4　母音/a/における計算結果

参考文献

[1] Sondhi, M. and Schroeter, J., "A hybrid time-frequency domain articulatory speech synthesizer", IEEE Trans. Acoust.,Speech and Signal Process.,ASSP-35(7):955-967,1987.

[2] Titze,I.R., "Nonlinear source-filter coupling in phonation;Theory", J.Acoust.Soc.Am.,123(5):2733-2749,2008.

[3] Kaburagi T., "On the viscous-inviscid interaction of the flow passing through the glottis", Acouct.Sci.&Tech.,29(2):391-404,2009.

[4] 鏑木, 「発声における音源-フィルタ相互作用に関する検討」,2009 年秋季音講論集,3-2-2,2009

[5] Pelorson, Hirschberg, van Hassel, Wijnands, Auregan, "Theoretical and experimental study of quasisteady-flow seperation within the glottis during phonation. Application to a modified two-mass model", J.Acoust.Soc.Am.,96(6):3416-3431,1994.

GPUを用いた流体ー構造体連成解析法の構築とその音声生成シミュレーションへの適用*

◎山本和彦, 鏑木時彦 (九州大)

1 背景と目的

有声音の音源波である声門波は、弾性体である声帯と流体である呼気流の複雑な相互作用によって生成される。左右の声帯は呼気流によって駆動され大変形を伴った運動をし、またそれによって呼気流の流路が動的に変形する。さらに、声門流は高レイノルズ数の流れであり、また声帯間の衝突も考慮する必要がある。従って、声門波の生成の過程をシミュレーションすることは最も難しいマルチフィジックス問題の一つであると言える。

本研究では, 声門波生成の過程を直接的にシミュレートするために、声帯を粘性抵抗を含む直交異方性弾性体として、声門流を圧縮性熱流体としてモデル化し、それぞれラグランジュスキームである MPS(Moving Particles Semi-Implicit) 法、オイラースキームである FDLB(Finite Difference Lattice Boltzmann) 法を用いて支配方程式を離散化したのちに、両者を連成させる。さらにスーパーコンピュータや PC クラスタを用いずに一台の一般的なパソコンでインタラクティブに実行するため、計算のほとんどを Shader(NVIDIA Cg, GLSL, and NVIDIA CUDA) を用いて GPU(Graphics Processing Unit) 上に効果的に実装し、加えて同時に、残りの計算も CPU 側で OS 依存スレッド関数を用いて実行しその計算もまた OpenMP を用いて並列化する。

2 流体—構造体連成解析手法

ここでは、まず、一般的に流体ー構造体の動的な連成運動解析を実現する方法を構築する。

2.1 構造体解析手法

MPS(Moving Particles Semi-Implicit) 法は本来非圧縮性流体の解析手法として越塚ら [1] によって提案された粒子ベースの計算手法である。後にこの手法は Song ら [2] によって等方性弾性体のシミュレーションへ拡張された。本研究では以下の直交異方性弾性体の運動方程式に MPS 法を適用し、構造体の運動を表現する。

$$\rho\frac{\partial^2}{\partial t^2}\vec{\Psi} = \Delta\, E\vec{\Psi} + \nabla(divE\vec{\Psi} - \nabla E\vec{\Psi}) + \eta\Delta\frac{\partial}{\partial t}\vec{\Psi} \tag{1}$$

ここで、$\vec{\Psi}$ は変位、 ρ, E, η はそれぞれ密度、弾性テンソル、粘性率である。弾性テンソルは、各軸方向のヤング率とポアソン比を用いて表される。MPS 法では式 (1) の微分演算子を対応する以下の粒子間相互作用モデルに置き換えることによって離散化を行う。

$$\Delta\phi_i = \frac{2d}{\kappa n_0}\sum_{j\neq i}\Big((\phi_j-\phi_i)w(|\mathbf{r}_j-\mathbf{r}_i|)\Big) \tag{2}$$

$$\nabla\phi_i = \frac{d}{n_0}\sum_{j\neq i}\left(\frac{\phi_j-\phi_i}{|\mathbf{r}_j-\mathbf{r}_i|^2}(\mathbf{r}_j-\mathbf{r}_i)w(|\mathbf{r}_j-\mathbf{r}_i|)\right) \tag{3}$$

$$\nabla\cdot\mathbf{u}_i = \frac{d}{n_0}\sum_{j\neq i}\left(\frac{(\mathbf{u}_j-\mathbf{u}_i)\cdot(\mathbf{r}_j-\mathbf{r}_i)}{|\mathbf{r}_j-\mathbf{r}_i|^2}w(|\mathbf{r}_j-\mathbf{r}_i|)\right) \tag{4}$$

上記の 3 つのモデルはそれぞれ上から、Laplacian、Gradient、Divergence の微分演算子に対応している。ここで、κ,、d、n_0 はそれぞれ発散をおさえるための定数、次元数、初期粒子数密度である。w は重み関数と呼ばれ、

$$w(r) = \begin{cases} \dfrac{r_e}{r} - 1 & (0 \le r \le r_e) \\ 0 & (r_e < r) \end{cases} \tag{5}$$

で表される。ここで、r は 2 つの粒子の距離、 i と j は粒子のインデックス、r_e は他の粒子の影響を考慮する影響半径である。

2.2 流体解析手法

流体に関しては圧縮性熱流体としてモデル化し、FDLB(Finite Difference Lattice Boltzmann) 法-D3Q39Model を用いて離散化する。FDLB 法は Lattice Boltzmann 法に差分スキームを導入した手法で、ここでは、2 段階 Runge-Kutta 法を時間差分に、3 次の風上差分を空間差分に用いた。粒子速度の格子点上でのアンサンブル平均である局所速度分布関数 f_i の時間発展式は

---

* Construction of Solid-Fluid Coupling Method using GPU and Apply to the Simulation for the Process of Glottal Wave Generation
by Kazuhiko Yamamoto, Tokihiko Kaburagi (Kyushu University)

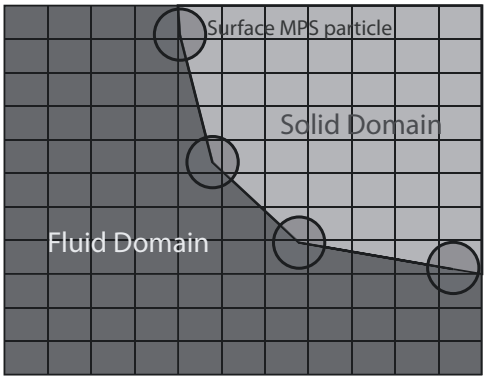


Fig. 1　流体と構造体の境界の表現方法.
構造体表面の粒子群から三角形の集合で表される境界表面を抽出する。粒子は一様な格子で表される流体領域の内部を自由に移動するのでこの境界表面は多くの場合、格子点上には存在しない。

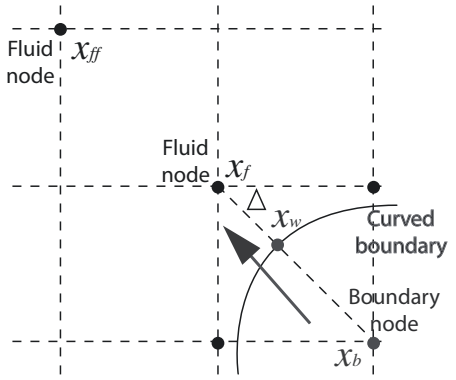


Fig. 2　流体領域での移動曲面境界の扱い.
流体格子の中間に位置するような境界でも、境界を挟んだ２点での格子にて境界条件を適用することによって連続的に取り扱うことが可能である。

$$\frac{\partial f_i(t,r)}{\partial t}+\frac{\partial}{\partial r_\alpha}c_{i\alpha}f_i(t,r)-\frac{Ac_{i\alpha}}{\tau}\frac{\partial(f_i-f_i^{(0)})}{\partial r_\alpha} = -\frac{1}{\tau}[f_i(t,r)-f_i^{(0)}(t,r)] \quad (6)$$

のように表される [4]。ここで、t、r はそれぞれ時間、空間座標、A は定数で、 添字 i、α はそれぞれ粒子の移動番号と方向、 $c_{i\alpha}$ は粒子の速度ベクトルである。また、τ は緩和時間で $f_i^{(0)}(t,r)$ は局所平衡分布関数と呼ばれ、圧縮性 Navier-Stokes 方程式を満たすように決定される。左辺第三項は高レイノルズ数流れを扱う場合の数値計算安定化のための負の粘性項 [3]、右辺は衝突項と呼ばれ、粒子同士が衝突することによって平衡状態へと近づいていく過程を表す。
局所平衡分布関数は最大 3 次の速度の多項式で記述され、

$$f_i^{(0)}=F_i\rho(1-2Bc_{i\alpha}u_\alpha+2B^2c_{i\alpha}c_{i\beta}u_\alpha u_\beta+Bu^2- 2B^2c_{i\alpha}u_\alpha u^2-\frac{4}{3}B^3c_{i\alpha}c_{i\beta}c_{i\gamma}u_\alpha u_\beta u_\gamma) \quad (7)$$

のように表される。ここで、ρ, u and e はそれぞれ、流体の密度、速度、内部エネルギーである。これらのパラメータは一つ前のタイムステップでの局所速度分布関数から計算される。係数 F と B に関しては内部エネルギーに依存して定められる。

2.3　構造体ー流体相互作用

流体と構造体の相互作用問題を解くために、ここでは FDLB 法と MPS 法を連成させる方法を提案する。まず、構造体の最も外側に位置する MPS 粒子を表面粒子と定義し、表面粒子を結んで三角形の集合で表される表面を抽出する (Fig.2)。この表面は FDLB 法で計算される流体領域における境界となる。流体領域において複雑な移動曲面境界を扱うため、ここでは Mei らの境界の扱い [4] に修正を加えることで FDLB 法の圧縮性熱流体モデルに適用する。この方法では、Fig.3 に示されているように、境界が曲面で構造格子の接点からずれていたり、シミュレーション中に移動する場合でも連続的に扱うことが可能となる。この方法では境界面を挟んだ 2 接点 x_f, x_b の局所速度分布関数を以下のように修正する。

$$f_i(t,x_b)=(1-z)f_i(t,x_f)+zf_i^*(t,x_b)+2B_q\rho e_i\cdot u_\omega \quad (8)$$

ここで f_i^* は仮想的な境界上での局所平衡分布関数、u_ω は境界の移動速度、z は境界面と粒子の速度ベクトルとの交差位置によって決定する変数である。また B_q は FDLB 法の局所平衡分布関数を導出する際に計算される係数、e_i は粒子速度単位ベクトルである。ここで、f_i^* の定義は以下のようになっている。

$$f_i^*=F_i\rho(A_q+B_qe_i\cdot u_{bf}+C_q(e_i\cdot u_f)^2+D_q(u_f)^2 +E_q(e_i\cdot u_{bf})(u_f)^2+F_q(e_i\cdot u_f)^3). \quad (9)$$

ここで、A_q から F_q は Eq.(7) と同じ係数、右辺の第三項は衝突項である。u_{bf} と z に関しては、
$\triangle > 1/2$ のとき、

$$u_{bf}=(1-\frac{3}{2\triangle})u_f+\frac{3}{2\triangle}u_\omega \quad and \quad z=\frac{2\triangle-1}{\tau+1/2} \quad (10)$$

$\triangle < 1/2$ のとき、

$$u_{bf}=u_{ff} \quad and \quad z=\frac{2\triangle-1}{\tau-2} \quad (11)$$

である。u_{bf} は仮想的な境界の移動速度を表しており、また、x_b は流体の速度を表している。u_f 、u_{ff}、

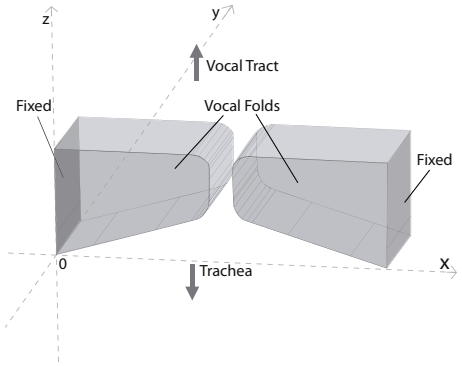


Fig. 3　声帯形状の作成.
声帯は2次元的な運動をするものとし、上図のような形状を用いた。

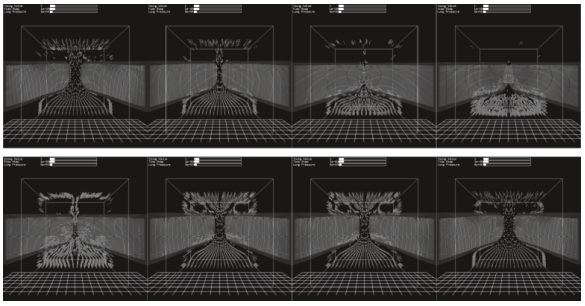


Fig. 4　声門波生成過程のシミュレーション結果.
声帯形状の時間変化と声門流のパーティクルトレースを表す。

u_w はそれぞれ x_f、x_{ff}、x_w の各節点での境界の移動速度である。
以上の境界条件は1時間刻みをさらに細かく分割して、全ての速度ベクトル (局所速度分布関数の分割ベクトル) が必ず一つ分の格子ごとに進むものとして適用していくことになる。

一方、構造体に作用する流体力は面に対して働くので、ここでは表面粒子1つ1つに働く力は表面力をその表面を構成する粒子に線形に分散させたものとする。まず、流体の離散化粒度に比べて大きな表面三角形の場合にはその三角形を細かく分割する。この分割された小三角形の重心位置の流体の圧力を小三角形の代表値とし、面積を掛けて小三角形にかかる力の大きさを求める。この小三角形からもともとの表面三角形を構成していた3つの粒子への寄与は粒子位置と三角形の重心までの距離とさらに直線を延ばして対辺とぶつかるまでの距離の比 w に応じて線形に分散させたものとして

$$F_{ia} = -(1-w)F_i(\vec{r})\vec{N} \tag{12}$$

と表すことができる。ここで $\vec{r}$ は表面上の位置ベクトル、$\vec{N}$ は表面の法線ベクトル、$F_i(\vec{r})$ は小三角形にかかる流体力の大きさである。また、表面上の任意の点における流体の物理変数の値は近傍の FDLB 構造格子の格子点における値をラグランジュ補間することによって求める。結果的にもとの三角形全体からの寄与は、再分割された全ての三角形からの力を足し合わせることで求めることができる。

2.4　実装

本研究では, 計算を一般的な PC でインタラクティブに実行可能にするために GPU(Graphics Processing Unit) を用いる。GPU は本来3 D グラフィックスの処理に特化して設計されたハードウェアであるが、近年その優れた並列処理能力を汎用計算にも利用する GPGPU(General Purpose Computation on GPUs) という分野が確立しつつある。GPU を汎用計算にも利用することは低コスト、高性能、将来性、一般性などの利点がある。本研究では, 計算を一般的な PC でインタラクティブに実行可能にするために GPU(Graphics Processing Unit) を用いる。ここでは MPS 法の計算を NVIDIA CUDA(Computed Unified Device Architecture) を用いて GPU 上に実装した。一方 FDLB 法の計算はボクセル化のテクニックを利用するために NVIDIA Cg を用いて GPU 上に実装した。Cg 計算部分から CUDA 計算部分へのデータの受け渡しには PBO(Pixel Buffer Object) を用いた。これにより、データをメインメモリにダウンロードすることによる転送時間をなくすことができる。FDLB 法において境界のボクセルを生成するために Wei ら [5] によって提案された動的境界生成法を利用する。この方法では深度剥離の考え方を応用し、半ば自動的に境界のボクセル化を行う。これによって僅かな前設定を行うことにより、障害物や壁境界をオフスクリーンレンダリングするだけで任意の境界条件を適用することが可能となる。加えて MPS 粒子同士の衝突検出と衝突力の計算は CPU 側に実装し、OS 依存スレッド関数を用いて GPU 側と同時に実行する。ここではさらに CPU 側の計算を OpenMP を用いて並列化している。

3　シミュレーション結果

以上の理論を用いて声帯を構造体、呼気流を流体として声帯振動のシミュレーションを行った。シミュレーションには声帯に約 60000 個の MPS 粒子、声門流に約 120000 構造メッシュから成る FDLB 法を用い、声帯形状には Fig.4 のような形状を用いた。実行環境はグラフィックボード1枚を搭載した1台の PC で行い、10 回の計算ループにつき1回の画面へのレンダリングを行った結果、約 60FPS の速度を得た。ここで、時間刻みは $10^{-5}[s]$ とした。正確なベンチマークは行っていないものの、CPU のみでの実装時に比べて 400 倍を超える高速化倍率が得られているものと考えられる。これによりインタラクティブに力学モ

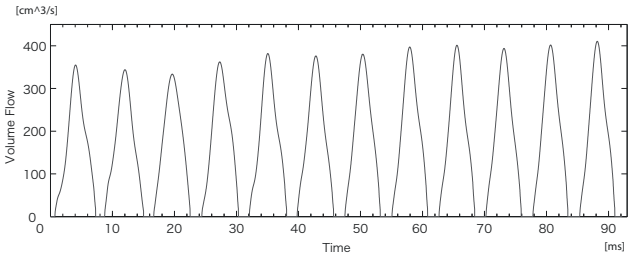


Fig. 5　生成された声門波.

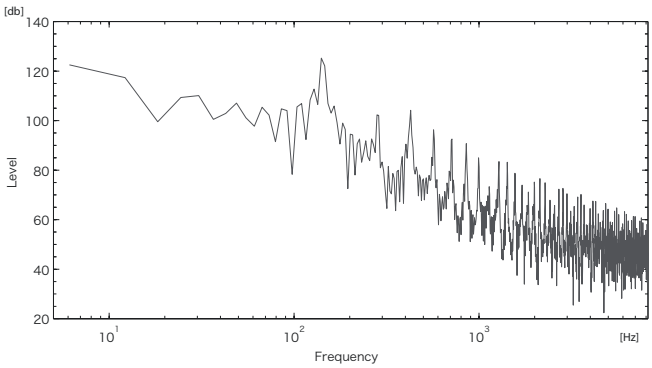


Fig. 6　生成された声門波の周波数特性.

デルのパラメータを操作しながらのシミュレーションが可能となった。さらにシミュレーションの結果は、三角波状の声門波や声帯上下間の運動の位相差など、過去の研究と定性的に良い一致が得られた。可視化されたシミュレーション結果を Fig.5、生成された声門流体積速度を Fig.6 に示す。Fig.6 では、横軸は時間 $[ms]$、縦軸は体積速度 $[cm^3/s]$ を表している。この図からは立ち上がりが急で、前方に傾きを伴った声門波独特の三角波上の波形が見て取れる。また、Fig.7 に生成された声門波の周波数特性を示す。Fig.7 では、横軸は対数スケールで周波数 $[Hz]$、縦軸は相対レベル $[dB]$ を表している。ここで、呼気流の流入条件は声門下圧 8000$[dyn/cm^2]$ を条件として局所平衡分布関数を設定することによって行い、結果的にレイノルズ数は最大 2500 から 3000 程となった。

4　まとめ

本研究では声門波の生成過程を直接的にシミュレートするために FDLB 法と MPS 法を連成する方法を提案した。その結果、大変形をする弾性体と高レイノルズ数の流れからなる複雑な連成物理問題をラグランジュスキームとオイラースキームを併用して計算する柔軟な手法を構築できた。さらに GPU と CPU 双方を用いて効果的な並列実装を行うことによって、従来莫大な計算時間を要したこの計算は、一般的なパソコンでもインタラクティブな速度で動作する。

文献

[1] Moving-Particles Semi-implicit Method for Fragmentation of Incompressible fluid, Kosizuka et al., Nucl. Sci. Eng., 1996, p421-434

[2] Dynamic Analysis of Elastic Solids by MPS Method, Song et al., Machine Society of Japan, 2005, p16-22

[3] Direct simulation of fluid dynamic sounds by the finite difference lattice Boltzmann method, Tshutahara et al, WIT Press, 2007, p3-12

[4] Lattice Boltzmann method for 3d-flows with curved bounndary, Mei et al., Journal of Comp. Phys, 2000, p680-699

[5] GPU-Based Flow Simulation with Complex Boundaries, Wei et al., GPU Gem 2, 2003, Chapter 47

マイクロホンアレーによる雑音除去音声の品質評価*

線形遅延和アレーの最適化の検討

◎吉國信太郎　水町光徳　二矢田勝行（九工大）

１．はじめに

先行研究[1]では、線形遅延和アレーにおける 8 本のマイクロホンを不等間隔で配置し、帯域に応じて使用するマイクロホンのペアを変更する方法を採用した 8ch 3^n 間隔 DS アレーを提案したところ、アレー全体としての SNR 改善率は 8ch 等間隔 DS アレーと比較して、同等かそれ以上の性能があることが分かった。しかし、各帯域で使用するマイクロホン数の低下により音声帯域における雑音除去性能が低下してしまうという問題があった。そこで本稿では、全ての帯域で常に 8 本のマイクロホンを使用し、帯域に応じてマイクロホン間の重みを調整することでアレーの最適化を図り、客観評価・主観評価ともに高性能なマイクロホンアレーと処理アルゴリズムの開発を目的とする。当面、アレーを大画面テレビ等に取り付けて、テレビ電話等での利用を想定しているので、目的方向は正面方向とし、アレーの総幅を約 90 cm としている。またインターネット回線を利用して G.722[2]形式で音声を伝送することを想定し、300 ～ 7000 Hz に帯域を制限する。

２．遅延和アレーの最適化

2.1　重みなし遅延和アレーの問題点

Fig.1 に、本研究で使用する 8ch 等間隔 DS アレー(DS アレー)、8ch 2^n 間隔 DS アレー(2^n アレー)[3]、8ch 3^n 間隔 DS アレー(3^n アレー)、のマイクロホン配置図を示す。次に、目的方向が 0° の場合の DS アレーの指向特性を Fig.2 に示す。理想的な指向特性は、Fig.2 の 0° 付近で全周波数に対して感度(Gain)が 0 dB となり、それ以外の方向では全周波数に対して-∞dB となるのが理想的である。しかし、そのような指向特性を持つアレーは有限個のマイクロホンでは実現不可能である。Fig.2 から分かるように、DS アレーでは、信号の入力角度が±30° より外側で、周波数が 3 kHz 以上の領域で空間エイリアシングが発生してしまい、目的方向以外の信号まで強調されてしまう。マイクロホン間隔を狭くすると、空間エイリアシングが発生する周波数は高くなるが、指向性が緩く(主ローブ幅が大きく)なってしまう。このように、エイリアシングが起こらない周波数と主ローブ幅の間にはトレードオフの関係が成り立ち、単一の等間隔 DS アレーではそれらを両立させることは難しい。

2.2　マイクロホン間の重み調整

Fig.3 にアレー間の重みを「11444411」とした場合、Fig.4 にアレー間の重みを「13322331」とした場合の 3^n 間隔 DS アレー(いずれも目的方向は 0°)の指向特性を示す。Fig.3 と Fig.4 から分かるように、マイクロホン間の重みを調整することにより、マイクロホンの配置を変更することなく、指向特性に変化を持たせることが可能となる。そこで本稿ではこれを上手く利用し、周波数帯域を多数に分割して、各々の帯域において各マイクロホン間の重みを調整することによって、所望の角度に死角が形成されるようにアレーの最適化をして、雑音除去性能の向上を図る。

Evaluation of speech distortion using noise-reduced speech
-Study on the optimization of channel-dependent weight for linear delay-and-sum beamformer -
by Shintaro YOSHIKUNI, Mitsunori MIZUMACHI, Katsuyuki NIYADA (Kyushu Inst.Tech)

2.3　最適化の手法

ここでは具体的なアレーの最適化の手法について述べる。まず、各マイクロホンの重みは最高値が 5 までの自然数とし、マイクロホン間の重みの組み合わせを指向特性が左右で線対称になるように注意しながらランダムに作成する。本稿では、重みが 2 つの数値(例: AAABBAAA, A は 1 固定, B は 2~5 で変動)となる組み合わせを 56 種類(14×4)、重みが 3 つの数値(例: ABBCCBBA, A/B/C:1~4 で変動, A≒B≒C)となる組み合わせを 26 種類(13×3)、重みが全て 1 のものが 1 種類の計 83 種類の組み合わせを作成した。重みの最高値を 5 までとしたのは、1 つのマイクロホンの重みが大きすぎると、重みが小さい他のマイクロホンを使用しない場合との変化が小さくなってしまうからである。次に、作成した全パターンの重みにおけるアレーの指向特性を計算する。そして、1 Hz ~ 8000 Hz を 100 Hz ごとに約 80 の帯域に分割し、各々の帯域で所望の角度においてゲインが最小となる重みの組み合わせを選択する。これを繰り返すことにより、所望の角度のみ全ての帯域でゲインが最小となるアレーが完成する。尚、マイクロホン間の重みの切り替え部分には周波数方向のスムージング処理を入れて、周波数方向に不連続にならないように考慮している。

ここで、死角を 70° で最適化した DS アレーの指向特性を Fig.5 に、2^n アレーの指向特性を Fig.6 に、3^n アレーの指向特性を Fig.7 に示す。Fig.5,6,7 から分かるように、Fig.3 や Fig.4 にみられた 70° 方向へのサイドローブの影響が小さくなっていることが予想できる。死角を 70° に設定した理由は、3^n アレーで死角の形成が一番上手くいっている角度であり、各アレーの指向特性を見比べた場合に違いがはっきりと分かるからである。以後は、最適化後のアレーの雑音除去性能を見ていくことにする。

3．雑音除去音声の品質評価

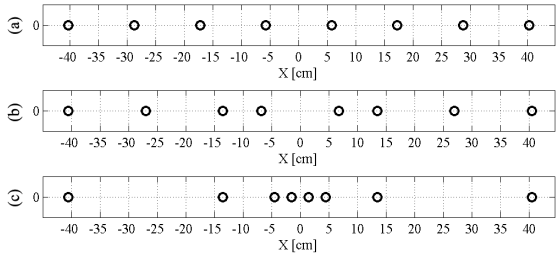
3.1　実験条件

最適化による雑音除去性能の向上具合を比較するため、**2.3** で示した方法により「簡易防音室(残響時間 0.4 s)」で最適化を行ったアレーを用いて、Tab.1 の実験条件の下、処理前の信号と各アレーで処理後の信号を客観的に評価する。ここでは、雑音方向は既知とし、雑音方向が死角となるように調整した。処理方法は、Fig.1 で示したマイクロホン配置による、Fig.2 の特性を有する最適化してない DS アレーによる処理、Fig.5～Fig.7 の特性を有する最適化した各アレーによる処理が 3 種類。更に先行研究と比較するため、帯域により使用するマイクロホンを切り替えるタイプの 8ch 3^n アレー(以後、旧 3^n アレー)による処理の計 5 種類である。評価方法は、G.722 に規格を合わせて 300 Hz ~ 7000 Hz に帯域を制限した信号対雑音比(SNR)を採用した。具体的には、各アレーでの処理後信号の SNR から処理前信号の SNR を引いた SNR 改善量を求めた。

3.2　実験結果 （客観評価 ：SNR）

雑音信号がホワイトノイズで観測信号の SNR が-15 dB の場合の各アレーによる SNR 改善量を Fig.8 に示す。尚、本稿では色の都合上、最適化してないアレーによる SNR 改善量は DS アレーのみを示す。Fig.8 より今回提案した方法により最適化した 3^n アレーは、全ての方向で「最適化していないアレー」「旧 3^n アレー」よりも高い雑音除去性能を持つことが分かった。

次に、全方向に対する平均的性能を比較する為、各アレーの SNR 改善量の全方向平均値を雑音の種類ごとに Fig.9 に示す。Fig.9 より様々な雑音信号に対して、今回提案した最適化アレーは「最適化していないアレー」「旧 3^n アレー」よりも全体的な雑音除去性能が高いことが分かった。



(a) DS アレー　(b) 2n アレー　(c) 3n アレー

Fig.1　線形遅延和アレーのマイクロホン配置

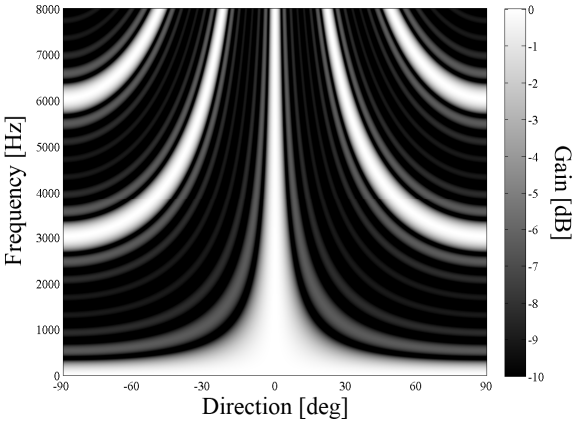


Fig.2 DS アレーの指向特性

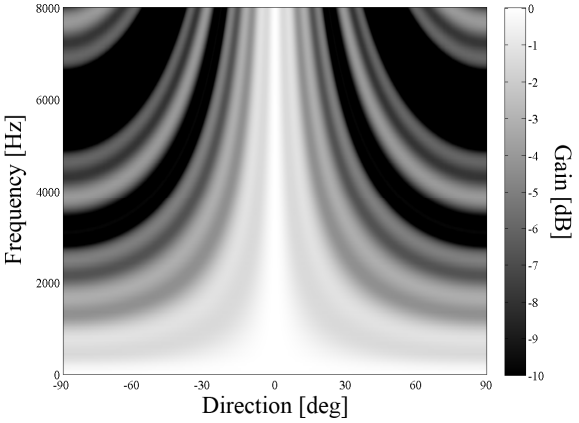


Fig.3 3^n アレー(重み:11444411)の指向特性

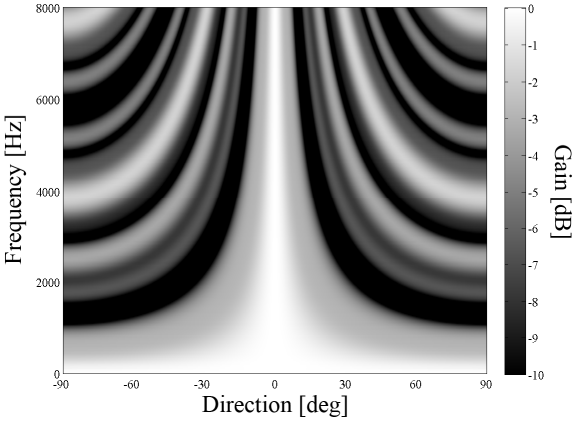


Fig.4 3^n アレー(重み:13322331)の指向特性

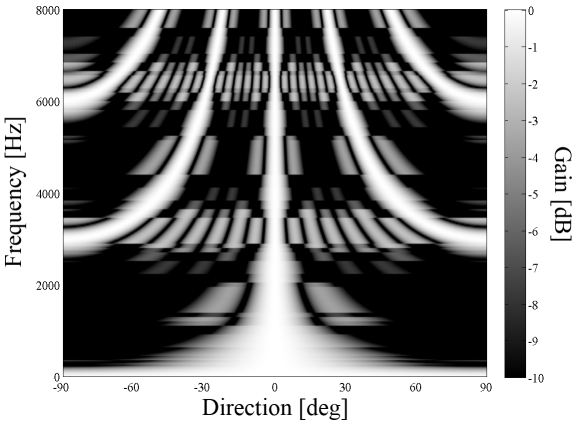


Fig.5 最適化後の DS アレーの指向特性

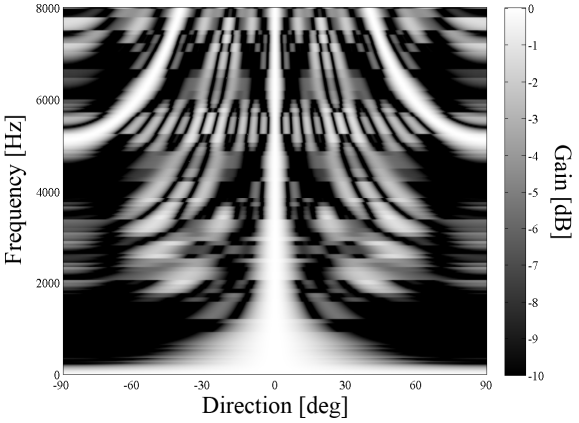


Fig.6 最適化後の 2^n アレーの指向特性

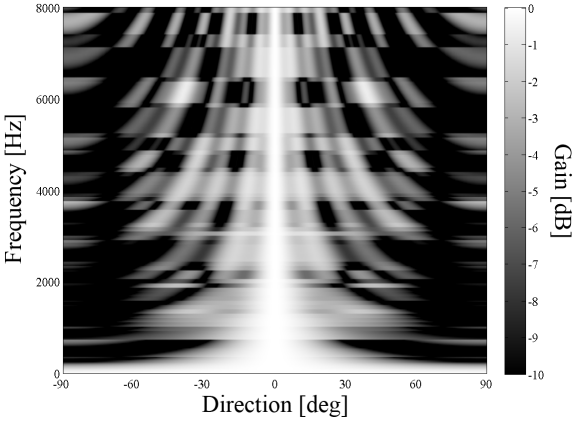


Fig.7 最適化後の 3^n アレーの指向特性

Tab.1　実験条件

目的信号	数字読み上げ音声 (英語・女性)
雑音信号 (特徴･ｴﾈﾙｷﾞｰが集中している帯域)	工場×3 種類 A：定常・3kHz 以下 B：非定常・1kHz 以下 展示会場 (定常・人の声・500Hz 以下) ホワイトノイズ (定常・全帯域)
目的方向	正面方向 (0°)
雑音(死角)方向	10°～80° (10°きざみ)
処理の種類	Non-process, DS, 3n_old (Optimized) DS, 2n, 3n
観測信号の SNR	-15 [dB]

４．まとめ

今回提案した最適化遅延和アレーは最適化していない遅延和アレーに比べて、客観的に見ると雑音の到来方向に関係なく雑音除去性能が高いことが分かった。しかし、マイクロホンの配置レベルで見ると雑音の種類によってどの配置がベストなのかははっきりとしていない。また、目的信号が正面方向からずれた場合の影響を今後調べる必要がある。更に伊藤憲三らの研究[4]により、人間が音を聴く場合、SNRが高いからと言って必ずしも聴き取りやすいわけではないことが分かっている。よって今後は、更に重み係数最適化アルゴリズムを改良し、アレーの雑音除去性能の向上を目指すと共に、主観評価による評価実験も行っていきたい。

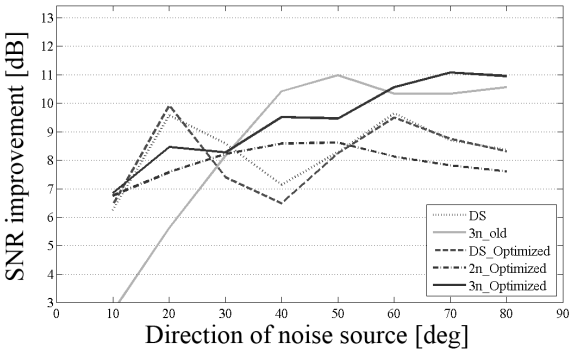


Fig.8 各アレーでの SNR 改善量(ホワイトノイズ)

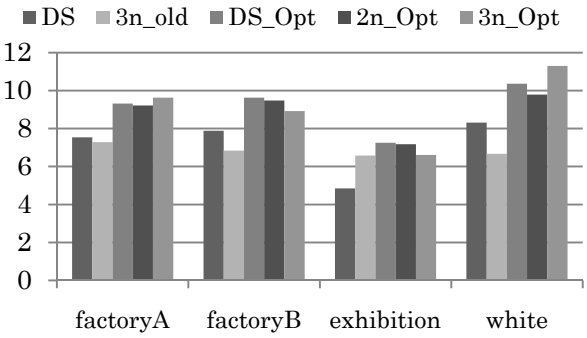


Fig.9 各アレーでの SNR 改善量 (全方向平均値)

参考文献

[1] 吉國信太郎　他, "マイクロホンアレーによる雑音除去音声の品質評価," 信学技報, vol. EA2008-97, pp. 75-80, Nov. 2008.

[2] ITU-T recommendation P.862, February 2001

[3] J. L. Flanagan et. al., "Autodirective Microphone Systems," Acoustica, vol.73, pp.58-71, 1991.

[4] 伊藤憲三 他, "音声のディジタル符号化方式の客観的品質評価尺度の検討," 信学論(A), vol. J66-A, no. 3, pp. 274-281, 1983.

高齢者の「めりはりのない声」に対応する物理量の検討*

濵﨑健太, 原田大輔, 宮崎健, 水町光徳, 二矢田勝行(九州工業大)

1　はじめに

人の声は性別,個人性,年齢の違いなどによって特徴が異なる。これまで個人性の研究などは一般成人男女を対象として行われてきたが、高齢者音声の研究は少ない。現在、先進諸国では高齢化が重要な社会問題となっており、高齢者音声の特徴解析は重要なテーマである。高齢者音声を解析することにより、高齢者に頑健な音声認識システムの構築や高齢者音声を聴きやすく補正することなどが期待される。

高齢者音声の主な聴感的特徴として、「しゃがれ声」・「めりはりのない声」・「発話の遅さ」が挙げられている[1]。なお、「めりはりのない声」とは正しい調音ができないため、音声が曖昧に聞こえる現象である。これは加齢による調音器官の衰えにより調音が不明瞭になるのが原因と考えられる。

これまで我々は「めりはりのない声」に着目し、めりはり度を評価するために音素間での振幅スペクトル包絡の時間的な動きの大きさ(遷移量)と速さ(遷移速度)を提案し、めりはり度との相関を明らかにしてきた[2][3]。本稿では遷移量・遷移速度と聴感的めりはり度の関係をさらに明確化すると共に、これらの物理量と他の高齢者音声の特徴である「発話の遅さ」との関係、及び加齢との関係について検討する。

2　めりはりに対応する物理量

まず高齢者(65 歳以上)内での「めりはりのある声」と、「めりはりのない声」のスペクトル例を Fig.1 に示す。フォルマントの動き等

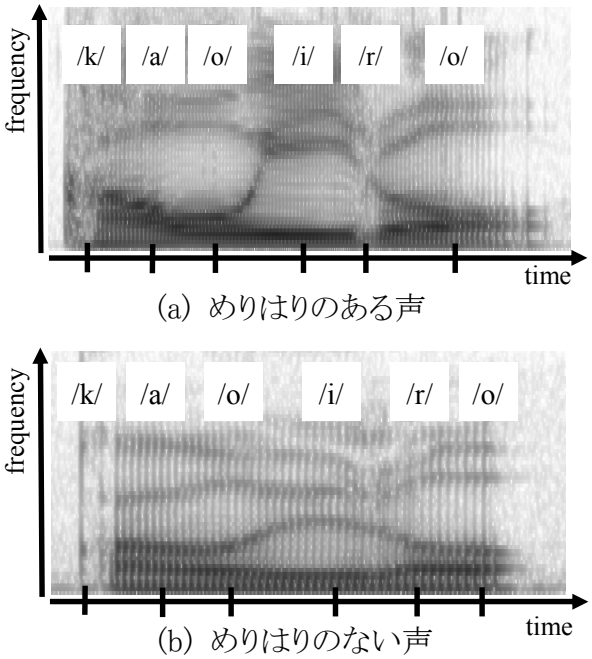


Fig. 1.　めりはりのある声とない声のスペクトル例（顔色 /kaoiro/）

に注意して両方を見比べると、「めりはりなし」の音声は「めりはりあり」の音声に比べ、スペクトルの時間的な動きが小さく、音素の区切りが不明確になっているのが読み取れる。多くのデータを参照し、我々はめりはりに対応する音響的特徴は音素間のスペクトルの時間的な変化形態にあると考えた。その特徴を定量化するために音素間のスペクトル「遷移量」と「遷移速度」をめりはり度に対応する物理量とした。

式(1)に音素間のスペクトルの距離を表す式を示す。Cは FFT ケプストラム係数、t1 と t2 は時間フレームを表している。

* Investigation on physical parameters related to non-brisk voice uttered by elderly people, by Kenta Hamasaki, Daisuke Harada, Takeshi Miyazaki, Mitsunori Mizumachi, and Katsuyuki Niyada(Kyusyu Institude of Technology).

$$dif_{\text{spectrum}} = \sqrt{\sum_{n=1}^{15}(C_n^{(t1)} - C_n^{(t2)})^2} \quad (1)$$

まず遷移量は 2 音素間でのスペクトル距離を表すため、式(1)の t1・t2 に各音素の特徴フレームを代入したものを用いる。特徴フレームとしては母音・摩擦音など音素区間が定常的なものは中心点を、破裂音など非定常な音素については目視で定めた最もその音素の特徴を表す部分とした。次に遷移速度は式(1)の t1・t2 間を 40msec で固定し、音素境界付近においてスキャンした値を算出し、その区間での最大値を充てた。算出区間は音素境界フレームの前後 3 フレームとする。

実験評価には音韻バランス単語 543 単語のデータベースを使用した。分析条件はサンプリング周波数 24kHz、フレーム幅 20ms、フレームのシフト幅 10ms である。

3 高齢者音声の特徴の定量化

3.1 めりはり度合い

高齢者音声の聴感的なめりはり度合いを求めるために聴取実験を行った。実験方法は音声を被験者に提示し、5 段階でめりはり度の評価を付加してもらう方法を用いた。それを全被験者で平均化し、被評価話者ごとに1つの印象値を算出する。値が大きいほど、その話者はめりはりがないと判断されたことになる。被評価話者は 60 歳以上の男性 36 名で、50 語の単語発話を連結させたものを被評価話者ごとに提示した。被験者は 20 歳代男女各 10 名であり、同一条件で 5 回評価させた。この聴取実験により聴感的なめりはり度合いが話者ごとに付加される。

3.2 話速

話速を表す量として、単位時間に発声しているモーラ数(mora/sec)を用い、分析単語である 543 語の平均話速を話者ごとに付加した。対象話者はめりはり度を付加した高齢者 36 名である。また比較のため 20～60 歳代での一般成人話者 43 人の話速も求めた。

4 遷移量・遷移速度とめりはり度との関係

まず聴感的なめりはり度との関連について考察する。曖昧な評価話者を除くために、聴取実験を行った 36 人の話者から、めりはり度の上位 6 人を「めりはりあり」、下位 6 人を「めりはりなし」の話者として抽出した。そして最もめりはりがあるグループと仮定して 20 代男性 6 名を「20 代成人」として加え、3 グループで比較した。Table.1 に各話者グループの平均めりはり度,平均年齢,平均話速を示す。

Fig.2 にめりはり度で分類した各グループの遷移量・遷移速度を算出し、比較した結果を示す。横軸は対象とする音素群結合で、上の音素群から下の音素群へ遷移することを示す。縦軸は各グループのスペクトル距離の平均値である。Fig.2 よりまず遷移量においては「高齢者めりはりあり」と「20 代」にはあまり違いがないが、「高齢者めりはりなし」のグループの遷移量は全ての結合において他のグループより小さいことが読み取れる。従って、遷移量は聴感的なめりはり度との関係が強いと言える。次に遷移速度を比較すると、「高齢者めりはりあり」は「20 代」よりも小さく、「高齢者めりはりなし」はさらに小さくなっている。遷移速度は聴感的めりはり度に加えて年齢にも関連した物理量であると言える。一方 Table1 より、各グループ間の話速を比較すると、「20 代」は高齢者よりも話速が速いが、高齢者間ではあまり差異がないことがわかる。

以上を総合すると次のように言える。

- 遷移量は聴感的めりはり度との関係が強い。
- 遷移速度は聴感的めりはり度と年齢の両方に関係している。

そして次のような仮説が成り立つ。

Table 1. めりはり度別のグループ情報

	20 代成人	めりはりあり	めりはりなし
めりはり度	–	1.96	3.68
話速[mora/sec]	5.71	4.41	4.69
年齢	25	72.17	69

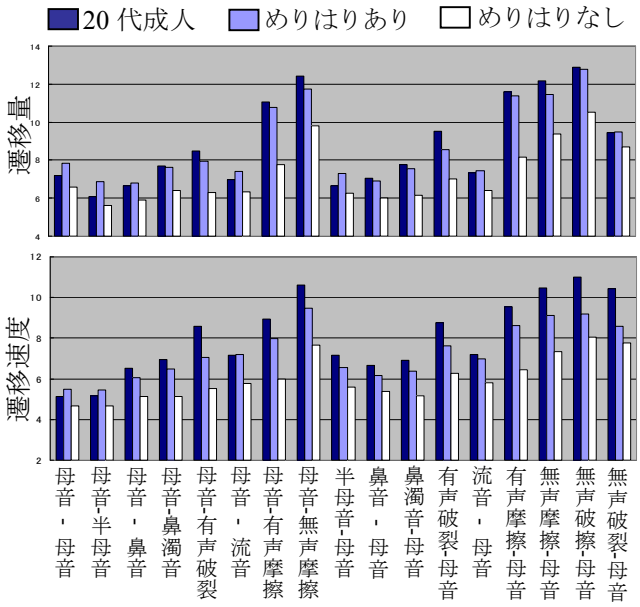


Fig. 2. めりはりと遷移量・遷移速度の関係

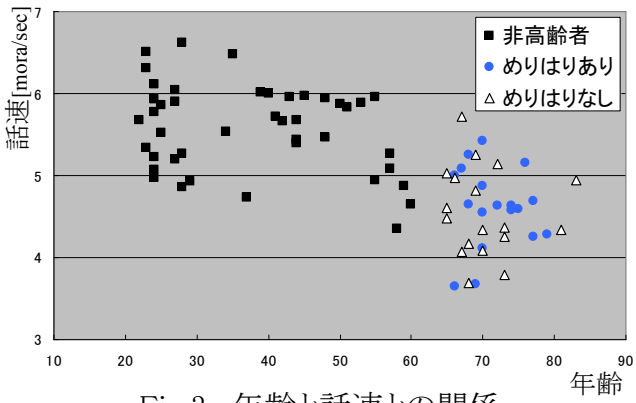


Fig.3. 年齢と話速との関係

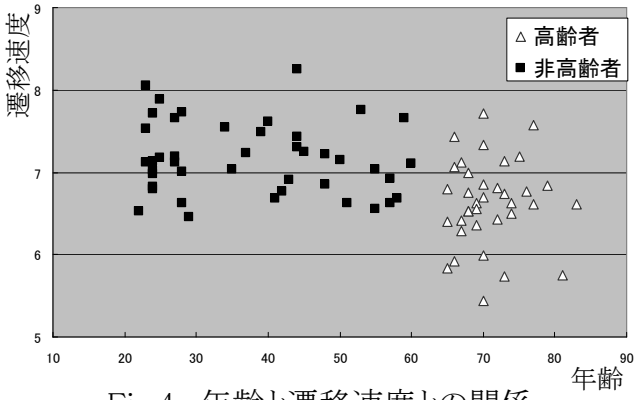


Fig.4. 年齢と遷移速度との関係

- 「高齢者めりはりあり」の遷移速度は遅いが、発話速度を遅くすることによって遷移量を大きくし、めりはり度を確保している。一方、「めりはりなし」の話者は遷移速度がさらに遅いため、話速を遅くしても十分な遷移量を確保できていない。

5 めりはりに関与する特徴量間の関係

第 4 章では、特徴的な話者のみを用いて全体的なめりはりの傾向を調査したが、本章では比較対象を絞り、評価話者数を増やすことにより前章の仮説について検討していく。

まず加齢と共に話速がどのように変化するかを調査した。Fig.3 に年齢と話速との関係を示す。横軸は年齢で、縦軸は話者ごとの平均話速である。まためりはり度との関係も調査するため、高齢者をめりはり度で 2 分し、上位半分を「めりはりあり」、下位半分を「めりはりなし」とした。Fig.3 より高齢者は明らかに非高齢者よりも話速が低下していることが読み取れる。また高齢者のめりはり度の有無では話速にほとんど差がないということが分かった。従って、加齢より話速は低下するが、高齢者間では話速とめりはり度には関連があまり無いといえる。

次に加齢と遷移速度には相関があるのかについて調査した。Fig.4 に年齢と遷移速度との関連を示す。横軸は年齢で、縦軸が話者ごとの平均遷移速度である。Fig.4 より個人差はあるものの、年齢と共に遷移速度は次第に小さくなる傾向が見られる。また第4章で用いた高齢者めりはりなしの話者 6 人の遷移速度は他の話者に比べて小さかったので、特に高齢者でも遷移速度が小さい話者が典型的なめりはりがない話者と判断されているといえる。

最後に高齢者は非高齢者に比べゆっくり発話することで遷移量を大きくしているのか

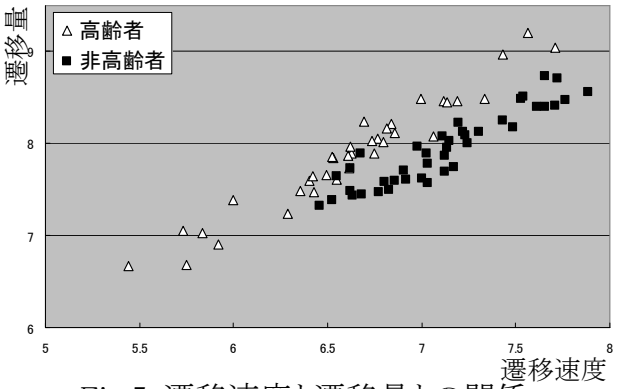


Fig.5 遷移速度と遷移量との関係

について調べた。Fig.5 に遷移速度と遷移量の相関関係のグラフを示す。横軸は話者ごとの平均遷移速度、縦軸は話者ごとの平均遷移量である。Fig.5 より非高齢者,高齢者共に遷移速度が大きくなれば、遷移量もほぼ比例して大きくなっていることが読み取れる。次に同程度の遷移速度で非高齢者と高齢者を比較すると、高齢者の遷移量の方が大きくなる傾向がある。遷移量・遷移速度および話速の関係、および Fig.2 のめりはり度との関係を総合的に見ると、高齢者は発話速度を遅くすることで遷移量を大きくし、めりはり度を確保していると考えることができる。しかし、一部の高齢者は遷移速度が非常に遅いため十分な遷移量を確保できない。これらの話者が「めりはりのない」話者ということになる。

以上の検討より、加齢によって音素間のスペクトル遷移速度が低下する傾向があり、その結果スペクトル遷移量が小さくなることがわかった。遷移量は話速を遅くすることによって改善できるが、高齢者の中には十分改善できない話者もいることがわかった。このように第4章の仮説を裏付ける結果が得られた。

6 まとめ

本稿では高齢者のめりはりを表す物理量として提案した音素間のスペクトル遷移量及び遷移速度について、「聴感的めりはり度」,「年齢」,「話速」の関係でその特徴を調査した。

その結果、まず遷移量はめりはり度と強い相関があることがわかった。これは音素間の遷移が大きいということは各音素の調音がしっかり行われているということに対応しているといえる。また遷移速度は加齢と共に遅くなる傾向が得られた。従って、加齢によって調音器官の動きが緩慢になるため、高齢者は急激なスペクトル変化に対応しづらくなると推測される。最後にめりはりに関与する特徴間の調査によって、高齢者はゆっくり発声することで遷移速度が多少遅くても十分な遷移量を確保しているということがわかった。しかし、それでも十分な遷移量を確保できない遷移速度の遅い話者が「めりはりのない」話者と知覚される。

今後の課題としては、上記の結論をさらに裏付けるデータの蓄積、しゃがれ声など他の高齢者の声の特徴との関係を調べることなどが挙げられる。

謝辞

本研究の一部は日本学術振興会科研費(No.19560387)の助成を受けて行われた。

参考文献

[1] 宮崎健 他, “高齢者音声を印象付ける聴覚的特徴に関する検討”, 音講論(秋), 283-286,2008.

[2] K.Hamasaki et al,”Investigation on acoustic feature to characterize ill-articulated voice uttered by elderly people”,Proc.NCSP,pp53-56,2009.

[3] 原田大輔 他, “高齢者音声の減り張りのなさとスペクトルの時間遷移の関係”, 電気関係学会九州支部,03-2P-08, 2009.

雑音環境における
ウェーブレット変換圧縮を利用した音声認識*

○渡壁　亨，緑川洋一，秋田昌憲（大分大）

1　はじめに

普段生活している環境で音声を録音すると、高い確率で雑音を含んでしまう。生活環境で雑音がないことがほぼ無いからである。この雑音は音声認識を行う上でとても不利な材料をなる。人間が発する以外の周波数帯域ではフィルタによって取り除くことが出来るが、同じ周波数帯になるとそうはいかない。そこで、ウェーブレット変換の圧縮作用を使い、特徴を抽出し、認識率の向上が出来ないかと考えた。

2　ウェーブレット変換

2.1 ウェーブレット変換

ウェーブレット変換には、連続系ウェーブレット変換と、離散値系ウェーブレット変換[1] [2] がある。今回使用したのは後者である。離散値系ウェーブレット変換はパワー概念があり、高速変換が可能である。その為、離散値系ウェーブレット変換はデータ圧縮やエネルギー解析等に適していると言われている為である。また、離散値系ウェーブレット変換に使うデータは 2 のべき乗個である必要がある。これによりデジタル測定器との相性がとてもよいと言われている。

2.2 基底

ウェーブレット変換には基底があるが、今回はハール基底とドビュッシー基底を使用した。ハール基底は 2 のべき乗個のデータベクトルを和と差にわける最も単純な基底である。和と差の概念にはそれぞれ積分を微分に対応するが、積分や微分には重みつきで行う場合がある。そのウェーブレットの基底がドビュッシー基底である。ドビュッシー基底は重みつきの積分演算、重みつきの微分演算の概念を導入し、高次の係数を使ったウェーブレット変換にするものである。本実験ではこのハール基底とドビュッシー基底の 2 つを用いた。また、ドビュッシー基底の基底関数は大きな変化を観察するため 10 次を用いた。

2.3 データ圧縮の原理

この離散値系ウェーブレット変換は与えられたデータの特徴を局所的に集める性質を持つ。絶対値の大きなスペクトルの部分を残し、他をゼロにすることで、原データの特徴を残したまま圧縮することが出来る。これがウェーブレット変換圧縮の原理である。圧縮の方法として、スペクトルの絶対値の大きい順にある個数を残す方法と、スペクトラムの特定部分を残す方法がある。本実験では後者を用いる。

3　認識実験

3.1 使用した音声

本実験では ichi、ni、san、yon、go、roku、nana、hati、kyu、zero の 10 個の数字音声を 1 セットとし、1 人 3 回ずつ、男性話者 8 人に孤立発音したものを用いる。これに雑音を加えることにより、雑音環境下の発声音源とした。雑音の種類はピンクノイズ 0 dB , 10 dB、自動車ノイズ 0 dB , 10 dB、これに無雑音を加えた 5 種類とする。このデータを使い、ウェーブレット変換圧縮を行う。

3.2 実験方法

まず、データが 2 のべき乗個である必要がある為、今あるデータに無音のデータを加え、2 のべき乗個にする。その後ウェーブレット変換を行う。ウェーブレット変換後は隣同士のデータの差、その差

* The speech recognition using Wavelet transform compression in the noise environment.by Toru, Watakabe and Yoichi, Midorikawa and Masanori, Akita (Oita University).

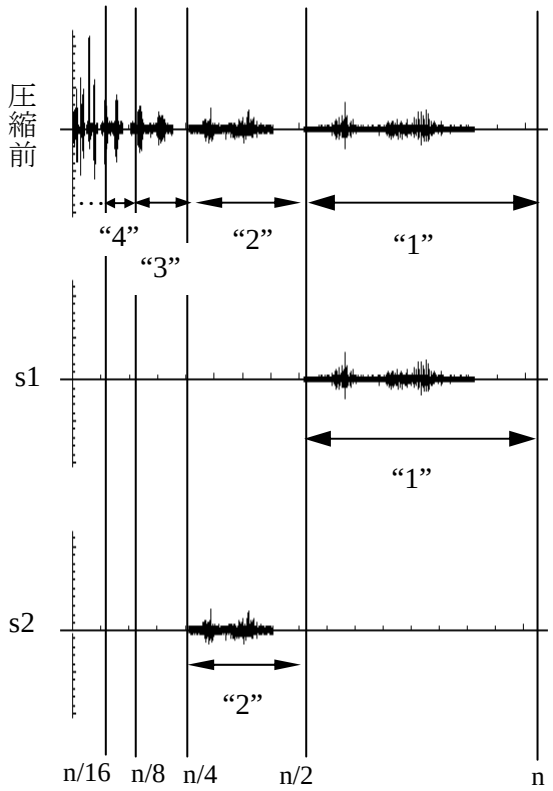


図 1　圧縮した場合のウェーブレットスペクトル(上:圧縮前、中:s1 で圧縮後、下:s2 で圧縮後)

の隣との差、さらにその差の隣との差…という形をグループ毎に後ろから順に並んだ形となり、先頭は最後の 2 つのデータの和を示している。この最初の段階のグループ、つまり、ウェーブレットスペクトル上では 1 番最後のグループを”1”のグループとし、2 段階目、ウェーブレットスペクトル上では最後から 2 番目のグループを”2”のグループとして表現する。

まず、ある特定の部分のみを残した圧縮を行う。圧縮の種類を表す方法として、最初に”s”を付け、1 段目の部分だけを残し、他をゼロにしたものを”s1”、2 段目の部分だけを残し、他をゼロにしたものを”s2”とし、”s7”までを行った(図 1)。

次にある特定の部分のみを取り除いた圧縮を行う。こちらの圧縮の種類を表わす方法として、最初に”w”を付け、その後ろに圧縮するグループの数字を並べることによって表わすこととする。例えば、2 段目と 4 段目だけをゼロにした場合は”w24”と圧縮の種類を表現する。ゼロにする部分としては”1”～”5“の 5 段階として、いろいろな組み合わせでデータをゼロにして認識を行う(図 2)。また、特殊な形として、”w15a”、” w16a”、” w17a”の 3 パターンも行う。こちらは”1”のグループと、”5”より大きな数字の部分を取り除いたパターンを”w15a”と表わすことに

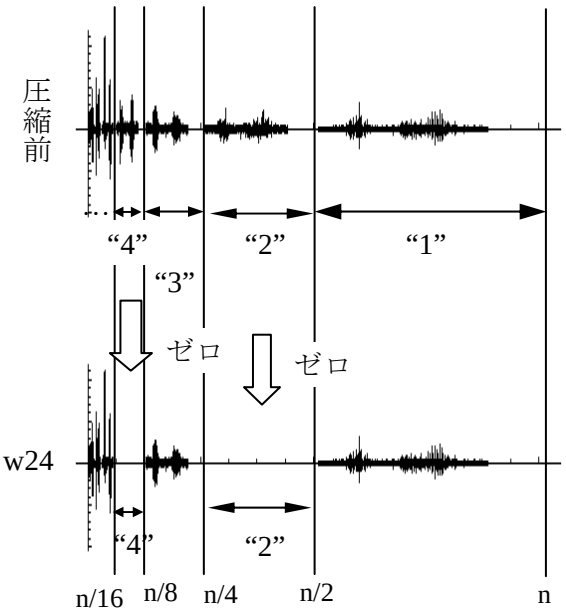


図 2　w24 で圧縮した場合のウェーブレットスペクトル(上:圧縮前、下:圧縮後)

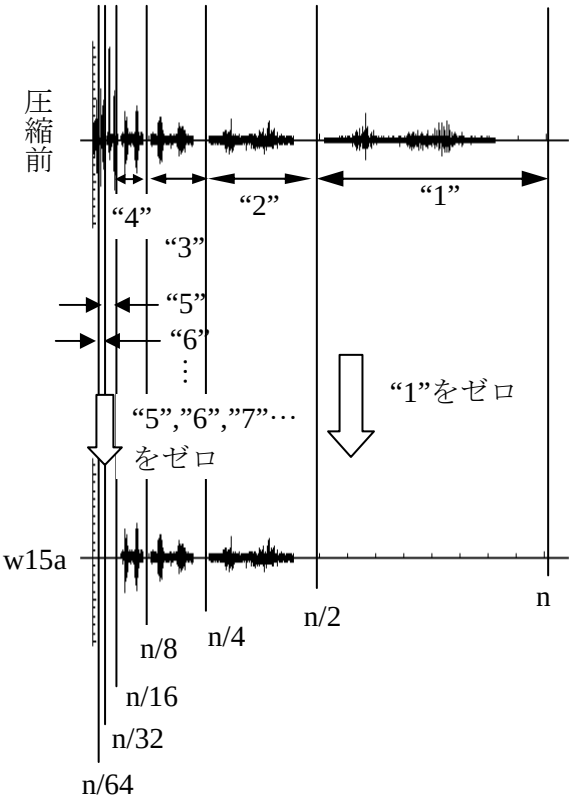


図 3　w15a で圧縮した場合のウェーブレットスペクトル(上:圧縮前、下:圧縮後)

し、”w16a”では”1”と”6”より大きな数字の部分を取り除いたパターン、”w17a”では”1”と”7”より数字の大きい部分を取り除いたパターンとする(図 3)。

このようにして圧縮されたデータをウェーブレット逆変換し、改良ケプストラム法によるスペクトル包絡の抽出[4] [5] [6]を行う。認識の方法として、同じ圧縮方法による無雑音データを基準とし、不特定話者として無雑音を含めた 5 種類の雑音を比較し、認識率を求める。

3.3 ドビュッシー基底

ドビュッシー基底に関してはすべてのデータを一括してウェーブレット変換することが出来なかったため、元のデータをウェーブレット変換出来る 2 のべき乗個のデータ数に区切り、最後の区切りのみ無音のデータを加え、2 のべき乗個にする。その後それぞれウェーブレット変換を行い、圧縮を行う。その後、それぞれウェーブレット逆変換を行い、すべてのデータを再び 1 つに合わせ、ウェーブレット変換圧縮とする。そのデータを用いてデータ音声認識を行う。この方法を行うと、一括でウェーブレット変換した場合より、ウェーブレットスペクトルの最後の部分がなくなる事になる。しかし、ハール基底の結果より、”w16a”がよいという結果が出ている点から大きな差は出ないであろうと考えたため、この方法を用いる。

4　認識結果

4.1 特定の部分のみを残す場合

表 1、表 2 に残す場所による認識率の違いを示す。この時使用した認識次数は 1-25 次を使用した時である。

“s1”が無雑音で最も良かった。ピンクノイズ 0 dB、10 dB、自動車ノイズ 0 dB は”s4”が最も良いという結果となった。自動車ノイズ 10 dB では”s3”が最も良い結果となった。次点も含めて考えると全体的に”s3”、”s4”の圧縮パターンがよくなる傾向が見られた。このことにより、認識に必要な部分は”3”や”4”の部分にあるのではないかと考える事が出来る。

表 1　特定の部分だけを残した場合の認識率(1)

	s1	s2	s3	s4
無雑音	86.75	86.94	79.88	54.86
ピンクノイズ 0 dB	13.63	10.75	12.86	23.06
ピンクノイズ 10 dB	28.85	25.85	43.91	44.23
自動車ノイズ 0 dB	19.03	11.57	15.73	19.60
自動車ノイズ 10 dB	40.62	27.56	45.65	42.98

表 2　特定の部分だけを残した場合の認識率(2)

	s5	s6	s7
無雑音	21.11	17.76	19.29
ピンクノイズ 0 dB	16.63	12.70	11.63
ピンクノイズ 10 dB	17.58	12.14	11.67
自動車ノイズ 0 dB	16.13	12.26	10.95
自動車ノイズ 10 dB	18.87	12.48	11.77

4.2 ハール基底による圧縮

まず、”1”~”5”までの組み合わせによる認識実験を行った。その結果と、特定の部分のみを残す場合の結果を踏まえ、”1”と”5”以降を圧縮、”1”と”6”以降を圧縮、”1”と”7”以降を圧縮のパターンを加えた。ハール基底で最も良かった圧縮の種類と認識率を表 3 に示す。この時使用した認識次数も 1-25 次を使用した時である。

表 3　最も良かった圧縮の種類と認識率(ハール基底)

雑音の種類	認識率(無圧縮)[%]	圧縮の種類	認識率[%]
無雑音	95.14	無雑音	95.14
ピンクノイズ 0 dB	15.86	w16a	31.15
ピンクノイズ 10 dB	28.31	w16a	68.91
自動車ノイズ 0 dB	24.29	w16a	34.40
自動車ノイズ 10 dB	54.03	w2345	74.48

雑音環境で最も良かったのは自動車ノイズ 10 dB の圧縮パターン”w2345”で 74.48[%]となった。認識率の向上は約 20[%]である。その他は”w16a”の圧縮を行ったデータの認識率が高いことが伺える。向上率で見ると、ピンクノイズ 10 dB が約 40[%]の向上となり、大きな効果があった事が言える。ピンクノイズ 0 dB、10 dB、自動車ノイズの 0 dB では約 15[%]、約 15[%]、約 10[%]の向上となった。これにより、0 dB では約 30～35[%]程度まで認識率が上がったことが分かる。また、この時の w16a の自動車ノイズ 10 dB は 73.21[%]となっている。今回最大の認識率であった自動車ノイズ 10 dB の 74.48[%]に迫る勢いである。全体としては”w16a”が最も効率よく認識率の向上が出来る圧縮パターンであると言える。

このことにより音声の特徴は”2”、”3”、”4”の部分に多くある可能性があり、また、”1”や”5”より数字が大きい部分に雑音の特徴が多く含まれているのではないかと考えられる。

4.3 ドビュッシー基底による圧縮

ドビュッシー基底で最も良かった圧縮の種類と認識率を表 4 に示す。この時使用した認識次数も 1-25 次を使用した時である。

表 4　最も良かった圧縮の種類と認識率 (ドビュッシー基底)

雑音の種類	認識率 (無圧縮)[%]	圧縮の種類	認識率[%]
無雑音	94.33	w17a	97.00
ピンクノイズ 0 dB	15.46	w15a	20.75
ピンクノイズ 10 dB	27.92	w1235	49.48
自動車ノイズ 0 dB	24.15	w2	28.19
自動車ノイズ 10 dB	54.03	w2	56.61

雑音環境で最も良かったのは自動車ノイズ 10 dB の 56.61[%]であった。しかし、このノイズは元々認識率が高い方だったため、向上率としては約 2.5[%]に留まっている。今回最も認識率が向上したのはピンクノイズ 10 dB で約 22[%]の向上が見られた。他は 5[%]程度の向上しかなかった。この結果から”2”や”5”が取り除かれることにより認識率の向上があると言える。

しかし、全体的にハール基底に対して認識率の向上が見られなかったことから基底関数が 10 次と大きすぎたのではないかと思う。また、無雑音の認識率が少しだが向上している。今までは無圧縮の場合が最も良かったが、重み付けの影響で無圧縮よりも良かった結果を踏まえると、重み付けは音声部分に有効に働くことが言える。

5　まとめ

ハール基底全体の結果から”2”～”5”の部分に音声の特徴が多く含まれている可能性を示した。しかし、無雑音では認識率が落ち込んでいるため、他の部分にも音声の特長が含まれていることは確かである。ドビュッシー基底では向上はするものの、基底関数 10 次では重み付けが強すぎる為か、ハール基底ほど認識率の向上は見られなかった。しかし、無雑音環境での音声認識が少しながら向上したことは、重み付けの有効性を示しているものと考えられる。今後、基底関数を変えることによる認識率向上にも取り組んでいきたいと思う。

参考文献

[1] チャールズ K、チュウイ著 桜井明・新井勉共訳，“ウェーブレット入門”，東京電機大学出版局，1993.
[2] 斉藤兆吉，“Mathematica によるウェーブレット変換”，朝日書店，pp.1-33，1996.
[3] 新居康彦・大崎正巳，“音声処理と DSP”，啓学出版，pp.140-145，1989.
[4] 今井聖・阿部芳春著，”改良ケプストラム法によるスペクトラム包絡の抽出”, 電子情報通信学会論文誌 J62-A4，pp.217-223
[5] 超川常治, ”信号解析入門”, 近代科学社, pp.120-130
[6] 古井貞熙，”音響・音声工学”，近代科学社，pp.115-123

発話訓練のための音声プロソディのリアルタイム推定法とその表示方式の提案*

小糸陽介　坂田聡 上田裕市 (熊本大院 · 自然科学研)

1　はじめに

聴覚障害者は，自己の音声を聴覚フィードバックすることの困難さから，発話が不明瞭となる．我々は，このような話者の発声発話訓練において，音声を視覚イメージに変換して発話者へのフィードバックを可能にする研究の一環として音声可視化手法を提案し，リアルタイム表示システムを構築した [1].

本研究では，その統合化音声画像 (Fig.1) をベースに，構音速度の表示機能を加えたプロソディ情報に特化したリアルタイム表示システムの開発を目的としている．構音速度表示機能の追加により，運動性発話障害者やパーキンソン病患者に対する構音速度の訓練補助 [2] や，日本語の特徴であるモーラ等時性によるモーラ型リズムの習得における訓練補助ツールとしての利用を想定している．

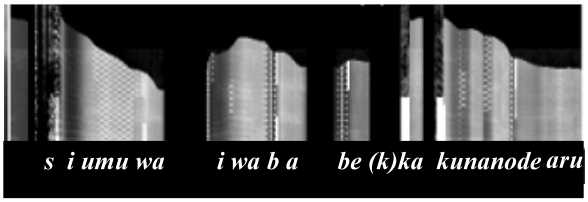


Fig. 1　An example of the integrated visual speech displayed by our real-time speech visualization system

2　音声プロソディの推定

リアルタイム表示するプロソディ情報として用いる，構音速度，有声/無声，またアクセントやイントネーション訓練のためのピッチ，インテンシティ(声の強弱) を推定する．

2.1　音声特徴量の推定エンジン

先に提案された音声特徴ベクトル推定エンジン [3](Fig.2) を用いて，音声のプロソディ情報を推定する．この推定エンジンは，フォルマント周波数や基本周波数などの音声特徴量から母音性や鼻音性などの音素特徴量，母音，子音の音素距離までの階層構造を持つ特徴ベクトルを，分析フレーム周期 (10ms) でリアルタイムに推定できる．

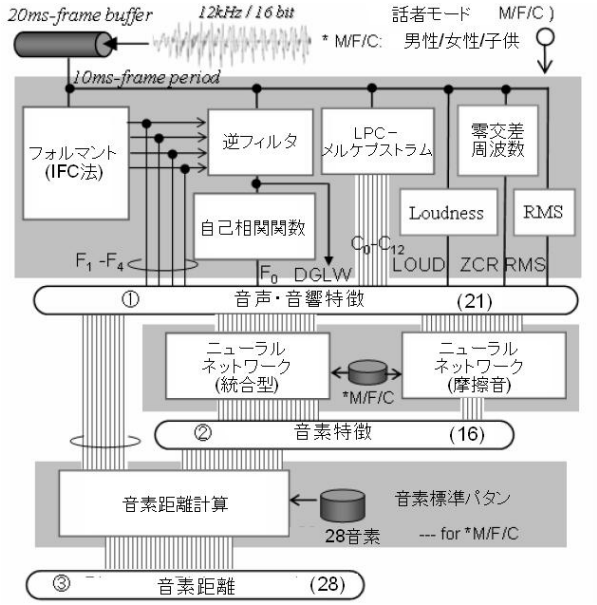


Fig. 2　Block diagram of a software engine for estimating speech feature

2.2　プロソディ要素の推定

表示するプロソディ要素として，推定エンジンより得られる音声特徴量から基本周波数，実効値を用いる．実効値は正規化し，表示ではインテンシティ情報として用いる．

また各分析フレームにおける音素特徴量 (ニューラルネット出力値) のうち 4 種 (母音性 (VOW)，有声性 (VOI)，無声性 (UNV)，無音性 (SIL)) を，弁別的素性とする 3 群 (子音，母音，無音) に分類する (Table.1)．母音性，有声性がある場合は母音，有声性のみある場合，もしくは無声性のみある場合は子音，無音性のみある場合は無音とする．これを基に，表示するプロソディ要素の有声/無声を推定する．

---

* A real-time estimation of prosodic features for speech training and a proposal of its visual representation.　by KOITO, Yosuke and SAKATA, tadashi and UEDA, Yuichi (Kumamoto University graduate School of science and Technology)

Table 1 Discrimination of phonemic feature by distinctive feature analysis

VOW	VOI	UNV	SIL	discrimination
1	1	0	0	Vowel
0	1	0	0	
0	0	1	0	Consonant
0	0	0	1	Silence

2.3 構音速度の推定

構音速度は，音素特徴から Fig.3 のフローチャートに示す分割方式 [4] を用いて推定する．モーラ境界を検出し，間の時間をモーラ長としてシフトバッファに格納する (Fig.4)．バッファ内に格納されたモーラの数を，バッファ内のモーラ長合計 (フレーム) で除した値を構音速度として用いる．バッファサイズ (時間) を変化させることにより，短い音節単位 (モーラ速度) から長い文節単位 (発話速度) での構音速度推定を考察する．

Fig.5 に発話「北風と太陽が力比べをしました」の音声波形，速度推定結果，および視察に基づく速度測定値を示す．構音速度はモーラが変わるタイミングで更新されるこ

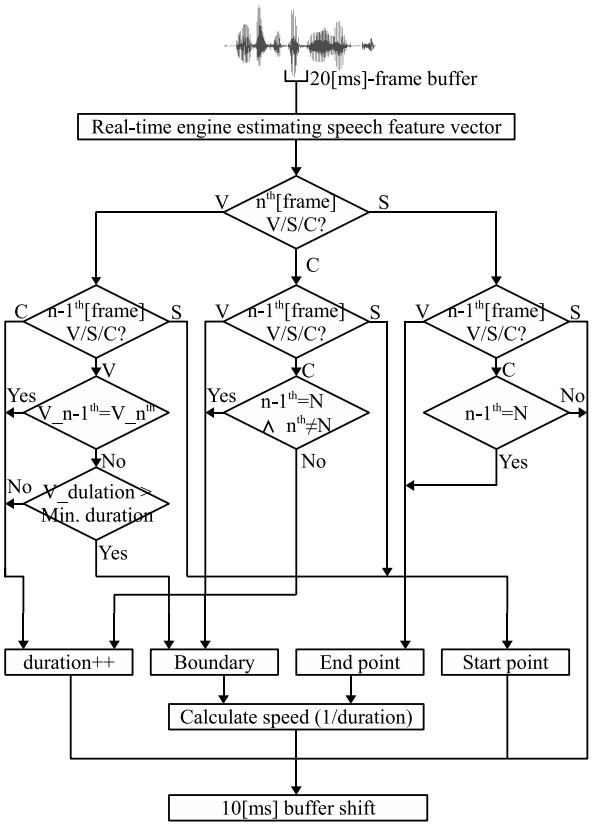


Fig. 3 Flow chart for estimating the speech rate based on phonemic features

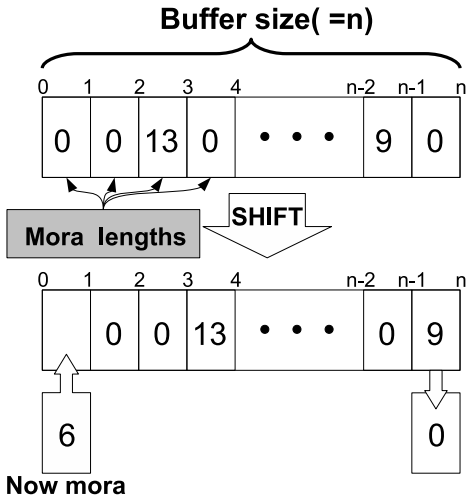


Fig. 4 Illustration of shift buffering to store the mora length estimates

とから，同期をとるため視察結果を 0.1s 遅く表示している．速度推定のバッファサイズが 0.1s，1s の場合において，視察による測定値を追従できている．これによりリアルタイムに表示することが可能であると考えられる．

3 特殊モーラへの対応

Fig.5 に示すように，速度推定に関して追従はできるが個々の値には差が生じている．この原因としては，特殊モーラや半母音，拗音などの存在が挙げられる．ここで，特殊モーラによる速度推定誤差とその対策を検討する．

3.1 音素特徴量 (NAS) の利用

日本語は CV モーラ構成 (母音のみ/子音-母音) で作られており，本システムでのモーラの分割はそれに基づいている．そのため特

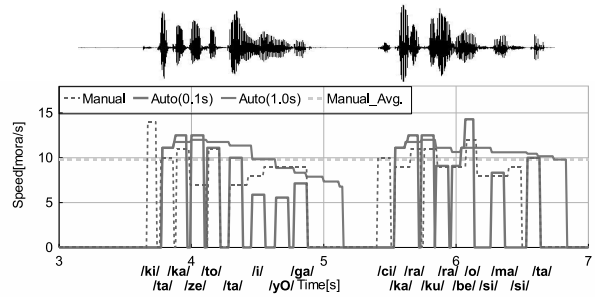


Fig. 5 An example of the speech rates estimated manually and automatically

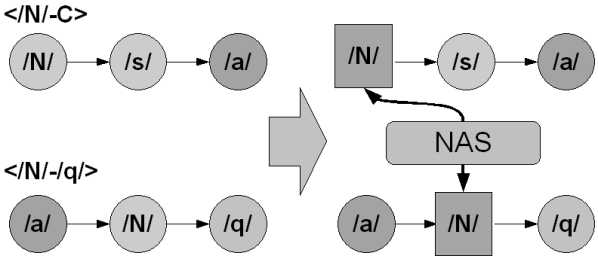


Fig. 6 Illustration of discriminating the syllabic nasals by use of "nasality index"

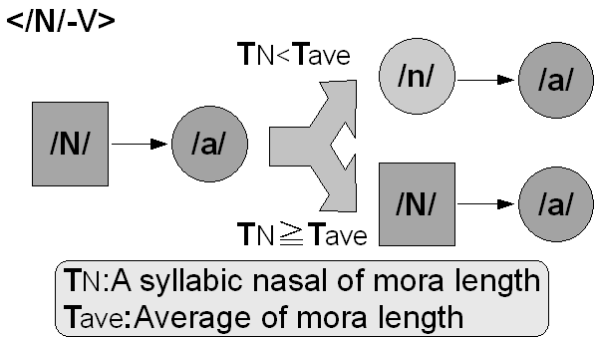


Fig. 7 Illustration of discriminating the syllabic nasals by use of the average mora lengths

殊モーラ (撥音/N/，長音/H/，促音/Q/) は，モーラの境界の検出に影響し，推定誤差が生じる場合があった．

子音/撥音を区別するため，音素特徴量のひとつである鼻音性 (NAS) を判別条件に加える．これにより撥音を子音と誤って判定することを抑えられ，子音前の撥音が子音に吸収される (例:saNkai)，無音前の撥音を子音と誤って判別しモーラ長がリセットされる (例:mikaN) といった問題をそれぞれ解消できる．Fig.6 上に前者の例，下に後者の例を示す．

しかし，撥音-母音 (例:reNai) と鼻音-母音 (例:ame) のどちらも撥音と判断する恐れがある．一方，撥音に鼻音が続く場合 (例:saNma) は鼻音性が継続して表れるため，正しく分割ができない．

3.2 平均モーラ長による音素特徴の判定

速度推定から得られる平均モーラ長 (構音速度の逆数 [秒/モーラ]) を用いた分割法を検討する．推定される個々のモーラ長と平均モーラ長を比較することで，鼻音/撥音/撥音-鼻音の判別を行う．推定モーラ長が平均モーラ長より短い場合は鼻音のみ，長い場合は撥音と判別し，後者はモーラ境界と判断する (Fig.7)．その後も鼻音性が持続する場合，以降の音素特徴は Fig.6 の鼻音性を考慮しない．これにより撥音ではなく，子音 (/m/,/n/) として判別することができる．同様に長音にも本手法を適用する．同一の母音が長時間続く際，モーラ長が平均モーラ長より短い場合は単一母音，長い場合は長音 (/H/) として判別する．

3.3 鼻音性，平均モーラ長を用いた音素特徴の判別例

Fig.8 に，男性話者による単語音声「カレンダー」モーラ長検出例を示す．モーラ長を検出するタイミングをモーラ境界として判断する．特殊モーラに対応していない場合では,「カレ」の後の/N/が/d/に吸収され，更に/a/が長音として判断されないため一塊になっている．

先に述べた特殊モーラ (撥音，長音) に対応させたアルゴリズムでは，/N/を検出でき，また長音も分割できていることが分かる．しかしながら平均との比較により，最後の長音は 1 モーラ分多く分割されている (kareNdaHH) ことも確認できる．

以上のことから，このアルゴリズムでは特殊モーラのうち，撥音，長音の検出が可能であり，これにより速度推定の誤差を低減できる．しかし話者の構音速度によって，閾値として用いる平均モーラ長は大きく影響されることが考えられる．

4 リアルタイム表示システムの構築

Fig.9 に，システム表示方式の例を示す．推定した音声プロソディから，上画面に構音速

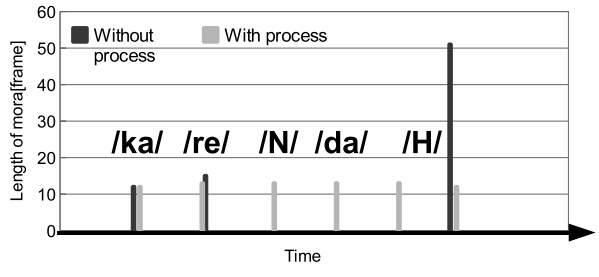


Fig. 8 An example of the detected mora boundaries and those mora lengths of a word speech(/kareNdaH/ uttered by a male)

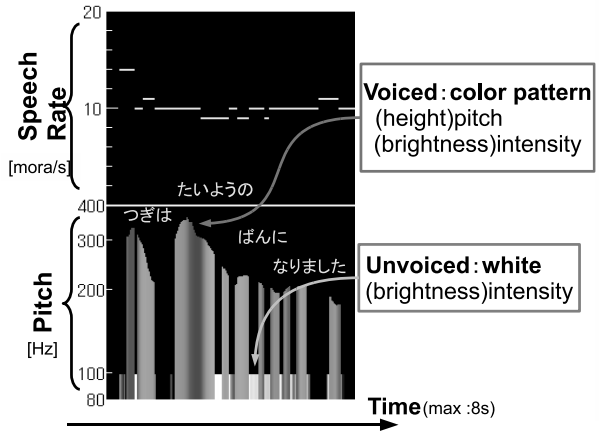


Fig. 9　Display mode for representing the speech prosodic feature

度，下画面にピッチと音声の色彩パタンを表示する．

下画面では，音声要素から得られる有声/無声を用いて色彩を決定する．有声 (母音) の場合，統合化音声画像の手法から，そのフォルマント周波数で決まる特有の色の帯を表示する．無声 (子音) の場合，白色単色を表示する．また，この両者の色彩パタンに関して輝度 (0.0 ～ 1.0) をインテンシティにより調節する．パタンの高さに対応するピッチ表示では，無声については一定 (100[Hz]) とする．

Fig.10 に構音速度の推定結果，有声/無声および音声ピッチのリアルタイムフロー表示例を示す．図は健常女性話者の発話表示例である．横軸は時間軸 (8 秒) であり，パタンは画面右から左方向に電光ニュース式に流して表示する．

上枠は構音速度 (最大 20[モーラ/秒]) であり，平均化時間長 (0.1 ～ 10[s]) を調節することで短いモーラ単位から長い文単位での構音速度を表示する．下枠は有声の色彩パタンとピッチ，インテンシティを表示する (図は「次は太陽の番になりました．太陽は雲の間から顔を出し，暖かな日差しを送りました．」[s] の例)．縦軸はピッチ (対数軸，最大 400[Hz]) により決定する．

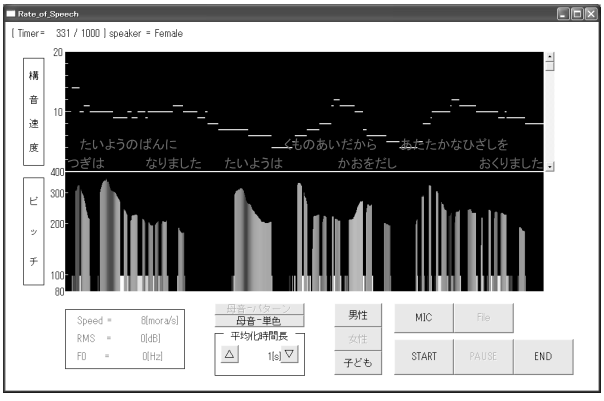


Fig. 10　A real-time tool for visualizing the prosodic feature

5　まとめ

リアルタイム推定される音声特徴ベクトルを用いた，構音速度等の音声プロソディの推定法を提案した．また音素特徴判定において，誤差の要因となる特殊モーラの判定アルゴリズムを考察した．推定された音声プロソディ要素のリアルタイム表示方式を紹介した．

今後は音素特徴推定における推定誤差の定量化，それに基づくアルゴリズムの改良，またイントネーションの訓練等を目的とした音声プロソディの表示形式の改良を行う予定である．

謝辞

本研究の一部は，平成 20·21 年度科学技術融合振興財団研究助成 (FOST) の補助を受けた．

参考文献

[1] 上田裕市他,“リアルタイム音声画像化処理に基づく発話訓練システムの構築,” 信学技報, WIT2007- 104, pp.79-84, 2008

[2] 伊藤元信他,“運動性発話障害の臨床-小児から成人まで-,” インテルナ出版, 2007

[3] 上田裕市他,“音声応用システムのためのリアルタイム音声特徴推定エンジンの構築,”信学技報, SP2008-67, pp.61-66, 2008

[4] 坂田聡他,“発話学習のためのプロソディ特徴量のリアルタイム推定法,”日本音響学会春季研究発表会講演論文集 (CD-ROM) 3-4-5, pp.499-500, 2009.03

母音音声の色彩表現に基づいた母音構音訓練における視覚的音韻規準に関する検討*

富田翔　米倉達郎 坂田聡 上田裕市 (熊本大院 · 自然科学研)

1　はじめに

先行研究 [1] において，聴覚障害児の発声訓練における視覚フィードバックとして音声画像表示システムが提案された．特に，フォルマント周波数の写像変換に基づく母音の色彩表現では，話者の年齢や性別に依存しない母音の音韻知覚が期待できる．しかし，学習の観点から，健常音声と自己音声との色の違いを確認する指標として規準となる母音色彩の明示が必要とされている．本稿では，母音の色彩表示に関して，健常音声領域を定め，母音調音訓練における視覚的音韻規準の提供を目的とする．さらに，母音色彩の視知覚の観点から色弁別特性を調べ，母音認識実験を行い，母音色彩を用いる構音訓練の可能性を検討する．

2　母音色彩における音韻性の客観的評価

2.1　音声分析

音声試料には，電子協共通音声データベースの成人男女各 75 名の発声による日本語 5 母音 (各話者 4 回発声，計 3000 試料 (= 2×75×5×4)) を用いた．記録フォーマット (48kHz-16bit, WAV 形式) を我々の分析フォーマット (12kHz-16bit, DAT 形式) に変換後，視察により各音声区間を切り出した．音韻不明瞭の試料と低 SN 比の試料を除く 2828 試料 (男声 1446 試料，女声 1382 試料) について，音声ベクトル推定エンジン [2] を用いて，音声特徴パラメータ群 (フレーム周期 10ms) を抽出した．主要な音声特徴量であるフォルマント周波数の抽出は，逆フィルタ制御法 (IFC 法)[3] による．

2.2　フォルマント空間における母音分布

各試料のフォルマント周波数 ($F_1 \sim F_3$) として，定常部 3 フレーム平均値を用いた．全試料の第 1~3 フォルマント周波数次元でのフォルマント空間での各母音の分布を Fig.1 に示す．男女声が混在するため母音カテゴリの重複が大きいのが確認できる．また男女声間での分布の違いも見て取れる．

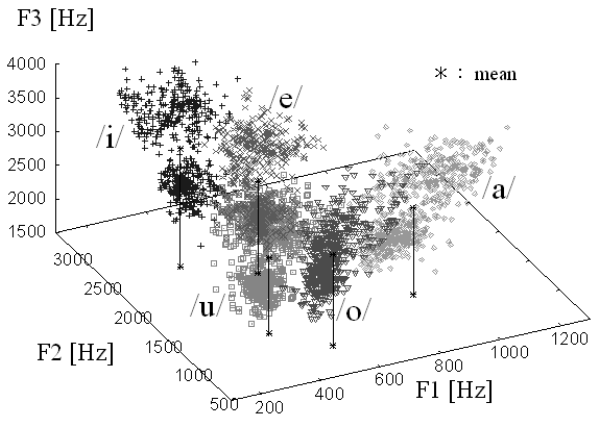


Fig. 1　Vowel category distributions in 3D-formant space(MF-group : 2828 vowels).

2.3　RGB 空間における母音分布

フォルマント情報から色彩情報への変換には，式 (1) を用いる．

$$R = \alpha\frac{5F_1}{F_3},\ G = \alpha\frac{3F_3}{5F_2},\ B = \alpha\frac{F_2}{3F_1} \qquad (1)$$

ただし，α は定数，R, G, B は赤，緑，青の三原色成分である．

Fig.1 のフォルマント空間を RGB 空間に写像した結果を Fig.2 に示す．RGB 色空間として見ると，各母音は概略的に，/i/ : 青領域，/o/ : 緑領域，/a/ : 黄 ~ 橙領域，/u/ : 青 ~ 緑の中間領域 (シアン)，/e/ : 赤 ~ 青の中間領域 (マゼンタ) にそれぞれ分布する．また，フォルマント空間 (Fig.1) に比べて各母音分布がまとまっており，母音音韻性に関する正規化が期待できる．

2.4　母音カテゴリ分離度の定量的評価

各空間における母音分布の分離度と集約度を定量的に評価するために，5 母音群の群

---

*Investigation of a Visual Phonemic Criterion in Articulation Training using the Colored Representation of Vowel Sound. by TOMITA, Kakeru, YONEKURA, Tatsurô, SAKATA, Tadashi, and UEDA, Yûichi(Graduate School of Science and Technology, Kumamoto University).

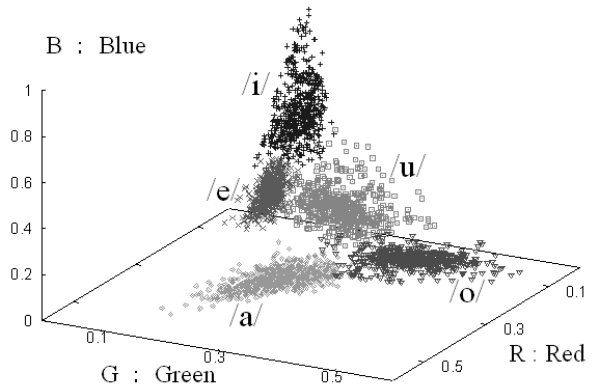


Fig. 2 Vowel category distributions in RGB-primary color space(MF-group : 2828 vowels).

間分散と群内分散の比として，式 (2) で定義される母音分離度 $D(D_{F123},\ D_{RGB})$[4] を用いた．

$$D = \sqrt{\frac{\frac{1}{5}\sum_{k=1}^{5}(m_k - m)^T(m_k - m)}{\frac{1}{N}\sum_{k=1}^{5}\sum_{i=1}^{n_k}(x_{k,i} - m_k)^T(x_{k,i} - m_k)}} \quad (2)$$

ただし，$x_{k,i}$ を母音 k 群の標本 i として

$$m_k = \sum_{i=1}^{n_k}\frac{x_{k,i}}{n_k},\ m = \sum_{k=1}^{5}\frac{m_k}{5},\ N = \sum_{k=1}^{5} n_k \quad (3)$$

が成り立つ．Table.1 に各空間での母音分離度を示す．男声群 (M 群) に比べて，女声群 (F 群) の D_{RGB} が D_{F123} より若干低下している ($D_{RGB}/D_{F123} < 1$) が，男女声が混在する条件下での正規化の良さという観点からは，男女混合群 (MF 群)D_{RGB} が優れている ($D_{RGB}/D_{F123} > 1$) ことがわかる．

Table 1 Vowel category separation in formant- and RGB space.

	D_{F123}	D_{RGB}	D_{RGB}/D_{F123}
Male(M)	2.862	3.428	1.198
Female(F)	2.770	2.468	0.891
M & F(MF)	1.608	2.777	1.727

3　母音色彩知覚における視覚的評価

3.1　HSV 表色系における母音分布

Fig.2 の RGB 空間について，色知覚の観点から考察するために，次式により，色相 (Hue)· 彩度 (Saturation)· 明度 (Value) 次元を持つ HSV 表色系に変換して母音分布を求めた．

$$V = \max \quad (4)$$

$$S = (\max - \min)/\max \quad (5)$$

$$H = \begin{cases} 60(G-B)/ & \text{if } R = \max \\ 60(2+(B-R)/\ \ ) & \text{if } G = \max \\ 60(4+(R-G)/\ \ ) & \text{if } B = \max \end{cases} \quad (6)$$

max=$Max\{R,G,B\}$, min=$Min\{R,G,B\}$,　=max−min

ただし，($0 \le$ R, G, B ≤ 1)，($0 \le$ S, V ≤ 1)，($0 \le$ H $\le$ 360deg) である．Fig.3 に，男女声 2828 試料を HS 平面上にプロットしたものを示す．色相次元 (H=0∼360deg：偏角成分) より，2.3 節で述べた各母音の色領域が確認できる．特に，/a/, /o/, /i/分布の集約性と，/u/と/e/(特に/u/) の拡散性が特徴的である．

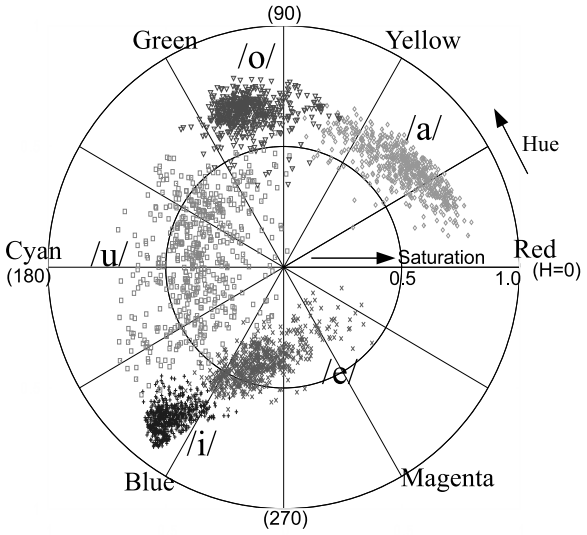


Fig. 3 Vowel distributions in HSV color model(HS-plane).

3.2　母音分布を考慮した色相帯規準

このように，多数話者の 5 母音は，そのフォルマント周波数に応じて特有の色度領域に存在するが，日本語単母音では存在し得ない領域 (赤など) が存在する．このような母音色彩分布の局在傾向は，母音構音訓練において色パタンをフィードバックする場合，異常構音の検知や目標とする構音を利用者に直感的に明示するという意味で有用である．

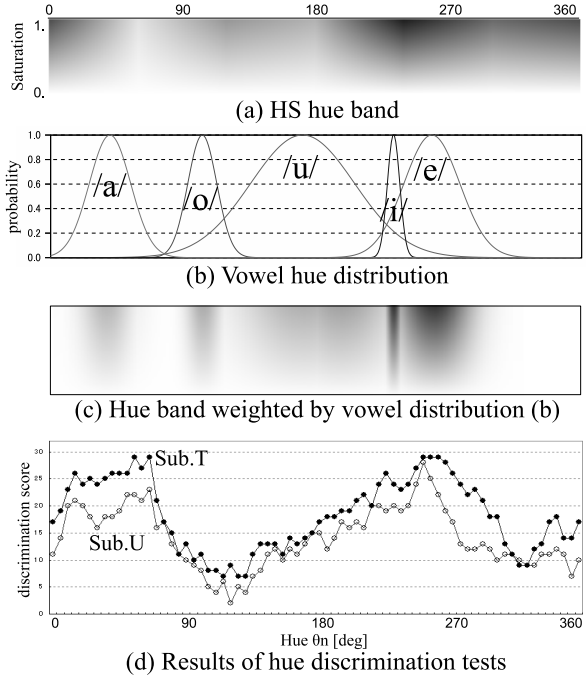


Fig. 4 Vowel distribution on the HS hue band (a)∼(c) and results of hue discrimination tests.

このような視覚的な母音音韻性に関する規準を定めるために，日本語 5 母音の分布確率を考慮した色相帯を生成した．Fig.4 に，(a) 色相帯 (横軸：H=0∼360deg，縦軸：S=0 ∼ 1)，(b) 各母音の色相確率分布 (Fig.3 において，母音毎の Hue 値の平均値と標準偏差から正規分布で近似) をそれぞれ示す．Fig.4(c) は，(a) の色相帯を (b) の確率分布で重みづけて表示したもので，各母音群で表示し得る色相範囲が明示されている．

3.3 母音色彩分布と色弁別能力

Fig.4(c) の日本語母音特有の色度分布と視覚における色弁別能力との関係について調べる為に色知覚に関する対比較実験を行った．

3.3.1 実験試料

実験試料 (色票) は 72 段階の色相からなる合成パタンである．HSV 表色系において，等彩度 (S=0.75)，等明度 (V=1.0)，色相 (H=0∼360deg で 5deg 間隔) の 72 通りである．色相 θ_n(n=1∼72) のペア試料として，±20deg(θ±5, 10, 15, 20deg) の 9 試料 (θ_n を含む) を用いた．各ペアについて順序を考慮した 72×9×2=1296 組のランダムリストを準備した．

3.3.2 実験方法

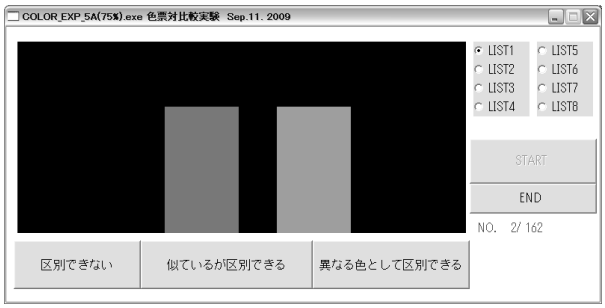


Fig. 5 PC-based visual experimental tool of hue discrimination.

実験には Windows ベースで作成した Fig.5 に示すツールを用いた．実験は 8 セッション (8×162 組) 実施した．各色票ペアは，Fig.5 において，画面右から左方向へ電光ニュース式に流れ，その色の違いを 3 段階 (0 : 区別できない，1 : 似ているが区別できる，2 : 異なる色として区別できる) で判断して対応するボタンを押す．パタンは中央で停止するが，ボタンを押すとパタンは消失し，次の色票ペアが流れる．モニタには 27inch 液晶モニタ (LG Electronics Japan(株)) を用いた．事前に色校正ソフト (”Spyder3”;Datacolor.Inc) を使って，ガンマ特性指数 2.2・色温度 6500K に設定した．被験者は，研究室の室内照明状況下で，画面から約 1m の位置から画面上のパタンを正視した．被験者は，色覚正常の成人男性 2 名 (20 歳代 1 名，50 歳代 1 名) である．評価値は，θ_n 毎に応答の評価点数 (0, 1, 2) の合計点 (0∼32) である．評点が高いほど，その色相の識別能力が高いことを意味する．

3.3.3 実験結果

Fig.4(d) に色相 θ_n の識別評価値を示す．被験者共通の傾向として，赤紫系 (H=300∼330) と緑 ∼ 青緑系 (H=90∼120) の弁別能力は低く，橙 ∼ 黄系 (H=20∼60) と青 ∼ 紫系 (H=240∼270) は比較的高い傾向にある．また，個人差として Sub.U において加齢による弁別力の低下傾向を認めることができる．

3.4 視覚的母音正規化の検証

母音色彩分布に関する定量的検討について，色知覚による母音認識実験を行った．

3.4.1　実験試料

試料は，(異常試料を含む) 全 3000 試料を発声回数毎 (NO.1∼NO.4) にランダムに並べてリスト (5 リスト ×150 試料=750 試料) を作成し，計 20 リストとした．

3.4.2　実験方法

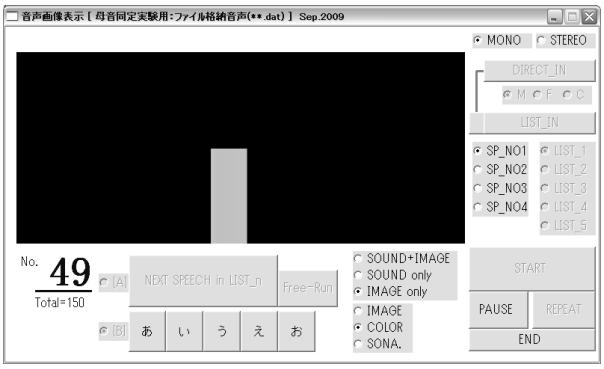


Fig. 6　PC-based visual experimental tool of vowel recognition.

Fig.6 のツール表示画面において，電光ニュース式に単独表示される母音パタンを見て，その音韻をボタンにより回答する．ボタン押下後に次のパタンを表示する．全 20 セッションの所要時間は約 2 時間である．実験条件は，3.3 節の識別実験と同様である．被験者には，母音色彩知覚の経験がない大学生 1 名 (Sub.S) を加えた．Sub.T と Sub.S は，母音と色彩の対応として，Fig.4(c) の色相帯を参照しながら判定した．結果については，正常試料 (2828 試料) についてのみ集計した．

3.4.3　実験結果

Fig.7 に，3 被験者の母音同定率を示す．被験者別の平均認識率は (Sub.U：95.8%，Sub.T:96.2%，Sub.S:88.9%) であった．未習熟の被験者 Sub.S では，/u/→/i/，/a/→/o/ への誤判定が比較的多く見られた．Sub.U と Sub.T では，女声群の/i/→/e/の誤答が見られたが，両者の判断には個人差はほとんど見られず，安定した母音判定が行われ，色彩判定規準が視覚的認識において良好に機能していることがわかった．

4　まとめ

音声可視化手法としての母音色彩表現において，RGB 色空間での母音分布の集約性の

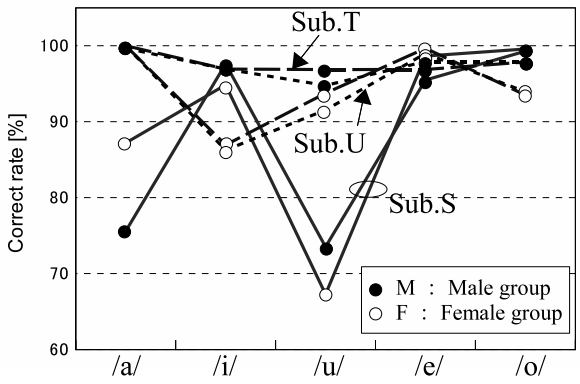


Fig. 7　Results of the colored vowel recognition tests.

良さの客観的評価を行った．さらに，日本語母音の規準色相帯を生成し，母音認識実験により色知覚における音韻正規化を確認した．今後，色弁別の測定結果を踏まえた色補正処理を加えると共に，幼児·児童音声における規準作成，(障害音声を含む) 母音声質と色度の関係などについて検討していく予定である．

謝辞　本研究の一部は，平成 20-21 年度科学技術融合振興財団 (FOST) 研究助成の補助を受けた．

参考文献

[1] 上田裕市 他，”リアルタイム音声画像化処理に基づく発話訓練システムの構築”，信学技報 WIT2007-104，pp.79-84，2008．

[2] 上田裕市 他，”音声応用システムのためのリアルタイム音声特徴推定エンジンの構築”，信学技報 SP2008-67，pp.61-66，2008.

[3] Akira Watanabe，”Formant Estimation Method Using Inverse-Filter”，IEEE TRANS. on Speech and Audio PROC.，Vol.9，pp.314-326，2001.

[4] 三好義昭 他，”2 段階標本選択線形予測法による高ピッチ音声の分析”，信学論 (A)，70(8)，pp.1146-1156，1987．

音声生成過程に基づく音声合成
－重畳モデルによる声道形状変化シミュレーション実験－*

中島　邦久，緒方　公一（熊本大）

1　はじめに

著者らは，人間の音声生成機構を再現するためにSondhiとSchroeterにより提案された声道シミュレータ[1]を計算機上に構築し，Javaによる GUI シミュレーションシステムの開発を進めている．声道形状の時間変化には，縦続一次系関数を用いて滑らかな変化を実現している[2,3].

著者らは，声道による調音運動は大局的な母音連続の調音に，子音の局所的な調音が重畳されており，それらが縦続一次系関数を用いて表現できることを示した[4]が，本稿ではこの重畳モデルを用いて/edade/の合成を行い，運動の重畳タイミングを変えた場合の合成音声への効果についてシミュレーションにより検討した．

2　音声合成システム

本音声合成システムは，Sondhi とSchroeterにより提案された声道シミュレータを基に作成された声道モデルによるものである．著者らの声道モデルは，20個の等長直円筒音響管を縦続接続することによって近似的に形状を表現している．各部位での面積A_n $[n = 1\sim20]$を変化させることによって声道形状を制御している．

縦続一次系関数を用いることで，調音運動の軌跡を良好に近似できるという報告があり[4]，本システムでは時間的な声道形状変化を表現する手段として，音響管の直径の変化，及び各パラメータの時間変化に縦続一次系関数を適用している．Fig.1は，縦続一次系関数を音響管の変化に適用した例であり，時刻t_1に直径d_1である音響管が，時刻t_2に直径d_2へ変化したときの様子を模式的に表現したものである．この音響管の動きが，時刻t_1に入力された縦続一次系関数のステップ応答に相当する．

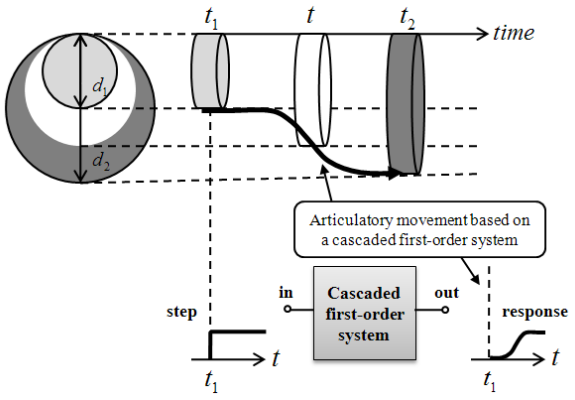


Fig.1 : Change in the area of one acoustic tube as a function of time.

一方，本システムの声帯は2質量モデルによって表現されており，各パラメータによって韻律制御を行っている．これらにも縦続一次系関数のステップ応答を利用して，時間変化を表現している．

3　子音重畳による調音運動の表現

本論文では，声道による調音運動を母音と子音の 2 つに区別し，大局的な母音-母音の連続的な調音運動に，局所的な子音の調音運動が重畳するというモデルに基づいて/VCV/という音節の表現を行った．例えば音節/eda/は，連続母音/ea/という調音運動に，閉鎖子音である/d/が局所的に重畳したものとして取り扱う．

3.1 子音の重み関数

子音/d/は，舌が硬口蓋に接触し声道に閉鎖を形成する．本システムの声道モデルで

*Speech synthesis based on a speech production mechanism　－Simulation of changes in vocal tract shape by superposition model－

By K. Nakashima and K. Ogata (Kumamoto University).

は，閉鎖位置が 20 個の音響管のうち声門から 18 番目の音響管に対応する．閉鎖部周辺の音響管はその影響を受けるが，閉鎖部から離れるにつれその影響が小さくなると考えられる．

本稿では，閉鎖部からの子音の影響が対称であり，閉鎖部から十分遠い音響管に対してほぼゼロであると仮定して，影響の度合いを正規分布でモデル化した．すなわち，Eq. (1)，(2)で表される最大値を 1 に標準化したn番目の音響管の重み関数$w(n)$ $[n = 1\sim20]$を用いた$(\mu = 18)$.

$$w(n) = f(n)/f_{\max} \qquad (1)$$

$$f(n) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(n-\mu)^2}{2\sigma^2}\right\} \qquad (2)$$

なお，著者らが導出した平均的な声道断面積関数[5,6]を参考に$\sigma^2 = 1$とした．

3.2 子音調整機能アルゴリズム

母音変化に重畳する子音の時間的な影響を，2 つの縦続一次系関数のステップ応答の重ね合わせで Fig.2 のように表現する．各ステップ応答の時定数，次数を固定し，子音持続時間，閉鎖時刻を指定することにより，入力時間とステップ応答の大きさを調整し，閉鎖時刻に本システムの閉鎖閾値$(A_{18} = 0.016)$ を下回るような組み合わせを選択した．

4 シミュレーションによる音声合成実験

これまでに述べた重畳モデルを用いて音声合成を行った．合成音は音形/edade/を用い，舌調音運動の計測[4]によって得られた時定数を参考に用いた．合成音声のサンプリング周波数は 20kHz である．母音/eae/は固定して扱い，2 つの子音/d/を重畳した．第 1 子音/d/の重畳と第 2 子音/d/の重畳は，互いが影響しない程度，時間的に離れた状態で，母音遷移の区間中における子音の閉鎖時刻を 25ms ずつ変化させ，それぞれの子音で 8 パターン，計 16 パターンの音声を合成した．

Fig.3 は，/edade/の母音，子音の各入力時刻とそれに伴う 18 番目の声道直径の時間変化を示したものの例である．16 通りの実験条件を，第 1 子音の閉鎖時刻を遅く設定した合成音から便宜的に 1-1，1-2…1-8 と番号付けし，同様にして第 2 子音についても 2-1，2-2…2-8 と番号付けしており，図では 1-6 の例を示している．上部に四角を伴う垂直線は，子音の入力時刻であり，これは Fig.1 のt_1に相当する．同様に下に四角を伴う垂直線は，母音の入力時刻を示している．1-6 の例では下側の四角に 2 で示す母音/a/の入力時刻を基準にして，上側の四角に 2 で示す第 1 子音/d/の閉鎖にむけての入力時刻を-113.8ms に設定したものとなっている．

このシミュレーション実験の評価は，合成音声と実音声のホルマント周波数の時系列データの傾向の比較，知覚的な印象，及びシミュレータ上の声道形状と舌調音運動の計測データとの比較によって行う．

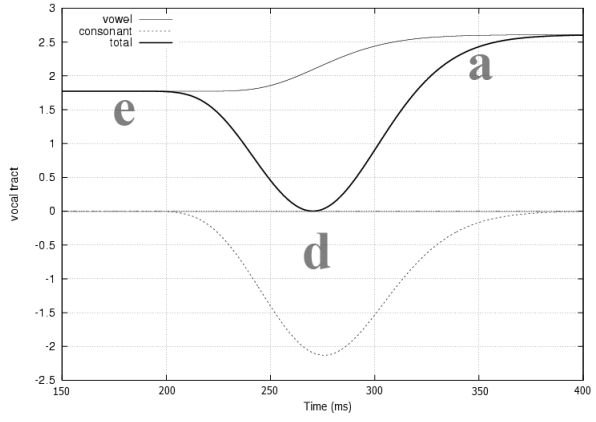


Fig.2 : An example of superposition.

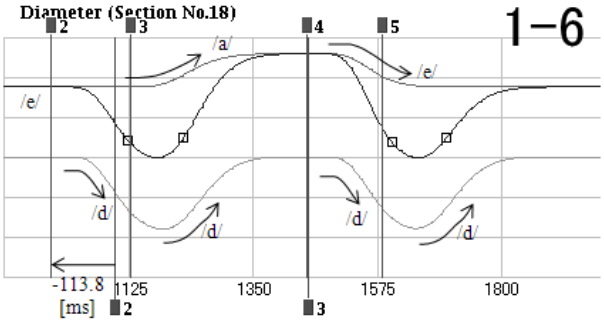
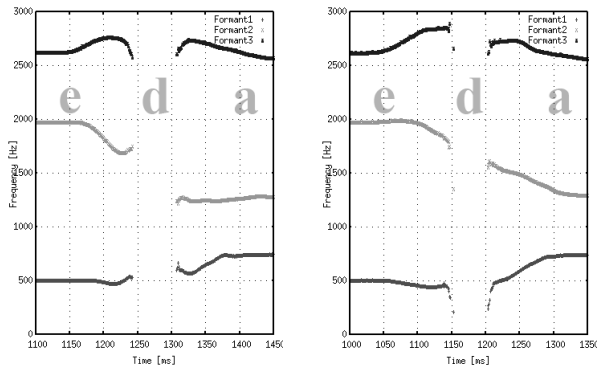


Fig.3 : Change in the diameter of the 18th acoustic tube as a function of time.

Table1：Difference of input time between consonant /d/ and successive vowel, and evaluation score for synthesized speech by 10 listeners.

No.	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8
difference of input time (ms)	+13.7	-5.4	-27.2	-53.7	-84.0	-113.8	-139.9	-163.7
evaluation score	0.9	1.1	0.6	1.3	1.8	1.8	2.0	1.6
No.	2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8
difference of input time (ms)	+59.8	+31.2	-1.7	-41.8	-88.7	-112.5	-114.1	-127.0
evaluation score	0.4	1.1	2.0	1.9	1.8	1.7	1.7	1.3



(a) 1-2　　(b) 1-6

Fig.4：Loci of the first three formant frequencies for /eda/ in synthesis speech /edade/.

5　実験結果と検討

Table1 に各合成音声の重畳子音と後続母音との入力時刻t_1の差と，簡易聴取実験の評価値を示す．ここでは，成人 10 名に対し，合成音声/edade/を「/edade/と聞こえる」，「/edade/と聞こえない，もしくは音に違和感がある」,「どちらともいえない」の 3 段階で評価してもらい，それぞれ 2，0，1 点と置き換えた場合の平均評価値を示している．以下，ホルマントの時間変化の傾向とともにこれらの検討をしていく．

5.1 第 1 子音/d/

Fig.4 に 1-2 と 1-6 を例として，/eda/の部分のホルマント周波数の時間変化を示す．実音声を分析した結果では，一般的に/da/の子音発音後，第 1 ホルマントは後続母音/a/のホルマントへ向かって上昇遷移し，第 2 ホルマントは下降遷移するという傾向がある[7].

子音の入力時刻が後続母音/a/における時刻付近に存在する 1-1～1-4 は第 1 子音発音後の第 1，第 2 ホルマントともに上昇遷移する傾向を示した(Fig.4(a))．評価値も比較的低い数値であり，/ebade/と知覚されることが多かった．これは，1-1～1-4 の入力時刻が遅かったために適切な調音運動が再現されず，結果として/ba/のホルマント遷移に近い値となったためと考えられる．

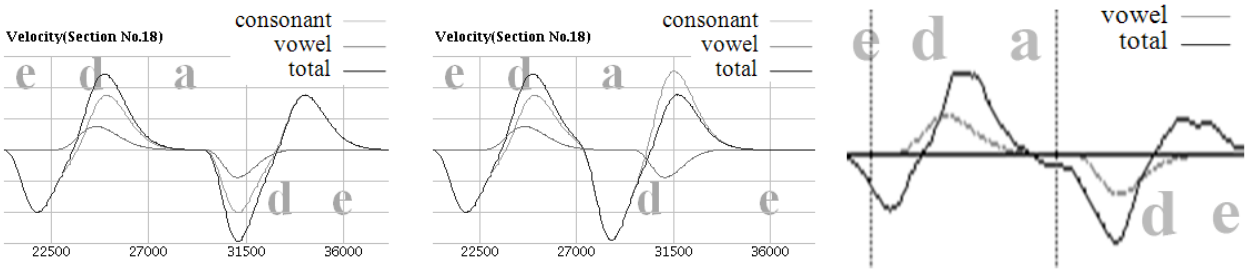
子音の入力時刻が後続母音/a/よりも比較的早い 1-5～1-8 のホルマント周波数は，実音声と似た傾向を示した(Fig.4(b))．聴取実験による評価値も 2.0 に近い値を示し，/edade/と聞き取れていることがわかる．

これらの結果より，/e/から/a/にかけての母音遷移での子音/d/の重畳のタイミングは，後続母音/a/の入力時刻よりも子音/d/の入力時刻のほうが早いほうが妥当であると考えられる．このことは，実際の舌運動をステップ応答で近似した際の傾向[4]と一致している．

5.2 第 2 子音/d/

一般的に/de/の子音発音後のホルマントは，後続母音/e/に対して第 1，第 2 ホルマント両方が上昇遷移する傾向がある[7]．2-1 から 2-8 まで第 2 子音発音後のホルマントの変化の傾向は，8 つすべての合成音で一致した．

一方，聴取実験の評価値は，2-1，2-2 が特に低く，2-3，2-4 でほぼ 2.0 となりそ



(a) 2-3　　(b) 2-8　　(c) articulatory data

Fig.5：Velocity pattern of the vocal tract in the simulation and observed one.

こからタイミングが早くなるにつれ，徐々に減少している．聴取時の感想を考慮すると，2-1，2-2 で評価値が低くなった理由としては，/edaede/と知覚されたことであった．これは，第 2 子音/d/の重畳のタイミングが遅すぎたためと考えられる．また，2-4 以降徐々に評価値が減少していく原因としては，タイミングのずれにより調音運動を適切に再現できなくなっていったためと考えられ，次に声道直径の変化速度に着目して検討する．

Fig.5 は後続母音/a/より前のタイミングで/d/が重畳されたもののうち，評価値の高い 2-3 と低い 2-8 の声道直径の変化速度を表したグラフである．また(c)には測定により得られた舌調音運動の運動速度の時間変化を参考までに示す．ただし，(c)では/edade/における舌部の運動速度と推定された/eae/の速度を示している．(a)と(c)に関しては細かい部分違いは見られるものの，傾向としては非常によく似ている．これは，子音による重畳モデルが縦続一次系関数でステップ応答による近似が良好に行われていることを示唆している．

しかし，(b)と(c)との比較では，第 2 子音/d/の発音前(Fig.5(b)の 1350ms から 1575ms 付近)に母音の速度と母音と子音があわさった声道の速度の変化のタイミングがずれていることがわかる．つまり，第 2 子音の重畳のタイミングが早すぎたために調音運動が適切に再現できずに，合成音に違和感が生じたものと考えられる．

6　まとめ

本稿では，重畳モデルを用いて調音運動の良好な再現を行うことを目的に音声合成シミュレーション実験を行った．取り扱った音形/edade/の第 1 子音/d/では子音の入力時刻を後続母音の入力時刻より早く設定することで，適切な調音運動を再現することができ，合成音声も/edade/と知覚された．この傾向は実際の舌調音運動[4]と一致している．また，第 2 子音/d/の結果より，調音運動を再現するにあたり，その声道変化の速度まで考慮して入力時刻のタイミングを決定することで，より正確な音声合成を行えると考えられる．今後他の母音構成や閉鎖子音等で，同様の結果が得られるかどうかのさらなるシミュレーションが必要である．

謝辞

本研究の一部は科学研究費補助金((C)20560398)の援助によることを記し謝意を表する．

参考文献

[1] M. M. Sondhi and J. Schoroeter, IEEE Trans. Acoust., Speech & Signal Process., ASSP-35 (7), 955-967,1987.
[2] 緒方他，信学技報，SP2004-30, 7-12, 2004.
[3] 緒方，増矢，音響学会誌，vol.62, no.3, 199-207, 2006.
[4] 緒方，園田，音響学会誌，vol.55, no.3, 156-164, 1999.
[5] 緒方，大塚，音講論集（秋），165-166, 2006.
[6] K. Ogata and B. Yang, Proc. of 19th International Congress on Acoustics, CD-ROM CAS-03-010, 2007.
[7] P. Delattre, A. M. Liberman, and F. S. Cooper, J. Acoust. Soc. Am., vol.27, 769-774, 1955.

全帯域インパルス応答からの低周波数帯域IACC導出時の留意点について*

松本博樹，近藤善隆，末廣一美（日本文理大），今井佐智代，岩上知広(千葉工大)，
福島学(日本文理大)，柳川博文（千葉工大），黒岩和治(日本文理大)

１．はじめに

低周波域は波長が長いことからホームシアタやオーディオ機器ではモノラルで扱われている.しかし,D. Griesinger[1], W. Martens[2]らは音楽の録音再生には低周波用のチャネルが 2 つ必要であると述べている.K. J. Gabriel[3]らおよび, N. I. Durlach[4]らは両耳間相関係数の弁別を調べており，M.Morimoto[5]らは低い周波数の効果を調べているが,100Hz 程度以下の低周波の音については調べていない．著者らはこれまでに，音場の「拡がり感」は拡散音場の 2 点の音圧の相関係数と両耳間相関係数が近い時に顕著に感じられることが報告している[6]．これがどのような実音場で感じられるかを調べるには，低周波域の音響伝達特性を計測し，そこから両耳間相関係数を導出しなければならない.

音場の信号伝播現象はインパルス応答で表現されているため，両耳に相当する位置で計測したインパルス応答から IACC を求めることが出来る．インパルス応答推定は，線形時不変を前提とした間接計測で行われるが,「拡がり感」を感じるようなコンサートホールでは必ずしもこの前提条件を満足した条件で計測できるとは限らない．特に低周波数域では，計測信号に十分なパワーが得られない場合には，正しい応答を計測しきれていないことや，暗騒音を十分零に収束しきれない場合がある．しかし，未知の値を推定しているため，例えば平均回数から予想し分析するしかなく，結局何を分析しているのかわからない場合も生じる.

本稿では低周波域において IACC を導出する際に，不適切な信号を使うと結果がどうなるかについて報告することで，IACC 導出時における留意点を述べる.

２．拡散音場における両耳間相関係数

拡散音場において距離 r にある 2 点の音圧の相関係数は

$$\rho = \sin(kr) / kr \quad (1)$$

where, $k = \omega / c = (2\pi f) / c$,

で算出される.両耳相当の 距離 r = 0.32m で ρ を求めた値は両耳間相関係数に近似しておりその値を Fig. 1 に示す.

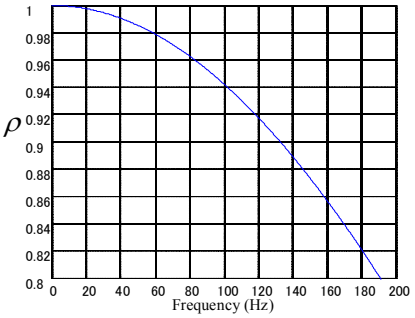


Fig.1 The theoretical correlation coefficients
(r = 0.32 m c=340m/s)

３．実音場インパルス応答からの導出

室容積 191lm^3 で残響時間が約 1 秒の扇状階段教室で計測したインパルス応答を Fig.2 に示す.図は縦軸に推定インパルス応答の振幅絶対値を dB で示し，横軸に時間を示している.Fig.2 はインパルス応答の立ち上がりが-60dB 程度得られていることから適切な計測が行えているように見える．また減衰パタンを見ても，直達音領域以降の減衰が指数関数的に減衰しており，ノイズフロア付近まで減衰しているように見える．全帯域で IACC を導出すると 0.44 であり，室容積や残響時間等の音響物理指標から妥当な IACC が導出さ

* The careful point in calculation of low frequency range IACC , by MATSUMOTO Hiroki, KONDOU Yoshitaka, SUEHIRO Kazumi (Nippon Bunri Univ.), IMAI Sachiyo, IWAKAMI Tomohiro (Chiba Inst. of Tech.), FUKUSHIMA Manabu (Nippon Bunri Univ.),YANAGAWA Hirofumi (Chiba Inst. of Tech.), KUROIWA Kazuharu (Nippon Bunri Univ.)

れていると判断できる．そこで，このインパルス応答に対して中心周波数が100Hz 程度の狭帯域通過フィルタを通し，その結果から IACC を導出する．IACC は過渡的両耳間相関関数（TRICC）の範囲がインパルス応答全体になった時点の値と同じである．このことから，ここでは過渡的両耳間相関関数を求めた．その結果を Fig.3 に示す．図は縦軸に TRICC を示し，横軸に TRICC 導出時の窓長を示す．

Fig.3 は直達音領域で音像が定位するため TRICC が低い範囲があり，その後複数方向から反射波が到来することで TRICC が高い値となる過程をあらわしており，妥当な推移のように見える．しかし，Fig.1 に示した拡散音場の IACC の予測値と比較すると，ここでは約 100Hz の狭帯域を使用していることから IACC が拡散音場でも 0.94 程度までしか低下しないにもかかわらず，0.78 まで低下してることがわかる．この原因を調べるために，狭帯域通過フィルタを通したインパルス応答を Fig.2 と同様に振幅絶対値を dB 表示として確認すると，Fig.4 のようになっており，直達音領域の S/N は十分確保されているものの，十分減衰するまで計測されていないことがわかった．そこで対象とする周波数帯域で十分減衰するまでのデータから再度 IACC を導出したところ，0.99 が得られた．

４．おわりに

低周波数域での IACC を導出するには，直達音領域で S/N を確認するだけでなく，IACC が何を導出しているのかを理解したうえで，それに大きく寄与する減衰パタンが正しく計測できることを確認することの重要性を改めて確認することができた．

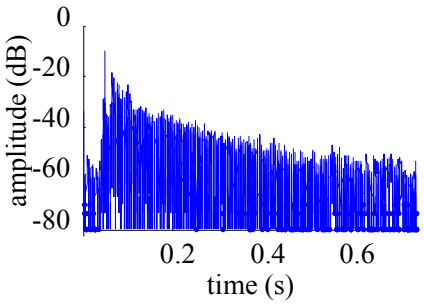


Fig.2 The estimated impulse response

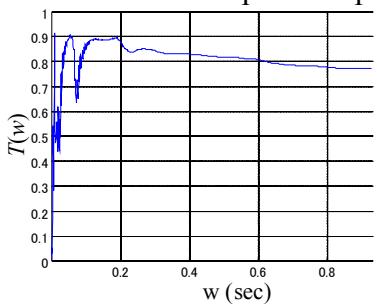


Fig.3 The TRICC (build up part) in low frequency range (narrow banded signal)

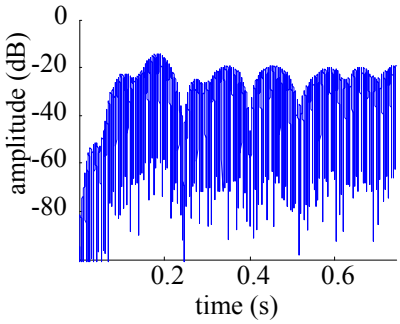


Fig.4 The estimated impulse response (narrow banded signal)

［参考文献］

[1] David Griesinger, "How Many Loudspeaker Channels are Enough? ", 109th AES convention, October 2000.

[2] William L. Martens, "The Impact of Decorrelated Low-Frequency Reproduction on Auditory Spatial Imagery: Are Two Subwoofers Better Than One?", AES 16th Internatuinal Conference on Spatial Sound Reproduction.

[3] K J Gabriel, et al. , "Interaural correlation discrimination: I. Bandwidth and level dependence", J. Acoust. Soc. Am. 69(5), May 1981, pp. 1394-1401

[4] N. I. Durlach, et al. , "Interaural correlation discrimination: . Relation to binaural unmasking", J. Acoust. Soc. Am. 79(5),May 1986, pp. 1548-1557

[5] M.Morimoto and Z.Maekawa, "Effects of Low Frequency Components on Auditory Spaciousness", ACOUSTICA, Vol.66, pp.190-pp.196, 1988

[6] H. Yanagawa, et al. , "Interaural Correlation Coefficients and their Relation to the Perception of Subjective Diffuseness", ACUSTICA, Vol. 71,1990, pp. 230-232

唇および口周辺の面積変化による数字音声認識の基礎的検討*

○小川佑輝，秋田昌憲，緑川洋一（大分大）

1　はじめに

これまでの研究では口唇画像、つまり唇の動きだけを使い認識実験を行ってきた。しかし、音声を発声する時、唇は非常に素早く細かな動きをしている。そのため、同じ話者でも 3 回の口唇画像の撮影で動きの差が生じてしまう。そこで、本実験では唇だけでなく、口周辺画像を用いることにより認識の向上になると考えた。

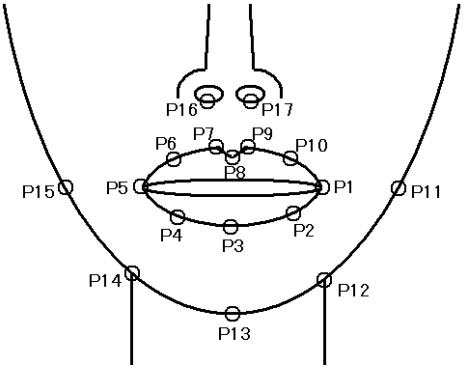


図 1　唇と口周辺の測定部位の設定

2　唇および口周辺画像とパターン認識

2.1 唇および口周辺画像

唇および口周辺画像とは、唇と口周辺を動画像で撮影することにより得られる画像情報である。本研究では、この画像情報より唇と口周辺の動きに着目し、動きの時系列による変化から特徴抽出を行うこととする[1]。

唇と口周辺の動きの変化は、座標の変化によって得られる。この際に、測定する点を自動追尾することにより座標データを取得する。

次に動画像からの唇と口周辺の特徴抽出を行う。それぞれ採取した動画像は、5 秒間で、150 フレームの画像変化より唇と口周辺の特徴を得ることになる。唇と口周辺の動きの変化を抽出するために、運動解析ソフトウェア Dipp-Motion 2D を用いた[2]。唇と口周辺の動きの抽出においては、図 1 に示すように唇は外周部に等間隔に 10 箇所を、あごは下あごに 5 箇所と鼻に 2 箇所の計 7 箇所を、唇とあごの合計 17 箇所を部位設定したのち自動追尾を行い、設定した部位を結んだ画像を作り出し、追尾して得られた座標データより面積変化量を得る[3]。

3　実験

3.1 使用したデータ

雑音が少ない静かな室内において 5 人の男性話者に数字音声を発声してもらい、デジタルビデオカメラを用いて唇および口周辺の動画像を撮影する。また、話者が発声した単語音声は 0/zero、1/ichi、2/ni、3/san、4/yon、5/go、6/roku、7/nana、8/hachi、9/kyu の数字音声で合計 10 個である。本実験では、それぞれ話者の 10 個の数字を 3 回ずつ撮影し、合計 150 個のデータを用いて唇および口周辺画像認識実験を行う。

3.2 実験方法

まず、1 回目を基準として 2 回目の 10 数字とのユークリッド距離を算出した後それぞれ比較して、その距離が最小の数字音声の認識結果とする。次に、1 回目を基準として 3 回目の 10 数字のユークリッド距離を算出し同様に比較する。このような認識を、次は 2 回目を基準にという具合に繰り返し行う。認識率は、基準の 0/zero ～9/kyu を正しく認識した数の合計を総データで除算して求める。話者一人分の総データ数を $10m$、正しく認識した数の合計を rec とすると、次式から認識率を求めることができる。

* Basic examination of figure voice recognition by changing the are of lip and around mouth, by Yuki OGAWA, Masanari AKITA and Yoichi MIDORIKAWA (Oita University).

$$認識率 = \frac{rec}{10m \cdot (m-1)} \times 100 \quad [\%] \quad (3\text{-}1)$$

なお本研究では、基準を含めた 10 数字を 3 回分用いたので $m=3$ である。この際に用いる特徴パラメータは、本研究で撮影した動画像から抽出した唇とあごの面積変化量を用いることとする。

次に認識箇所について検討する。図 1 に示すように、唇とあごの合計 17 箇所の部位から P1～P10 で囲まれた面積を唇として、もう一つは P11～P17 で囲まれた面積をあごとして認識箇所を決定する。

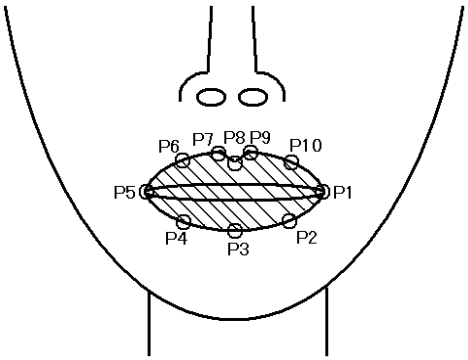


図 2　唇の認識箇所

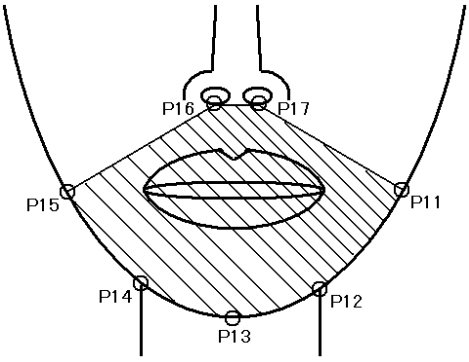


図 3　あごの認識箇所

3.3 分析範囲

本研究では 2 種類の分析範囲を用いて行う。

まず、面積変化の最大値を中心に $\pm x$ 間（$2x$）の分析範囲を指定し均等に n 分割して（$x+1$）フレームを分析する方法で行う。特徴量として、面積変化量を用いて最大値により正規化を行い、パターン認識により認識率を求める。分析範囲には $2x=40$ に指定して行う。

次に、音声波形の始端と終端の間を均等に n 分割して（$x+1$）フレームを分析する方法で行う。特徴量として、面積変化量を用いて最大値により正規化を行い、認識率を求める[4]。

4　認識実験

2 種類の分析範囲の方法により、唇、あごの面積変化について検討する。

図 4、5 では、面積変化の最大値を中心に $\pm x$ 間（$2x$）の分析範囲を指定し均等に n 分割して（$x+1$）フレームを分析する方法で、唇、あごについての認識結果を示す。

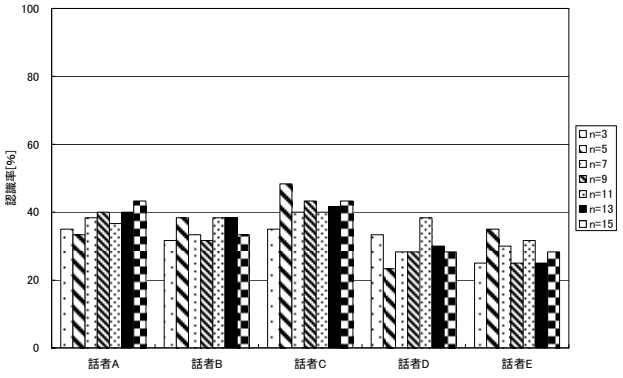


図 4　面積変化の最大値を中心に $\pm x$ 間（$2x$）の分析範囲を指定し均等に n 分割して（$x+1$）フレームを分析した結果（認識箇所は唇、$2x=40$）

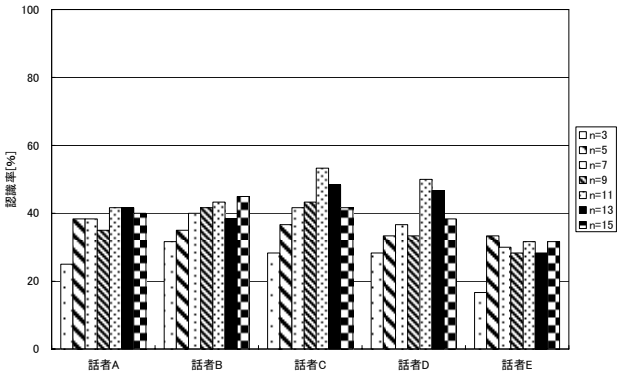


図 5　面積変化の最大値を中心に $\pm x$ 間（$2x$）の分析範囲を指定し均等に n 分割して（$x+1$）フレームを分析した結果（認識箇所はあご、$2x=40$）

図 6、7 では、音声波形の始端と終端の間を均等に n 分割して（x+1）フレームを分析する方法で、唇、あごについての認識結果を示す。

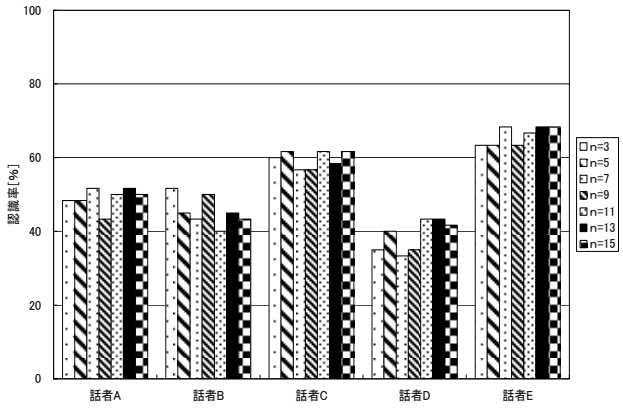


図 6　音声波形の始端と終端の間を均等に n 分割して（x+1）フレームを分析した結果（認識箇所は唇）

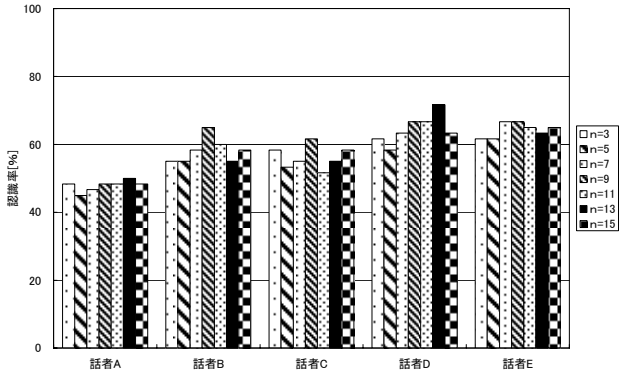


図 7　音声波形の始端と終端の間を均等に n 分割して（x+1）フレームを分析した結果（認識箇所はあご）

口から音声を発声させようとすると、唇やあごを動かし、面積を変化させる。しかし、同じ話者が同じ音声を発声させようとしても、毎回全く同じ動きをすることは非常に困難である。その為、同じ音声でも特徴パラメータにも違いが生じてしまう。この違いが、認識方法の差につながったものと考えられる。その例として、図 8、9 に話者 D の唇の特徴パラメータを示す。

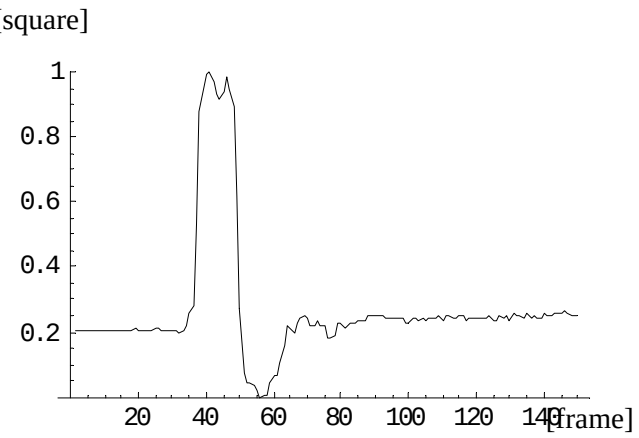


図 8　話者 D 1/ichi の 1 回目結果（唇）

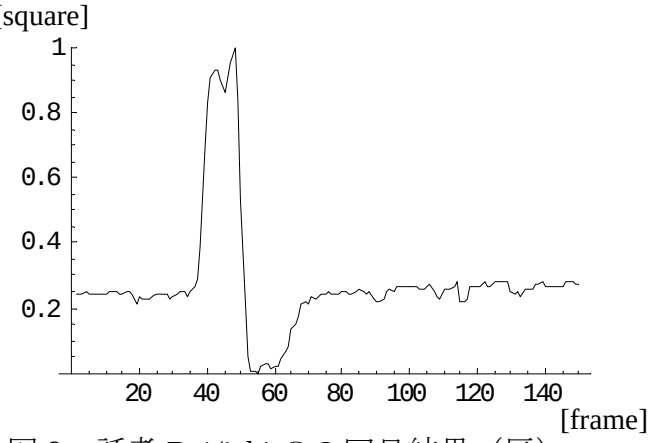


図 9　話者 D 1/ichi の 2 回目結果（唇）

この唇の特徴パラメータを比較すると、最大値が同じではないことが分かる。この違いが認識率の差につながったものと考えられる。

また、特に唇やあごの動かし方に癖の強かった話者 C は全体的に認識が高いが、動かし方にはっきりとした癖を見ることのできない話者 D では、分割数を増加させても大きな認識の変化は表れなかった。話者の癖の動きが認識率の差に影響しないような方法の検討が必要になる。

次に、音声波形の始端と終端の間を n 分割して（x+1）フレームを分析する方法について検討する。

図 6、7 の結果を見てみると、最大値を利用した方法に比べると全体的に認識率が向上している。これは、音声波形から始端と終端を決めていることが要因であると考えられる。その例として、図 10 に話者 D 1/ichi の音声波形を示す。

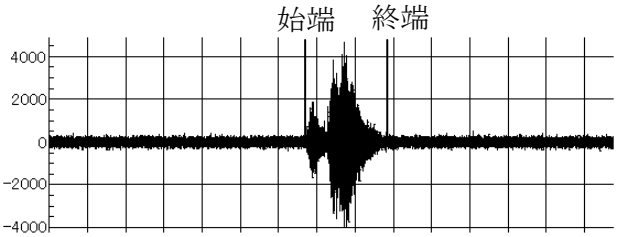


図 10　話者 D 1/ichi の音声波形

しかし、今回は音声を使わずに画像のみで認識を行うようにしている為、音声波形の始端と終端を目測で行っている。これに関しては、音声波形も含めた認識方法の検討が必要であると考えられる。

5　まとめ

本研究では、音声認識の雑音環境下での認識率の低下に対して、画像情報による認識を行うことにより改善を図るという目的のもと、唇および口周辺画像を用いることが有用であるか検討を行ってきた。

分析範囲の決定方法においては、面積変化の最大値を中心に$\pm x$間（$2x$）の分析範囲を指定し均等に n 分割して（$x+1$）フレームを分析する方法、音声波形の始端と終端を決定し、その範囲を均等に分割して分析する方法の 2 種類を行ってきた。その結果、特徴パラメータが安定してない話者は高い認識率が期待できない為、分析範囲は音声波形から求める方法が最適だと考えられる。

また、唇とあごの認識率を比較すると、あごの認識率が向上していることから、これから認識箇所についても検討して行かなければならない。

本研究において認識に関しては画像データのみを用いて行っているものであるので、音声、画像の両方のデータを交えての認識を行うことができるようになれば、本研究としての本来の目的は達成されるものと考える。

参考文献

[1] 南 敏、中村 納共著「画像工学 －画像のエレクトロニクスー」コロナ社、pp131-143 (1989)
[2] 運動解析ソフトウェア「Dipp-Motion 2D User's Manual Ver.3.13」ディテクト (2003)
[3] 南 敏、中村 納共著「画像工学 －画像のエレクトロニクスー」コロナ社、pp131-143 (1989)
[4] 古井 貞熙著「新音響・音声工学」近代科学社、pp175-179,216-224 (2006)

母音性音素を用いたスペクトル強調法の検討*

○吉田亮平　秋田昌憲　緑川洋一（大分大学）

1　はじめに

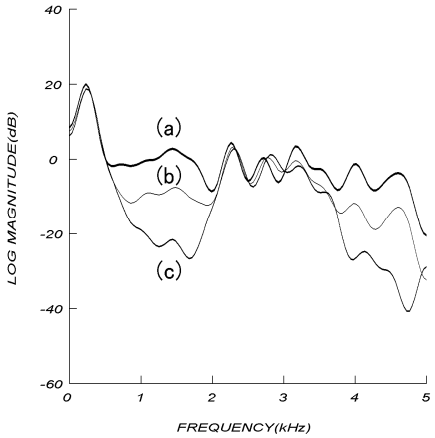
雑音環境下では音声の認識率の低下は必然である。それは雑音が音声に重畳されると，低レベル部でのレベル上昇を起こして，スペクトル包絡全体が平滑化し谷が埋もれてしまい，元の音声の周波数特徴が失われ，認識率低下の原因となる。

雑音環境下での認識率の改善は，これまでいろいろな方法で行われてきた。一般的には，スペクトルピーク強調法(1) が用いられている。また他に時変周波数軸変換法(2)というものがある。これは音声の特徴が母音部，つまり低周波数領域や有声部によく表れることから，周波数軸変換によって低周波数領域を強調し，認識率の改善を図るものである。さらに，スペクトル包絡の埋もれた低レベル部の谷を回復するスペクトル規則変形法(3)もある。これらの方法を使うことで認識率は改善できるが，認識率をよくするために過度な強調を行い，スペクトルが乱れ元の音声データの原型すら分からなくなることもある。またそんな過度な強調によって偶然に認識率が向上することもある。さらに，しきい値の設定によって誤付加が生じることもあるなど問題があり十分でない。そこで本論文では，男性５人の５母音を利用して，母音の特徴を表す母音平均スペクトルを作成し，これらを用いて雑音重畳により平坦化されたスペクトル包絡を変形させ，有声部の特徴を強調し認識率の向上を試みる。母音平均スペクトルを用いる試みは一部でなされている(4)が, その検討は十分でないため再検討し，さらに母音の特徴だけでは補えない欠落した谷を回復するために，スペクトル規則変換法(3)を併用しての実験も行った。

2 雑音を付加した数字音声認識

2.1　雑音付加音声の周波数の特徴

音声に雑音が付加されると，図１のように低レベル部のレベル上昇によりスペクトル全体が平坦化し，スペクトル包絡上で音声の特徴である谷の部分が欠落し，認識率の低下の原因となるのである。谷の欠落は，いわゆる母音の特徴の欠落ともつながるのである。



(a) Pink 0 dB　(b) Pink 10 dB　(c)Clean
図１　スペクトル包絡図

また, 雑音重畳音声は SN 比 0 dB, 10 dB のピンクノイズ及び自動車ノイズの計４種類のノイズを付加した音声を用いている。雑音を付加した音声の無変化状態の認識率を以下の表に示す。標準パターンは無雑音とする。

表１　無雑音を標準にした時の各認識率

	Pink		Mobile	
Recognition rate[%]	0 dB	10 dB	0 dB	10 dB
	15.9	28.3	24.3	54.1
Average[%]	30.6			

2.2　スペクトル規則変形法

音声の有声部のスペクトルにおける欠落した谷の規則による回復を図る。雑音

* Examination of the spectral emphasizing method using spectral parameters of vocalic phonemes, by Ryohei YOSHIDA , Masanori AKITA and Yoichi MIDORIKAWA (Oita University).

が重畳された母音音声音素のスペクトル包絡の変形の特徴を考慮し、２つの谷を補充する。本実験では、しきい値を TH1，TH2，TH3，TH4，TH5 と設定し変形を行った。以下にしきい値を用いて，谷の付け方を示す。

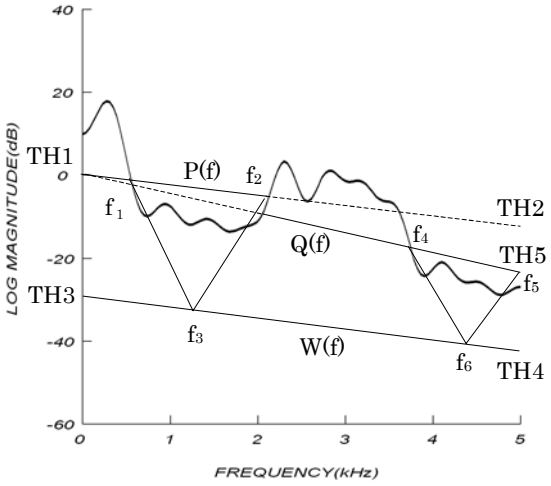


図２　スペクトル規則変形法

2.3　母音平均スペクトル

母音平均スペクトルは，母音である /a/,/i/,/u/,/e/,/o/のデータを用いて，各々母音のスペクトル包絡の時間変化図から，それぞれスペクトルの乱れが少ない有声部の中央付近より１フレームを抜き出し，ケプストラム次数ごとに平均を求めることで得られる。

得られた値を母音の平均ケプストラム *cm(i)*とし，元の音声ケプストラムを *c(i)*，強調後のケプストラムを *co(i)*とすると，

$$co(i) = c(i) + H \cdot cm(i) \quad (1 \le i \le 25) \quad (1)$$

で表され，*i* はケプストラム次数，*H* は強調係数である。なお，今回の実験では母音平均スペクトルは元の音声スペクトルの有声部のみに付加することとする。

3　実験方法

3.1　母音平均スペクトルを用いた音声強調

雑音重畳部では，母音性音素のスペクトル包絡の谷の部分が埋まって，スペクトル包絡全体が平坦化される。そこでそれを補正する意味で，母音平均スペクトルを用いて，スペクトル包絡の傾斜強調を行う。雑音を付加した認識率の低い音声データに母音平均スペクトルを，ケプストラム次数ごとに足し，スペクトル包絡を変形することで，雑音で失われた音声の特徴を強調する。

母音平均スペクトルは，母音のスペクトル包絡の時間変化図からそれぞれ有声部中央付近の１フレームを抜き出し，ケプストラム次数ごとに平均を求めることで得られる。また，実験は無雑音を標準として認識を行った。今回使用した音声データは，男性５人の５母音である /a/,/i/,/u/,/e/,/o/計 25 データをもとに作成し，音声データは雑音を数字音声に擬似的に付加したものを使用した。作成した母音平均スペクトルは，５母音,/a/,/i/,/u/,/e/,/o/と母音の音響的特徴が近い/a,o/,/i,e/の８種類を作成し，雑音を付加する前の音声を基準として，それぞれに母音平均スペクトルを合成した雑音音声を使い実験を行う。図３はその８種類のうちの１つ，５母音の母音平均スペクトルの包絡図である。

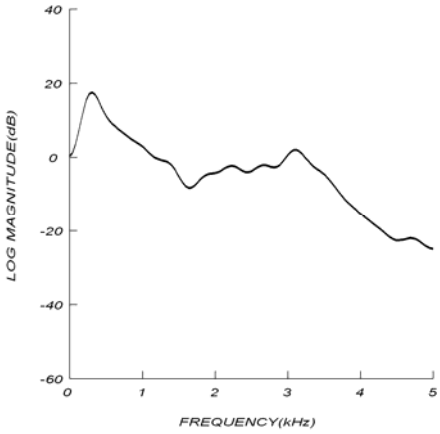


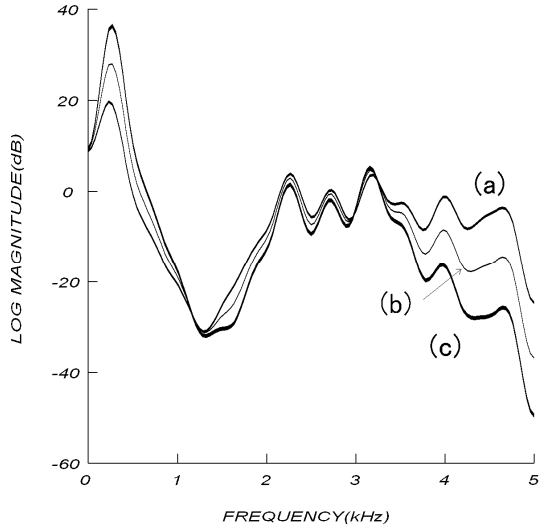
図３　５母音平均スペクトル
(男性５人の５母音：計 25 データ)

3.2　スペクトル規則変形法との併用

母音の特徴だけでは補えない欠落した谷を回復するために，スペクトル規則変形法との併用を考えた。雑音が付加された音声に母音平均スペクトルを用いて変形させていたが，今回の実験は元の音声データにスペクトル規則変形法を用いて，先に音声の谷の部分を強調させ，それに母音平均スペクトルを用いて変形させる。また認識において標準パターンを雑音の

ないデータに谷を付加させたものとしている。これは雑音が重畳して谷が埋もれた音声データに，先に谷を強調して少しでも谷を回復させ，母音を足すことによって，認識率の改善に最適かどうかの検討である。

今回谷付けした音声データに用いたしきい値を以下の表２に示す。また，無雑音に谷を付けたものを標準パターンとし，それぞれの認識したものを同様に以下に示す。図４はピ ンクノイズ 0 dB を付加した音声データに，しきい値(TH1=0，TH2=-10，TH3=-30，TH4=-40，TH5=-20)で谷付けし，５母音の母音平均スペクトルを足したものである。



(a) H=0 (b) H=0.5 (c) H=1
（使用母音データ： 5 母音）
（付加されている雑音：Pink 0 dB）
図４　谷付け後母音平均スペクトルによる変形例

4　実験結果・考察

4.1　母音平均スペクトルを用いた認識実験

図５は８種類の母音平均スペクトルにおいて，最も認識率に改善が見られた強調係数を（）内に示し，認識率の平均値を示した。また，無変形での認識率はClean(0)で示している。平均はピンクノイズ，自動車ノイズそれぞれ 0 dB, 10 dB の平均値である。５母音平均スペクトルの強調が一番改善され約 10(%)の上昇である。しかし，/a/,/i/,や/a,o/,/i,e/の母音は認識率は上がっているものの，改善は著しくない。またすべての母音平均スペクトルに言えることは，自動車ノイズの認識率の改善がどれも乏しかったことである。

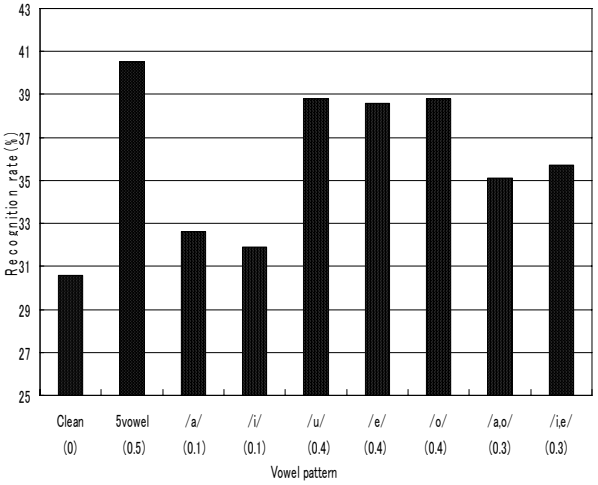


図５　各母音平均スペクトルを付加した時の認識率の比較

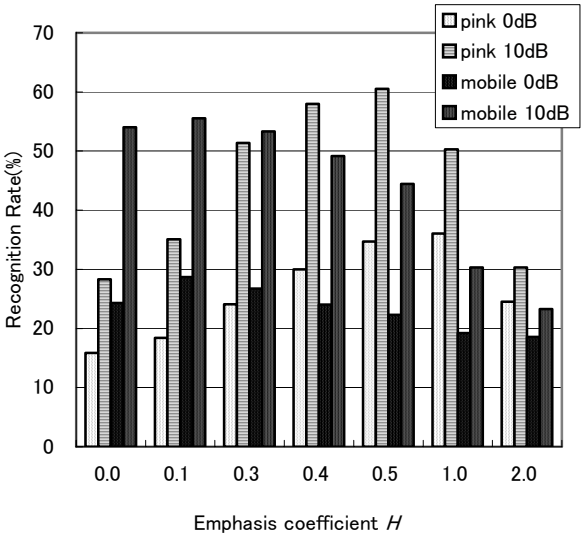


図６　強調係数変化時のノイズ別認識率（５母音平均）

ノイズ別に最も認識率が良かった５母音（*H*=0.5）においても，自動車ノイズに関してはほとんど改善が見られていない。さらに，自動車ノイズ 0 dB においては認識率の変化がほぼないに等しく，10 dB においても，強調係数を上げていくにつれて認識率は低下していく傾向にある。これはどの母音の平均スペクトルを用いても同様で，0 dB はほとんど変化はなく，10 dB についてはほとんど下がる一方であった。これは第１、２ホルマントの間の谷に母音平均スペクトルがうまく重な

ってないことが原因の一つとして考えられる。逆に，ピンクノイズに関しては，強調係数を上げていくにつれて認識率が上がる傾向がある。これはピンクノイズが自動車ノイズに比べて，高周波領域でのレベル上昇が大きいため，変形を行うことで高周波領域でのレベルの下降による強調効果が高いからであると考える。

4.2　スペクトル規則変形法との併用を用いた認識実験

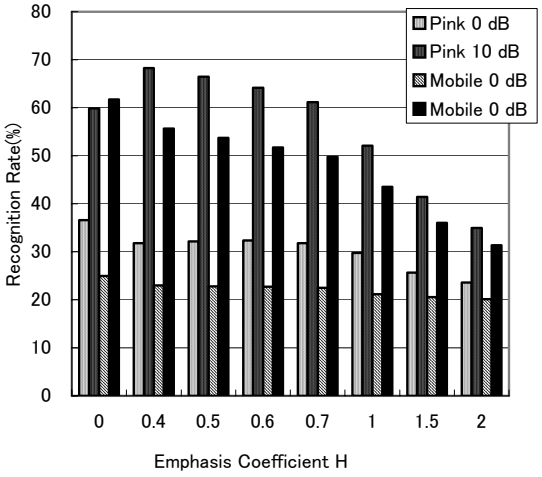


図７　強調係数変化時のノイズ別認識率（谷付け＋５母音平均）

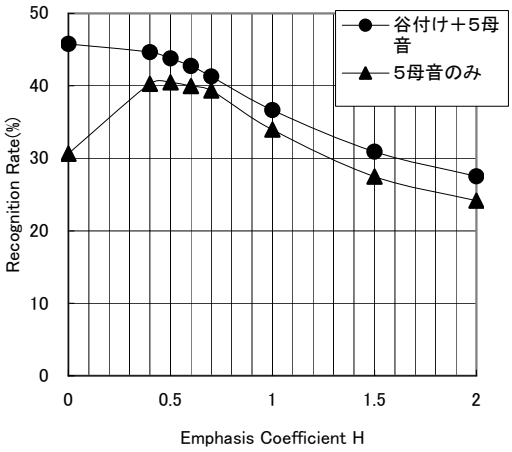


図８　強調係数変化時の平均認識率

図７よりノイズ別に分けるとピンクノイズ0 dBと自動車ノイズ0 dBはほぼ認識率の変化がない。また残りの二つのノイズも認識率の改善が見られない。谷付けのみで母音を足さない（*H*=0）音声の認識率の方が高いことから，母音平均スペクトルを足すことで，谷の位置がずれる原因になってしまったと考える。また図８は一例として，５母音平均スペクトルのみとスペクトル規則変形法を併用した場合の認識率の平均値の比較である。

図８からわかるように，５母音のみの付加に比べて谷付けと母音平均スペクトルを併用した方が認識率は向上している。しかし，併用した方だけでみると，母音平均スペクトルを足していくにつれて，認識率が低下していくのがわかる。これは今回用いたしきい値によって，スペクトル包絡に谷の付き方が悪く，それにさらに母音による変形により，標準パターンとはほど遠い形になったと考える。

5 まとめ

今回母音平均スペクトルを８種類作ったが，ピンクノイズ，自動車ノイズ共に５母音の母音平均スペクトルを足すことが，認識率の改善に一番効果があった。しかし，母音平均スペクトルだけでは，認識率の改善に限界があると考える。そこで，スペクトル規則変形法や他の変形方を併用することが認識率の改善につながると考える。スペクトル規則変形法においては，しきい値を変えることで，よりよい谷付けの出来たスペクトル包絡を用いて，母音を足すことによって，自動車ノイズも認識率の改善があるのではないかと考える。

参考文献

(1) 秋田，有川：“スペクトル強度軸の非線形変換による雑音環境音声認識”，音響学会春季講演論文集，3-Q-4，pp.217-223，1999
(2) 秋田，玉井，緑川：“時変低域強調を用いた音声認識における雑音環境への対応”，信学技法，EA96-61，1996
(3) 秋田，大倉：“雑音環境におけるスペクトル変形回復の一方法”，電気通信学会技術研究報告，EA95-57，1995
(4) M.Akita：“emphasis of the spectral feature parameters for signals polluter with noise using cepstral coefficients”，Journal of Technical Physics, Vol.39,3-4,pp.377-384,1998

入眠予兆のための体内音測定*

○坂口正和，秋田昌憲，緑川洋一（大分大）

1　まえがき

近年、疲労状態や居眠りによる機械、自動車等の運転による事故が多発している。現代の社会でこの問題は非常に重要な課題となっている。その防止対策の一つの入眠予兆に関して、脈拍、呼吸、心拍数、脳波など様々な分野で広く研究がすすめられている[1]。

最近の研究により人体の脳波や脈波、筋電図等と入眠の間の相関関係を示す例が見られるようになってきている[2]。しかし、これらは人体に大掛かりな装置を取り付けておこなう方法であり、実用化することは困難である。実用化するには、このシステムの簡易化が必要になってくる。

本研究室では、体内音の音響的信号から得られるパラメータを利用して入眠予兆信号の検出ができないか検討してきた[3]。測定は、肉伝導マイクロホンを用い、胸、わき腹に取り付けて行った。着座時の臀部における圧電センサ信号の測定も同時に行い、その測定結果と照らし合わせマイクロホンによる信号の検出を行った。従来の研究で、入眠区間でケプストラムの時間変化にしきい値を定めると、しきい値を超える点数が増加することがわかった。しかし、測定データが少なく、入眠予兆信号の傾向がはっきりしていなかったので、測定回数や被験者を増やし特徴を検出する。

ケプストラムに変換したデータから、スペクトル包絡を描いたが、これでは入眠予兆の検出が不十分だった。

そこで、ケプストラムを時間変化にまとめてからしきい値を定め、そのしきい値をこえたものを入眠予兆信号としてとらえ検討することにする。また、ケプストラムパラメータを用いたセンサ信号の特徴抽出におけるケプストラム次数の選択のしかたによっても特徴の現れ方が違うのでその比較についても行う。

2　測定方法および解析方法

2.1 測定方法

非可侵の生体反応を抽出する方法として、着座時の臀部における圧電センサ信号の情報が有用であるという報告がなされている。臀部の圧電センサ信号は、有色雑音信号と類似した波形形態を示しており、また前述の報告で信号の周波数解析を行った場合、低周波数領域の特徴が呼吸状態に関係することが示される。そのため、センサの信号処理に本方法を適用して、低周波数域部の特徴変化を強調することが有用ではないかと考えられる

雑音の少ない部屋で、圧電フィルムセンサ（縦 155mm、横 18.5mm、厚さ 55 μm）を取り付けた椅子に着座し、丸型のマイクロホンを左胸部に喉元から 70mm の胸側 100mm の位置に、左腹部の喉元から 70mm の胸側 370mm の位置に取り付け測定した。マイクロホンには 100Hz の LPF を通している。

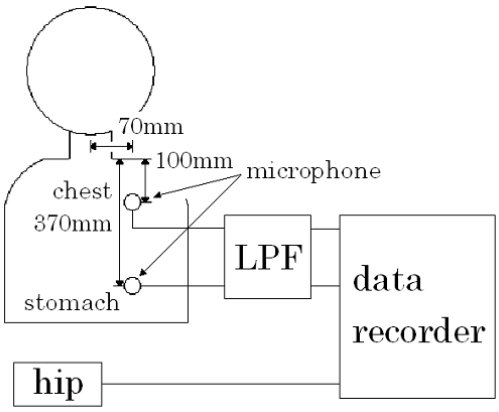


図 1　マイクロホンの装着箇所

* The measurement of the sound signals in the human body for prediction of sleep in sleep-wake state, by Masakazu SAKAGUCHI, Masanori AKITA and Yoichi MIDORIKAWA.(Oita University)

図 2　測定に使用したイス

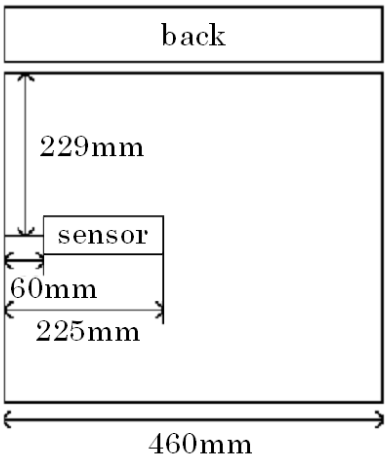


図 3　圧電センサの位置

ここでは、男性 3 名で合わせて 9 回の着座入眠実験を行った。なお、サンプリング周波数は 200Hz である。測定時間は 40 分間とした。

測定時には、被験者が体勢を変えたり、動いたりした場合の時間をメモするための観測者をつけた。入眠区間を決定する方法としては、測定中に被験者の体が無意識に動いたり、頭が下がったり、呼び掛けに対して反応が無かった点、測定後の聴取によって入眠地点を推定する方法を用い、臀部の測定結果と照らし合わせ検討する。

2.2 解析方法

200Hz でサンプリングされたデータについて、音声信号と同様フレーム長およびフレーム周期 1.28 秒で改良ケプストラム分析を行い 25 次のケプストラムに変換する[4]。そして、それらのスペクトル包絡の時間変化を描く。

スペクトル包絡の時間変化のままではわかりにくいので、次にケプストラムの時間変化について考える。ケプストラムの時間変化にしきい値を定めそのしきい値を超えたものを入眠予兆信号とする。

3　ケプストラムによる特徴パラメータ

ケプストラムにより入眠予兆の有無を判断するときは、ν フレームの i 次のケプストラムを $c_\nu(i)$ とすると、ケプストラムの和 P_ν は

$$P_\nu = \sum_{i=n}^{m} c_\nu(i) \qquad (1)$$

となる。

以下に、ケプストラム次数の選択（n, mの値）のしかたによってケプストラムの和が表す部分が変化する例を示す。

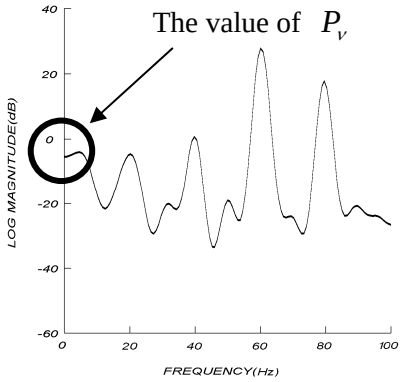


図 4　スペクトルの原包絡（n=0,m=25）

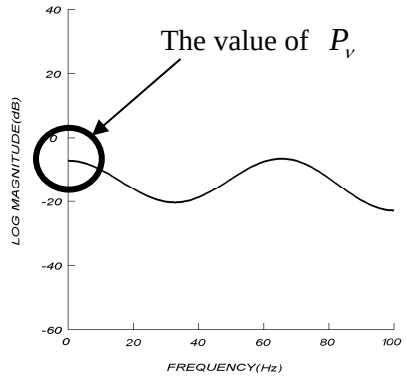


図 5　スペクトル包絡（n=0,m=3）

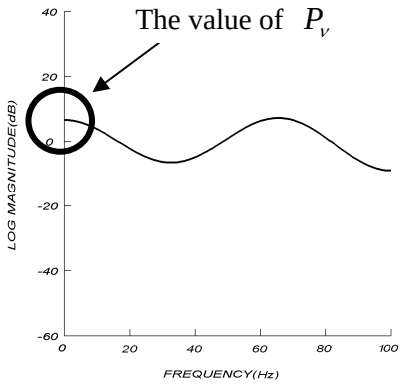


図 6　スペクトル包絡（n=1,m=3）

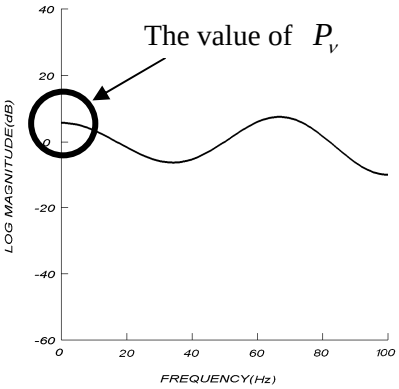


図 7　スペクトル包絡（n=1,m=4）

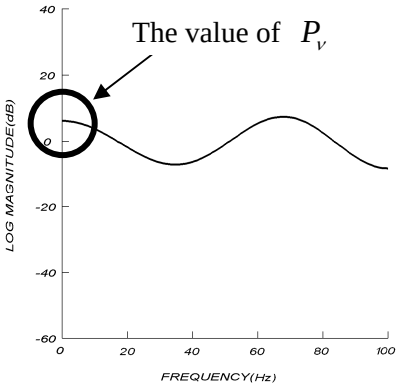


図 8　スペクトル包絡（n=3,m=4）

次に、しきい値を定めて入眠予兆信号とする一例を示す。データは入眠が確認された被験者 S の臀部のデータ（n=1,m=3）を用いた。このデータは 750 フレーム付近で入眠が確認された。入眠区間のスペクトル包絡の時間変化（図 9）をみてみると 0～60Hz の低周波域でスペクトル包絡の形が大きく曲がっている部分がある。この区間のケプストラム時間変化（図 10）をみてみると、突起部があらわれていることがわかる。ここに着目し、しきい値を定め特徴を検討する。図 10 のしきい値を 1 とすると、この区間の点数は 4 である。これらを全区間で算出する。100 フレームを 1 ブロックとして表すことにする。

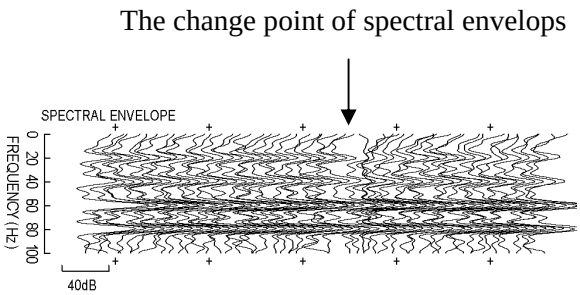


図 9　スペクトル包絡の時間変化
（750~800 フレーム）

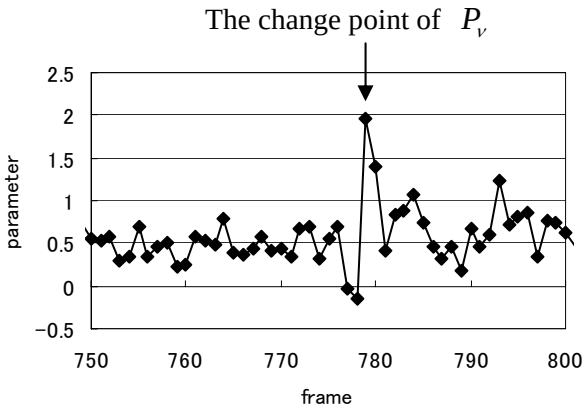


図 10　ケプストラム P_ν の時間変化
（750~800 フレーム）

この臀部のデータにしきい値を 0.6 と定めたとき、全区間でのケプストラムの和がしきい値を超えた回数のグラフを図 11 に示す。図 11 において、図 10 で示した区間は 7.5～8block となる。（━ は入眠区間を表す）

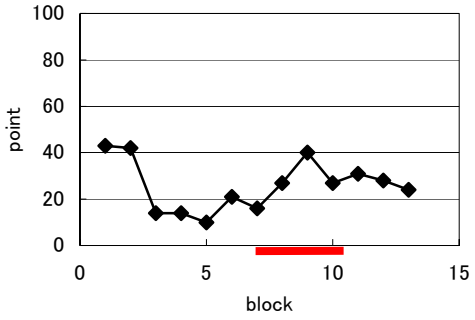


図 11　ケプストラムの和がしきい値を超えた回数
（被験者 S の臀部のデータ）

4　実験結果

ケプストラムの和にしきい値を定め、しきい値を超えた点数をまとめた結果、入眠区間や入眠直前でしきい値を超える点数が増加している結果をいくつか得ることができた。逆に、入眠区間でも変化がないものもあった。ケプストラム次数の選択のしかたによっても特徴の表れ方違っていた。

今回は､3 人で実験を行い 9 個のデータを得た。そのうち 5 個のデータが入眠を確認したデータである。入眠予兆信号がでているかどうかについては、入眠区間中の 1 ブロックでしきい値を超えた点数が前の 1 ブロックより 8 個以上増えた時を予兆信号がでているとした。

入眠を確認した 5 データについて、ケプストラム次数の選択のしかたによって特徴のでかたが違うので表1にまとめる。

表 1　ケプストラム次数の選択のしかたによる特徴の違い

cepstrum order	n=0,m=0	n=1,m=1	n=0,m=3
chest	3	3	3
stomach	2	1	0
hip	2	2	2
cepstrum order	n=1,m=3	n=1,m=4	n=3,m=4
chest	3	4	2
stomach	3	4	2
hip	2	2	2

覚醒時に関しては、入眠区間と同じような特徴がある部分があるデータをいくつかみることができた。体が少し動いた場合でも、ケプストラムの和がしきい値をこえた点数が増え同じような特徴がでたと考えられる。また、入眠時に比べしきい値を超えた点数の推移が少なかった。

5　まとめ

ケプストラムの時間変化において入眠予兆の有無を判断することに関しては、ケプストラム次数の和の組み合わせを様々に変化させて検証した。ケプストラムから入眠の予兆区間や眠気が続く区間では、ケプストラムの時間変化に突起部が現れる部分があった。この突起部をしきい値によって定め、予兆区間と普通の信号の区間にわけ、傾向をまとめた。その結果、ケプストラム次数の選択のしかたとしては n=1,m=3､n=1,m=4 の部分で特徴がよくでていた｡0 次を含んでいるとケプストラムの時間変化にノイズなどにより特徴がわかりにくくなるので､0 次は含まないほうがいいと考えられる。逆に 4 次や 5 次、10 次など高次数部を選ぶと特徴はあまりみられず、検討しにくい結果となった。

また、胸は腹部に比べ予兆の信号が出ている箇所があり、今後も測定をしていく価値はあると思われる。

今後の入眠予兆の研究の方針としては、測定の精度を高め、被験者、データ数を増やすのはもちろんのこと、マイクロホンの場所の検討、低次数部ケプストラム次数の選択に注目して解析を試みてみる必要があると考えられる。

参考文献

[1] 前田慎一郎，落合直樹，小倉由美，榎芳美，藤田悦則，村田幸治，亀井勉，上野義雪，金子成彦 “臀部からの生体信号の簡易計測法”, 第 37 回日本人間工学会　中国･四国支部大会講演予稿集，pp.8-9 (2004)

[2] 藤田悦則，小倉由美，落合直樹，安田栄一，土居俊一，村田幸治，亀井勉，上野義雪，金子成彦，“指尖容積脈波情報を用いた長時間着座疲労の簡易評価の開発”，人間工学 Vol.40 No5, pp254-263 (2004)

[3] 秋田昌憲，緑川洋一，“選択的スペクトル平滑化の信号処理への応用”，信学技法，EA2005-71，pp.19-24(2005)

[4] 今井聖、阿部芳春「改良ケプストラム法によるスペクトル包絡の抽出」信学論,J63-A,4 pp217-223(1974)

入眠予兆のための周波数信号処理の基礎的検討*

○兼近達也，秋田昌憲，緑川洋一（大分大）

1　はじめに

現在の複雑な社会状況で、車、電車等運転のための安全性確立のために、疲労や入眠の予兆を行うことは、重要な研究課題の一つとなっている。このため、居眠り防止技術の開発が様々な分野で広く研究されている[1]。

また、近年の研究により人体の心拍数、脈波、脳波等と入眠との相関関係を示す例が見られるようになってきている [2]。しかしこれらは実用するには不便であるので、本研究室では文献[3]で示されているように、体内音の音響的信号から得られるパラメータを利用して入眠予兆信号の検出ができないか検討している。

得られた信号は処理の過程でケプストラムに変換し、そのパラメータを入眠予兆の信号として扱うことを考える。また低周波数部分に入眠予兆の特徴があるという報告が様々な文献にも記されているので、そのことから低周波数部分の強調を行い、さらに高周波数部分を平滑化することによって、入眠予兆にどのような変化があるか比較・検討する。

以上のような実験、解析から、音声信号の入眠予兆のための周波数処理の基礎的検討を行う。

2　スペクトル選択平滑化法

スペクトル選択平滑化法[4]とはスペクトル軸の非線形評価とリフタリングを用いて、低周波数域のスペクトルピークの形状を保存しながら高周波数のスペクトルローカルピークを平滑化する方法である。

Oppenheim，Johnson の再帰式[5]により直線周波数で評価された最小位相ケプストラム c(m)を非直線周波数上で評価されたケプストラム $\tilde{c}$ (m)に変換する。ただしαは周波数変換係数でα>0 の時、低周波数域が拡大する。

$$
\begin{aligned}
&w_0^{(m)} = c(m) + \alpha w_0^{(m+1)} \\
&w_l^{(m)} = (1-\alpha^2) w_0^{(m+1)} + \alpha w_1^{(m+1)} \\
&w_j^{(m)} = w_{j-1}^{(m+1)} + \alpha\left[w_j^{(m+1)} - w_{j-1}^{(m)}\right] \\
&(j = 2,3\ldots,m) \\
&\tilde{c}(i) = w_i^{(0)}
\end{aligned}
\qquad (1)
$$

この周波数軸の非直線性は、式(2)のオールパスフィルタの周波数特性である式(3)と同じとなり、サンプリング周波数が 10kHz の場合、α=0.35 のときに人間の心理尺度であるメル周波数軸、α=0.5 のときにバーク周波数軸と大体一致することが知られている。

$$
\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \qquad (2)
$$

$$
\tilde{\Omega} = \Omega + 2\tan^{-1}\left[\frac{\alpha \sin\Omega}{1 - \alpha\cos\Omega}\right] \qquad (3)
$$

このようにして低周波数が拡大された非線形ケプストラムを有限次数でリフタリングし、高周波数域のローカルピークを平滑化する。

図１と図２に平滑化前のスペクトルと平滑化後のスペクトルの比較を示す。図１～図３に使用したデータは被験者 A が実験を行った時のもので、左胸部のマイクロホンから得たデータである。なお１フレームは 1.28 秒である。

図１から分かるように低周波数域の包絡形状は同一で、高周波数域の細かいローカルピークだけ抑圧・平滑化されている様子がわかる。

図２はスペクトル包絡の時間変化を表したものである。上が選択平滑化処理なしで周波数変換係数を 0.5 とした場合、下

* Basic examination of signal processing on frequency domain for detecting sleep in sleep-wake state, by Tatsuya KANECHIKA, Masanori AKITA and Yoichi MIDORIKAWA (Oita University).

が周波数変換係数を 0.5 とし、その後のリフタリング次数を 10 とし選択平滑化を行った場合を示す。この図から、低周波数域を単に強調しただけではスペクトル包絡の明確な特徴変化はわからないが、矢印の部分のように選択平滑化処理で高周波数域の包絡を平滑化することにより、スペクトルの平坦化がはっきりと分かるようになっている。

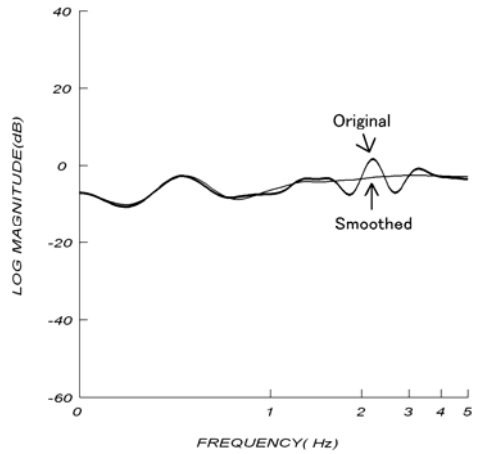


図１　スペクトル選択平滑化によるスペクトル包絡の変化 (Order:25, α =0.5,Liftering order:10)

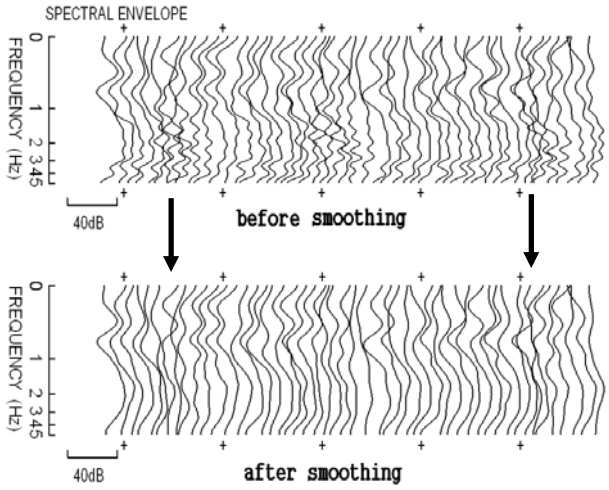
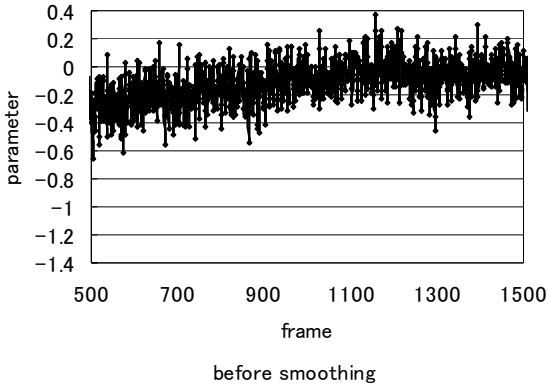


図２　スペクトル選択平滑化によるスペクトル包絡の時間変化(被験者 A の左胸部から得た 1100~1150 フレーム間のデータ, 上図：Order:25, α =0.5,下図：Order:25, α =0.5,Lifteringorder:10)

図３に平滑化前と平滑化後のケプストラムの時間変化を示す。この図から平滑化を行うとパラメータの変化の増減が顕著に表れるようになっていることが分かる。



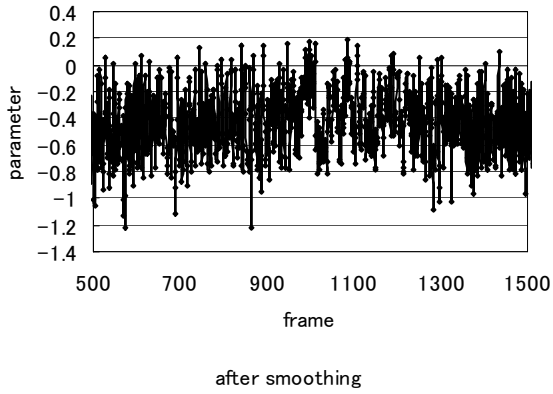


図３　ケプストラムの時間変化(被験者 A の左胸部から得た 500~1500 フレーム間のデータ, 上図:Order:25, α =0,下図:Order:25, α =0.5,Lifteringorder:10)

3　入眠予兆特徴検出のための圧電センサ信号･マイク信号特徴抽出実験

3.1 測定方法について

本測定は、文献[3]と同様にして行った。また測定データは全部で 9 回行われ、本報告ではそのうちの 1 つについて解析を行った。そのデータは被験者 A によって行われたもので、40 分の測定のうち 15 分～22 分で入眠の確認がされている。

3.2 解析方法について

200Hz でサンプリングされたデータについて、改良スペクトラム分析[6]を行い、25 次のケプストラムを求めて、その後選択平滑化処理で特徴強調を行う。
なお、ケプストラムを求めて閾値を定め、閾値を越える点数から入眠予兆の特徴ではないかと捉える方法については文献[3]と同様である。

3.3 結果・考察

図５に右臀部のセンサ信号について周波数軸変換、スペクトル選択平滑化を行っていない場合のケプストラムの和が閾値を超えた回数を表すものを示す。これから分かるように入眠区間において閾値を超えた回数の変化があまりなく、特徴が表れてないと考えられる。

図６は同信号において、周波数軸変換係数を 0.5、その後のリフタリング係数を 10 としたものである。図５と比較すると分かるように 8~9 ブロックの区間において本報告において特徴と定めたものが表れていることが分かる。なお 1 ブロック=100 フレームである。

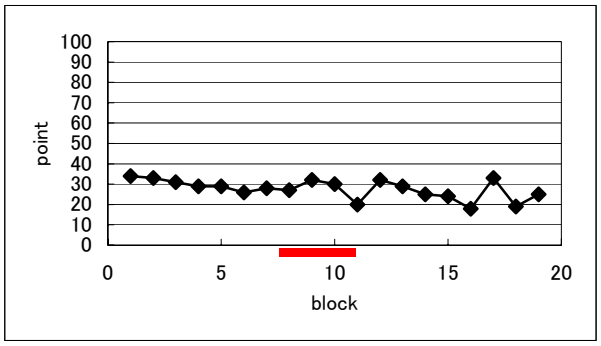


図５　ケプストラムの和が閾値を超えた点数(被験者 A の右臀部から得たケプストラム次数の選択が 1 次のみの場合のデータ)

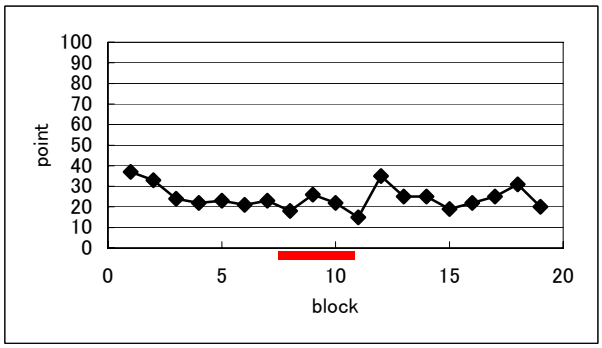


図６　ケプストラムの和が閾値を超えた点数
(被験者 A の右臀部から得たケプストラム次数の選択が 1 次のみの場合のデータ, α =0.5,liftering order:10)

表１　左胸部、左わき腹、右臀部についてケプストラム次数の選択、選択的平滑化による特徴の有無

	平滑化	1 次のみ	0次～3次の和
左胸部	なし	○	×
	あり	○	×
左わき腹	なし	×	○
	あり	○	○
右臀部	なし	×	○
	あり	○	○
	平滑化	1次～3次の和	1次～4次の和
左胸部	なし	○	○
	あり	○	○
左わき腹	なし	○	○
	あり	○	○
右臀部	なし	×	○
	あり	○	○
	平滑化	3次～4次の和	4次～5次の和
左胸部	なし	○	×
	あり	○	×
左わき腹	なし	○	×
	あり	○	×
右臀部	なし	×	○
	あり	○	○

表１に各測定部、ケプストラム次数の選択、選択的平滑化による特徴の有無についてまとめた。これから、左わき腹、右臀部のケプストラム次数が１次のみの場合、そして右臀部のケプストラム次数が３次～４次の和の場合において、スペクトル選択平滑化を行った時に特徴が表れたという結果になった。

このことからスペクトル選択平滑化を行い、周波数軸変換によって低周波数域を強調することが入眠予兆のための周波数処理において有効である可能性が認められる。

4　まとめ

本報告では、ケプストラムの周波数変換操作と、リフタリングを用いて、信号のスペクトルの特定周波数帯域を平滑化する選択平滑化法を扱い、入眠予兆の特徴抽出のための周波数信号処理の基礎的検討を行なった。

選択平滑化法により高周波数域を平滑化し、さらに周波数軸変換を行うことで、入眠予兆であると考えられる特徴が表れやすくなるということが示された。これに関しては、周波数軸変換係数やリフタリング次数の選択によって、まだ検討課題が多いと考えられる。

参考文献

[1] 前田慎一郎，落合直樹，小倉由美，榎芳美，藤田悦則，村田幸治，亀井勉，上野義雪，金子成彦，“臀部からの生体信号の簡易計測法”, 第 37 回日本人間工学会　中国･四国支部大会講演予稿集，pp.8-9 (2004).

[2] 藤田悦則，小倉由美，落合直樹，安田栄一，土居俊一，村田幸治，亀井勉，上野義雪，金子成彦，“指尖容積脈波情報を用いた長時間着座疲労の簡易評価法の開発”, 人間工学 Vol.40 No.5, pp.254-263，(2004).

[3] 坂口正和，秋田昌憲，緑川洋一，“入眠予兆のための体内音測定”，日本音響学会九州支部 2009 学生のための研究発表会（投稿予定），pp.1-4 (2009).

[4] 秋田昌憲，緑川洋一，“選択スペクトル平滑化の信号処理への応用”，信学技法 EA2005-71，pp.19-24 (2005).

[5] A.V.Oppenheim and D.H.johnson. “Discrete Representation of Speech”, Proc. IEEE, 60, pp.681-691, (1972).

[6] 阿部芳春，今井聖，“改良ケプストラム法によるスペクトル包絡の抽出”, 信学論（A），Vol.J62-A, 4, pp.217-223. (1979).

[7] 秋田昌憲，“音声認識に用いるメルケプストラム算出法の評価”，信学論（A），Vol.J72-A, 10, pp.1695-1696. (1989).

[8] 今井聖，“音声信号処理”, 森北出版株式会社，pp.148-160,pp.169-174 (1996).

泣き声を用いた乳児の情動推定のための有効な韻律的特徴の検討*

道脇慎司†、山内勝也††、松永昭一††、山下優†、篠原一之†††
(†長崎大院・生産科学研　††長崎大・工　†††長崎大・医)

1　はじめに

乳児は甘えているときやお腹が空いているとき、眠いときなど様々な理由で泣く。それは気持ちを言葉で伝えることができないためである。対して、両親は乳児がなぜ泣いているのかを考え、泣き止ませようと試みるだろう。しかし、乳児の泣いている原因を理解することは容易なことではなく、意思がうまく伝わらない場合、乳児と両親ともにストレスを抱えることになる。そこで、乳児の情動表出の客観的測定技術の開発が進めば、両親は乳児に対し的確な対処をすることができ、乳児が感じる不快感も軽減できる。さらに、親子間の円滑なコミュニケーションが行われることで、乳児虐待などの社会問題の解決にもつながるのではないかと考える。

問題解決の手段の一つとして、泣き声を用いた乳児の情動推定が研究されている [1]。また、我々はこれまでの研究で、泣き声データを多数用意し、HMM(隠れマルコフモデル) を用いた最尤法に基づき作成した泣き声音響モデルを用いて統計的に情動推定を行ってきた [2, 3]。その際、音響特徴量としてスペクトル包絡情報を示す MFCC(メルケプストラム係数) を用いていた。しかし、情動に関する情報がスペクトル包絡情報だけに含まれるとは限らない。そこで、泣き声の韻律的特徴を分析して、韻律的特徴から乳児の情動の推定手法を検討する。

一般に韻律とは、発話の中で観測される「音の強さ、長さ、高さやそれらの変化、ポーズの置き方など」と定義されている [4]。成人の発話音声における韻律的特徴の中には感情や個人性などの情報が多く含まれており、乳児の泣き声から情動を推定する際も、韻律的特徴の利用が有効な可能性があると考えられる。

2　泣き声データベース

本研究に利用する泣き声データベースについて概説する。本データベースは、泣き声データ、情動評価表、音響特徴ラベルにより構成される。

2.1　泣き声データ

泣き声データは乳児の母親がデジタル IC レコーダを用いて、日常生活における乳児の泣き声を収録した。収録期間は協力者に依るが、数日から数ヶ月に渡り、各データは 30 秒程度である。乳児は月齢 8〜13ヵ月 (平均月齢：10.6ヵ月) の 23 名 (男児 11 名、女児 12 名) であり、計 402 データの収録を行った。また、収録時のサンプリング周波数は 44.1kHz であり、16kHz にダウンサンプリングした泣き声データを用いてデータベースを作成する。録音に際しては、IC レコーダを乳児の顔からおよそ 30cm 程の位置に持ち、IC レコーダをなるべく動かさないように依頼した。また、テレビなどの周囲の音はできるだけ録音されないような収録環境を依頼した。今回、使用したデータは比較的データ数の多い乳児の 3 名 (145 データ) を対象とした。

2.2　情動評価表

収録した泣き声データの情動に関しては、乳児の母親が泣いている状況や経験などから判断して評価した。その際、Table. 1 に示す 10 個の情動から単一または複数の情動を選択し、1 点 (ほとんどその情動ではない) から 5 点 (とてもその情動に感じる) までの 5 段階で評価値をつけるよう求めた。

Table. 1 に示した 10 個の情動の中でも、「恐れ」や「驚き」、「排泄」、「痛み」と評価された泣き声データは少なく、それらの情動は推定困難であると判断し、逆に「甘え」、「怒り」、「空腹」の 3 情動は母親に評価された回数が多く、重要な情動であると判断した。また、Satoh らによりそれら 3 情動は互いに結び付きが弱く、独立した情動因子であるという結果が得られている [3]。よって、本研究では「甘え」、「怒り」、「空腹」の 3 情動を推定対象の情動とした。

「甘え」、「怒り」、「空腹」の 3 情動に対し、母

*A Examination of Effective Prosodic Feature for Presumption of Emotion in Infant's Cries. by S. Michiwaki†, K. Yamauchi††, S. Matsunaga††, M. Yamashita† and K. Shinohara††† (†Graduate School of Science and Technology, Nagasaki University ††Faculty of Engineering, Nagasaki University †††Depertment of Translational Medical Sciences, Nagasaki University)

Table 1　母親による情動評価において評価対象とした情動

甘え	怒り	空腹	悲しみ	恐れ
驚き	眠い	排泄	不快	痛み

Table 2　サンプルデータ数

	乳児 A	乳児 B	乳児 C
甘え	14	13	9
怒り	14	14	10
空腹	8	11	7

親の評価値が 4 点以上であるものをその情動のサンプルデータとした。Table. 2 に、各情動のサンプルデータ数を乳児ごとに示す。ただし、それら 3 情動に 4 点以上の評価値がつけられているデータはサンプルデータから除外する。

2.3　音響特徴ラベルの付与

泣き声データの韻律的特徴を調査する際に、泣き声データの音響的特徴に対応させて、複数のセグメントを定義し、泣き声データに対してラベル付与を行った。一覧を Table. 3 に示し、セグメントラベルの付与例を Fig. 1 に示す。

3　泣き声に含まれる韻律的特徴

本研究では、泣き声データの観察から韻律的特徴として、泣き声の高さと泣き声を構成する各セグメントの時間割合に着目した。

3.1　泣き声に含まれるピッチ情報

成人の発話音声における感情分析でピッチは大きな影響を及ぼすが、乳児の泣き声のピッチも情動に関する情報を含んでいる可能性がある。その可能性が最も高いと考えられる泣き声音セグメント (cry) を対象とし、ピッチを抽出した。

ピッチの抽出には、京都大学音声メディア研究室で井本和範氏が作成したプログラムを利用した。使用したピッチ抽出プログラムは、線形予測分析 (LPC) 結果に基づく予測残差波形の自己相関関数を計算し、その相関値を基にして各フレームの基本周波数値を抽出している。大局的な声道特性をよく反映する LPC 分析予測値を波形値から指し引いた予測残差を用いることで、より声道特性を反映した特徴を用いている。また、相

Table 3　使用するセグメント

セグメント名	意味
cc	無音
noise	雑音
cry	泣き声音 (エーンなどの区間)
suu	息継ぎ音
ho	咳音
qu	クーイング (クー、アーなど)
ver	喃語 (言葉にならない発声)
unaru	唸り音

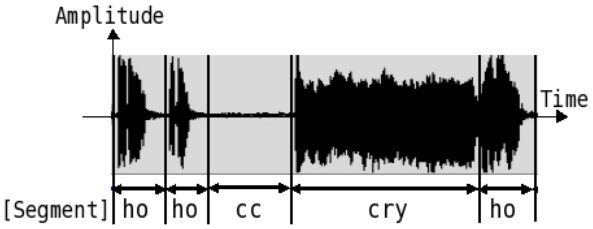


Fig. 1　セグメントラベルの付与例

関値の高い上位 5 候補を抽出し、相関値が低い場合は、第二候補以降の値でも連続性を保つ基本周波数値を正解として抽出している。

抽出する際、フレーム長 20ms、フレーム周期 5ms、窓関数にハミング窓を用いた。各情動の泣き声データから抽出されたピッチ平均値と標準偏差を Table. 4 に示す。

Table. 4 より、乳児 A に関しては「甘え」と比較して「怒り」、「空腹」のピッチの平均値がわずかに高いが、標準偏差が大きく有意な差ではない。乳児 B に関しては「甘え」、「怒り」に比べ「空腹」のピッチ平均値が高くなっている。しかし、乳児 C においては 3 情動ともほぼ同じピッチ平均値であった

以上のような結果から、ピッチ平均値からは情動による各乳児共通の特徴を得ることができず、乳児ごとにみても情動によりピッチ平均値に大きな差があるとは言えない。よって、ピッチ平均値を用いて情動推定を行うことは困難であると言える。

Table 4　泣き声音 (cry) のピッチ平均値 (標準偏差)[Hz]

	乳児 A	乳児 B	乳児 C
甘え	413(124)	406(110)	425(113)
怒り	450(127)	404(123)	415(108)
空腹	443(111)	498(119)	417(101)

3.2　各セグメントの時間割合

乳児の泣き声データを構成するセグメントの数や長さはデータによって異なる。ここでセグメントの時間割合 R を次のように定義する。

$$R_{segment} = \frac{T_{segment}}{T_{total}} \tag{1}$$

ここで、$T_{segment}$ は 1 データ中のあるセグメントの合計時間とし、T_{total} は 1 データの合計時間とする。

時間割合の平均値が最も大きい 3 つのセグメントである泣き声音 (cry)、無音 (cc)、咳音 (ho) の時間割合平均を Table. 5〜 7 に示す。

Table. 5 より泣き声音 (cry) は「甘え」、「空腹」に比べ「怒り」の時間割合が大きい傾向であることがわかる。Table. 6 を見ると、無音 (cc) では「怒り」に比べ「甘え」、「空腹」の時間割合が大きくなっている。実際、「怒り」の泣き声データは間をおかず激しく泣いている場合が多く、それがこれらの指標に反映されていると言える。

次に Table. 7 の咳音 (ho) をみると、「甘え」、「怒り」に比べ「空腹」の時間割合が大きい。乳児にとっての食事はミルクや離乳食が主であり、喉が渇いていると考えられることから咳音 (ho) の時間割合が増えたのではないかと予想できる。

このように泣き声データを構成する各セグメントの時間割合には、情動によって違いがある可能性が示唆された。

4　情動推定実験

4.1　実験手順

本実験では、泣き声音 (cry)、無音 (cc)、咳音 (ho) のセグメントの時間割合を利用し、「甘え」、「怒り」、「空腹」の 3 情動の推定実験を行った。

以下に情動推定の手法を示す。

1. 人手で付与したセグメントラベルを用いて泣き声データ中の泣き声音 (cry)、無音 (cc)、咳音 (ho) セグメントの時間割合を求める。それら 3 つのデータをベクトル空間に表したものを、その泣き声データの時間割合ベクトル e_i とする。

$$e_i = (R_{cry}, R_{cc}, R_{ho}) \tag{2}$$

2. e_i を各次元ごとに正規化する。ここで AVE は平均値を、$STDEV$ は標準偏差を表す。

$$R'_{cry} = \frac{R_{cry} - AVE(R_{cry})}{STDEV(R_{cry})} \tag{3}$$

$$R'_{cc} = \frac{R_{cc} - AVE(R_{cc})}{STDEV(R_{cc})} \tag{4}$$

$$R'_{ho} = \frac{R_{ho} - AVE(R_{ho})}{STDEV(R_{ho})} \tag{5}$$

$$e'_i = (R'_{cry}, R'_{cc}, R'_{ho}) \tag{6}$$

3. 評価データの正規化された時間割合ベクトル e'_i と、それ以外のデータの正規化された時間割合ベクトルとのユークリッド距離を求める。

4. 各情動ごとに評価データの近傍の k 個のデータを見つける。評価データとその k 個のデータとの距離の平均が最も小さい情動を推定結果とする。

Table 5　泣き声音 (cry) の時間割合平均 (標準偏差)[%]

	乳児 A	乳児 B	乳児 C
甘え	20.8(12.5)	14.2(6.6)	32.1(14.6)
怒り	43.7(13.9)	40.3(22.2)	50.8(17.8)
空腹	24.6(7.5)	19.7(10.8)	32.5(20.3)

Table 6　無音 (cc) の時間割合平均 (標準偏差)[%]

	乳児 A	乳児 B	乳児 C
甘え	35.2(9.9)	48.0(10.9)	23.7(7.9)
怒り	22.4(8.0)	26.9(12.0)	15.2(6.4)
空腹	37.5(10.2)	43.2(13.9)	29.0(14.1)

Table 7　咳音 (ho) の時間割合平均 (標準偏差)[%]

	乳児 A	乳児 B	乳児 C
甘え	14.1(7.7)	1.5(1.7)	4.8(2.9)
怒り	9.7(5.2)	2.5(2.9)	4.3(1.3)
空腹	25.1(6.6)	4.3(2.4)	8.7(3.4)

Table 8　情動推定結果と母親の判定の一致数 ($k = 1$)

	乳児 A	乳児 B	乳児 C
甘え	9/14	9/13	1/9
怒り	8/14	8/14	7/10
空腹	4/8	6/11	4/7

Table 9　情動推定結果と母親の判定の一致数 ($k = 6$)

	乳児 A	乳児 B	乳児 C
甘え	11/14	9/13	5/9
怒り	10/14	9/14	8/10
空腹	4/8	6/11	5/7

5　結果と考察

3 名の乳児の泣き声に対して、それぞれ 2 節に示すデータ数で $k = 1$〜6 とし、情動推定を行った。

本手法による $k = 1$ のときの推定結果と母親の判断結果の一致数を Table. 8 に、最も正解率が高かった $k = 6$ のときの一致数を Table. 9 に示す。

全体の正解率は $k = 1$ のときに 56%だが、$k = 6$ のときは 67%であった。このことから、適切な k の値を選択することが重要なことがわかる。また、3 名の乳児ともにある程度の正解率が得られたことから、乳児の情動推定に各セグメントの時間割合を用いることは有効的であることが示された。

6　おわりに

本稿では、乳児の情動推定のために韻律的特徴の分析をし、泣き声データを構成する各セグメントの時間割合が有効である可能性を示した。また、これまでの乳児の情動推定に関する研究では、泣き声データを解析する際、フレームごとに特徴量を抽出していた。しかし、本研究ではフレームより大きいセグメント単位で情動推定を行い、ある程度の正解率を得ることができた。

今回は人手で付与したセグメントラベルを用いて各セグメントの時間割合を求めたが、実際には泣き声データのどの区間にどのセグメントが対応するかわからない。そのため、今後セグメンテーション方法について検討していく。また、引き続き情動推定のために有効な韻律的特徴の調査を進める。

参考文献

[1] Kaoru Arakawa：Recognition of Babies's Cries from Frequency Analyses of Their Voice Classification Between Hanger and Sleepiness, Proc. of ICA2004,pp.1713–1716 (2004)

[2] 坂口清起、山下優、松永昭一、宮原末治、西谷正太、篠原一之：泣き声による乳児の情動識別のためのラベル付与、日本音響学会春季講演論文集、pp.367–368 (2006)

[3] Noriko Satoh, Katsuya Yamauchi, Shoichi Matsunaga, Masaru Yamashita, Ryuta Nakagawa, Kazuyuki Shinohara、Emotion Clustering Using the Results of Subjective Opinion Tests for Emotion Recognition in Infants' Cries, Proc. of Interspeech 2007, pp.2229–2232 (2007)

[4] 日本音響学会編、"新版音響用語辞典 "、コロナ社 (2003)

FPGAを用いたFFTケプストラム係数の抽出法の検討*

辻恭志† 山内勝也†† 山下優† 松永昭一†† 小栗清††
(† 長崎大院・生産科学研 †† 長崎大・工)

1 はじめに

近年はロボットを始めとする自動制御システムの開発が飛躍的に進んでおり, 入力インタフェースとしての音声認識技術が注目されている. また近い将来には情報家電の普及が現在以上に進むと考えられ, その点においても機器操作の敷居を下げる意味で音声認識システムは非常に重要なシステムとなり得る. しかし, 現在使われているシステムは主にソフトウェアによる認識であり, 小型の情報機器に搭載させるには大きさの面で不向きである. そこでFPGA(Field Programmable Gate Array)を始めとするリコンフィギャラブルシステムを利用することで, その問題の解決を図る研究がなされている.

リコンフィギャラブルシステムとはハードウェアを設計データによって動的に作り替えることができるシステムであり, ハードウェアの高い処理能力とソフトウェアの柔軟な構成を併せ持つのが特徴である. これによりユーザ自身で回路構成情報を書き換えることで何度でもハードウェアの再構成が可能となる. また, アプリケーションごとにハードウェア構成を最適化できるので, 並列処理, 可変データ幅, 多様な演算器構成, データの流れの最適化が可能となり, 高速化を図ることができる.

音声認識システムをFPGAに実装させる研究のうち, HMM(Hidden Markov Model) 等を用いた認識に関する研究は行われているが [1], 認識に用いる特徴分析に関しては注目度が低い. 回路の規模, メモリ容量が限定されているFPGA上の音声認識において, 特徴分析の精度とそれに伴う回路規模への影響を把握することは重要な意味を持つ. 従来のFPGAを用いた研究では量子化ビット数8ビットの音声を使用した特徴分析が主流であったが, 今回は16ビットの音声を処理できるシステムを構築した. 本研究では, 音声の特徴分析としてFFT(Fast Fourier Transform) ケプストラム係数を用いて, 量子化ビット数8ビットの音声と16ビットの音声それぞれの処理に対応した回路実装と精度に関わる問題を検討する.

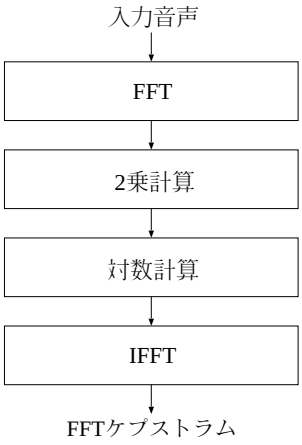


図1 特徴分析の流れ

2 回路設計

本研究では, 予め録音したWAV形式の音声データに対して, 1,024点FFTによるFFTケプストラムの算出を行う. FFTケプストラムは対数振幅スペクトルの逆フーリエ変換で定義され [2], 一般に音声認識に使用されるMFCC(Mel-Frequency Cepstrum Coefficient) と比較して求めるための回路が比較的小規模で作成できることから今回の回路で使用した. 本研究の特徴分析の流れを図1に示す.

作成する回路の中で, 中心となるFFT回路, 一般的なものとは異なるアルゴリズムを採用した対数回路についての実装アルゴリズムを以下に示す.

2.1 FFT回路

離散時間信号 $x(n)$ の N 点FFT, 及びIFFT(Inverse FFT) は以下の式で表すことができる.

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot W_N^{nk} \quad (k = 0, 1, \cdots, N-1) \tag{1}$$

* An examination of the FFT cepstrum coefficient extraction method on FPGA board, by TSUJI Yasushi†,YAMAUCHI Katsuya††, YAMASHITA Masaru†, MATSUNAGA Shoichi††, and OGURI Kiyoshi†† (†Graduate School of Science and Technology, Nagasaki University †† Faculty of Engineering, Nagasaki University)

$$x(n) = \frac{1}{N}\sum_{n=0}^{N-1} X(k)\cdot W_N^{-nk} \quad (2)$$
$$(n = 0, 1, \cdots, N-1)$$

ただし, W_N^{nk} は回転因子であり,

$$W_N^{nk} = \exp^{-j(\frac{2\pi}{N})nk} \quad (3)$$

とする.

本研究では radix-4 バタフライ演算を用いて FFT を計算する. これは N 点からなる入力信号列 $x(n)$ を 4 分割し, $N/4$ 点の FFT4 つにする手法であり, これを繰り返すことで最終的に 4 点の FFT に帰着する. 例として $N = 64$ の場合を考える. 64=4^3 であるため, (1) 式は次のように展開できる.

$$X(k) = \sum_{n_0=0}^{3}\sum_{n_1=0}^{3}\sum_{n_2=0}^{3} x(n)\cdot W_{64}^{nk} \quad (4)$$
$$k = k_0 + 4k_1 + 16k_2$$
$$(k_0, k_1, k_2 = 0, 1, 2, 3)$$
$$n = n_0 + 4n_1 + 16n_2$$
$$(n_0, n_1, n_2 = 0, 1, 2, 3)$$

各 $\sum$ の加算個数は 4 個で, 3 つの $\sum$ からなる式に展開できる. この 1 つの $\sum$ が 1 ステージの処理となり, この例では 3 段で全体の処理を行うことができる. 本研究では 1,024 点 FFT を用いるので, この場合 1,024=4^5 なので, 5 段の処理で FFT を求めることができる.

また, (4) 式から更に展開すると,

$$X(k) = \sum_{n_0=0}^{3}\sum_{n_1=0}^{3}\{(x(\alpha) + (-j)^{k_0}x(\alpha+16) + (-1)^{k_0}x(\alpha+32) + j^{k_0}x(\alpha+48))W_{64}^{\alpha k_0}\}W_{16}^{\alpha(k_1+4k_2)} \quad (5)$$
$$(\text{ただし}, \alpha = n_0 + 4n_1)$$

を得る. この式は, 16 点ずつ離れた 4 点の値を入力として計算を行っているが, この計算はシフトレジスタによってバタフライ回路への入力を制御することで実現できる. その様子を図 2 に示す. 回転因子の値は, 三角関数の演算のコストが高いため, データを ROM に格納し, それを呼び出している.

また, IFFT 回路は (1)(2) 式の比較で分かるように, 異なる点は指数部の符号のみであるため,

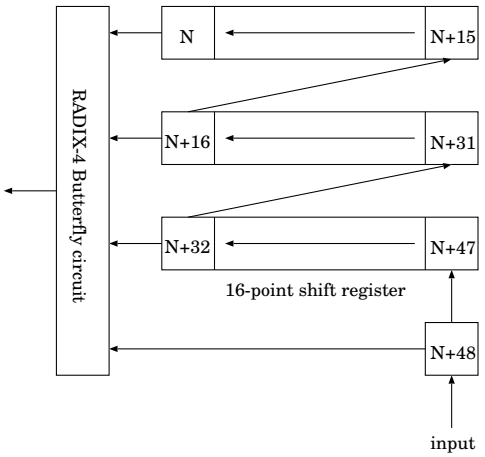


図 2　シフトレジスタによる入力制御ブロック図 (in radix-4 64point FFT)

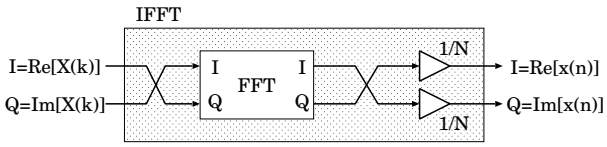


図 3　IFFT 回路ブロック図

図 3 に示すように FFT 回路を用いて IFFT 回路を実現することができる. そのため, 構成が単純で回路面積が小規模に収まるという点で FFT ケプストラムを用いることは非常に有効といえる.

2.2　対数回路

対数計算には級数展開や連分数展開を利用した方法があるが [3], 多量の計算とクロックを使用するため望ましくない. そのため, LUT(Look Up Table) を利用して対数を求める.

パワースペクトルを $S(k)$ とすると, 対数の計算は $\log|S(k)|$ で求められる. このとき, 対数の整数部は $|S(k)|$ の中で上位ビットから調べて初めて 1 が出てくるビット位置である. この対数の計算を図示したものが図 4 である. 図 4 の場合, 整数部の値は N となり, 小数点以下の値は整数部となるビットから上位 8 ビットを抜き出し, その部分を LUT で参照し, 値を求める. この手法だと少量のメモリを使用するだけで, 1 クロックでほぼ正確な対数を求めることが可能となる.

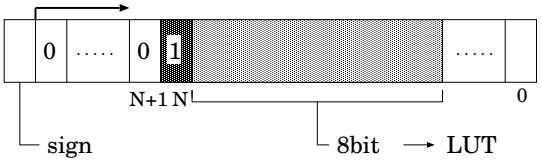


図 4　対数計算図

3 実装条件

上記の設計を用いて量子化ビット数 8 ビットの音声特徴分析回路 (以降 8 ビット特徴分析回路と呼ぶ), 16 ビットの音声特徴分析回路 (以降 16 ビット特徴分析回路と呼ぶ) を作成し, FPGA にて実装検証を行った. また, 求めた特徴量を使用し, 音声認識ソフトウェア HTK を用いて音声認識精度の検証を行った. 本章では実装条件について記す.

3.1 FPGA ボード

実装において使用した FPGA ボードは Altera 社製 Cyclone デバイス EP2C35F484C8 である. 仕様は以下の通りである.

- Total Pins ··· 322
- Total Logic Elements ··· 33,216
- Total Memory Bits ··· 483,840

なお, Total Pins は FPGA の入出力等の接続に使われるピンの数, Total Logic Elements はデバイスの規模を表すロジックエレメントの数, Total Memory Bits は FPGA 上で使用できるメモリブロック量を表す.

3.2 音声データ

今回の実験で使用する音声データの仕様を以下に記す.

- サンプリング周波数 ··· 16kHz
- 収録人数 ··· 男性 5 名
- 音声データ内容 ··· “ゼロ (0)” から “キュウ (9)” までの 10 単語

音声データ数については, 各々で同じ単語をそれぞれ 5 回収録し, 合計で 250 音声を用意した. また, 量子化の精度について, 音声データを予め量子化ビット数 16 ビットで録音し, それを変換することで, 16 ビットのものと 8 ビットのものを用意した. その 2 種類について FPGA を用いて特徴分析を行い, その結果を比較する.

3.3 特徴量

今回の実験で用いる特徴量の仕様を以下に記す.

- 特徴量 ··· FFT ケプストラム
- FFT 点数 ··· 1,024 点
- フレーム長 ··· 65ms
- フレームシフト ··· 32.5ms

この仕様に基づいた特徴量を FPGA に実装した回路で使用して求め, 検証を行う.

3.4 音声認識

認識実験には音声認識ソフトウェア HTK を用いる. 音響モデルの仕様は以下の通りである.

- モデル ··· 単語 HMM
- 状態数 ··· 3
- 混合数 ··· 1
- 音声特徴量 ··· 12 次ケプストラム

3.3 節の特徴量を用いて認識を行う. なお, 認識方法は録音した 250 音声のうち 200 音声を学習データに使用し, 残り 50 単語を認識に使用する, という方法を 5 回繰り返すことで全音声の認識を行った.

4 結果 · 考察

前章の実装条件で得られた実験結果, 考察について本章で述べる.

4.1 回路性能

実装した 8 ビット特徴分析回路, 16 ビット特徴分析回路の回路性能を表 1 に記す. Logic Elements は回路の実装に使用したロジックエレメント数, Memory Bits は使用したメモリ量, Actual Time が最大動作周波数を表す. 最大動作周波数とは, 回路がどの程度高速な動作を行うかを示す値である. 最大動作周波数が高い程高速な回路であることを示す.

表 1 から分かる通り, 8 ビット特徴分析回路と 16 ビット特徴分析回路の必要なロジックエレメント数の差異は 3,221 であった. これは本研究で使用している EP2C35F484C8 の総ロジックエレメント数の約 10%にあたる. また, 最大動作周波数は 8 ビット特徴分析回路と比較して 16 ビット

表 1 回路性能

	8 ビット 特徴分析回路	16 ビット 特徴分析回路
Logic Elements	9,673	12,894
Memory Bits	159,104	166,592
Actual Time	24.00MHz (41.667ns)	20.28MHz (49.30ns)

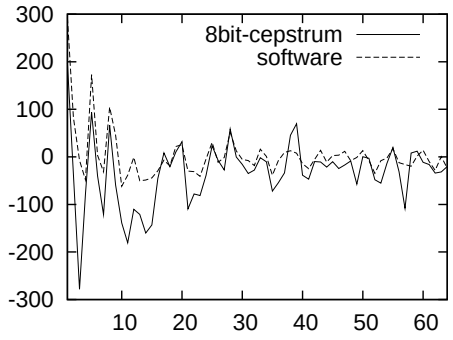


図 5　特徴量比較 (8 ビット特徴分析回路, ソフトウェア)

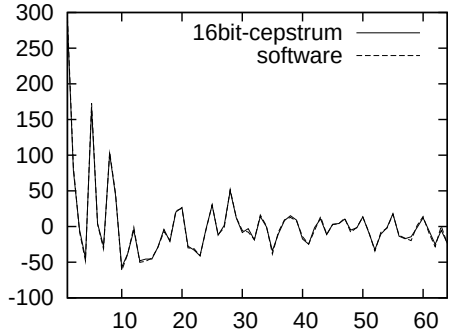


図 6　特徴量比較 (16 ビット特徴分析回路, ソフトウェア)

特徴分析回路は約 20%低下という結果になった. 1 フレームの特徴分析にかかるクロック数の違いはほとんどなかったため, 速度についても同程度の差であると考えられる.

4.2　特徴量精度

“イチ” という音声の第 1 フレームについて, 3.3 節に従って求めた特徴量と, 倍精度 (符号 1 ビット 指数 11 ビット 仮数 52 ビット) でソフトウェアを用いて計算した特徴量を比較した. 図 5 に 8 ビット特徴分析回路で求めた特徴量と比較したものを, 図 6 に 16 ビット特徴分析回路で求めた特徴量と比較したものを示す.

16 ビット特徴分析回路で求めたケプストラムはソフトウェアによる特徴量と比較して誤差がほとんど見られないのに対し, 8 ビット特徴分析回路で求めたケプストラムでは, 値に誤差が見られるのが分かる. ただし, FFT 計算によりスペクトルを求めた時点ではその誤差は大きいものでなく, パワースペクトルを求める際の対数計算で値を切り捨てる必要があったため, その誤差が影響したものだと思われる.

表 2　音声認識精度

特徴量	認識数/全音声
ケプストラム (16bit)	201/250　(80.4%)
ケプストラム (8bit)	104/250　(41.6%)

4.3　音声認識精度

8 ビット特徴分析回路, 16 ビット特徴分析回路で求めたケプストラムを用いて認識を行った結果を表 2 に記す. 特徴量内の括弧は使用した音声の量子化ビット数である. 音声認識の精度については, 16 ビット特徴分析回路によるケプストラムが 8 ビット特徴分析回路と比較して 40%程度高い精度を得ることができた. 今回の認識実験では使用した音声に関して「ゼロ」から「キュウ」までの 10 単語のみ, 男性のみ, 使用する音声が合計で 250 音声, という制限があったために偏りが出た可能性はあるが, 16 ビット特徴分析回路によるケプストラムを用いて実用的な認識を行うことは可能だと考えられる. 今後, 学習データを増やした上で, かつ大語彙での認識実験による評価が求められる.

5　まとめ

本研究では, FPGA を用いて FFT ケプストラムを特徴量とした特徴分析回路の実装について検討した. 更に量子化ビット数が 8 ビットの音声, 16 ビットの音声それぞれに対応した回路を実装し, 回路規模, 特徴量 · 音声認識の精度に関する評価を行った. 結果, 8 ビット特徴分析回路と比較して, 16 ビット特徴分析回路は回路規模が 10%, 処理時間が 20%ほど増加したが, 音声認識精度は 40%程度の上昇が得られた. 今後の課題として, 大語彙での音声認識実験, 回路規模を削減した特徴分析回路の構築が挙げられる.

参考文献

[1] 中谷正吾, 山内宗, 梶原信樹, “HMM の対数型アルゴリズムの FPGA 上へのマッピング,” 情報処理学会全国大会講演論文集, 73-74, 1996.
[2] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄, “音声認識システム, ” オーム社, 2001
[3] 奥村晴彦, “C 言語による最新アルゴリズム辞典, ” 技術評論社, 1991.

膜鳴楽器の音響振動連成解析に関する研究

荒木陽三 柳平直徳 鮫島俊哉 (九大・芸工)

1　はじめに

理想的な円形膜振動の固有モード関数は，ベッセル関数を含んでおり，本来調和的な部分音構造をもたない。しかし，ティンパニやタブラなどの膜鳴楽器は，整数比に近い調和的な部分音構造をもつことが知られている [1, 2]。ティンパニは半球状のケトルをもつ膜鳴楽器である。ティンパニが調和的な部分音をもつ要因としては，ケトル内に閉じ込められた空気や膜の曲げ剛性の影響などがあるとされている。一方タブラは密度が均一でない膜をもつ膜鳴楽器であり，膜の密度を変化させることによって整数比に近い部分音を作り出している。

また，これまで膜鳴楽器について様々な解析的研究がされてきているが，振動場や音場の形状や境界条件において現実に則した厳密なものではなかった。一方近年，計算機の高性能化により，室内音響学などの分野で有限要素法や境界要素法といった数値解析手法を用いたより厳密な解析が盛んに行われてきている。

本研究では，調和的な部分音構造をもつ膜鳴楽器の中でもティンパニをとりあげる。ティンパニが調和的な部分音をもつ要因をより明らかにし，さらに，調和的な部分音をもつためのより最適なケトルの形状を見つけることを目的とする。今回は，そのための解析手法の定式化を行った。膜振動の解析には固有モード展開を用い，音場の解析にはケトルの形状をより厳密に扱うため，境界要素法を用いる。また，この解析手法の妥当性を確認するため，実測結果との比較を行った。

2　解析手法

Fig.1 のような，膜と音場からなる解析モデルを考える。音場には直角座標を，円形膜には極座標を用いる。円形膜は周辺を単純支

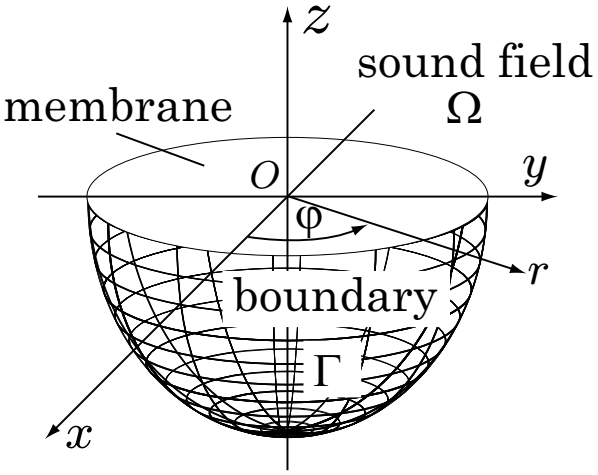


Fig. 1　解析モデル

持されているとし，今回はケトルの振動は考えず，剛であるとして扱う。膜振動場の解析に固有モード展開，音場の解析には法線方向微分型境界要素法を用いて，それらを力と速度の関係により連成することを考える。

2.1　音場の解析

音場の解析には，法線方向微分型境界要素法を用いる。境界 Γ 上を除く領域 Ω 内の任意の受音点 p での速度ポテンシャル　は，Kirchhoff の積分方程式の形で次式のように表すことができる。

$$(p) = \int_{\Gamma} \left(\frac{\partial}{\partial n} \quad \frac{\partial}{\partial n} \right) dS + Q \qquad (1)$$

Q は点音源の強さ，　はグリーン関数であり，受音点の座標を $p = (x', y', z')$ としたときに次式を満たす関数である。

$$\nabla^2 \quad + k^2 \quad = \quad \delta(x \quad x', y \quad y', z \quad z') \quad (2)$$

ここで障害物が非常に薄いと仮定し，Terai の手法 [3] を用いる。この場合，Fig.2 のように境界面 Γ は障害物の表側 Γ_1 と裏側 Γ_2 の和として表すことができ，それぞれの面の法線ベクトルについて $\boldsymbol{n_1} = \quad \boldsymbol{n_2}$ が成り

A study on vibro-acoustic analysis of membranophones. by ARAKI Yozo, YANAHIRA Yoshinori, SAMEJIMA Toshiya (Kyushu University)

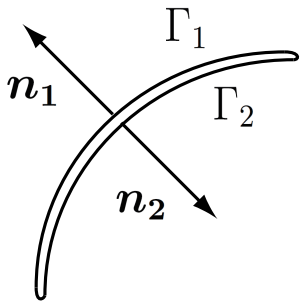


Fig. 2　薄い障害物の概念

立つ。このことと速度ポテンシャル表裏差 $\tilde{}=\ (\Gamma_2)\quad(\Gamma_1)$ を用いると，式 (1) は次式のようになる。

$$(p)=\int_\Gamma\left(\tilde{}\frac{\partial}{\partial n}\qquad\frac{\partial\tilde{}}{\partial n}\right)dS+Q \tag{3}$$

ただし，$\boldsymbol{n}=\boldsymbol{n_2}$，$\Gamma=\Gamma_2$ とおいた。また，境界 Γ 上の点 p と境界を挟んだ裏側の点 $\hat{p}$ での速度ポテンシャルの関係は，速度ポテンシャル表裏差を用いて，次式のように表される。

$$\begin{aligned}&\frac{1}{2}\ (p)+\frac{1}{2}\ (\hat{p})\\&=\int_\Gamma\left(\tilde{}\frac{\partial}{\partial n}\quad\frac{\partial\tilde{}}{\partial n}\ \right)dS+Q\end{aligned} \tag{4}$$

右辺，積分内第 2 項の $\partial\tilde{}/\partial n$ は粒子速度の表裏差であり，今，障害物が非常に薄いと仮定しているので障害物の表側と裏側で粒子速度は等しいと考えられ，$\partial\tilde{}/\partial n=0$ となる。さらに式 (4) を境界上の受音点 p での法線方向で微分すると，

$$\frac{1}{2}\frac{\partial\ (p)}{\partial n_p}+\frac{1}{2}\frac{\partial\ (\hat{p})}{\partial n_p}=\int_\Gamma\tilde{}\frac{\partial^2}{\partial n_p\partial n}dS+Q\frac{\partial}{\partial n_p} \tag{5}$$

となる。ここで，$\partial\ (p)/\partial n_p$ は点 p における粒子速度に等しいため，これを $c(p)$ とおくと，

$$\int_\Gamma\tilde{}\frac{\partial^2}{\partial n_p\partial n}dS=\frac{1}{2}c(p)+\frac{1}{2}c(\hat{p})+Q\frac{\partial}{\partial n_p} \tag{6}$$

となり，速度ポテンシャル表裏差と Γ 上の粒子速度の関係式として表される。境界 Γ を N 個の要素に離散化し，i 番目の要素の法線を n_i とすると次式となる。

$$\sum_{j=1}^{N}\int_{\Gamma_j}\frac{\partial^2}{\partial n_i\partial n}dS\tilde{}_j=\frac{1}{2}c_i+\frac{1}{2}c_{\hat{i}}+Q\frac{\partial}{\partial n_i} \tag{7}$$

これより，粒子速度 c_i，$c_{\hat{i}}$ が求まれば速度ポテンシャル表裏差を求められることになる。さらに，これによって求められた速度ポテンシャル表裏差を式 (3) に代入することによって，音場 Ω 内の任意の受音点における速度ポテンシャルを求めることができる。

2.2　膜振動場の解析

膜振動場の解析には，固有モード展開法を用いる。円形膜振動の方程式は，

$$\begin{aligned}\rho_M\frac{\partial^2\zeta}{\partial t^2}=&P\Big[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\zeta}{\partial r}\right)+\frac{1}{r^2}\left(\frac{\partial^2\zeta}{\partial\varphi^2}\right)\Big]\\&+P_a(r,\varphi,t)+k\rho_M\frac{\partial\zeta}{\partial t}\end{aligned} \tag{8}$$

と与えられる。ここで ζ は膜の変位，ρ_M は面密度，P は張力，P_a は膜に作用する単位面積当たりの加振力，k は制動係数である。円形膜の固有モード関数 N_{nm} は，

$$N_{nm}=J_n\left(\frac{nm}{a}r\right)\cos n\varphi \tag{9}$$

と書ける。$_{nm}$ は，n 次ベッセル関数の m 番目の零点，a は円形膜の半径である。式 (8) の時間項を $e^{j\omega t}$ として調和振動を仮定し，固有モード関数 N_{nm} を用いると，変位 ζ は，

$$\begin{aligned}&\zeta(r,\varphi)\\&=\sum_{n=0}^{\infty}\sum_{m=1}^{\infty}\frac{\int_0^a\int_0^{2\pi}P_aN_{nm}rd\varphi drN_{nm}(r,\varphi)}{\rho_MM_{nm}(\omega_{nm}^2\quad\omega^2+jk\omega)}\end{aligned} \tag{10}$$

となる。ここで M_{nm} は規準化因数であり，

$$\begin{cases}M_{0m}=\pi a^2J_1^2(\ _{0m}) & (n=0)\\ M_{nm}=\frac{\pi a^2}{2}J_{n\ 1}^2(\ _{nm}) & (n>0)\end{cases} \tag{11}$$

である。

2.3　音場と振動場の連成

音場と振動場それぞれについての方程式が得られたので，次にそれらの方程式を連立し

て解くことを考える。まず，膜面上の z 方向粒子速度は，その点における膜の振動速度と等しくなるため，

$$\frac{\partial \zeta}{\partial t} = \quad \frac{\partial}{\partial n_p} \tag{12}$$

という関係が成り立つ。また，膜の表側と裏側の粒子速度は等しいと考えられ，

$$c(p) = c(\hat{p}) \tag{13}$$

が成立する。次に，膜に加わる圧力 P_a について考える。音場内にある膜に加わる加振圧力は，膜の両面にかかる音圧差と機械的な加振圧力で与えられる。機械的な加振圧力を f とし，膜面上の法線の向きを z 軸の正の方向にとると，膜に加わる加振圧力は，

$$P_a = f \quad \rho \frac{\partial \tilde{}}{\partial t} \tag{14}$$

となる。これを式 (10) に代入し，膜の振動速度 $j\omega\zeta$ を式 (6) に代入すると，

$$\begin{aligned} &\int_\Gamma \tilde{} \frac{\partial^2}{\partial n_p \partial n} dS \\ =&\frac{j\omega}{\rho_M} \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} \Big[\frac{\int_0^a \int_0^{2\pi} f N_{nm} r d\varphi dr}{M_{nm}(\omega_{nm}^2 \quad \omega^2 + jk\omega)} \\ &\frac{j\omega\rho \int_0^a \int_0^{2\pi} \tilde{} N_{nm} r d\varphi dr}{M_{nm}(\omega_{nm}^2 \quad \omega^2 + jk\omega)} \Big] N_{nm}(r, \varphi) \end{aligned} \tag{15}$$

となる。なお，音場内に膜以外には音源はないものとする。境界を N 個の要素に分割し，そのうち膜の要素数を M 個とすると，式 (15) は，

$$\begin{aligned} &\sum_{j=1}^{N} \int_{\Gamma_j} \frac{\partial^2}{\partial n_i \partial n} dS \tilde{}_j \\ &\frac{\omega^2 \rho}{\rho_M} \sum_{j=1}^{M} \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} \frac{\int_{\Gamma_j} N_{nm} dS N_{nm}(i) \tilde{}_j}{M_{nm}({\omega_{nm}}^2 \quad \omega^2 \quad jk\omega)} \\ =&\frac{j\omega}{\rho_M} \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} \frac{\int_0^a \int_0^{2\pi} f N_{nm} r d\varphi dr N_{nm}(i)}{M_{nm}({\omega_{nm}}^2 \quad \omega^2 \quad jk\omega)} \end{aligned} \tag{16}$$

と書くことができる。式 (16) を全要素について立て，得られた連立方程式を解くことにより，各要素の速度ポテンシャル表裏差を求めることができる。求められた速度ポテンシャル表裏差を式 (3) に代入することにより，音場 Ω 内の任意の受音点における速度ポテンシャルを求めることができる。

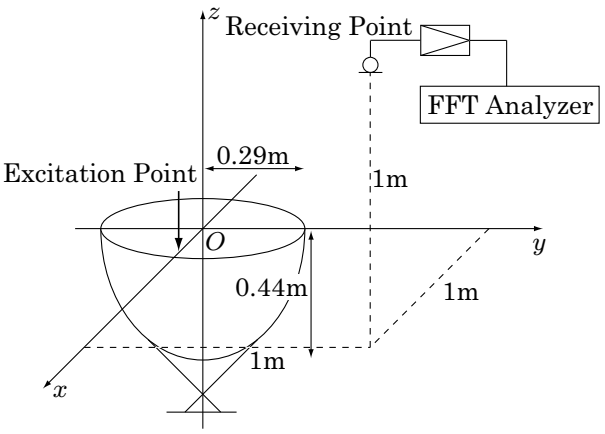


Fig. 3　計算と実測における受音点位置

3　実測との比較

本解析手法の妥当性を確認するため，実際にティンパニ (半径 0.29m，高さ 0.44m) のヘッドに，スティックにより加振圧力が単位インパルス関数となるように与え，放射音の周波数応答関数を測定し，本手法による計算結果と比較した。計算に用いた張力の値は，測定より得られた $(0, 1)$ モードの固有周波数を用いて逆算して求めた。Fig.4〜Fig.7 はそれぞれ，Fig.3 の座標系において，加振点を A(0.05, 0, 0)，B(0.10, 0, 0)，C(0.15, 0, 0)，D(0.20, 0, 0) としたときの受音点 $(0, 0, 1)$ での周波数応答関数である。実線が本手法を用いて計算より得られた周波数応答関数，破線が実測より得られた周波数応答関数である。計算に用いた膜の密度は $0.3\mathrm{kg/m^2}$，張力は $7000\mathrm{N/m}$ である。

Fig.7 は，計算結果と実測結果において比較的よい一致を示している。また，加振点を膜の中心から外側に移していくにつれ，励起されるモードの数が増える傾向が計算と実測ともに見られる。誤差の原因として，今回計算においては膜振動の減衰や曲げ剛性などを考慮していないこと，また，計算条件と実測条件が厳密には合致していないことなどがあげられる。

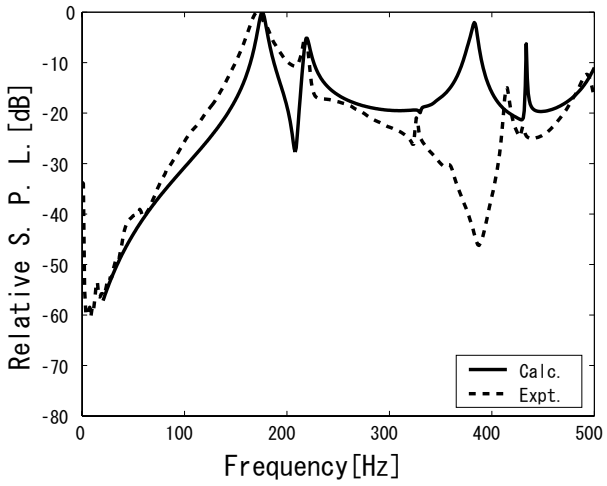


Fig. 4　加振点を A(0.05, 0, 0) としたときの周波数応答関数

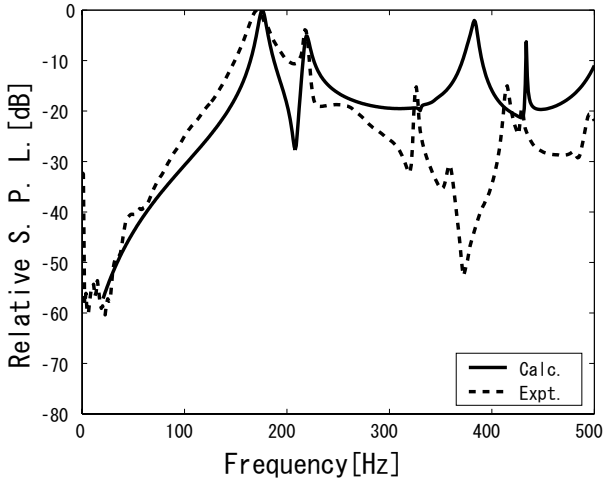


Fig. 5　加振点を B(0.10, 0, 0) としたときの周波数応答関数

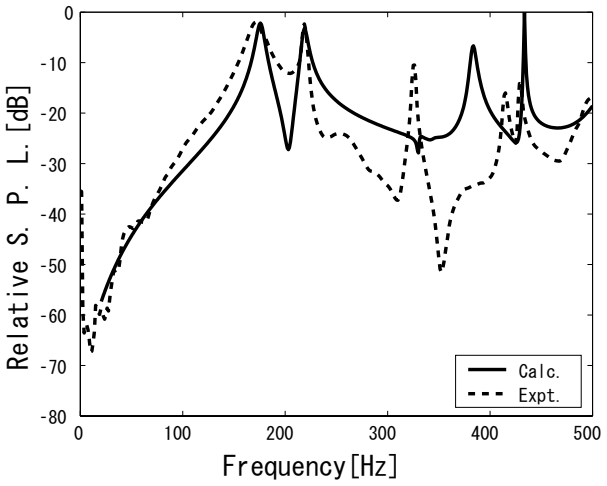


Fig. 6　加振点を C(0.15, 0, 0) としたときの周波数応答関数

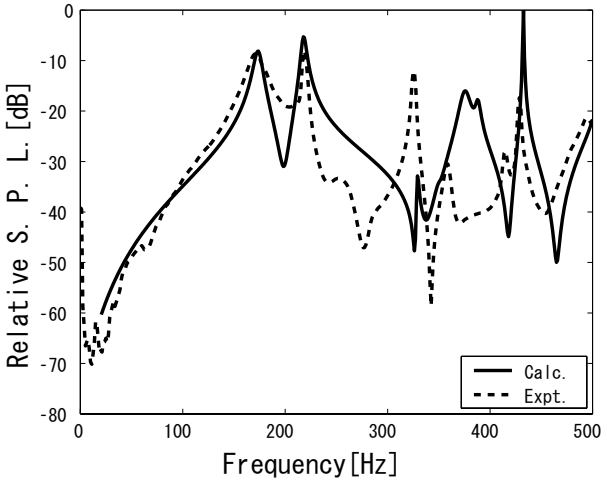


Fig. 7　加振点を D(0.20, 0, 0) としたときの周波数応答関数

4　まとめ

ティンパニを円形膜と音場でモデル化し，膜振動にモード展開，音場に境界要素法を用いた解析手法の定式化を行った。また，この解析手法による計算結果の妥当性を確認するため，実測結果との比較を行った。

今後の課題として，まず解析精度の向上が求められるであろう。具体的には，膜振動の減衰や曲げ剛性などを考慮した解析などがあげられる。また，この解析手法を用いてティンパニが調和的な部分音構造をもつ要因をより詳細に調べ，より最適なケトルの形状を提案できるであろうと考えられる。

参考文献

[1] R. S. Christian, R. E. Davis, A. Tubis, C. A. Anderson, R. I. Mills, and T. D. Rossing, "Effects of air loading on timpani membrane vibrations," J. Acoust. Soc. Am. 76(5), 1336-1345 (1984)

[2] G. Sathej, R. Adhikari, "The eigenspectra of Indian musical drums," J. Acoust. Soc. Am. 125(2), 831-838 (2009)

[3] T. Terai, "On culculation of sound fields around three dimensional objects by integral equation methods," The Journal of Sound and Vibration, 69, 71-100, (1980)

H_∞ 制御理論に基づく 2 入力 2 出力系の逆フィルタ設計

竹下　真（九州大学）　鮫島俊哉（九州大学）*

1　はじめに

音場再生における逆フィルタの設計手法として最小自乗法による近似や離散フーリエ変換による近似を用いた手法がある。しかし, 最小自乗法では係数長を十分に長くしても誤差がある値以下にはならない, 離散フーリエ変換を用いた手法では逆フィルタの係数長を元のインパルス応答の 8〜16 倍にしなければならない, 遅延の許されない応用には向かないなどの問題点がある. そこで, これらの問題点を改善するべく先行研究 [1] として H_∞ 制御理論による逆フィルタ（本研究では H_∞ 逆フィルタと呼ぶ）の設計が提案された. 先行研究では, 図 1 のトランスオーラルシステムに対して H_∞ 逆フィルタの設計が行われ, 2 入力 2 出力のシステムに対して逆フィルタ設計が可能であること, 最小自乗法に基づく逆フィルタの制御には及ばないものの音場再生を行う上で十分な制御効果を得ることが示された. 本研究では,H_∞ 逆フィルタの設計のためのモデルマッチング問題に対して線形行列不等式（LMI）という解法を適用した. 計算規模を低減するために式変形を行い, 1 入力 1 出力のシステムの計算を 4 回行うことで 2 入力 2 出力系の逆フィルタを設計する. また,H_∞ 逆フィルタと最小自乗法に基づく逆フィルタとの制御効果について比較, 検討を行い,H_∞ 逆フィルタの有効性について考察を行う.

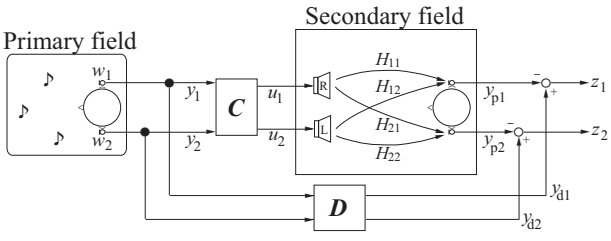


Fig. 1　トランスオーラルシステム

2　H_∞ 制御理論

図 2 に示すシステムを考える.$\boldsymbol{D}(z)$ は目標関数,$\boldsymbol{P}(z)$ は制御対象,$\boldsymbol{C}(z)$ はこのシステムで設計される制御器である. モデルマッチング問題とは誤差 $\boldsymbol{z}$ がある評価関数の下で最小の値となる制御器 $\boldsymbol{C}(z)$ を求める問題である.

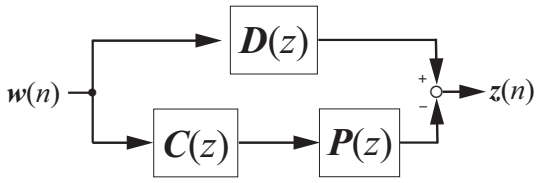


Fig. 2　モデルマッチング問題のブロック線図

図 3 は H_∞ 制御問題の形式で表されるモデルマッチング問題のブロック線図である. ただし,$\boldsymbol{D}$ は目標関数,$\boldsymbol{P}$ は制御対象,$\boldsymbol{C}$ は制御器,$\boldsymbol{w}$ は外乱,$\boldsymbol{y}$ は観測出力,$\boldsymbol{u}$ は制御入力,$\boldsymbol{z}$ は制御量である.$\boldsymbol{w}$,$\boldsymbol{z}$ 間の伝達関数 $\boldsymbol{\Phi}(z)$ について,

$$\| \boldsymbol{\Phi}(z) \|_\infty < \gamma \tag{1}$$

を満たすような制御器 $\boldsymbol{C}$ を求めることにより, 制御量 $\boldsymbol{z}$ を小さく保つことが制御の目的となる. また,$\boldsymbol{G}$ は一般化プラントと呼ばれ次のような状態方程式で表される.

$$\begin{cases} \boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}_1\boldsymbol{w}(k) + \boldsymbol{B}_2\boldsymbol{u}(k) \\ \boldsymbol{z}(k) = \boldsymbol{C}_1\boldsymbol{x}(k) + \boldsymbol{D}_{11}\boldsymbol{w}(k) + \boldsymbol{D}_{12}\boldsymbol{u}(k) \\ \boldsymbol{y}(k) = \boldsymbol{C}_2\boldsymbol{x}(k) + \boldsymbol{D}_{21}\boldsymbol{w}(k) + \boldsymbol{D}_{22}\boldsymbol{u}(k) \end{cases} \tag{2}$$

本研究では, 図 3 のブロック線図を図 1 の 2 入力 2 出力のトランスオーラルシステムに適用し, 式 (1) を基に H_∞ 逆フィルタをモデルマッチング問題の解として求めた. また, この時, 式変換をし 2 入力 2 出力のシステムに対して一度に解を求めるのではなく 1 入力 1 出力のモデルマッチング問題を 4 回解くことで H_∞ 逆フィルタを求めた. その数学的解析法として LMI を用いた.LMI を用いた理由は,H_∞ 標準問題として解いた場合と比べて, 可解条件の制約がないため, 制御器の設計が容易になるためである.

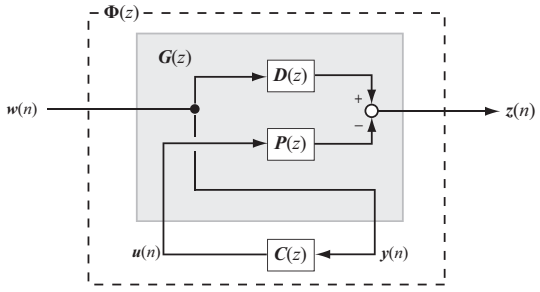


Fig. 3　H_∞ モデルマッチング問題のブロック線図

* Inverse filter design for two-input two-output system based on H_∞ control theory.　by TAKESHITA ,Makoto (Kyushu University)and SAMEJIMA, Toshiya (Kyushu University)

3　提案する手法

図 1 において $\boldsymbol{H}$ はスピーカから耳までの伝達関数を,$\boldsymbol{C}$ は逆フィルタを,$\boldsymbol{D}$ は目標とするモデルを表している. この時, モデルマッチング問題の式を変換することで,2 入力 2 出力のシステムの逆フィルタ設計を 1 入力 1 出力のシステムの設計を 4 回行うことで設計できるようにする [2].

$$\boldsymbol{H}(z)=\begin{bmatrix}H_{11}(z) & H_{12}(z)\\ H_{21}(z) & H_{22}(z)\end{bmatrix}$$

$$\boldsymbol{C}(z)=\begin{bmatrix}C_{11}(z) & C_{12}(z)\\ C_{21}(z) & C_{22}(z)\end{bmatrix},\boldsymbol{I}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

とし, このシステムの逆フィルタ $\boldsymbol{C}(z)$, 伝達関数 $\boldsymbol{H}(z)$, 目標関数 $D(z)$ が満たすべき式は,

$$\boldsymbol{I}\cdot D(z)\cdot z^{-\Delta}=\begin{bmatrix}H_{11}(z) & H_{12}(z)\\ H_{21}(z) & H_{22}(z)\end{bmatrix}\begin{bmatrix}C_{11}(z) & C_{12}(z)\\ C_{21}(z) & C_{22}(z)\end{bmatrix} \tag{3}$$

となる. 式 (3) において,Δ はシステムの因果性を保証するモデリングディレイを表している. この式 (3) 中の伝達関数 $\boldsymbol{H}(z)$ に対して,$\boldsymbol{H}(z)$ の逆行列 $\boldsymbol{H}(z)^{-1}$ を両辺にかける.$\boldsymbol{H}(z)$ の逆行列は,

$$\boldsymbol{H}(z)^{-1}=\frac{1}{H_{11}(z)H_{22}(z)-H_{12}(z)H_{21}(z)}\begin{bmatrix}H_{22}(z) & -H_{12}(z)\\ -H_{21}(z) & H_{11}(z)\end{bmatrix} \tag{4}$$

より, 式 (3) は, 式 (5) に変換できる.

$$det(\boldsymbol{H}(z))\begin{bmatrix}C_{11}(z) & C_{12}(z)\\ C_{21}(z) & C_{22}(z)\end{bmatrix}=\begin{bmatrix}H_{22}(z) & -H_{12}(z)\\ -H_{21}(z) & H_{11}(z)\end{bmatrix}D(z)\cdot z^{-\Delta} \tag{5}$$

$det(\boldsymbol{H}(z)) = H_{11}(z)H_{22}(z) - H_{12}(z)H_{21}(z)$ であり,$\boldsymbol{H}(z)$ の行列式である. この式 (5) から C_{11} ~ C_{22} の各逆フィルタを一つずつモデルマッチング問題に適用する. 図 3 の $\boldsymbol{P}$,$\boldsymbol{D}$ を, 式 (5) より $\boldsymbol{P} = det(\boldsymbol{H}(z)),\boldsymbol{D} = H_{22}(z)\cdot D(z)\cdot z^{-\Delta}$ と置くことで逆フィルタ C_{11} の設計が可能になる.C_{12},C_{21},C_{22} に対しても同様である.

4　計算機シミュレーションによる検討

図 1 のシステムにおいて H_∞ 逆フィルタと最小自乗法で求めた逆フィルタについて計算機シミュレーションを行い, 制御効果の比較, 検討を行う. この時, 制御器の規模 (自由度) が等しくなるようにした. 先行研究では, 制御器の規模をパラメーターに制御効果の違いを比較していた. 本研究では, 制御器の規模を 200 次 (最小自乗法の逆フィルタではタップ数 4× 50[point]) とした場合のモデリングディレイの違いによる制御効果の違いに着目する.

4.1　シミュレーション条件

MIT のデータベース [3] の HRIR を制御対象とした. 測定状況は図 4 に示すように, ダミーヘッドの中心からスピーカまでの距離が 1.4m, スピーカ間隔が 60 度, 仰角が 0 度である. サンプリング周波数は 2450[Hz] としてシミュレーションに用いた. 測定状況の対称性から, 図 1 における H_{22} は H_{11} と,H_{12} は H_{21} と同じ伝達特性であると考え,(左のスピーカ～左耳までの HRIR)(左のスピーカから右耳までの HRIR) のみを使用した. 図 5 にシミュレーションで用いた HRIR の周波数応答を示す. 伝達関数のモデル化手法には, 特異値分解法を用いた. 目標関数には, カットオフ周波数 100[Hz],1000[Hz] のバンドパスフィルタを用い, モデリングディレイの値を 8[point]～17[point] まで変化させた. また, 原音場にお

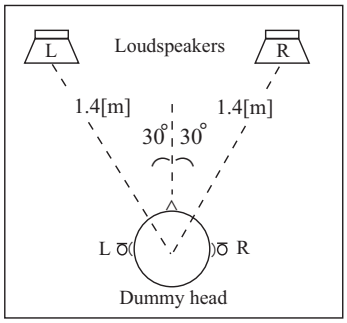


Fig. 4　使用した制御対象の測定状況

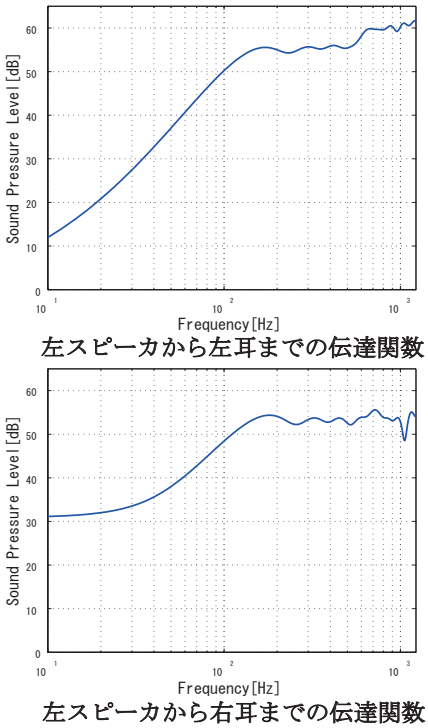


Fig. 5　制御対象 (HRIR) の周波数応答

いて $\boldsymbol{w}(z) = \{0, 1\}^T$ を入力した場合を考える. 再生音場においては, 左耳では逆フィルタにより等化された周波数応答特性が, 右耳ではクロストークキャンセル効果が得られていることが望ましい.

4.2 結果とおよび考察

図 6 と図 7 にモデリングディレイの長さを 11[point] に設定した時の, 制御後の周波数応答を示す. この時, 左耳を実線, 右耳を破線で表している.H_∞ 逆フィルタが最小自乗法に基づく逆フィルタに比べてクロストークキャンセルがとれていることがわかる.

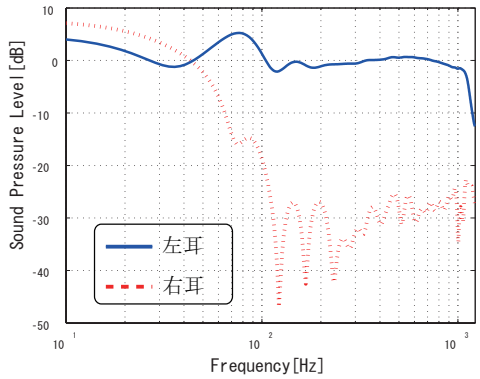


Fig. 6 制御後の周波数応答 (H_∞ 理論, モデリングディレイ=11[point])

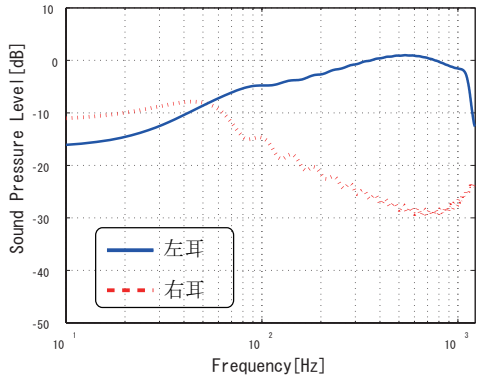


Fig. 7 制御後の周波数応答 (最小自乗法, モデリングディレイ=11[point])

そこで, モデリングディレイを 8[point]〜17[point] に変化させた場合の制御効果の違いを評価関数を用いて検証する.

評価関数 J は標準偏差により周波数応答の平坦さを表す評価関数である [4].J は次式で表される.

$$J = \left[\frac{1}{N} \sum_{i=N_L}^{N_H} (20 \log |Y_2(\omega_i)| - r)^2 \right]^{1/2} \qquad (6)$$

ただし,

$$r = \frac{1}{N} \sum_{i=N_L}^{N_H} 20 \log |Y_2(\omega_i)|$$

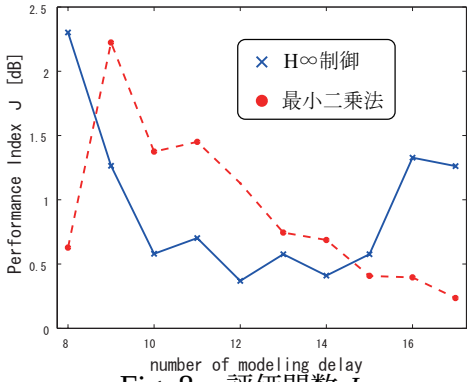


Fig. 8 評価関数 J

である。ここで,$Y_2(\omega_i)$ は逆フィルタによる左耳での制御後の周波数応答である. また, 目標関数であるバンドパスフィルタの周波数応答が平坦である帯域を逆フィルタの制御帯域とするため, N_L は目標関数の低域カットオフ周波数に対応するポイント,N_H は高域カットオフ周波数に対応するポイント,$N = N_H - N_L + 1$ である. J の値が小さいほど周波数応答が平坦であることを示す. 図 8 に評価関数 J による結果を示す. ここでも, モデリングディレイが 9[point]〜14[point] まで周波数応答の平坦さで最小自乗法より優位なことがわかる.

β はクロストークキャンセル量を表す評価関数である [5].β は次式で表される.

$$\beta = 20 \sum_{k=N_L}^{N_H} \frac{1}{k} \log \frac{|Y_2(k)|}{|Y_1(k)|} \Bigg/ \sum_{k=N_L}^{N_H} \frac{1}{k} \qquad (7)$$

$Y_2(k)$ は左耳での制御後の周波数応答,$Y_1(k)$ は右耳での制御後の周波数応答であり, 左耳に対する右耳のクロストークキャンセル量を示す. β の値が大きいほどクロストークキャンセルが実現できている. β の値が 20[dB] 以上あれば聴感上十分な音場再生ができると言われている. 図 9 に評価関数 β による結果を示す. 最小自乗法と比べてモデリングディレ

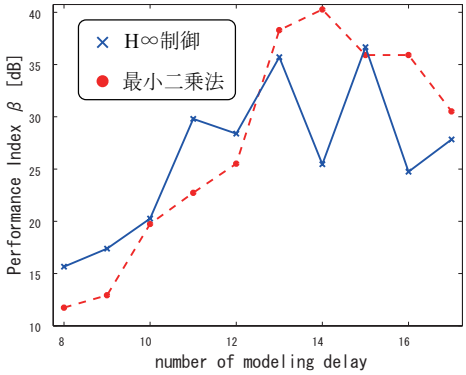


Fig. 9 評価関数 β

イの短い 8[point]〜12[point] で H_∞ 逆フィルタのクロストークキャンセル量が勝っていることがわかる.

$\|H\|_\infty$ は二つの周波数応答の差の最大値を表す評価関数である. 次式で定義される.

$$\|H\|_\infty = \max_{\omega_{N_L} \leq \omega \leq \omega_{N_H}} |Y_d(\omega_i) - Y_r(\omega_i)| \quad (8)$$

まず, 再生音場の左耳での周波数応答がどれだけ等化されているかを $\|H\|_\infty$ で評価する. 式 (8) の中の $Y_d(\omega_i)$ を目標関数の周波数応答,$Y_r(\omega_i)$ を左耳での周波数応答とし評価する. この時, 値が小さいほど目標関数との差が小さいことになる. 図 10 に評価関数 $\|H\|_\infty$ による結果を示す.J,β 同様にモデリングディレイが短い 8[point]〜14[point] の範囲で制御効果と目標関数の差が最小自乗法よりも優れている.

次に, 制御後の左耳と右耳の周波数応答の差を $\|H\|_\infty$ で評価する. 式 (8) の中の $Y_d(\omega_i)$ を左耳の周波数応答,$Y_r(\omega_i)$ を右耳の周波数応答とする. $\|H\|_\infty$ の値がクロストークキャンセル量の最大値を表している. 図 11 に評価関数 $\|H\|_\infty$ の結果を示す. 全体的に,H_∞ 逆フィルタが勝っていることがわかる.

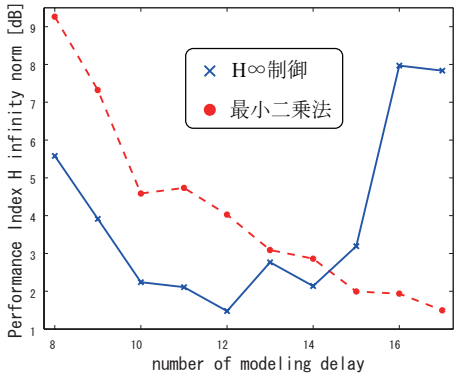


Fig. 10　評価関数 $\|H\|_\infty$(目標関数と左耳の周波数応答の差の最大値)

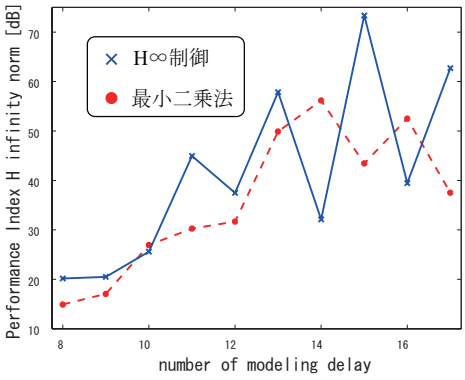


Fig. 11　評価関数 $\|H\|_\infty$(左耳と右耳の周波数応答の差の最大値)

今回のシミュレーションの結果から H_∞ 逆フィルタは, 最小自乗法に基づく逆フィルタに比べて同等かそれ以上の制御効果を挙げていることがわかる. モデリングディレイを変えてシミュレーションを行った結果, 特に,H_∞ 逆フィルタは最小自乗法による逆フィルタに比べ, 遅延が短い場合でも制御効果を得ることを確認した. ただ, モデリングディレイが長くなった場合に, 最小自乗法と比べて劣った制御性能を示している., 今回のシミュレーションは制御器の規模を 200 次と設定して行っていることから, 一般化プラントの中のモデリングディレイの占める割合が大きくなりモデル化誤差を生んだことが原因だと考えられる. モデリングディレイが長くなった場合は, モデル化誤差を小さくするために制御器の規模を大きくすることで制御効果を高めることができる.

5　まとめと今後の検討

本研究では, スピーカを用いた音場再生において H_∞ 制御理論を用いた 2 入力 2 出力の逆フィルタの設計を行い, 制御効果を最小自乗法と比較した. 逆フィルタ設計時に, 式変換を行い 1 入力 1 出力のモデルマッチング問題を線形行列不等式 (LMI) で解くことで, 計算規模を低減した.

計算機シミュレーションの結果より, 最小自乗法に基づく逆フィルタに比べても同等な制御効果を得ることができることがわかった. また, 評価関数による検討により遅延に対して H_∞ 逆フィルタが優位な性質を持っていることが確認された.

今回のシミュレーションは低次の伝達関数に対して行うことで逆フィルタの設計に H_∞ 制御理論を用いた場合の制御効果を検証することができた. これからは, より高次な伝達関数に対して H_∞ 逆フィルタの設計を行っていきたい.

参考文献

[1] 藤田亮太, “ H_∞ 制御理論による逆フィルタの設計とその音響問題への適用 ”, 九州大学修士論文,2009
[2] Sang-Myeong Kim and Semyung Wang , “ A Wiener filter approach to the binaural reproduction of stereo sound ”J.Acoust.Soc.Am. 114(6),pp3179-3188,2003
[3] http://sound.media.mit.edu/KEMAR.html
[4] P.M.Clarkson,et. al.,“ Spectral,Phase,and Transient Equalization for Audio Systems ”J.Audio Eng.Soc 33(3),pp127-132,1985
[5] 浜田晴男, “ 基準音収音・再生を目的とするOrthostereophonic System の構成 ”, 日本音響学会誌,vol.39,no.5,pp.337-348,May 1983.

動的圧縮型ガンマチャープフィルタを用いた音場評価法に関する検討

松本悠希, 鈴木正博, 尾本章 (九大芸工)

1 はじめに

近年, 音場を評価するためには音響物理指標が一般的に用いられており, 残響時間,Clarity(C), inter-aural cross correlation(IACC) などが代表的なものとして挙げられる. 特に残響時間に関しては, 減衰の初期から読み取る“初期減衰時間 (Early Decay Time)”が主観的な残響感と強い相関関係にあると言われている.

本研究は, 室の純粋なインパルス応答に動的圧縮型ガンマチャープフィルタ“Dynamic compressive gammachirp filter”[3](以下 dcGC) を適応し, そこから得られる残響曲線及び減衰時間に着目する. 特に従来の EDT との比較を通して, 室内音場における聴覚の特性に即した残響感の評価手法の基礎的検討を行うものである.

2 動的圧縮型ガンマチャープフィルタによる分析方法

2.1 動的圧縮型ガンマチャープフィルタ

Patterson らは基底膜の振動の生理学的データとノッチドノイズ法による心理学的データより導かれたガンマトーンフィルタと呼ばれる聴覚フィルタの聴覚モデルへの導入を行った. この聴覚フィルタは, ガンマ分布関数によって描かれるエンベロープと正弦波の搬送波で構成され, 基底膜の振動を近似している. ガンマトーンフィルタのインパルス応答を (1) 式に示す.

$$g_t(t) = at^{n_1\ -1}\exp(\ \ 2\pi b_1 ERB_N(f_{r1})t)\exp(j2\pi f_{r1}t + j\varphi_1). \quad (1)$$

ここで, a は振幅, n_1 と b_1 はガンマ分布関数によるエンベロープを決定するパラメーターであり, f_{r1} は聴覚フィルタの中心数波数. φ_1 は初期位相角を表している. また, ガンマトーンフィルタの次数, つまり n_1 を 4 とすると聴覚フィルタとして有効な近似が得られる. $\mathrm{ERB_N}(f_r)$ は聴覚フィルタの等価矩形帯域幅であり, (2) 式で定義される.

$$\mathrm{ERB_N}(f_r) = 24.7(4.37f_r(kHz) + 1). \quad (2)$$

また Patterson と入野は, 人間の初期聴覚系の処理形態は時間–スケール表現であると仮定し, その最小不確定性をみたすフィルタ関数であるガンマチャープ関数 ((3) 式) を導出した.

$$g_c(t) = at^{n_1\ -1}\exp(\ \ 2\pi b_1 ERB_N(f_{r1})t)\exp(j2\pi f_{r1}t + jc_1 \ln t + j\varphi_1). \quad (3)$$

(1) 式のガンマトーンフィルタとの違いは $jc_1 \ln t$ の項であり, 周波数変調項 c_1 と時間の自然対数 $\ln t$ の積により時変型のガンマチャープフィルタが定義されている。このガンマチャープ聴覚フィルタは心理物理実験の結果にガンマトーンフィルタより適合しやすいという成果を得ている [2]. (1),(3) 式から, ガンマトーンはガンマチャープにおいて $c = 0$ の特別な場合であることがわかる. つまり, 前述した一連の聴覚フィルタは, 計算理論として最適なフィルタ関数の形を有しながら, 生理学的および心理学的な実験データの近似を可能とするものである.

さらに, 聴覚の圧縮特性を模擬するために, 入力信号のレベルに依存してガンマチャープフィルタの利得を制御する $h_c(t)$ を定義し, (3) 式の $g_c(t)$ と畳み込むにより dcGC を実現する.

$$g_{cc}(t) = a_c\ \ g_c(t) * h_c(t). \quad (4)$$

$h_c(t)$ は周波数領域において高域通過型のフィルタであり, $h_c(t)$ の中心周波数を入力信号の

Method of Acoustic Evaluation by using Dynamic Compressive Gammachirp Filter. by MATSUMOTO,Yuki, SUZUKI,Masahiro and OMOTO,Akira(Faculty of Design/Kyushu University)

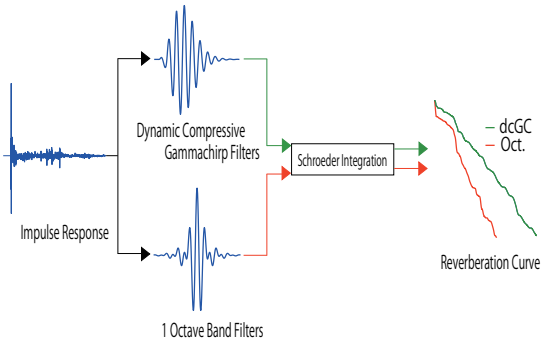


Fig. 1　Block diagram of analysis system

レベルに依存性を持たせ, 変化させることで, 非線形性を有する聴覚圧縮特性のシミュレートが可能な dcGC が定義される.

2.2　分析方法

この研究の基本的な分析方法は, インパルス応答に対してフィルタを用いて信号を分析し, 逆自乗積分を用いて残響曲線を求めるものであり, 手法としては従来法と同じある. ただし, 分析においてはオクターブバンドフィルタと 2.1 節でその概要を記述した dcGC の 2 つのフィルタを用いている. Fig. 1 に分析システムの概要図を示す.

第一にインパルス応答のフィルタリングを行う. 本研究では, スタジオとコンサートホールのインパルス応答を評価対象とする. 分析に用いるフィルタは前述したように 2 種類である. フィルタの中心周波数は 500 Hz, 1k Hz, 2k Hz とした. また, 聴覚フィルタによる分析経路においては, 前処理として等ラウドネス曲線に基づいた設定された FIR 型のフィルタによるフィルタリングを行っている. こうして一つのインパルス応答に対して 2 つのフィルタリングされた信号が作成される. その信号に対してそれぞれ逆自乗積分を行うことで, 一つの分析対象のインパルス応答から 2 種類の残響曲線が作成される. この二つの残響曲線に対する比較　検討を行う.

3　スタジオのインパルス応答を用いた検討

分析対象としたインパルス応答の測定は 5.1 ch サラウンドシステムを備えたスタジオにて行われた. 残響時間は 0.12 sec から 0.14 sec 程度である. 本研究ではセンタースピーカーを音源としたインパルス応答を分析して導出した残響曲線について検討を行った. 2.1 節で述べたように, 聴覚フィルタのインパルス応答の形は入力信号のレベルに依存しては時的に変化する. そのため, 入力信号のレベルの違いが聴覚フィルタの出力から得られる残響曲線の形を変化させ, そこから読み取られる減衰時間も変化する. この影響を観測するために, 一つのインパルス応答から振幅値が定数倍された数種類のインパルス応答を分析した. Fig. 2 にフィルタの中心周波数が 500 Hz, 1000 Hz, 2000 Hz として分析した残響曲線を示す. 図中の 50 dB, 80 dB は瞬時最大レベルを 50 dB, 80 dB としたインパルス応答を dcGC で分析した残響曲線であり, Oct. はオクターブバンドフィルタによる分析結果である. 図中にはそれぞれの残響曲線に直線近似を行い読み取った 60 dB 減衰するまでの減衰時間を記してある. 残響曲線における極初期の減衰 (0〜2 msec 付近) は, フィルタ自身の立ち上がりによる影響であるため評価の対象とはしない. また測定における S/N 比の問題により, 定常状態から 0.1 sec までを評価対象としている. インパルス応答のレベルの違い毎に, オクターブバンドフィルタと聴覚フィルタによる残響曲線をそれぞれ比較すると下記のような傾向が伺える.

聴覚フィルタは入力信号の dB 値を音圧レベルと対応づけられるよう設計されているので, 瞬時最大レベル 80 dB のインパルス応答を分析した結果は, 瞬時最大音圧レベル 80 dBSPL のインパルス応答を聴取した条件として検討することができる. オクターブバンドフィルタによる初期の残響曲線において比較的直線近似しやすい区間の回帰直線と, dcGC による残響曲線の回帰直線はほぼ平行となっており, 減衰時間が近づいている傾向が伺える. また, オクターブバンドにおいて回帰直線をフィッティングさせた区間に注目すると, 初期 5 dB から 10 dB の減衰区間となっており,EDT の算出範囲に近い. この傾向は 3 つの周波数帯域で現れており, 80 dBSPL という音圧レベルが主観聴取実験で採用される基準音圧レベルに近いことを考慮すれば, dcGC において確認される減衰時間が残響感と対応している可能性は十分あるといえる.

さらに, 50 dB と 80 dB の残響曲線からそ

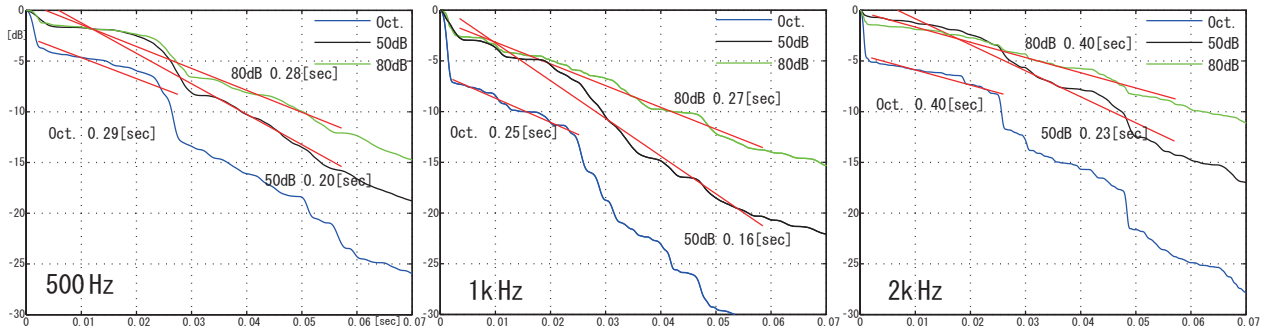


Fig. 2　Reverberation Curves at the studio

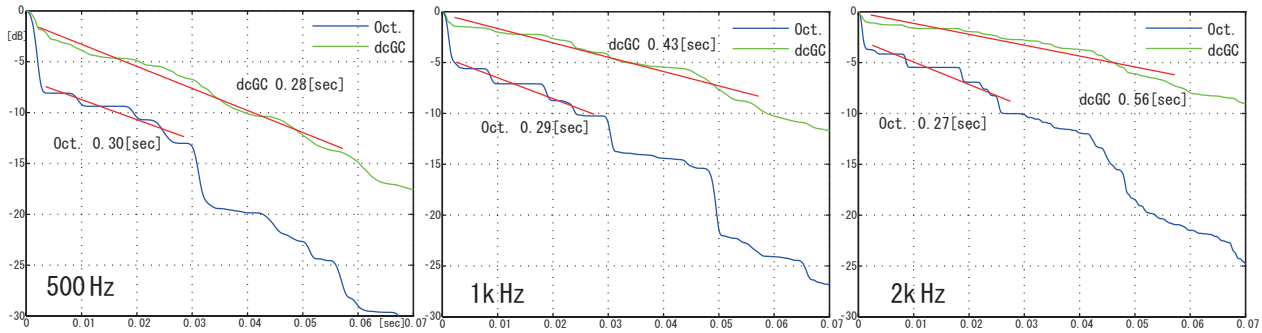


Fig. 3　Reverberation curves of sparse impulse response

れぞれ読み取った減衰時間に着目すると, 50 dB から 80 dB のインパルス応答のレベル増加に伴い減衰時間の増加が全周波数帯において観測された. その増加率は 500 Hz で 40%, 1000 Hz, 2000 Hz で約 70%であった. 以上のことから, インパルス応答のレベルの変化に応じた残響感の変化を考慮した評価を行える可能性を示したと言える.

4　反射音密度が疎なインパルス応答を用いた検討

比較的デッドな空間に対する聴覚フィルタを用いた分析法の効果を確認するために, 反射音成分が少ないインパルス応答を分析した残響曲線を算出した. 反射音成分の少ないインパルス応答は, 実際に測定されたインパルス応答の極値のみを抽出した後, 一定のエネルギー以下の反射音成分を, 優勢な反射音成分に加算することで作成した (Fig. 4). こうして作成されたインパルス応答は, 元のインパルス応答と比べて, 全体のエネルギーは保ちながらも反射音の密度が異なる. そのため, それぞれのインパルス応答から観測される残響時間はほぼ同値となった. 3 節で分析対象としたインパルス応答の反射音密度を変

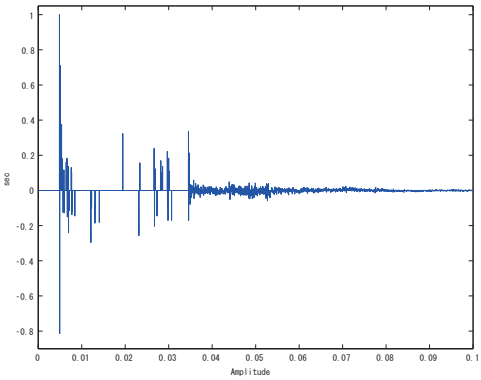


Fig. 4　Sparse impulse response

化させたインパルス応答から算出した残響曲線を Fig. 3 に示す. 図示するのはインパルス応答の瞬時最大レベルを 80 dB とした場合である.

Fig. 2 と Fig. 3 を比較すると, 反射音密度が疎なインパルス応答を dcGC で分析した残響曲線から観測される減衰時間は, オリジナルのものと比べて上昇している. この観測結果を説明する仮説として, 反射音密度の変化による平均自由行路の変化が考えられる. 反射音密度が疎なインパルス応答は, 元のインパルス応答に比べて平均自由行路が長くなり, インパルス応答が測定される音場の空間が広くなったと見なすことができる. すな

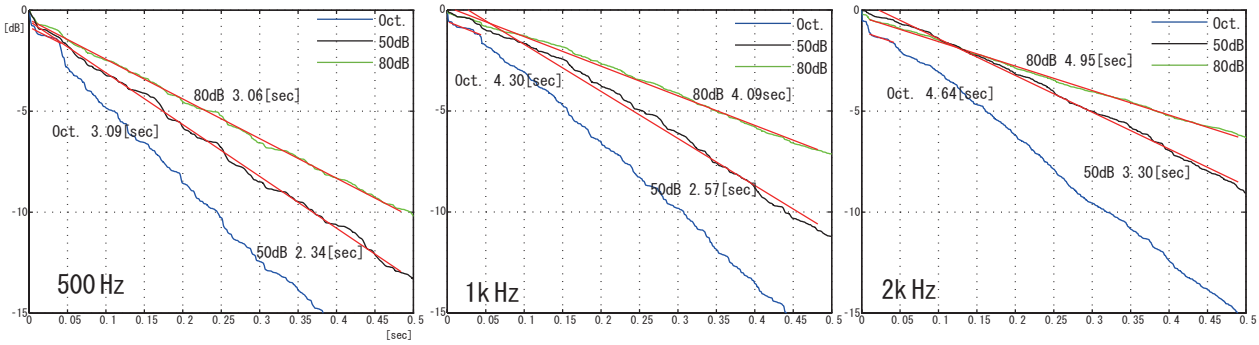


Fig. 5　Reverberation Curves at the hall

わち, 物理的な残響時間が等しくても, 聴覚フィルタを用いた分析により音場の空間の広さに伴う残響感の程度を評価が可能となるかもしれない.

また, Fig. 3 のオクターブバンドを用いた残響曲線を見ると, 時間 20 msec 付近に曲線が平行な区間が存在しており, その区間は最小自乗法を用いた直線近似との誤差が大きくなっている. 対して聴覚フィルタから算出された残響曲線は比較的滑らかであり, 近似した直線もオクターブバンドフィルタの場合に比べて誤差が小さく減衰時間を読み取りやすい. 音線法のような反射音が連続的に到来するインパルス応答をシミュレーションしにくい場合でも, 聴覚フィルタを用いた残響曲線の算出により残響感を考慮した音場のシミュレーションを行える可能性がある.

5　ホールのインパルス応答を用いた検討

今回は先行研究 [1] で用いられたインパルス応答を用いて分析を行った. インパルス応答を測定したホールは残響時間 1.5〜2.5 sec のシューボックス型のコンサートホールである. 任意の一席にて測定されたインパルス応答を分析した結果を Fig. 5 に示す. 3 節と同様に, 入力インパルス応答のレベルを変化させ, 残響曲線に与える影響を観測した.

インパルス応答の瞬時最大レベルが 80 dB の場合の 2 種類の残響曲線を比較した結果, オクターブバンドフィルタによる残響曲線の初期 10 dB ではなく初期 50〜100 msec から読み取れる残響曲線と, dcGC の残響曲線が類似した傾向示すことを確認した. 対して, インパルス応答のレベル変化による減衰時間の変化は, スタジオのインパルス応答を分析した場合と同様に, レベル増加に伴う減衰時間の増加が観測された.

6　おわりに

本研究では, 聴覚の特性を考慮した動的圧縮型ガンマチャープフィルタによる残響減衰過程の評価を行った. この聴覚フィルタの導入により, 聴覚知覚として現れる残響感をインパルス応答という物理量から分析できる可能性を示した. さらには, 残響感を考慮した, すなわち聴覚の特性に即した音場評価を実行できる可能性を提示することができた.

今後は, 残響感に焦点をあてた聴覚フィルタの分析結果と主観評価の対応関係を, 心理実験などを行いながら検討を重ねていきたいと考えている.

参考文献

[1] Taeko Akama and Akira Omoto. Selection of receiving of positions suitable for evaluating acoustical parameters. *Acoustical Society of Japan*, No. 1-6-13(2007,9).

[2] T. Irino. A computional theory of the peripheral auditory system. *Technical report of IEICE*, Vol. 95, No. 140, pp. 23–30, 1996.

[3] T. Irino and R. D. Patterson. A dynamic compressive gammachirp auditory filterbank. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 6, pp. 2222–2232, 2006.

近接 2ch マイクによる距離推定における適用範囲の調査*

伊田匠，近藤善隆，野田裕(日本文理大学)，阿部宏樹，岩上知広(千葉工大)，末廣一美，福島学(日本文理大学)，柳川博文(千葉工大)，黒岩和治(日本文理大学)

１．はじめに

対象物までの距離を計測する技術は基礎的かつ重要な技術である．特にユビキタスネットワーク実現には「現実世界の把握」が「適切な処理の実行」に必要不可欠であり，その需要はますます高くなっている[1]．しかし，工場等ではカメラを用いて正確に距離および位置を合わせる技術があるが，一般家庭での利用を考えるとプライバシの問題から「映像」に対する嫌悪感があり，ホームセキュリティでも宅内に監視カメラを設置することが難しいのが現状である[2]．このため，プライバシに配慮可能かつ簡便な装置で実現可能な距離推定手法が求められる．

映像以外の情報を用いた距離推定手法として，赤外線や超音波を用いるものが実用化されているが，これらは指向性が高く，対象方向以外が検知できないため，広い範囲をカバーするにはアレイ状に配置するか，センサを回転させることが必要となる．

この問題を解決するには比較的波長が長く，広範囲をカバーし，死角の出来にくい可聴域の音波を利用することが考えられる．マイクロ波のレーダで用いられる定在波を利用した距離推定手法を可聴域で実現する技術が提案されその有効性が示されている[3]．この手法は特に近距離における距離推定を主な目的としており，音源近傍で観測することで反射物までの距離推定を可能としている．

一方著者等は可聴域の音響伝送特性であるインパルス応答から風の流れといった映像情報では捉えることの出来ない事象を検出する手法を提案している[4]．この手法は，定在波を用いた距離推定手法と原理式で類似しており，音響伝送特性からも距離推定が可能であると考えられる[5]．著者等はこれまでにこの検証を行い，移動する反射物を捕らえることに成功している．また，厳密に音響伝送特性を推定する必要がないことから，近傍収録信号を音源信号に代用しての距離推定の可能性も報告している[6]．

しかし，これらでは距離推定が可能であることは報告しているが，実際にどの程度の適用範囲なのかについての検討が十分ではなかった．そこで本報告ではこれまでに提案している距離推定手法がどの程度の適用範囲であるかについて調査した結果を報告する．

本報告では，観測点として音源から 2 点を対象とし，それらの計測結果を比較することで適用範囲について検討を行う．

２．実験条件

音源信号を白色雑音とし，ラウドスピーカから放射した信号を,Fig.1 に示す条件で設置したステレオマイク（M1，M2）で収録し，それぞれのインパルス応答を推定する．ここではサンプリング周波数を 44100Hz とし,量子化ビット数を 16bit とする．

*An investigation on the range of the distance estimation using the side-by-side 2ch microphone, by IDA Takumi, KONDOU Yoshitaka, NODA Hiroshi (Nippon Bunri Univ.), ABE Hiroki, IWAKAMI Tomohiro (Chiba Inst. of Tech.), SUEHIRO Kazumi, FUKUSHIMA Manabu (Nippon Bunri Univ.), YANAGAWA Hirofumi (Chiba Inst. of Tech.), KUROIWA Kazuharu (Nippon Bunri Univ.)

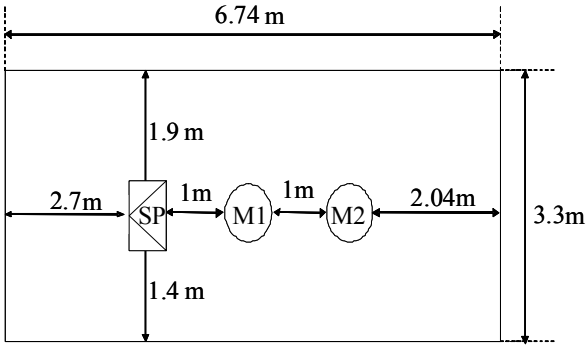


Fig. 1 Configuration of Loud Speaker, Stereo Microphone in a laboratory

信号を $x_p(n)$ とし，ラウドスピーカ信号を p=s，M1 観測信号を p=1R，1L，M2 観測信号を p=2R，2L と表記することとする．但し，n は離散時間に相当するサンプル番号．これらの信号から推定したインパルス応答を，$\tilde{h}_{SO}(n)$ とし，S を音源，0 を観測信号とする．$\tilde{h}_{SO}(n)$ は S=$x_s(n)$ のとき，0=$x_{1R}(n)$，$x_{1L}(n)$，$x_{2R}(n)$，$x_{2L}(n)$ の 4 種類となり，各マイクで観測した信号を S としたとき各 3 種類となる．総計で 16 種類の $\tilde{h}_{SO}(n)$ が得られる．但し，ステレオマイクを音源方向に水平に設置しても，ごく僅かに距離が異なれば，非因果な推定値が存在するため，実際には 10 種類の $\tilde{h}_{SO}(n)$ となる．また 1L と 1R，2L と 2R はほぼ同位置とみなせるため，調査対象を $\tilde{h}_{s1L}$，$\tilde{h}_{s2L}$，$\tilde{h}_{1L1R}$，$\tilde{h}_{2L2R}$，$\tilde{h}_{1R2R}$ の 5 種類の距離スペクトルとする．

推定インパルス応答のパワースペクトルに定在波による変調が現れるため，パワースペクトルをフーリエ変換することで距離スペクトルを求める．Fig. 2 に離散時間で 5 だけ必要な距離に反射物のある条件でパルス信号を出した場合の模式図を示す．図は横軸に離散時間に相当するサンプル番号を示している．この信号のパワースペクトルは Fig. 3 となり，パワースペクトルの周期が反射物までの距離に対応していることが確認できる．このパワースペクトルをフーリエ変換すると，Fig. 4 に示す通り，パワースペクトルの周期，すなわち反射物までの距離に関係するパラメータを抽出することが出来る．Fig. 4 に示したものを距離に対応するスペクトルから導出したものとして距離スペクトルと呼ばれている．実測の場合は，音響伝達特性として複数の反射音が記録されること，距離および反射によって減衰すること，暗騒音が混入する，ことから推定距離が複数計測されることが予想される．

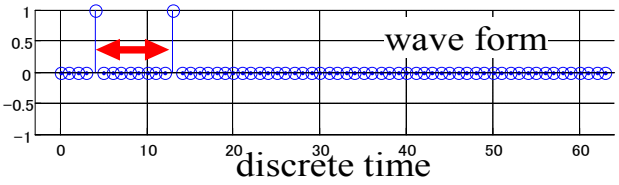


Fig. 2 A wave that indicate measured source signal and reflected signal

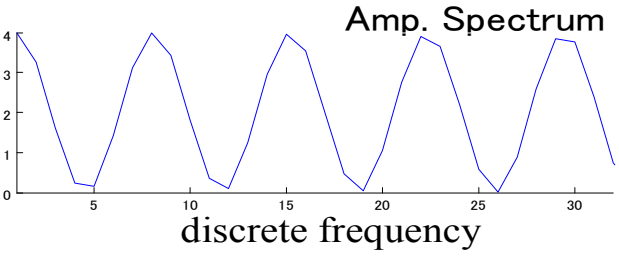


Fig. 3 The power spectrum of Fig. 2

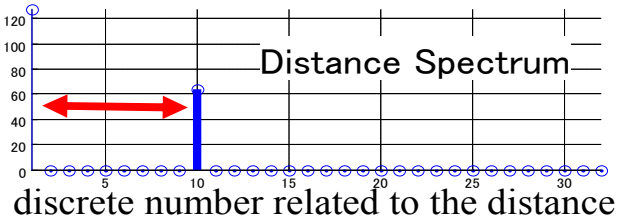


Fig. 4 The distance specrtum from Fig. 2

3．計測実験

計測マイクには SONY 製ステレオマイク（ECM-MS957）を使用した．各条件で求めた距離スペクトルを Fig. 5 から Fig. 9 に示す．図は横軸に推定距離に相当する離散番号を示しており，縦軸にパワースペクトルに現れる変調強度を示している．変調強度が強い，すなわち縦軸の値が大きいほど，顕著な反射であると判断し，推定候補値を黒矢印，予測値を白矢印で図中に示す．予測値と推定候補値が一致

していると白矢印が消え，推定候補から外れていると残る．推定候補は推定距離の値から上位 3 個としている．

Fig.5 に音源信号と M1 観測信号から求めた距離スペクトル $\tilde{h}_{s1L}$ を示す．垂直方向の反射物までの往復距離 6m が求まると推測する．Fig.5 より予想通り距離 2.5m，6m，8m に反射物が検出されていることが確認出来た．

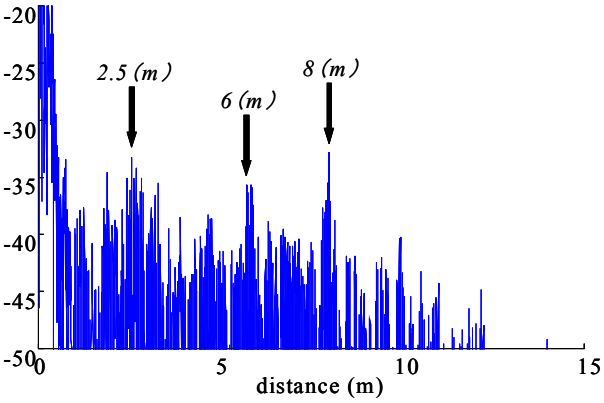


Fig.5 The Distance Spectrum from $\tilde{h}_{s1L}$

Fig.6 に音源信号と M2 観測信号から求めた距離スペクトル $\tilde{h}_{s2L}$ を示す．垂直方向の反射物までの往復距離 4m が求まると推測する．Fig.6 より予想に反して，2m，8m，10m に反射物が検出されている．これは，$\tilde{h}_{s2L}$ の音源位置からマイクまでの距離が $\tilde{h}_{s1L}$ と比べて遠いため，反射が多く検出され本来の検出対象が埋もれていると考えられる．

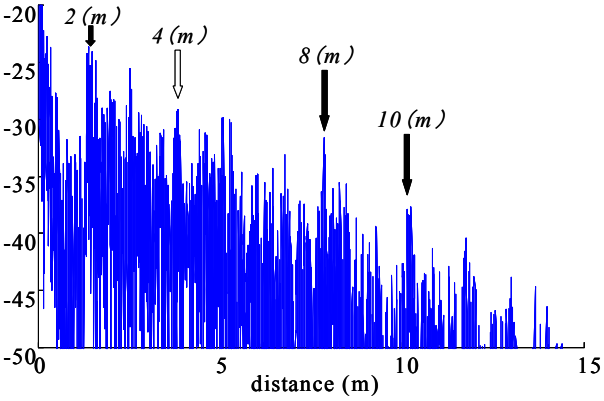


Fig.6 The Distance Spectrum from $\tilde{h}_{s2L}$

Fig.7 に M1 観測信号の L，R から求めた距離スペクトル $\tilde{h}_{1L1R}$ を示す．M1 の位置を基準に垂直方向の反射物までの往復距離 6m が求まると推測する．Fig.7 は予想通り距離 1m，3m，6m に反射物が検出されていることが確認出来た．

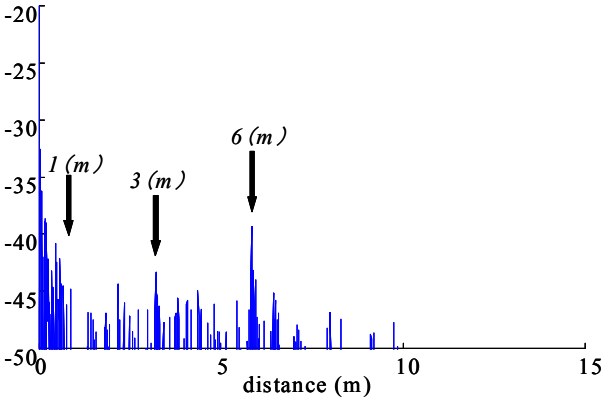


Fig.7 The Distance Spectrum from $\tilde{h}_{1L1R}$

Fig.8 に M2 観測信号の L，R から求めた距離スペクトル $\tilde{h}_{2L2R}$ を示す．M2 の位置を基準に垂直方向の反射物までの往復距離 4m が求まると推測する．Fig.8 は予想に反して距離 2m，5m，8m に反射物を検出している．

Fig.5 と同様に，$\tilde{h}_{2L2R}$ の音源位置からマイクまでの距離が $\tilde{h}_{1L1R}$ と比べて遠いため，反射が多く検出され本来の検出対象が埋もれていると考えられる．

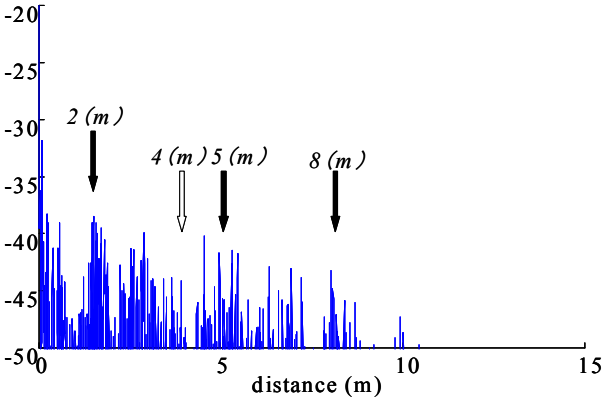


Fig.8 The Distance Spectrum from $\tilde{h}_{2L2R}$

Fig.9 に M1，M2 観測信号から求めた距離スペクトル $\tilde{h}_{1R2R}$ を示す．M2 の位置を基準垂直方向の反射物までの往復距離 4m が求まると推測する．Fig.9 は予想に反して距離 2m，4.8m，8m に反射物を検出している．これは，$\tilde{h}_{1R2R}$ のマイク間の距離が $\tilde{h}_{1L1R}$

に比べて遠いため反射が多く検出され本来の検出対象が埋もれていると考えられる．

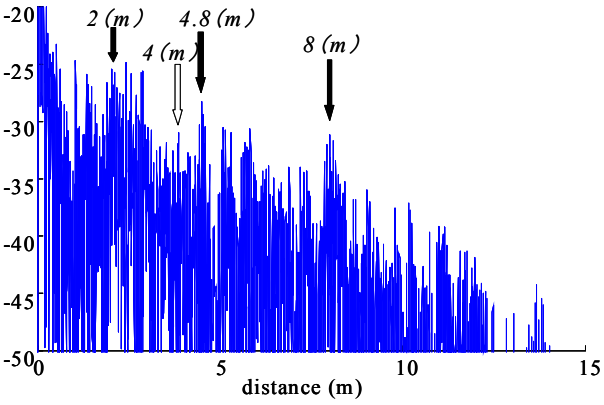


Fig.9 The Distance Spectrum from $\tilde{h}_{1R2R}$

Fig.5 から Fig.9 結果から音源位置からマイク位置までの距離が 1m である Fig.5,Fig.7 では,反射物までの距離の予測値が推定値の中に含まれているが音源位置から距離が 2m の Fig.6,Fig.8 では予測値が推定値の中で他の推定値よりも小さな値となっていた．また，マイク間距離が 1m である Fig.9 でも同様に予測値が推定値の中で他の推定値よりも小さな値となっていた．推定距離を推定候補の大きな値から選出するとなると，推定候補の中で誤推定値の方が大きくなり，本当に求めたい推定距離が候補から外れてしまうことが予測される．このため，本データからは音源位置から反射音を観測する マイクを 1m 以内に設置するか，推定値精度を向上し，誤推定値の値を小さくすることで正しい推定距離が得られるようにしなければならないことがわかった．

４．おわりに

本稿では，因果律を満たす 10 種の距離スペクトルの中から 5 種類を用いて適用範囲の調査を行った．調査の結果から，有効な反射であると判断できる範囲は音源位置からマイク位置までの距離が 1m 程度以内，マイク間の距離が近接であるときであることが実験からわかった．音源位置からマイク位置までの距離，マイク間の距離が遠くなる事で，推定したい方向以外からの多くの反射が入ることが考えられ，有効な反射であると判断する事が難しくなっていることがわかった．適応範囲を拡大するには，有効な反射を判断する指標を考えるか，目的反射音を抽出することが必要であることがわかった．

[参考文献]

[1] 福島学，岡本壽夫，”ユビキタスネットワーク実現のための不可視事象センシング技術の一検討”，日本文理大学，pp.43-50，第 35 号，第 1 号，2007

[2] 福島学，”一般家庭の警備を目的とした音場把握システムの研究”，研究報告集，セコム科学技術振興財団，第 21 巻，2002

[3] 上保 徹志 ，中迫 昇 ，大亦 紀光 ，板垣 英恵,”帯域雑音信号による複数対象物の距離推定”,電気学会論文誌. C，電子・情報・システム部門誌，128 巻，7 号，pp.1117-pp.1122 ,2008

[4] Manabu Fukushima, Hisao Okamoto, Hirofumi Yanagawa, ”AN INVESTIGATION of A SOUND FIELD MONITORING TECHNIQUE FOCUSING on THE DIFFERENCE in SHORT TIME ESTIMATED IMPULSE RESPONSE”, ASJ-ASK Joint Conference on Acoustics 2007, 210-1-210-4, 2007A

[5] 福島学，末廣一美，高山泰典，松本博樹,近藤善隆,阿部宏樹,柳川博文,黒岩和治，”インパルス応答の時間構造分析による距離推定に関する一検討”,日本音響学会 2009 年春季研究発表会講演論文集，3-P-14，2008

[6] 近藤善隆，伊田匠，阿部宏樹，岩上知広，末廣一美，福島学，柳川博文,黒岩和治，”近接 2ch 計測信号による距離推定に関する一検討”，日本音響学会 2009 年秋季研究発表会講演論文集，1-P-3，2009

近接2chマイクによる距離推定における推定精度向上に関する一検討*

近藤善隆,伊田匠,野田裕(日本文理大学),阿部宏樹,岩上知広(千葉工大),末廣一美,
福島学(日本文理大学),柳川博文(千葉工大),黒岩和治(日本文理大学)

1. はじめに

対象物までの距離計測は基本的かつ重要な技術である．特に，簡便かつ少ない素子数で距離を求める技術は，ユビキタスネットワークにおける現実事象を検出するためのセンサとして重要である．

著者等はこれまでにステレオピンポイントマイク程度に近接している 2ch マイク収録信号の片方を音源信号に見立てて音響伝送特性を推定し，そこから距離が推定可能であることを報告してきた[1]．これは，近傍反射音により振幅スペクトルに生じる特徴が検出可能であれば距離推定できることを意味している．しかし，正しい値が推定値の候補に埋もれてしまう事があるという問題がある[2]．

著者等はこれまでにクロススペクトル法を用いたインパルス応答推定において，ノイズレベル推定を同時に行う手法（DLR-CS 法）を提案している[3]．この検証として，減衰補正も行っている [4]．これを距離推定に応用することで，ノイズによる誤推定値の低減および推定値をより顕著化するための減衰補正が可能であると考えた．

本稿では，１）ノイズレベル推定によるノイズ対策，２）減衰補正，による距離推定精度向上を試みた結果を報告する．

2. ノイズレベル推定

音源信号 $x(n)$ がインパルス応答 $h(n)$ の系を経て観測すると，その信号 $y(n)$ は

$$y(n) = x(n) * h(n) = \sum_{p=0}^{M-1} x(n-p)h(p) \quad (1)$$

となる．但し，n は離散時間を示すサンプル番号，M はインパルス応答を示す信号系列のそうサンプル数．これから，$y(n)$ と $x(n)$ の相互相関関数 $\gamma_{xy}(\tau)$ を求めることで $h(n)$ の推定量である $\tilde{h}(n)$ が得られる．$\gamma_{xy}(\tau)$ のスペクトルがクロススペクトルであることから，クロススペクトル法では $\tilde{h}(n)$ のスペクトルである伝達関数 $\tilde{H}(k)$ を

$$\tilde{H}(k) = \frac{\overline{X^*(k)Y(k)}}{X^*(k)X(k)} = DFT\left[\frac{\gamma_{xy}(\tau)}{Px}\right] \quad (2)$$

で求める．但し，k は離散周波数に相当するサンプル番号，$\overline{X}$ は X の多数回平均，Px は $x(n)$ のエネルギー，τ はラグタイムに相当するサンプル番号．

しかし，$x(n)$ $y(n)$ を同数のサンプル数 N 個で使用すると $y(n)$ に $x(n)$ のレスポンスが完全に含まれず，τ 個づつサンプルが減少する．$\gamma_{xy}(\tau)$ で計算しても，この現象が発生するため，τ が大きくなるにつれ $\tilde{h}(n)$ が本来よりも早く減衰しているかのような歪みが生じる [3]．DFT の巡回性を考えて N サンプルでクロススペクトルを求めることを時間領域で考えると，Fig.1 のようになる．図の左に $\gamma_{xy}(\tau)$ を求める 2 信号を示し，横軸に τ を示している．すなわち，$\gamma_{xy}(\tau)$ は図の縦方向の和となることを示している．図の網掛け部が DFT の巡回性により巡回した成分が計算に含まれる部分を示しており，時間領域で計算した場合に一般に y(n)を 0 として計算する範囲である．$\gamma_{xy}(\tau)$ での計算でもクロススペクトル法でも網掛けの部分か

* Accuracy of the estimated distance spectrum using side-by-side set 2ch mic, by KONDOU Yoshitaka, IDA Takumi, NODA Hiroshi (Nippon Bunri Univ.),ABE Hiroki, IWAKAMI Tomohiro(Chiba Inst. of Tech.),SUEHIRO Kazumi, FUKUSHIMA Manabu (Nippon Bunri Univ.), YANAGAWA Hirofumi (Chiba Inst. of Tech.), KUROIWA Kazuharu (Nippon Bunri Univ.)

ら正しい計算結果が得られないにもかかわらず，Fig.1 の τ によらず同数の推定値が得られることを前提として平均操作を行う．このため，$\tilde{h}(n)$ の n が大きくなるほどに推定値が急速に減衰しているかのような推定値が得られる．

DLR-CS 法では，この問題を解決するために，$y(n)$ からレスポンスがはみ出る $x(n)$ を零として計算することで，Fig.2 に示すように一定範囲の推定値を同数とすることで歪みが生じない推定値を得る．

y(τ)	x(0)	x(0)y(0)	x(0)y(1)	x(0)y(2)	x(0)y(3)	x(0)y(4)	x(0)y(5)	x(0)y(6)	x(0)y(7)
y(τ+1)	x(1)	x(1)y(1)	x(1)y(2)	x(1)y(3)	x(1)y(4)	x(1)y(5)	x(1)y(6)	x(1)y(7)	x(1)y(0)
y(τ+2)	x(2)	x(2)y(2)	x(2)y(3)	x(2)y(4)	x(2)y(5)	x(2)y(6)	x(2)y(7)	x(2)y(0)	x(2)y(1)
y(τ+3)	x(3)	x(3)y(3)	x(3)y(4)	x(3)y(5)	x(3)y(6)	x(3)y(7)	x(3)y(0)	x(3)y(1)	x(3)y(2)
y(τ+4)	x(4)	x(4)y(4)	x(4)y(5)	x(4)y(6)	x(4)y(7)	x(4)y(0)	x(4)y(1)	x(4)y(2)	x(4)y(3)
y(τ+5)	x(5)	x(5)y(5)	x(5)y(6)	x(5)y(7)	x(5)y(0)	x(5)y(1)	x(5)y(2)	x(5)y(3)	x(5)y(4)
y(τ+6)	x(6)	x(6)y(6)	x(6)y(7)	x(6)y(0)	x(6)y(1)	x(6)y(2)	x(6)y(3)	x(6)y(4)	x(6)y(5)
y(τ+7)	x(7)	x(7)y(7)	x(7)y(0)	x(7)y(1)	x(7)y(2)	x(7)y(3)	x(7)y(4)	x(7)y(5)	x(7)y(6)
		γ(0)	γ(1)	γ(2)	γ(3)	γ(4)	γ(5)	γ(6)	γ(7)

Fig.1 The signals in $\gamma_{xy}(\tau)$ with traditional cross spectral technique

y(τ)	x(0)	x(0)y(0)	x(0)y(1)	x(0)y(2)	x(0)y(3)	x(0)y(4)	x(0)y(5)	x(0)y(6)	x(0)y(7)
y(τ+1)	x(1)	x(1)y(1)	x(1)y(2)	x(1)y(3)	x(1)y(4)	x(1)y(5)	x(1)y(6)	x(1)y(7)	x(1)y(0)
y(τ+2)	x(2)	x(2)y(2)	x(2)y(3)	x(2)y(4)	x(2)y(5)	x(2)y(6)	x(2)y(7)	x(2)y(0)	x(2)y(1)
y(τ+3)	x(3)	x(3)y(3)	x(3)y(4)	x(3)y(5)	x(3)y(6)	x(3)y(7)	x(3)y(0)	x(3)y(1)	x(3)y(2)
y(τ+4)	0	0	0	0	0	0	0	0	0
y(τ+5)	0	0	0	0	0	0	0	0	0
y(τ+6)	0	0	0	0	0	0	0	0	0
y(τ+7)	0	0	0	0	0	0	0	0	0
		γ(0)	γ(1)	γ(2)	γ(3)	γ(4)	γ(5)	γ(6)	γ(7)

Fig.2 The signals in $\gamma_{xy}(\tau)$ with DLR-CS

ここで，網掛け部分を推定値から分離することが出来れば，非因果な成分のみを抽出することとなる．物理現象は因果律を満たすことから，この成分は多数回平均操作によって零に収束すべき成分であると考えることが出来る．すなわち，この網掛け部分を分離抽出することで，推定誤差に含まれるノイズレベルを推定することが出来る．

Fig.2 の網掛部をみると，$y(n)$ のサンプルが大きく巡回が遅れれば分離可能であることがわかる．そこで，$x(n)$ $y(n)$ にさらに零を追加する．その結果が Fig.3 である．Fig.3 は Fig.2 の網掛部が分離出来ていることを示している．

Fig.4 に DLR-CS 法により求めた推定値を示す．Fig.4 は推定値の振幅絶対値を縦軸に dB で示し，横軸に離散時間に相当するサンプル番号を示している．Fig.4 は Fig.3 に示した通り，推定値の最後にノイズ成分を示す値が出ていることを示している．これを用いてノイズレベルを推定し，ノイズに埋もれている成分を距離推定で使用しないようにすることが可能であると判断する．

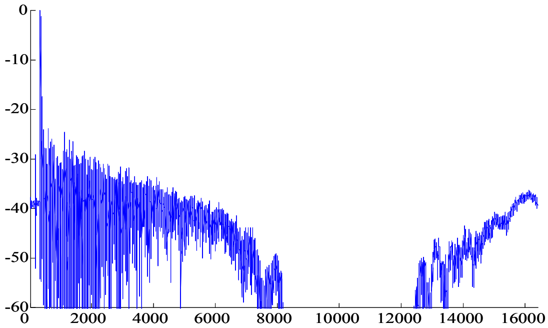


Fig.4 The result with zero added DLR-CS

3. 減衰補正

距離推定において反射波が減衰していると振幅スペクトルに現れる定在波の影響が現象し，距離スペクトルが検出しにくくなる．このため減衰を補正し距離スペクトルを算出することが必要である．$\tilde{h}(n)$ から残響曲線を求めることが出来る．しかし一般に減衰を補正することは補正

y(τ)	x(0)	x(0)y(0)	x(0)y(1)	x(0)y(2)	x(0)y(3)	x(0)y(4)	x(0)y(5)	x(0)y(6)	x(0)y(7)	0	0	0	0	0	0	0	0
y(τ+1)	x(1)	x(1)y(1)	x(1)y(2)	x(1)y(3)	x(1)y(4)	x(1)y(5)	x(1)y(6)	x(1)y(7)	0	0	0	0	0	0	0	0	x(1)y(0)
y(τ+2)	x(2)	x(2)y(2)	x(2)y(3)	x(2)y(4)	x(2)y(5)	x(2)y(6)	x(2)y(7)	0	0	0	0	0	0	0	0	x(2)y(0)	x(2)y(1)
y(τ+3)	x(3)	x(3)y(3)	x(3)y(4)	x(3)y(5)	x(3)y(6)	x(3)y(7)	0	0	0	0	0	0	0	0	x(3)y(0)	x(3)y(1)	x(3)y(2)
y(τ+4)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
y(τ+5)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
y(τ+6)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
y(τ+7)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		γ(0)	γ(1)	γ(2)	γ(3)	γ(4)	γ(5)	γ(6)	γ(7)	γ(8)	γ(9)	γ(10)	γ(11)	γ(12)	γ(13)	γ(14)	γ(15)

Fig.3 The signals in $\gamma_{xy}(\tau)$ with zero added DLR-CS

量が大きくなるとノイズを増幅することになるため，安易には補正を行うことができない．そこで，先に推定したノイズレベル以上の信号のみに対して補正を行うことを考える．

Fig.5 に減衰補正前の距離推定に用いる係数の絶対値を dB で示し，これを補正したものを Fig.6 に示す．Fig.6 より適切に補正できていることが確認できる．

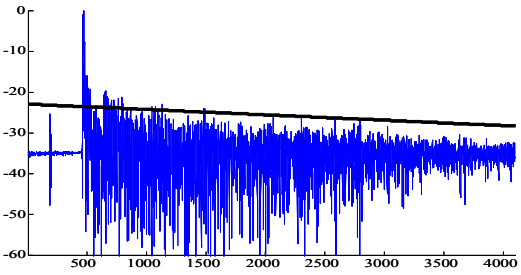


Fig.5 The data before amplitude revised

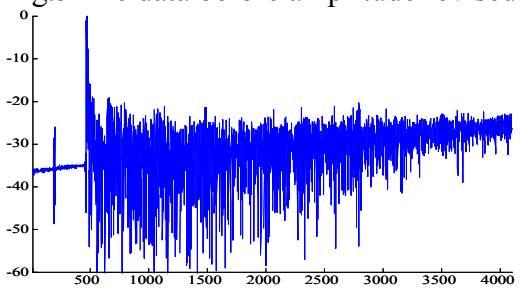


Fig.6 The amplitude revised data

4. 実環境実験による検証

音源信号を白色雑音とし，ラウドスピーカから放射した信号を，Fig.7 に示す条件で設置したステレオマイク（M1，M2）で収録し，それぞれのインパルス応答を推定する．ここではサンプリング周波数 44100Hz，量子化ビット数を 16bit とする．

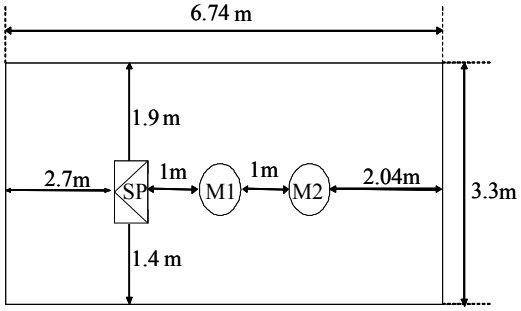


Fig.7 Configuration of Loud Speaker, Stereo Microphone in a laboratory

信号を $x_p(n)$ とし，放射信号を p=s，M1 観測信号を p=1R，1 L，M2 観測信号を p=2R，2L と表記することとする．但し，n は離散時間に相当するサンプル番号．これらの信号から推定したインパルス応答を，$\tilde{h}_{SO}(n)$ とし，S を音源，O を観測信号とする．$\tilde{h}_{SO}(n)$ は S=$x_s(n)$ のとき，O=$x_{1R}(n)$，$x_{1L}(n)$，$x_{2R}(n)$，$x_{2L}(n)$ の 4 種類となり，各マイクで観測した信号を S としたとき各 3 種類となる．総計で 16 種類の $\tilde{h}_{SO}(n)$ が得られる．但し，マイク観測信号を音源信号として使用する場合，ステレオマイクを音源方向に水平に設置しても，ごく僅かに距離が異なれば，非因果な推定値が存在するため，実際には 10 種類の $\tilde{h}_{SO}(n)$ となる．本稿では音源信号と観測信号を使って求めた $\tilde{h}_{s1L}$ と，観測信号 L，R を使って求めた $\tilde{h}_{1L1R}$．$\tilde{h}_{2L2R}$ の合計 3 種類を検討の対象とする．実験には SONY 製ステレオマイク(ECM-MS957)を使用した．

Fig.8 に音源信号と M1 観測信号から求めた距離スペクトルを示す．本来 6m に推定値がでるはずであるが，それ以外の推定値が得られていることがわかる．

Fig.9 にノイズ処理および減衰補正を行って求めた距離スペクトルを示す．Fig.9 は 6m の推定値のみが得られていることを示している．

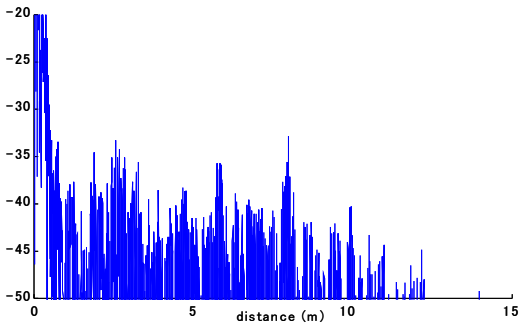


Fig.8 The distance spectrum with $\tilde{h}_{s1L}(n)$

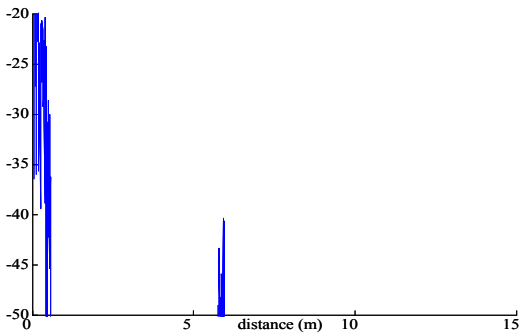


Fig.9 The distance spectrum with processed $\tilde{h}_{s1L}(n)$

これは音源信号を用いたものであり，もともとの $\tilde{h}_{s1L}(n)$ の推定精度が高いと考えられる.そこで,M1 のステレオマイクで観測した LR 信号より推定した結果で比較する．Fig.10 に $\tilde{h}_{1L1R}$ から求めた距離スペクトルを示す．推定距離として 1m および 6m が得られるはずであるが，Fig.8 と同様にそれ以外の値が出ている．

Fig.11 に Fig.9 と同様の処理を施した結果を示す．Fig.11 より，Fig.9 と同様に正しく推定が行えていることが確認できる．

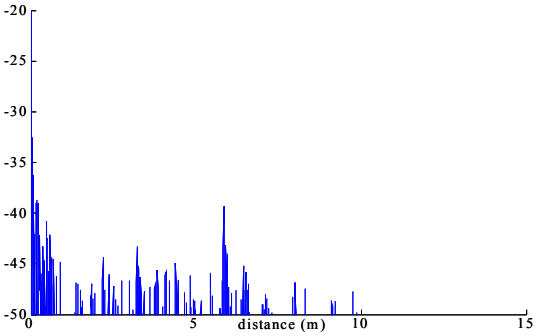


Fig.10 The distance spectrum with $\tilde{h}_{1L1R}$

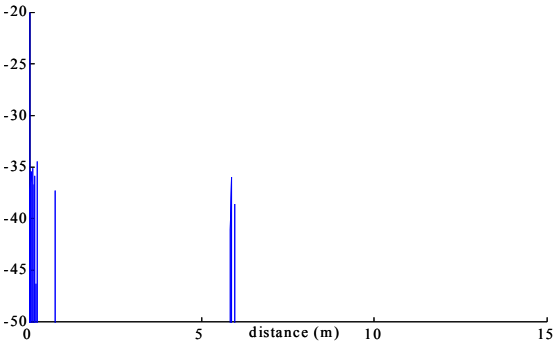


Fig.11 The distance spectrum with processed $\tilde{h}_{1L1R}$

さらに，音源から離れた M2 でも同様の実験を行った．処理前を Fig.12 に示し，処理後を Fig.13 に示す．

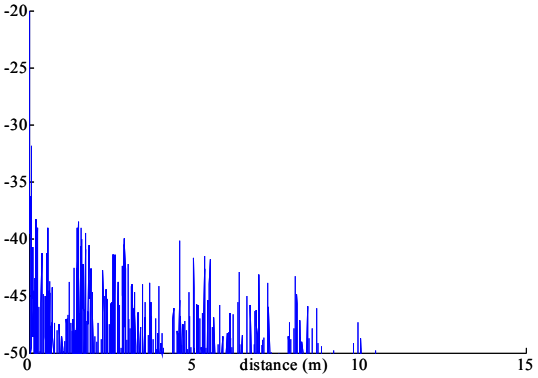


Fig.12 The distance spectrum with $\tilde{h}_{2L2R}$

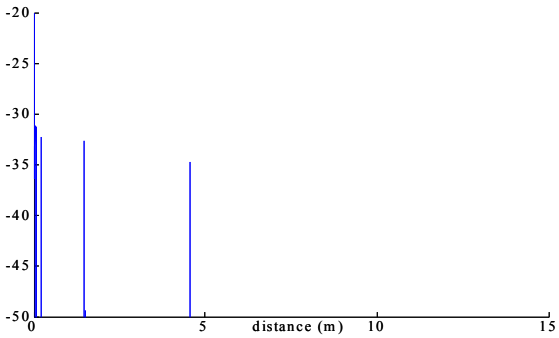


Fig.13 The distance spectrum with processed $\tilde{h}_{2L2R}$

5. おわりに

本稿では，近接して設置している 2ch マイクで観測した信号で距離を推定において，推定値のノイズレベルを推定し，ノイズ対策およびそれにもとづく減衰補正を施すことで推定精度を向上することを試みた．その結果，提案手法により適切な推定距離が得られることがわかった．

［参考文献］

[1] 近藤善隆，伊田匠，阿部宏樹，岩上知広，末廣一美，福島学，柳川博文，黒岩和治，”近接 2ch 計測信号による距離推定に関する一検討”，日本音響学会 2009 年秋季研究発表会講演論文集，1-P-3，2009

[2] 伊田匠，近藤善隆，野田裕，阿部宏樹，岩上知広，末廣一美，福島学，柳川博文，黒岩和治，”近接 2ch マイクによる距離推定における適応範囲の調査”，日本音響学会九州支部「学生のための発表会」，023，2009

[3] Manabu Fukushima, Hiroto Inoue, Ken'itiro Kamura, Hirofumi Yanagawa, Ken'iti Kido, "A method for the determination of noise factor in estimated transfer function - cross spectral technique by use of 1-0 and 1-000 windows -", Proc. of The 18th International Congress on Acoustics, pp.166-169, vol.25, no.2, 2004

[4] Manabu Fukushima, Takatoshi Okuno, Hirofumi Yanagawa, Ken'iti Kido, "Improvement of the Accuracy in Attenuation Constant Estimation using the Cross-Spectral Technique", J. IEICE (E), pp.626-633, Vol.E82-A, No.4, 1999

触覚空間定位に及ぼす聴覚刺激の影響*

野副幸臣　積山薫（熊本大）

1　はじめに

異なる感覚間の相互作用をみるクロスモーダル課題において、実験課題によっては、片方の感覚が、もう片方の感覚の作用を促進または妨害する効果により、感覚統合が起こる。そこには、ある感覚モダリティの優位性もうかがえる。

本研究では、聴覚と触覚のクロスモーダル作用について、触覚による左右弁別課題時に聴覚刺激を与えたとき、聴覚と触覚の刺激の呈示条件が一致または不一致であることで、触覚空間定位にどのような影響があるかを調べ、聴触覚間の関係性について検討していきたいと思う。

2　先行研究

聴覚と触覚の相互作用をみるために触覚刺激として左右の耳たぶに電気刺激を提示した実験がある[1]。判断の対象となる電気刺激は耳たぶの左右にランダムに提示され、同時に、妨害音として、頭部の左右後方45度の位置からスピーカーでノイズが提示された。スピーカーは、頭の中心から 20 cm の位置から提示する条件と 70 cm の位置から提示する条件とがあった。妨害音を無視し、電気刺激の左右弁別課題に集中するように教示したが、結果、電気刺激と妨害音の左右が一致しない場合に、反応時間とエラー率の増加がみられた。妨害音の干渉効果は、近くから提示された条件の方が強かった。

この実験により、聴覚と触覚間の相互作用が身体近傍空間において生じることが示唆された。また、この実験ではスピーカーを左右後方において実験しているが、前方から妨害音を提示した実験の結果と比較したところ、後方から提示した実験の方が成績が悪かったことから、聴覚と触覚間の相互作用は、頭部の後方の空間でより強く生じることが示唆された。

刺激の組み合わせが不一致であることで、触覚空間定位における聴覚による妨害効果がみられたが、本研究では、さらに聴覚のみ、触覚のみでの単一感覚刺激についての試行を行うことにより、聴触覚間相互作用による反応の促進または妨害効果がみられるかどうかについて詳細に検討することを目的とする。

3　実験

3.1 方法

実験参加者

正常な聴力を有する大学生８名（男性４名，女性４名）を対象とした。

刺激

触覚刺激となる空気刺激（0.1 気圧）は、左右の頬の横に設置されたチューブの先端（外径 3 mm）からランダムに提示された。空気刺激発生装置は、圧縮空気のボンベ、ボンベに取り付けた減圧器（B1-1 NR-1 G5 G-B1 N1）、電磁バルブ、減圧器と電磁バルブの開閉を駆動するリレー回路から構成され、リレー回路はデジタル入出力ボードを装着したコンピュータによって制御された。聴覚刺激はホワイトノイズを用い、参加者の頭部の左右後方 45 度の位置に設置されたスピーカー（KENWOOD）から提示された。スピーカー

*Influence of auditory stimulus for localization for tactile stimulus, by NOZOE, Yukiomi and SEKIYAMA, Kaoru (Kumamoto university).

はアンプ（ONKYO A-905TX）に接続されていた。スピーカーと参加者の頭の中心との距離は 20 cm であった。聴覚刺激と触覚刺激はそれぞれ 50 ms ずつ提示された。左右の判断には、参加者の足元に設置したフットスイッチを用いた。

手続き

空気刺激は左右の頬の位置に設置されたチューブからランダムに提示され、頭部後方の左右からランダムに妨害音が提示された。参加者は、空気刺激が提示された方向（左 vs 右）に関して、聴覚刺激の左右は無視し、触覚刺激の左右についてのみ、できるだけ正確に反応するように教示された。実験は、それぞれの参加者において、比較のため触覚刺激のみ提示し、左右を判断してもらう条件を 30 試行、聴覚刺激も加えた上で、触覚刺激の判断をしてもらう条件を 30 試行実施した。

3.2 結果

空気のみの条件(Air only),聴触覚一致条件 (Agreement), 不一致条件(Disagreement)それぞれの平均反応時間およびエラー率を Fig.1 に示した。

反応時間に関して分散分析を行ったところ、空気のみの条件と一致条件との間には有意差はなく、空気のみの条件と不一致条件との間 (F (1,7)=27.176, p <.005) および、一致条件と不一致条件との間(F (1,7)=52.523, p<.001)において有意差がみられた。

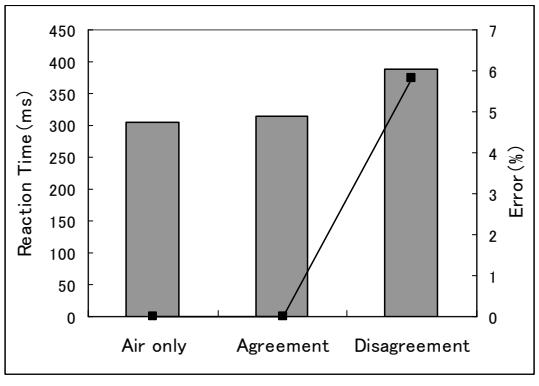


Fig. 1. Reaction Time of the Task.

3.3 考察

聴覚刺激と触覚刺激の提示方向が一致であるときよりも不一致であるときの方が反応時間は長く、エラー率も高かった。また、空気だけの条件と不一致刺激提示条件間でも同様の結果であった。

聴覚と触覚の刺激の提示方向が不一致であるという空間的違いが、触覚刺激の判断に影響を与えたことが伺える。

空気だけの条件と一致条件との間では優位な差はみられないものの、聴覚刺激が与えられた条件の方が反応時間が長くなっていることから、一致する方向からの聴覚刺激による、触覚判断への促進効果はなかったことが示唆される。

以上の結果は、電気刺激を用いた先行研究[1]の結果を支持するものである。

しかし、本実験では、刺激の提示部位が頬という、耳（聴覚刺激の受容器）から近い部位にしか触覚刺激を提示していないため、単に両刺激の提示部位の近さから生まれた妨害効果だった可能性も否定できない。聴触覚間の相互作用をみるためには、やはり触覚刺激を提示する身体部位を頬だけではなく、手や足などの部位にも提示してみて検討する必要があるだろう。また、触覚刺激提示の空間的位置についても、頭部の近傍空間である頬の横に付けた手に提示する条件と、膝の上などに置いた手に提示する条件などを設定して検討してみることで、両感覚の相互作用が、頭部の近傍空間に限って顕著なものであるのかどうかを明らかにすることができるだろう。

参考文献

[1] Kitagawa . N , Zampini . M , Spence . C , Audiotactile interactions in near and far space . *Experimental Brain Research* , 166 , 582-587 , 2005 .

音声基本周波数が音声時間波形狭帯域包絡線間相関による話者識別に与える影響の調査*

吉川浩司，武本良平，末廣一美（日本文理大），今井佐智代，岩上和広（千葉工大），
福島学（日本文理大），柳川博文（千葉工大），黒岩和治（日本文理大）

1　はじめに

情報機器の普及が利便性をもたらした反面，情報の流出等の問題を生じている．特に企業における機密情報の流出は企業の社会的信用にかかわる深刻な問題である．パソコンを含む情報機器の多くはログイン認証とログアウト処理により正当な情報の利用者のみが情報を利用するように設計されているが，文字コードであるIDやパスコードは盗難や紛失により容易になりすましが可能であるという問題がある．このため，例えば社員証や登録済み携帯電話等の常時携行するものを利用した本人認証装置を導入している企業が増加している．しかし，盗難の根本的問題解決とはなっていない．これを解決するためにバイオメトリクスを用いた本人認証が提案され実用化されている．例えば静脈認証は銀行 ATM 等でも導入されている．しかし，それらの多くは専用の認証装置を必要とすることと，認証のための利用者の負担が大きいたことが多くワンタイム認証であることが多い．

著者等は情報流出が問題となると考えられる携帯型情報端末であるノート PC にはマイクが標準装備されていることに着目し，マイクで収録可能なバイオメトリクスとして発話語に着目し，それを用いた本人認証技術の開発を目指している．

本手法は，音声時間波形の狭帯域包絡線に含まれる話者の特徴量を，狭帯域包絡線の帯域間相関係数から作られる狭帯域包絡線間相関係数行列（NECM: Narrowband Envelope Correlation Matrix）により抽出し，NECM が話者毎に相違する点に着目した話者識別である．また，狭帯域包絡線を短区間に分割して短区間毎のNECMを求めて平均を取ることで，発話した語に依存しない識別ができる可能性があることを明らかにしている．

しかし，狭帯域包絡線の相関分析区間長と識別率・頑強性の関係調査を行っている[1]が，被験者数が少ないことと，本手法では狭帯域分割した包絡線を短区間毎に分割するため相関分析区間長によってはフィルタの影響について調査が十分とは言えない．

本稿は話者識別に適した，1)音声の狭帯域包絡線の相関分析区間長，2)音声を狭帯域に分割するための狭帯域分割フィルタ長，それが何に起因しているのかについて検討を行った結果を報告する．

2　狭帯域包絡線間相関係数による話者識別

はじめに著者等が提案している話者識別システムにおける特徴量パラメータの抽出手法について述べる．

音声時間波形 $v(n)$ に対して中心周波数となる 1/4 オクターブバンドの狭帯域分割フィルタ $h_b(n)$ により b 帯域の音声時間

* "An investigation on the filter length in the text independent talker identification under noisy condition", by YOSHIKAWA Koji, TAKEMOTO Ryohei,SUEHIRO Kazumi (Nippon Bunri Univ.), IMAI Sachiyo, IWAKAMI Tomohiro (Chiba Institute of Technology), FUKUSHIMA Manabu (Nippon Bunri Univ.), YANAGAWA Hirofumi (Chiba Institute of Technology), KUROIWA Kazuharu (Nippon Bunri Univ.).

波形$v_b(n)$を得る．

$$v_b(n) \equiv \sum_{p=0}^{M-1} v(n-p)h_b(p) \qquad (1)$$

但し，bは帯域番号，nは離散時刻に相当するサンプル番号,Mは狭帯域分割フィルタ長．

フィルタの長さが短ければ，分割された信号に他の帯域の信号が混入することになり，フィルタの長さが長ければ振幅周波数特性はクロスオーバが小さくなるが，フィルタ特性が時間領域では長く影響することとなる．このため，話者識別に適したフィルタの長さが存在すると考えられる．

本手法は，安定して発話語の狭帯域包絡線を得るために，平均操作を行う．この音声切り出し区間（分析区間長）が長ければ発話した語の種類への依存性が高くなると考えられ，短い場合は音高に関連する音声基本周波数への依存性が高くなると考えられる．

本稿では，この 2 点について調査を行う．

3　相関分析区間長と狭帯域分割フィルタ長が話者識別に与える影響の調査

被験者は日本語を母国語とする 19～25 歳の男性 20 名，女性 11 名とする．発話語は約 1 秒の音声が収録されている音声コーパス(産業総合研究所:ETL-WD-I&II)を参考にして選出する．選出基準は，個人性情報が多く含まれている鼻音と単母音[2]を含む単語 40 単語とする．但し，鼻音のみの語や単母音のみの語は一般的に使われる語ではないため鼻音と単母音以外の音素も含む単語とする．単語一覧を Table 2 に示す．登録用に 20 秒，識別用に 1 秒の音声を用いる．このため，一人あたりの識別回数は 20 回となる．

収録は吸音処理を施した場所で，被験者が通常の速度で発話した音声を，サンプリング周波数 44100 Hz,16 bit 量子化，広帯域精密騒音計（小野測器 LA-5111）で行う．

Table 1 List of words for Registration / Recognition

Registration (20 words / Person)		Recognition (20 words / Person)	
adobeNcha-	niwaume	amyu-zumeNto	mineuchi
puremiasho-	nyu-tauN	anenyo-bo-	niuri
rokuamida	hohoemashi-	fiaNse	nekonadegoe
huroNtia	doroenogu	keana	maewatashi
itogoNnyaku	enuji-	naishuqketsu	nietagiru
inuzamurai	maNetsu	iwatsubame	moetatsu
yakiimo	oreseNgurahu	mitsuzoroi	otazunemono
hainyu-	biNgoomote	norikumiiN	kaotsunagi
maguneshiumu	omowazu	shimauchu-	shiogumori
unuboreru	omiotsuke	uNmakase	hesonoo

Table 1 の発話語を用いて，話者識別を行うために必要な相関分析区間長 L と狭帯域分割フィルタ長 M を考える．

基本周波数（F0）を利用した話者認識システムが提案されている[3]．本手法においても基本周波数が話者識別に重要であるならば，基本周波数に相当する相関分析区間長と狭帯域分割フィルタ長で正解率が高くなることが予想される．そこで，Table 1 の発話語を用いて被験者毎の基本周波数を調べる．基本周波数は音声時間波形のパワースペクトルを，フレーム長 23.2 ms（1024 サンプル）で，1/2 オーバーラップ，ハミング窓で求める．男女 1 名づつの分析例を Fig.1 と Fig.2 に示す．図上段は音声時間波形を縦軸に振幅，横軸に時間で示し，図下段は求めた基本周波数縦軸を縦軸に周波数および周期，横軸にフレーム番号で示す．図の発話語は「煮えたぎる」である．

Fig.1 から男性の基本周波数の平均が 132 Hz，Fig.2 から女性が 253 Hz であることがわかる．収録音声の基本周波数を求めたところ，男性の平均値は約 128 Hz，標準偏差が 21 Hz，女性の平均値は約 250

Hz，標準偏差が 28 Hz となった．これは，一般的な基本周波数の平均値が，男性で 125 Hz，女性で 250 Hz と報告されている[4]ことから妥当な収録音声であると判断する．そこで，相関分析区間長を男性であれば 125 Hz に相当する 8 ms，女性であれば 250 Hz に相当する 4 ms となることが予想されるため，これを基準に調査範囲を決めることとする．

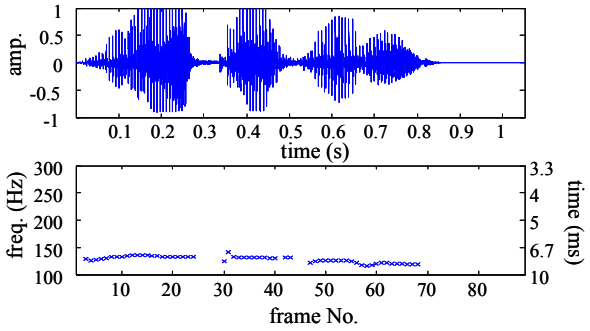


Fig.1 The wave from and F0 of a male voice(/nietagiru/)

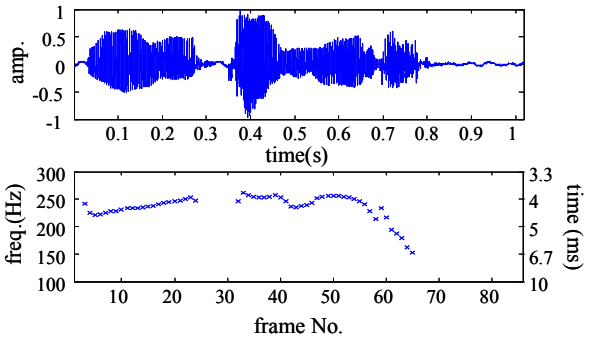


Fig.2 The wave from and F0 of a female voice(/nietagiru/)

一方，日本語音声の音節は概ね 1～3 の音素のまとまりで構成され，音節を部分的に分割した単位であるモーラの時間長は 1 モーラ当たり 90～250 ms 程度である．識別に音韻の継続時間やイントネーション等の韻律情報が必要であればモーラの時間長付近で正解率が最大となることが予想される．

これらのことから，相関分析区間長を

$$L(\mathrm{ms})=1,2,4,8,16,32,64,128,256$$

の 9 種類とする．L は男女共通である．

ここでは音声の狭帯域包絡線を短区間で区切り使用することとしているため，相関分析区間長 L によっては狭帯域分割フィルタ長が結果に影響を与えることが予想される．そこで，相関分析区間長の調査と併せて狭帯域分割フィルタ長の調査も行う．

男性の基本周波数が 125 Hz（8 ms）であることから，男性の狭帯域分割フィルタ長は

$$M\,(\mathrm{ms})=12,23,46,93$$

とする．女性の基本周波数が 250 Hz(4 ms)であることから，女性の狭帯域分割フィルタ長は

$$M\,(\mathrm{ms})=6,12,23,46$$

とする．これらの相関分析区間長 L と狭帯域分割フィルタ長 M を用いて正解率がどのように変化するかを調べる．調べた結果を Fig.3～Fig.6 に示す．Fig.3 と 4 は男性 20 名の平均正解率と標準偏差を示し，Fig.5 と 6 は女性 11 名の平均正解率と標準偏差を示す．Fig.3 と 5 の縦軸は平均正解率を示し，横軸は相関分析区間長，図中の記号は狭帯域分割フィルタ長を示している．Fig.4 と 6 の縦軸は標準偏差を示し，横軸および図中の記号は Fig.3 と同様である．

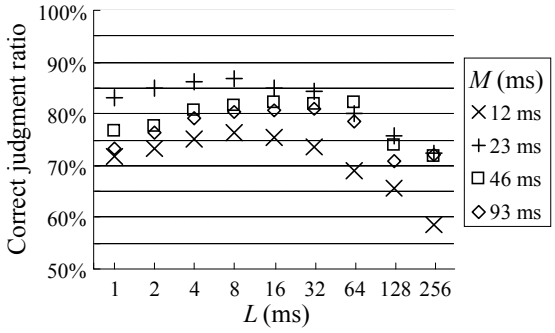


Fig.3 The relation among correct judgment ratio, filter length, and length for analysis (20 male subjects)

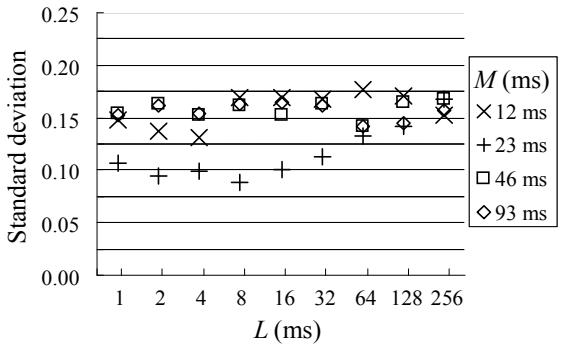


Fig.4 The relation among standard deviation, filter length, and length for analysis (20 male subjects)

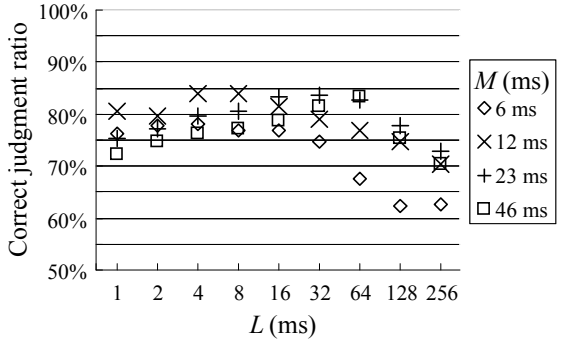


Fig.5 The relation among correct judgment ratio, filter length, and length for analysis (11 female subjects)

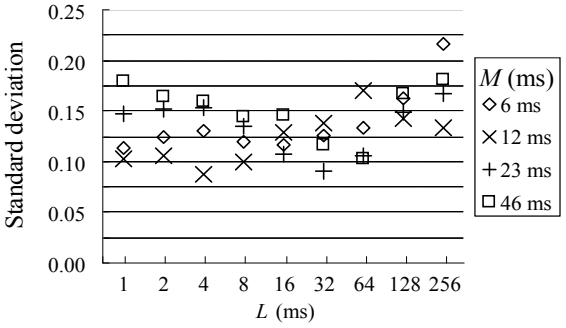


Fig.6 The relation among standard deviation, filter length, and length for analysis (11 female subjects)

Fig.3 と 5 から，男性は L=8 ms，M=23 ms で平均正解率が高くかつ標準偏差が低くなることがわかる．Fig.4 と 6 から女性は L=4 ms，M=12 ms で平均正解率が高くかつ標準偏差が低くなることがわかる．両者の最適値を比較すると，男女の基本周波数に相当する相関分析区間長と狭帯域分割フィルタ長であることがわかった．また，狭帯域分割フィルタ長が長くなると，長い相関分析区間長が必要になることがわかった．

4. おわりに

情報通信技術を利活用し安全で快適な生活をサポートするシステムを実現するために，利用者の生体情報に基づく認証が必要である．著者等は日常生活で良く使われる音声に着目し，音声時間波形を 1/4 オクターブバンドの狭帯域に分割した信号の包絡線から狭帯域包絡線間相関係数を求め，それを特徴パラメータとする話者識別システムの提案を行っている．

本稿では，話者識別に適した相関分析区間長と狭帯域分割フィルタ長を調査し，それが何に起因しているのかについて検討を行った．

その結果，男女の基本周波数に相当する相関分析区間長（男性 8 ms と女性 4 ms）と狭帯域分割フィルタ長（男性 23 ms と女性 12 ms）が話者識別に適していることがわかった．

文　献

[1]長尾優次，”狭帯域包絡線相関を用いた話者識別における包絡線算出手法・分析区間長と識別率・頑強性の関係に関する一検討”，日本音響学会秋季研究発表会，pp.767-768，2004

[2]竹内章司，粕谷英樹，城戸健一，”鼻音のスペクトルに及ぼす鼻副鼻腔の影響”，日本音響学会誌，pp.163-172，33 巻，4 号，1977

[3]古井貞熙，”声の個人性の話”，日本音響学会誌，Vol.51，No.11，1995

[4]斉藤，加藤，寺西，”音声の基本周波数の特性について”，音響学会誌，14，2，pp.111-116，1958

有色性雑音が音声時間波形狭帯域包絡線間相関による話者識別に与える影響の調査*

武本良平，吉川浩司，末廣一美（日本文理大），今井佐智代，岩上和広（千葉工大），
福島学（日本文理大），柳川博文（千葉工大），黒岩和治（日本文理大）

1　はじめに

スマートフォン等の高性能携帯型情報端末の登場は，付加価値の高い情報の携帯を可能にしている．これは，携帯電話装置盗難の危険があるだけでなく，蓄積されるまたは携帯端末を窓口としてアクセスできる情報そのものの盗難の危険が高くなったといえる[1]．このような情報そのものの盗難に対応するには適切なセキュリティが必要不可欠である．しかし，実際には誤操作防止のためのボタンロック機能や 4 桁程度の暗証番号でのセキュリティ程度が施されており，指紋認証装置を搭載している機種でも一度認証してしまうと，ユーザによる明示的なロック操作を行わないと保護されないのが現状である．このため，携帯電話等の携帯型情報端末の利便性を損なうことなく，情報そのものの安全を確保するためのセキュリティ技術が必要不可欠となっている．高いセキュリティを実現する方法として，利用者のバイオメトリクスを用いた本人認証技術が提案されている[2]．

本研究では，携帯電話型情報端末が持っているデバイスで利用可能でかつ，利用者に過度な負荷をかけず適時認証可能なバイオメトリクス認証として，発話語によらない話者識別システムの開発をおこなっている．このシステムは携帯電話の基本機能であるマイクによる音声収録を用いる．特に短時間発声の音声で識別することで，従来のパスコードの入力または指紋認証よりも簡便な手法としての確立を目指している．

著者等はこれまでに，狭帯域包絡線の帯域間相関係数から作られる狭帯域包絡線間相関係数行列（NECM: Narrowband Envelope Correlation Matrix）を用いた話者識別システムを考案している[3]．

NECM では，約 1 秒の発話音声で 9 割程度の識別率が得られ，追認証を逐次実施することで10割の正解率が得られること[4]，通常発声音声以外の例えば裏声でも認証可能なことを報告してきた．また，大学祭等の雑踏での公開実験でも認証可能であることを検証してきている．このため，外来雑音に対して頑強な手法であることは実験的に検証されている．しかし，日常生活でどの程度雑音に対して頑強であるかについて十分とはいいがたい．

そこで本稿では，利用場面を想定し，また日常生活で頻繁に発生し，かつ提案している識別手法のアルゴリズム的に識別率を低下させると考えられる要素を含む雑音を選び，それらの雑音強度（SNR）と識別率の関係を調査した結果を報告する．

2　狭帯域包絡線間相関係数による話者識別

著者等が提案している話者識別システムにおける特徴量パラメータの抽出手法について述べる．

---

*” An investigation on the external incomming noise factor in the text independent talker identification”, by TAKEMOTO Ryohei, YOSHIKAWA, Koji, SUEHIRO, Kazumi(Nippon Bunri Univ), IMAI　Sachiyo, IWAKAMI Tomohiro(Chiba Institute of Technology), FUKUSHIMA Manabu(Nippon Bunri Univ) and YANAGAWA Hirofumi(Chiba Institute of Technology), KUROIWA Kazuharu(Nippon Bunri Univ)

音声時間波形 $v(n)$ に対して1/4オクターブバンドの狭帯域分割フィルタ $h_b(n)$ により b 帯域の音声時間波形 $v_b(n)$ を得る．

$$v_b(n) \equiv \sum_{p=0}^{M-1} v(n-p)h_b(p) \qquad (1)$$

但し，b は帯域番号，n は離散時刻に相当するサンプル番号，M は狭帯域分割フィルタ長．

狭帯域分割フィルタは，FIR フィルタとし，狭帯域分割した音声時間波形 $v_b(n)$ からスペクトル $V_b(k)$ を

$$V_b(k) \equiv \sum_{n=0}^{N-1} v_b(n) e^{-j2\pi\frac{kn}{N}} \qquad (2)$$

但し，k は離散周波数に相当するサンプル番号，N は信号のサンプル数，
で求める．ここから $v_b(n)$ の片側スペクトルとなる解析的信号表現 $\tilde{v}_b(n)$

$$\begin{aligned}\tilde{v}_b(n) &\equiv V_b(0) + \sum_{k=1}^{N/2} V_b(k) e^{j2\pi\frac{kn}{N}} \\ &= |v_b(n)| e^{j\theta}\end{aligned} \qquad (3)$$

但し，$|X|$ は X の絶対値，θ は位相角，を求める．式(3)は $\tilde{v}_b(n)$ が包絡線と瞬時位相特性の積で表されていることと，$\tilde{v}_b(n)$ の絶対値から包絡線情報が得られることを示している．そこで狭帯域分割した音声時間波形の包絡線 $e_b(n)$ を

$$e_b(n) \equiv |\tilde{v}_b(n)| \qquad (4)$$

より求める．この包絡線をヒルベルト包絡線と呼ぶ．これをさらに振幅を dB 変換した $d_b(n)$ を

$$d_b(n) \equiv 20\log_{10}(e_b(n)/e_{\max}) \qquad (5)$$

但し，$e_{\max}$ は $e_b(n)$ の最大値，
で求める．

本手法では狭帯域包絡線の類似性を相関係数により求める．相関係数が，狭帯域包絡線の小さい振幅範囲で決まるのを防ぐため，$d_b(n)$ を－30dB で打ち切ることとする．$d_b(n)$ を相関分析区間長毎に分割し，短区間毎の $d_{b_1}(n)$ と $d_{b_2}(n)$ より帯域番号 b_1 と b_2 の狭帯域包絡線間相関係数 $\gamma_{b_1b_2}$ を次式により得る．

$$\gamma_{b_1b_2} \equiv \frac{1}{\sigma_{d_{b_1}}\sigma_{d_{b_2}}} \sum_{n=0}^{N-1} \left\{ d_{b_1}(n)d_{b_2}(n) - \overline{d_{b_1}}\,\overline{d_{b_2}} \right\} \qquad (6)$$

但し，σ_x は x の分散，$\overline{X}$ は X の平均，N は d_x の長さに相当するサンプル番号．

ここでは狭帯域分割数が39であるため，短区間毎の狭帯域包絡線間相関係数 $\gamma_{b_1b_2}$ は 39 行 39 列の行列となる．ここではこれを狭帯域包絡線間相関係数行列Γと呼ぶ事とする．

$$\Gamma \equiv \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \Lambda & \gamma_{1,39} \\ \gamma_{2,1} & \gamma_{2,2} & \Lambda & \mathrm{M} \\ \mathrm{M} & \mathrm{M} & \Lambda & \mathrm{M} \\ \gamma_{39,1} & \gamma_{39,2} & \Lambda & \gamma_{39,39} \end{bmatrix} \qquad (7)$$

本システムでは約 1000 ms の音声を用いるため相関分析区間長 L と分析時の平均回数 a の関係を

$$1000ms = a \cdot L \qquad (8)$$

とすることで，L によらず使用するデータ量が一定になるようにする．L =10 ms と設定したときは，1 単語につき 100 個のΓが生成されるため，100 単語用いたときと同様の効果が期待される．この平均操作により発話語によらない識別となることが期待される．

本システムでは，収録した音声情報から，平均操作を行った狭帯域包絡線間相関係数行列（NECM: Narrowband Envelope Correlation Matrix）Γを求め蓄積する．これに対して識別対象 X の音声情報から求めた NECM を求め，それと蓄積された NECM との相関係数 γ_{XA} を求める．本システムでは γ_{XA} の最大値 $\gamma_{\max}$ を識別候補とする．識別候補者 A が本人と一致した場合に正解とし，正解した回数から正解率 C_A を求める．本稿では，音声に環境雑音を加えた場合に，正解率 C_A がどのように変化するかを調べる．

3　環境雑音と正解率の関係の調査

実利用を考えると音声収録時に環境雑

音が混入することが考えられる．そこで，日常生活雑音下でも話者識別可能かどうかを調べる．

約1500単語収録されている音声コーパス（産業総合研究所：ETL-WD-I&II）を参考にして選出する．選出基準は，個人性情報が多く含まれている鼻音と単母音を含む単語40単語(約1秒／単語)とする．但し，鼻音のみの語や単母音のみの語は一般的に使われる語ではないため鼻音と単母音以外の音素も含む単語とする．単語一覧をTable.2に示す.登録用に20秒，識別用に1秒の音声を用いる.このため，一人あたりの識別回数は20回となる．収録は吸音処理を施した場所で，被験者が通常の速度で発話した音声を，サンプリング周波数44100 Hz，16 bit量子化，広帯域精密騒音計（LA-5111）で行う．

Table 2 List of words for Registration / Recognition

Registration (20 words / Person)		Recognition (20 words / Person)	
adobeNcha-	niwaume	amyu-zumeNto	mineuchi
puremiasho-	nyu-tauN	anenyo-bo-	niuri
rokuamida	hohoemashi-	fiaNse	nekonadegoe
huroNtia	doroenogu	keana	maewatashi
itogoNnyaku	enuji-	naishuqketsu	nietagiru
inuzamurai	maNetsu	iwatsubame	moetatsu
yakiimo	oreseNgurahu	mitsuzoroi	otazunemono
hainyu-	biNgoomote	norikumiiN	kaotsunagi
maguneshiumu	omowazu	shimauchu-	shiogumori
unuboreru	omiotsuke	uNmakase	hesonoo

環境雑音として電子協騒音データベースから，1）展示会場（ブース内），２）人ごみ，３）走行自動車内，の3種類を使用する．また，実環境で収録した，４）走行自動車内，５)走行自動車内(窓開)，６）繁華街，の3種類を使用する．雑音は約2分～5分程度あるため,帯域に偏りのある1秒程度を使用する．また，音声や楽曲等の周期的な雑音は包絡線情報に影響を及ぼすため，正解率が低下することが予想される．このため，3）展示会場から鐘の音が含まれる雑音，２）人ごみと６）繁華街から発話語が判別できる男性1名の音声と女性3名の音声，が含まれる雑音も使用する．2)，3)，8）の周期性を確かめるために，Fig.1 で 2),3),8)毎の音声時間波形，基本周波数，ヒストグラムを示す．それぞれ，上段の横軸に時間，縦軸に振幅，中段の横軸にフレームナンバー，縦軸に周波数，下段の横軸に周波数,縦軸に出現回数を示す.また，ノイズを類似度順とするため，雑音のパワースペクトルを，フレーム長 23.2 ms（1024 サンプル）で，1/2 オーバーラップ，ハミング窓で求める．結果を Fig.2 に示す.Fig.2 は縦軸に振幅を dB で示し，横軸は周波数を対数で示している．ここでは周波数の偏りが識別率に関連すると予想し,Fig.2 に示したパワースペクトルの類似度から雑音に順位をつける．類似度は相関係数で求める．雑音の種類と特徴を Table.3 に示す．

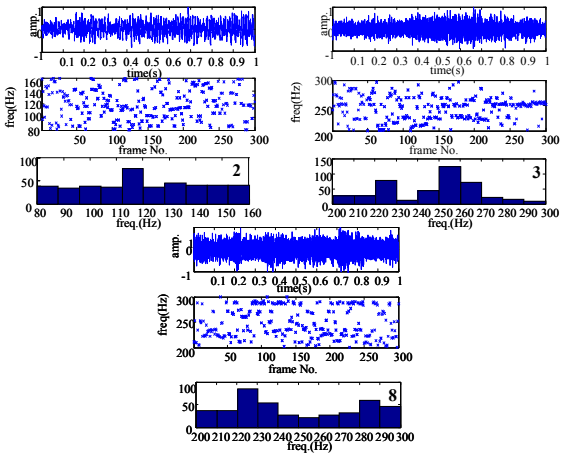


Fig.1 The noise wave form, F0 candidates, histgram of F0 (Upeer: wave form, Middle: F0 candidates, Lower: histgram of F0)

これらの雑音を音声に加え，擬似的に雑音状況下を作り出して調査を行う．話

者識別システムの利用方法を考えると話者情報登録時は収録環境を選ぶことができ，識別時は選ぶことができない場合がある．ここでは識別語のみに雑音を加えることとする．なお，音声に加える雑音の SNR は 20 dB，15 dB，10 dB，5 dB，0 dB の 5 種類とし，1 話者毎 1 単語毎に計算を行った．このため 1 話者につき合計 900 回（識別語 20 単語・雑音 9 種類・SNR 5 種類）の識別を行う．調べた結果を Fig.3 に平均正解率，Fig.4 に標準偏差，を示す．Fig.3，Fig.4 は縦軸に SNR を示し，横軸に Table.3 に対応する雑音の種類を示し，図中の色が白に近いほど高い値であることを示している．

Fig.3 より，(1)の走行車内騒音と（２）人ごみ（男性 1 名の音声）において SNR10dB までは正解率が 70%程度以上であることがわかる．（２）人ごみ（男性の音声）は発話語を聴き取ることが出来る程，明瞭に音声が混入していることから，雑音の周期性が必ずしも識別精度に影響を与えるとは言えないことがわかる．

４　おわりに

本稿では，著者等の提案手法が日常生活雑音下でどの程度の識別率を得ることができるかを調査した．

その結果，(1) 走行車内騒音，（２）人ごみ（男性の音声）において SNR 10 dB までは正解率が 70 %程度以上であることがわかった．（２）人ごみ（男性の音声）は発話語を聴き取ることが出来る程に明瞭に音声が混入していることから，雑音の周期性が必ずしも識別精度に影響を与えるとは言えないことがわかった．

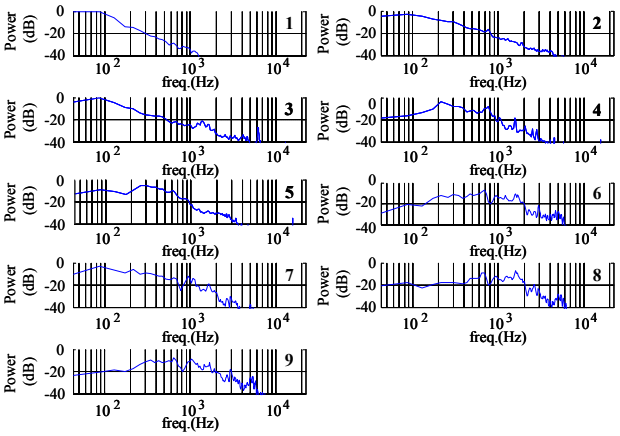


Fig.2 Power Spec of ambient noises

Table 3 The list of ambient noises

No.	種類	特徴
1	走行自動車内(窓閉)	窓を閉めて一般道を走行中(2000cc) (帯域に偏りあり)
2	人ごみ(男性1名の声)	男性1名の音声あり(語の判別可能) (周期的な雑音)
3	展示会場(ブース内)	BGM(鐘の音)あり (周期的な雑音)
4	展示会場(ブース内)	アナウンス(語の判別不可)や複数の物音 (帯域に偏りあり)
5	人ごみ	複数の音声あり(語の判別不可) (帯域に偏りあり)
6	繁華街	複数の音声(語の判別不可)や物音 (帯域に偏りあり)
7	走行自動車内(窓閉)	窓を閉めて安定した速度で走行中 (帯域に偏りあり)
8	繁華街(女性3名の声)	女性3名の音声あり(語の判別可能) (周期的な雑音)
9	走行自動車内(窓開)	窓を開いて走行中 (帯域に偏りあり)

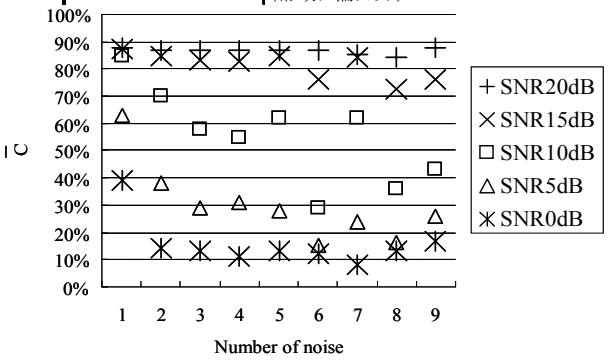


Fig.3 The correct judgment ratio for ambient noise (20 male subjects)

文　献

[1] 平成 19 年度版情報通信白書，総務省，2007 u-Japan ベストプラクティス事例集

[2] 鷲見和彦，” バイオメトリクスセキュリティ概論”，電子情報通信学会誌，Vo.89，No.1，pp.27-30，2006

[3] 長尾優次，”狭帯域包絡線相関を用いた話者識別における包絡線算出手法・分析区間長と識別率・頑強性の関係に関する一検討”，日本音響学会秋季研究発表会，pp.767-768，2004

[4] 末廣一美，吉川浩司，武本良平，近藤善隆，今井佐智代，岩上知広，福島学，柳川博文，黒岩和治，“実環境での発話語に非依存な話者識別に適したフィルタ長の検討” 2009 秋 1-P-2

ハノイでの航空機騒音に関する社会調査*

椎名知代 矢野隆 Nguyen Thu Lan（熊本大）西村強（崇城大）

1 はじめに

前世紀には航空機，自動車，鉄道などの大量輸送手段が発達し，人や物資の移動が便利になった反面，環境騒音，特に交通騒音は多くの国で社会問題化し，更に交通網の拡大によりグローバルな問題となっている． このことを背景に，ヨーロッパや北米の先進諸国では騒音についての社会調査や心理音響実験が数多く行われ，その成果は騒音政策に反映されてきた．[1] 一方で，これまで日本以外のアジア諸国では騒音に関する社会調査そのものがあまり行われておらず，騒音政策は欧米の調査結果をもとに議論されてきた． しかし，騒音の人々への影響を交通手段ごとに欧米と日本の間で比較すると，明らかな差異を確認している報告もある．

従って，欧米での成果をアジアに直接適用することは困難であり，アジアの騒音政策はそれぞれの地域で実施された社会調査に基づいて策定されるべきだと考えられる．特に著しい経済発展に伴い，深刻な騒音問題に直面しているベトナムをはじめとする東南アジアでの騒音政策を立案し，対策を講じるための調査データが早急に求められている．

当研究室では，2005 年と 2007 年にベトナムの 2 大都市のハノイ（人口約 400 万人）とホーチミン（人口約 800 万人）で道路交通騒音に関する社会調査を行い，ベトナムでの道路交通騒音の影響を調査した[2]．また，2008 年に行ったホーチミンでの航空機騒音の社会調査[3]と今回の調査結果により，空港周辺に住む人々への航空機騒音単独の影響と，航空機騒音と道路交通騒音との複合騒音の影響を把握する．本研究では 2009 年 8～9 月に行ったハノイ近郊のノイバイ空港周辺での航空機騒音に関する社会調査結果を報告するものである．

2 調査手法

2.1 社会調査

2009 年 8 月に航空機騒音及び航空機と道路交通の複合騒音についての社会調査をノイバイ空港周辺の 9 地区で行った．

このうち 7 地区は離着陸航路に沿った地域であり，航空機騒音に加え道路交通騒音の暴露の見られる地区，2 地区は空港の南に位置し，それぞれ幹線道路沿いの比較的交通量の多い地区と少ない地区を選定した．選定地区を Fig.1 に示す．

社会調査には，2 種類のアンケート調査票を用いた．選定した地区内の幹線道路沿いの居住者は，航空機騒音に加え道路交通騒音よる影響も考えられるため，複合騒音のアンケート調査票を，幹線道路から離れたエリアの居住者には航空機騒音のアンケート調査票をそれぞれ配布した．

各地区，各質問表毎に 100 世帯を対象に調査を行い，合計 1650 世帯のうち 1397 世帯（回答率 84.7%）となった．地区ごとの回収数，回収率を Table.1 に，各地区の集計結果を Table.2 に示す．

*Social survey on community response to aircraft noise in Hanoi, by SHIINA Tomoyo,YANO Takashi, and NGUYEN Thu Lan (Kumamoto university).

Fig. 1. Map of survey sites around the Noi Bai Airport

Table 1. Response number and rate on survey sites

aircraft noise survey

Site No.	1	2	3	4	5	6	7	8	9	total
Response number	96	89	100	99	76	99	88	90	87	824
Response rate(%)	96	89	100	99	76	99	88	90	87	91.6

combined noise survey

Site No.	1	2	3	4	5	6	7	8	9	total
Response number	99	70	53	27	67	–	81	77	99	573
Response rate(%)	99	70	53	54	67	–	81	77	99	76.4

Table 2. summary of some demographic factor

	items	aircraft noise survey	(%)	combined noise survey	(%)
Gender	Male	370	46.3 %	280	50.5 %
	Female	430	53.8 %	275	49.5 %
Age	20s	197	24.2 %	145	25.6 %
	30s	188	23.1 %	121	21.4 %
	40s	192	23.6 %	135	23.9 %
	50s	141	17.3 %	110	19.4 %
	60s	62	7.6 %	45	8.0 %
	≧70	33	4.1 %	10	1.8 %
Length of residence	0-5years	96	12.0 %	100	12.0 %
	5-10years	107	13.3 %	96	13.3 %
	10-20years	232	28.9 %	213	28.9 %
	20-50years	291	36.3 %	132	36.3 %
	≧50yeaes	76	9.5 %	20	9.5 %
Occupation	Employed	505	62.1 %	310	54.9 %
	Students, housewives, retired, and unemployed	308	37.9 %	255	45.1 %
response rate		817	90.8 %	572	84.1 %

アンケート調査表の主な調査項目は，回答者の社会的データ（年齢，性別，就業状況等）住居の居住形態，近隣の環境，騒音のうるささ反応，日常の行動のしやすさ，敏感さ，交通機関の使用頻度や印象等である．特に騒音のうるささについては，5 段階言語評価（全く…ない，それほど…ない，多少，だいぶ，非常に）と 11 段階数値評価（0[全く…ない]～10[非常に]）を用いて評価した．

2.2 騒音測定

2009年9月に航空機騒音測定を社会調査と同様の9地区，道路交通騒音測定及び道路交通量調査を第3，6地区を除く7地区で行った．

航空機騒音調査は9地区のそれぞれ地区の中で最も高い住居を選択し，その屋根又は屋上で騒音を1週間にわたり騒音計（RION NL-22）を用いて測定した．1週間測定した騒音データからグラフを作成し，騒音波形と離発着スケジュールから航空機騒音を特定し，航空機騒音の各L_Eを算出し，$L_{Aeq,24h}$量の長時間の騒音暴露指標を求める．

道路交通騒音は7地区の代表的な幹線道路沿いで24時間，騒音計（RION NL-22）を用いて測定した．また映像の撮影による道路交通量を道路交通騒音測定と平行して行った．

3　社会調査結果

3.1 航空機騒音のうるささ評価

航空機騒音のうるささ反応の5段階言語評価結果をFig.2に，11段階数値評価結果をFig.3に示す．

5段階言語評価では，航空機騒音に対しての評価を’非常にうるさい’，あるいは’だいぶうるさい’と回答した割合が航空機騒音調査で38.5%，複合騒音調査で21.2% 17%程度航空機騒音調査が複合騒音調査に比べ多い結果となった．同様に，11段階数値評価でも，-%Highly annoyedを示す8，9，10のいずれかに回答した割合が，航空機騒音調査で23.1%，複合騒音調査で17.9% と約5% 航空機騒音調査の結果が上回った．

この結果の要因としては暗騒音（道路交通騒音）レベルの影響が考えられる．複合騒音の影響については各国でいくつかの研究がなされており，高い暗騒音レベルの地域では，主となる騒音（この場合航空機騒音）に対するうるささ反応が低くなるという結果が出されており，今回の社会調査結果も，騒音レベルと比較していく必要がある．

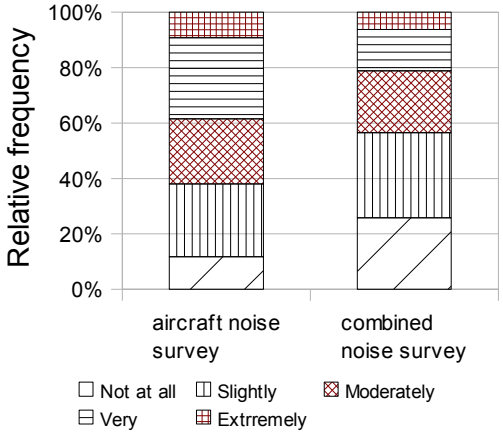


Fig.2.　Annoyance distribution with the 5-point verbal scale

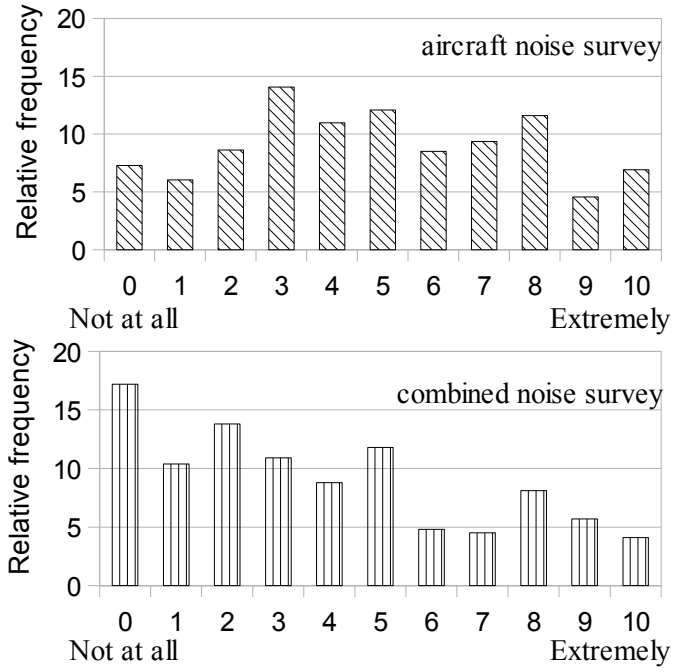


Fig.3. Annoyance distribution on the 11-point numeric scale

3.2 航空機騒音の人間の活動への影響

航空機騒音の人間の活動への影響結果をFig.4に示す.

航空機騒音調査では，航空機騒音がテレビ/ラジオの視聴時に’非常に邪魔’，’だいぶ邪魔’だと回答した割合が36.1% ，また電話時も28.5% と他の項目に比べ高く，航空機騒音が人が聴取する活動の妨げになっていることが分かる.

複合騒音調査では，うるささ評価の結果と同様に全体に航空機騒音調査の結果よりも’非常に邪魔’，’だいぶ邪魔’だと回答した割合が聴取活動時や休憩，睡眠時に約5～10%低い結果となった.

これに対し，読書などの集中時，窓を開ける頻度，振動などは航空機騒音，複合騒音によらず1～3% 程度の違いでほぼ同じ割合となった.

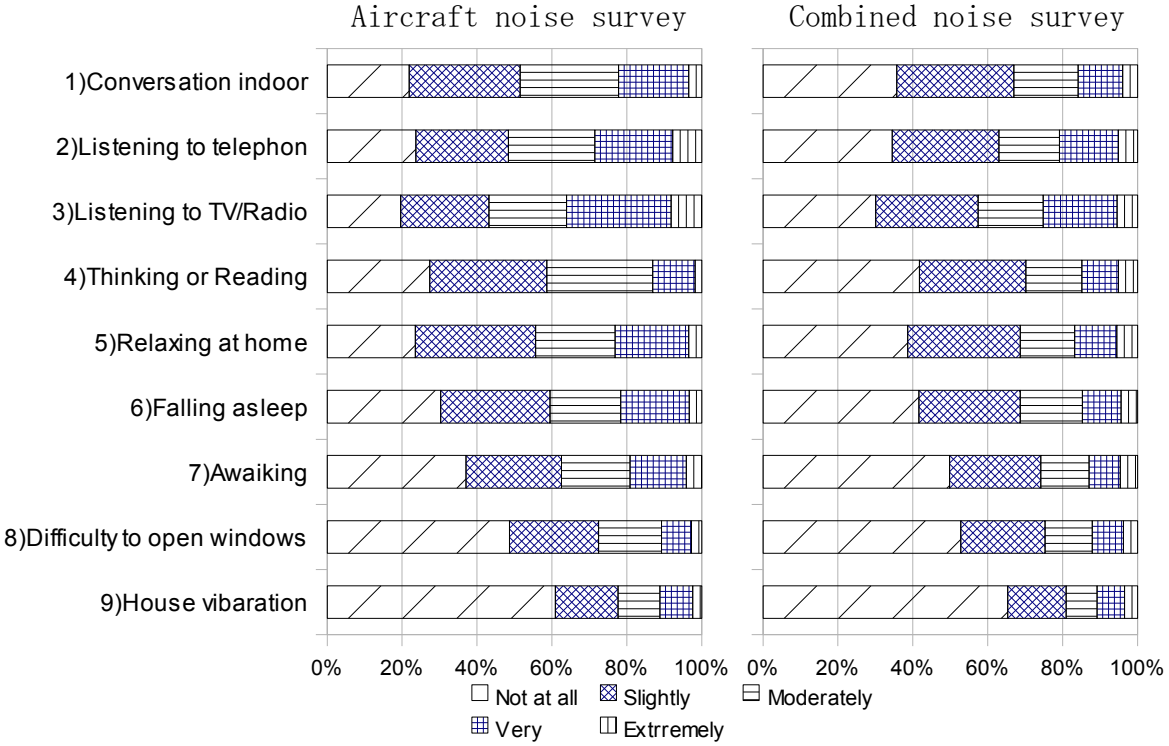


Fig.4 Disturbance level by aircraft noise in certain cases

4 まとめ

航空機騒音と複合騒音の社会調査結果の比較から，航空機騒音のうるささ評価，人間の活動への影響に音源により違いが生じる結果となった．発表では今回の社会調査の結果とあわせて物理特性である騒音レベルと比較し検討した結果を報告する.

参考文献

[1]H.M.E.Miedema, Exposure-response relationships for transportation noise, J.Acoust. Soc. Am. 104.3432-3445(1998)

[2]Hai Yen Thi Phan, T.Yano, et al., Row house and apartment residents’ reaction to road traffic noise in Hanoi, Proceedings of Inter-Noise 2007, Istanbul (2007).

[3]Nguyen Thu Lan, T.Yano,et al.,Social survey on community response to aircraft noise in Ho Chi Minh City, Inter noise 2009, Ottawa(2009)

[4]Evy Öhrström,et al.Annoyance due to single and combined sound exposure from railway and road traffic,J. Acoust. Soc. Am. 122 5,2642-2652,(2007)

現代作家が描く音環境のイメージの印象評価*

井坂幸大　岩宮眞一郎（九州大）

１．はじめに

近年、サウンドスケープの「音と人間とその聞かれた状況(コンテクスト)の相互作用」を重視するといった概念が広まったことで[1]、音の物理量や人間の心理的反応だけでなく、その音を聞く際の状況まで検討する研究が多く見られるようになってきた。

多くの視聴覚相互作用の研究により、音の印象は視覚情報によって大きく変化することが明らかになっている[2]、[3]。また、視覚情報だけでなく、聞く際の気温、天気の違いや[4]、[5]、現場と実験室との違い[6]、さらには、疲労度の違い[4]によっても音環境の印象が異なることも報告されている。従って、音の印象を正確に捉えるためには、音を聞く際の状況まで詳細に検討する必要があると考えられる。

そこで本研究では、音を聞く際の時刻、季節、心理状態などコンテクストが同時に含まれる「文章」を用いて、そこからイメージされる音環境の評価を行った。

２．実験Ⅰ

２．１　評価対象の選定

本研究に用いた評価対象は、2007 年以前の芥川賞受賞作 10 作品から選定した。作品を表 2-1 に示す。その中から、音環境、景観や空間、登場人物の心理描写等、状況が詳細に描かれている 33 文とした。

２．２　実験手法

音環境の印象評価は、評定尺度法を用いて行った。評価は「音環境」「景観・空間」「登場人物の心理状態」の 3 つについてそれぞれ 5 対を、さらに、音環境の総合評価（好き―嫌い）も含めた 16 対の両極尺度を用いて 7 段階評価で行った。また、本実験の被験者は 21～26 歳の九州大学の学生 15 名(男 11 名、女 4 名)である。

さらに、文章を「音」「時刻」「場所」「状況」というアイテムに分け、実験で得られた音環境の総合評価とアイテムと

表 2-1　本実験で用いた作品

受賞年	作者名	小説名
2003年下	金原ひとみ	蛇にピアス
2003年下	綿矢りさ	蹴りたい背中
2004年上	モブ・ノリオ	介護入門
2004年下	阿部和重	グランド・フィナーレ
2005年上	中村文則	土の中の子供
2005年下	絲山秋子	沖で待つ
2006年上	伊藤たかみ	八月の路上に捨てる
2006年下	青山七恵	ひとり日和
2007年上	諏訪徹治	アサッテの人
2007年下	川上未映子	乳と卵

表 2-2　音の総合評価と各評定尺度の相関関係

	両極尺度	相関係数
音環境	快い ― 不快な	** 0.936
	静かな ― うるさい	** 0.504
	迫力のある ― 物足りない	−0.325
	にぎやかな ― さびしい	−0.312
	自然な ― 人工的な	** 0.467
空間・景観	美しい ― 醜い	** 0.895
	開放的な ― 閉鎖的な	* 0.416
	明るい ― 暗い	0.184
	にぎやかな ― さびしい	−0.214
	自然な ― 人工的な	* 0.352
心理	安心できる ― 不安な	** 0.676
	落ち着いた ― 緊張する	** 0.650
	くつろいだ ― イライラする	** 0.770
	冷静な ― 興奮した	0.243
	充実した ― 空虚な	** 0.674

** : $p<.01$, * : $p<.05$

*Subjective ratings of image of soundscape represented by modern novelists, by ISAKA, Yukihiro and IWAMIYA, Shin-ichiro (Kyushu university).

の関係を調べた。

２．３　実験結果

本研究で用いた評定尺度は、優勢を 7、劣勢を 1 とし、被験者 15 名の平均値を用いて検討を行った。音環境の総合評価と各評定尺度の相関関係を表 2-2 に示す。また、音環境の総合評価と「くつろいだ―イライラする」の相関図を図 2-1 に示す。

表 2-2 から、音環境の総合評価は、空間・景観の評価の「美しい―醜い」や心理描写の評価の「安心できる―不安な」や「くつろいだ―イライラする」と強い相関があることがわかる。従って、音環境の総合評価は、音の印象だけでなく、空間や景観といった周囲の印象や描かれている登場人物の心理描写といったコンテクストも影響していると考えられる。特に、図 2-1 のように、心理描写に関しては強い正の相関が見られるものが多く、文章からイメージされる音環境の評価は、登場人物の心理描写が大きく影響しているものと考えられる。

また、文章を「音」「時刻」「場所」「状況」のアイテムに分け、各アイテムの音環境の評価への影響を見るため、音環境の総合評価の平均値を目的変数、各アイテムを説明変数として重回帰式を求める数量化Ⅰ類を行った。しかし、回帰式に有意性はみられず、また寄与率も低かったため、アイテムの音環境の評価への影響は確認できなかった。

３．実験Ⅱ

３．１　実験概要

実験Ⅰで明らかにできなかった、各アイテムのカテゴリーが音環境の総合評価に影響を及ぼすかを検討するため、各アイテムのみを切り出して、そこからイメージされる音環境の印象評価実験を行った。本実験で用いた各アイテム「場所」「季節」「時刻」「状況」のカテゴリーを表 3-1 に示す。

３．２　実験方法

実験は、呈示された４つのアイテムからその状況をイメージさせ、そこでの音環境の印象を評価させた。評価は音環境の総合評価のみで、7 段階評価で行った。また、イメージした環境でどのような音が聞こえてきたかを記述させた。評価は、一つのアイテムのカテゴリーの違いによる音環境の評価の差を調べるため、表 3-2 に示すように 3 グループに分けて行った。被験者は 3 グループとも 21～27 歳の九州大学の学生 14 名で、グループ①は男 11

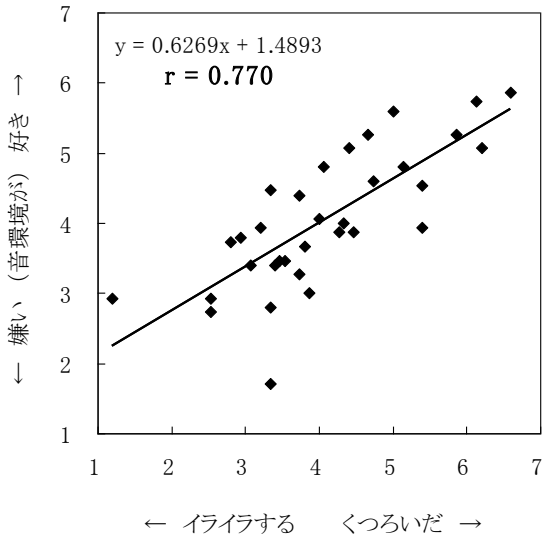


図 2-1　音の総合評価と「くつろいだ―イライラする」の相関図

表 3-1　各アイテムのカテゴリー

アイテム	カテゴリー		
場所	国道沿いのカフェテラス		
	田舎の縁側		
	にぎやかな商店街		
	静かな森の中		
季節	夏	秋	冬
時刻	昼間	夕方	夜
状況	会話中	読書中	休憩中

表 3-2　実験条件

グループ	①	②	③
場所	田舎の縁側		
季節	秋		
時刻	昼間	夕方	夜
状況	読書中		

グループ②は男 7 名、女 7 名、グループ③は男 14 名であった。なお、グループ間に被験者の重複はない。

３．３　実験結果

3 グループ間の音環境の総合評価の評定平均値を比較するため、一元配置の分散分析を行った。その結果、有意差が認められた 4 つの条件を表 3-3 に示す。有意差が認められた要因は、条件 1 では時刻、条件 2 では季節、条件 3 では時刻、条件 4 では状況であった（それぞれ $p<.05$）。また、「イメージした環境でどのような音が聞こえてきたか」の条件 2 の結果を図 3-1 に示す。

分散分析の結果から、音を聞く際の「季節」、「時刻」、「状況」といった一つのアイテムのカテゴリーが異なるだけで音環境の評価が異なる場合があることがわかった。これは、一つのカテゴリーの違いで、聞こえてくる音が大きく異なったことによるものと考えられる。しかし、条件2においては､両者ともに｢車の音」「客の話し声」「ＢＧＭ」「食器の接触音」など回答された音はほとんど同じであったにも関わらず、音環境の評価に有意な差がみられた。従って、条件 2 で季節のカテゴリーが「夏」の方が音環境の評価が低くなった要因は、泉らが報告していることと同様[3]、夏の蒸し暑いネガティブなイメージとの相乗効果によって生じたものと考えられる。

４．実験Ⅲ

４．１　実験概要及び実験方法

音環境の印象に、空間・景観や心理描写などが及ぼす影響をさらに詳細に検討するため、実験Ｉで用いた文章から、「音のみ」が描かれている文章、「音と空間」が描かれている文章を抜き出し、それらの音環境の総合評価を 7 段階評価で行った。次頁に示した文章の場合、「音のみ」は下線が記されている部分、「音と空間」は太字部分である。また、被験者は「音のみ」「音と空間」の評価ともに 21～27 歳の九州大学の学生 15 名（男 9 名、女 6 名）で、実験Ｉとの重複は 8 名である。

表 3-3　音環境の総合評価で有意差のみられた条件

条件	1		2		3		4	
場所	田舎の縁側		国道沿いのカフェテラス		静かな森の中		田舎の縁側	
季節	夏		秋	夏	秋		秋	
時刻	夜	昼間	夕方		昼間	夜	夜	
状況	休憩中		読書		休憩中		休憩中	会話中
音環境の総合評価	6.500	5.500	4.929	3.286	5.857	4.429	6.357	5.214

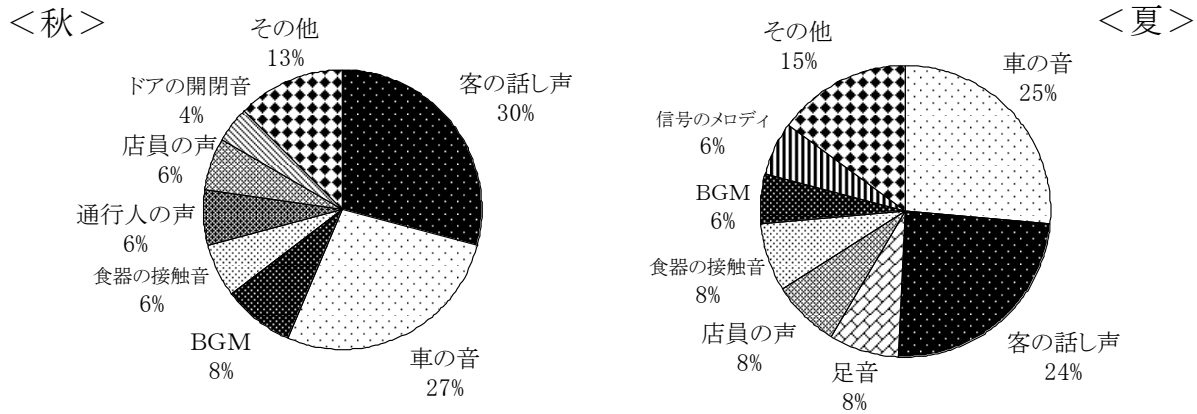


図 3-1　条件 2 における「イメージした環境で聞こえてきた音」の回答の比率

４．２　実験結果

実験Ｉでの音環境の総合評価と、本実験での「音のみ」と「音と空間」の結果の 3 条件の評定平均値を比較するため、一元配置の分散分析を行った。その結果、33 の文章のうち 11 の文章で有意差が認められた。従って、同じ音もしくは同じ音環境が呈示された条件の評価であっても、文脈によって印象が変化することがあることがわかった。

また、有意差が見られた文章は、ポジティブやネガティブな表現が含まれていることが多いことがわかった。空間の描写においては、「風が前髪を揺らし、上に広がる空は白く」、「昔なつかし、といったほのぼのとした言葉の似合う家」「庭に落ちている西日の色」や「どれだけ目を凝らし遠くを見ても、僅かな明かり一つ見ることができなかった」といった表現である。また心理の描写においては「のんびりといい気持ち」や「死にたいな、と思った」「一刻も早く逃れたい」といった表現である。これらポジティブな表現が含まれると音環境は良い評価になり、ネガティブな表現が含まれると音環境は悪い評価になったことから、「音環境と空間の評価」や「音の評価と描かれた登場人物の心理描写」の間には相乗効果があると考えられる。

店を出るともう外は陽が傾きかけていた。空気がさわやかで、むせかえりそうだった。電車に乗って、アマの家に向かう。**駅から家までの道、家族連れが多い商店街で、うるさい人々の声**に吐き気を覚えた。ゆっくり歩く私の足に、子供がぶつかった。私の顔を見て、素知らぬ顔をするその子の母親。私を見上げて泣き出しそうな顔をする子供。舌打ちをして先を急いだ。こんな世界にいたくないと、強く思った。とことん、暗い世界で身を燃やしたい、とも思った。

５．おわりに

以上の実験結果から、音環境の評価は聞く際の状況（コンテクスト）によって大きく異なることが明らかになった。特に、文章における音環境のイメージの印象評価においては、心理描写の影響が大きいことがわかった。

さらに、音の評価と、空間や登場人物の心理描写には相乗効果があることが示唆された。今後は、条件を増やし評価の差の生じ方の傾向を把握することで、相互作用に関してさらに詳細に検討していく必要がある。

参考文献

[1] 岩宮眞一郎,音の生態学―音と人間のかかわり―, コロナ社, p6, 2000
[2] 宮川雅充他, 音環境の印象に及ぼす視覚情報と聴覚情報の相乗効果に関する研究, 環境衛生工学研究, 第 15 巻第 3 号, pp187-191, 2001
[3] Hugo Fastl,Audio-visual interactions in loudness evaluation, Proc 18th International Congress on Acoustics vol.2, pp1161-1166, 2004
[4] 泉清人他, 現場と実験室における騒音評定の相違についての考察, 学術講演梗概集 D 環境工学, pp.145-146, 1986
[5] 藤原舞他, 生活環境の印象評価における視聴覚相互作用に関する研究 ―現場実験と実験室実験による研究―, 日本音響学会講演論文集, pp.771-772, 2007
[6] Yukiko YAMADA *et al*, Differences of Evaluated Values of Environmental Sounds on the Actual Spots and in a Laboratory, Proc Inter noise 08, p111, 2008

複写機稼働音の時間構造が音質に与える影響*

小野田伸一郎　星証人　高田正幸　岩宮眞一郎（九州大）
穂坂倫佳　大富浩一（東芝）

1　はじめに

機械製品の稼働音は製品のイメージに大きく貢献するものとして注目されている。以前は主に騒音レベルを下げる対策がとられてきたが、現在では不快でない音にする音質改善が重視されている[1]。

本研究の目的は、複写機の聴取印象の改善である。複写機の稼働音の時間構造に注目し、稼働音の発生タイミングを変化させることが聴取印象にどのような影響を与えているのかを明らかにする。

印刷時の複写機の主要な稼働音には、ピックアップローラーダウン音(PRD 音)、ピックアップローラーアップ音(PRU 音)、レジストローラー音(RR 音)、の 3 音が挙げられる。PRD 音の発生から次の PRD 音の発生までに用紙は 1 枚印刷され、これを 1 周期とし、この間に含まれる PRU 音や RR 音の発生タイミングを前後させた時の聴取印象への影響を検討した。

2　実験内容

2.1 PRU 音及び RR 音の影響

評定尺度法による印象評価実験を行った。刺激は 1 周期を 1000 ms とし、PRU 音と RR 音の発生タイミングを前後させて作成した。125 ms, 250 ms, 375 ms, 500 ms, 625 ms, 750 ms, の 6 箇所何れかの位置に PRU 音と RR 音を配置させ、全ての組合せで 36 個とした。評定尺度は Table. 1 に示す 7 段階の両極尺度 6 対を使用した。ヘッドフォンから、被験者の両耳に同一の刺激を 80 dB でランダムに呈示した。被験者は 16 名(男性 12 名、女性 4 名)。

Table. 1. Rating scale.

評定尺度	快い-不快な
快い-不快な	-
好き-嫌い	0.918**
ゆったりした-せわしない	0.501**
ふらついた-安定した	-0.853**
リズミカルな-リズミカルでない	0.811**
高級感のある-安っぽい	0.614**

相関係数は, **: 1%水準で有意

各尺度について刺激ごとに平均評定値を求めた。全ての尺度間で高い相関があったため、各尺度を代表して「快い-不快な」の尺度の平均評定値を Fig. 1 に示す。また「快い - 不快な」の刺激の平均評定値を従属変数とし、PRU 音、RR 音それぞれの発生位置を要因として分散分析を行った。その結果、RR 音の主効果が有意であった（$p<.01$）。従って RR 音の位置の変化、つまり PRD 音と RR 音の間隔の変化が「快さ」の印象に影響を与えると言える。PRU 音の主効果は見られなかった。

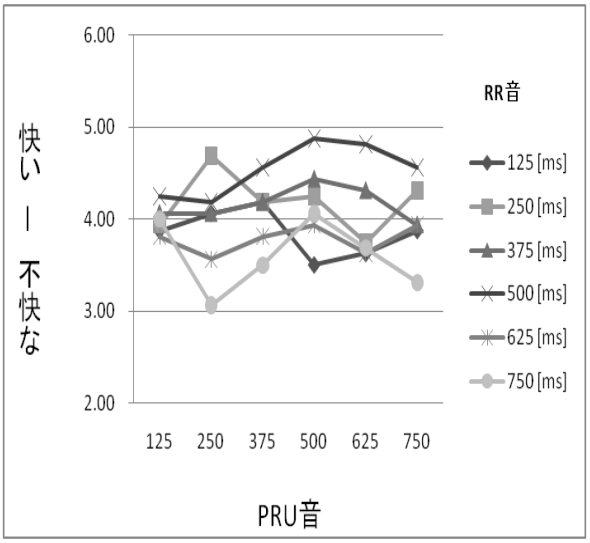


Fig. 1. Average ratings of pleasantness of each sound stimulus

*Effect of time structure of operating sounds of a copy machine on the sound quality of machinery noise, by ONODA,Shin-ichiro, HOSHI, Akito, TAKADA, Masayuki, IWAMIYA, Shin-ichiro (Kyushu university), HOSAKA, Rika, OTOMI, Koichi (TOSHIBA).

また、Fig.1 の平均評定値から RR 音の発生タイミングが 250 ms ～500 ms の時に比較的快く判断される傾向が見られた。

2.2 RR 音の最適な発生位置

RR 音の最適な発生タイミングを確認するため、一対比較法による印象評価実験を行った。刺激は RR 音の発生タイミングが、0 ms, 250 ms, 333 ms, 416 ms, 500 ms, 583 ms の 6 個用いた。被験者には 6 個の試験音からランダムに選ばれた基準となる音(基準音)と、対比して評価する音(評価音)について、「基準音に比べ評価音がどれだか快いか」を 7 段階で判断させた。被験者は 17 名(男性 13 名、女性 4 名)。

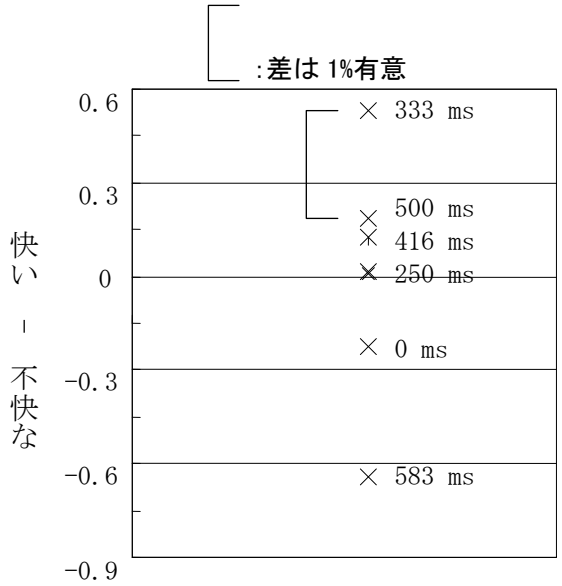


Fig.2. Average ratings of pleasantness of each sound stimulus in the case of paired comparison experiment

刺激の平均評定値を Fig.2 に示す。Fig.2 より RR 音の発生タイミングが 333 ms の時に最も「快い」ことが確認できた。

以上より、周期が 1000 ms 時の RR 音の最適な発生タイミングは、周期の 1/3 である 333 ms であることが分かった。

2.3 異なる周期による RR 音の影響

1 周期の長さが 1000 ms の時の RR 音の最適な発生タイミングを確認することができたが、1000 ms 以外の時に RR 音の発生タイミングが 1/3 の位置が良いのか、333 ms の位置が良いのかの確認をしていないことから、異なる周期において、RR 音の位置はどちらが快いかを検討した。

1 周期が 750 ms, 1250 ms, 1500 ms, 1750 ms, 2000 ms, の時の RR 音の発生タイミングが、周期の 1/3 の位置の場合と 333 ms の場合のどちらが快いかを判断させた。被験者は 15 名(男性 11 名、女性 4 名)。

カイ二乗検定の結果、1 周期が 1750 ms, 2000 ms, の時には RR 音が 333 ms の場合が快いと判断された ($p<.05$)。また、有意差はないが 1250 ms, 1500 ms, においても RR 音の位置が、周期の 1/3 の場合よりも 333 ms の場合の方が快いと判断される傾向が見られた。周期が 750 ms の時には反対に RR 音の位置が 1/3 の方が快いと判断される傾向が見られた。

以上より、周期による RR 音の発生タイミングと聴取印象との関係では、RR 音の発生タイミングが周期の 1000 ms より短い場合は 1/3 が快いとされ、1000 ms より長くなると 333 ms が快いとされる可能性が考えられる。

3 まとめ

本研究の結果、複写機の主要な稼働音の聴取印象には、PRD 音と RR 音の間隔の寄与が大きいことがわかった。また、その間隔は、周期が 1000 ms より短いときは RR 音の発生タイミングが周期の 1/3 の位置にある時に快く、周期が 1000 ms より長くなると、333 ms の位置にある時に快い傾向が見られた。今後は 1000 ms 以外の周期の PRD 音、RR 音の間隔の聴取印象への影響をより詳細に検討する。

参考文献

[1] 穂坂倫佳, “家電製品の低騒音化技術と音質改善,” 日本音響学会誌, 62(10), pp. 744-749(2006)

聴覚を考慮した音場評価手法に関する研究

鈴木正博・尾本章 (九州大)

1　はじめに

室内音響、特に音場に関する分野では、これまでに多くの研究がなされており、そのための指標は従来から数多く提案されてきた。これらの指標はインパルス応答をもとに算出され、音場を物理的な側面から客観的かつ定量的に扱うことを可能にした。

しかし、従来この分野では、人の聴覚に関する特性などは直接的に考慮されてきていない。最終的に音場の中で音を聴くのが人であることを考えると、考慮の余地があると考えられる。そこで、本研究では人の聴覚特性を考慮することで、新しい観点から音場評価をとらえることができないかという試みを行ってきた。

本報告は、新たに構築した評価手法をもとに、現在までに行った検討の結果をまとめたものである。

2　本研究での評価手法

本研究では、人の聴覚を模擬した聴覚フィルターをもとに、新たな評価手法を構築した。そして、これを基にして、様々な検討を行っている。

2.1　聴覚フィルター

本研究では、人間の聴覚を模擬した処理方法を導入している。具体的には聴覚フィルター [1] を用いたもので、これは聴覚末梢系をシミュレートするフィルター群である。これを用いることで、音を聴いたときの内耳の基底膜上でのスペクトル表現が得られる。音サンプルを分析処理する段階では、まずこのフィルターを通して出力を得た。

2.2　評価手法の概要

評価手法の検討対象は、２つの音場の類似度とする。２つの部屋のインパルス応答を音楽に畳み込んだ信号を用い、これを聴覚フィルター群 (16 フィルター, 中心周波数 50-16,000Hz) に通し、その出力をもとに差を求めるという評価手法を新たに構成した。

差の算出については、各フィルター各時間毎に対応するもの同士でまず振幅のレベル (dB) の絶対差をとり、その値を各フィルター×時間の全体にわたって平均する、という方法が最も主観評価との対応がよいことが現在までに分かっている。[2]

2.3　実験：主観評価との比較

聴覚フィルターを使った分析結果と主観評価実験での判断とを比較した。その結果、聴覚フィルターの出力をもとに２つの信号の差を求めると、主観評価と高い相関が得られることが分かった。本研究の実験を模式的に表したのが次の Fig. 1 である。

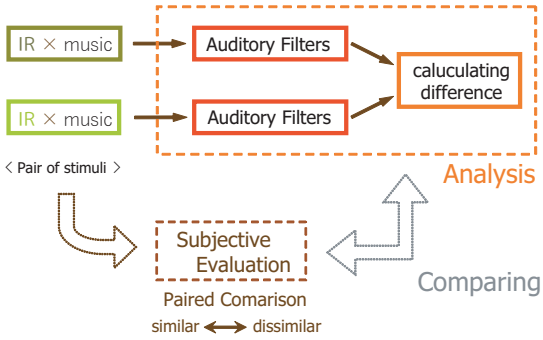


Fig. 1　実験の模式図

2.4　主観評価との対応

結果の一例を Fig. 2 に示す。これは主観評価と本研究の評価手法による評価の対応関係を示すものである。横軸方向に主観評価、縦軸方向に評価手法の評価をそれぞれとって、１種類の音楽についての結果をプロットしている。図における１つひとつの点は、部屋同士の組み合わせに対応する。

今回４種類の音楽を用いたが、どの音楽についても主観評価と分析による評価のあいだに高い相関が得られた。相関係数は約 0.8〜

Method of Acoustic Evaluation based on Auditory Perception by SUZUKI Masahiro and OMOTO Akira, (Kyushu University)

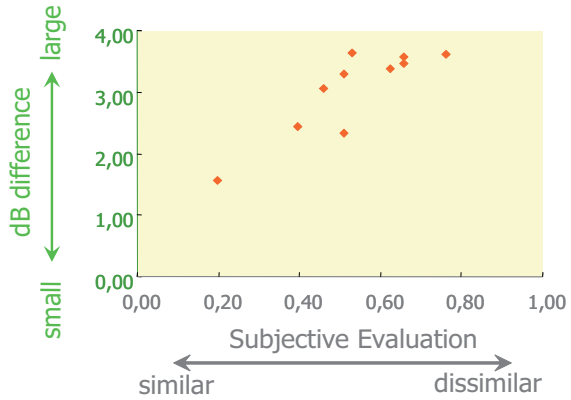


Fig. 2　主観評価と分析による評価の対応例

0.9 である。つまり、主観評価においての判断と、この分析で評価したものとがよく対応していると考えられる。

ここで、異なる種類の音楽すべてについて高い相関になった、というのはひとつのポイントである。

3　評価手法についての検討

本研究で用いる評価手法では、周波数と時間、音場の特性と音楽など複数の要因が絡みあっている。そこで、これらの要因を１つずつ切り分け、何を評価しているかについての分析を詳しく行った。

3.1　周波数領域と時間領域における評価

まず、周波数と時間に関する分析を行った。小規模音場では、カラレーションとよばれる周波数領域での歪みが生じることが知られている。上述の実験では小規模音場を対象としていたため、これは周波数領域での評価を行っているのではないかという仮説を立てた。

そして、評価手法に変更を加え、周波数領域もしくは時間領域のみの評価を行うように変えた上で、それぞれについてオリジナルと同様に主観評価との対応関係を調査した。

検討の結果、仮説に反して、周波数領域よりもむしろ時間領域の影響が大きいことが示唆される結果が得られた。また追加検討によって、時間変動の差のみからでも、同様の評価を行うことが可能であることが分かった。

3.2　音源による評価の違い

続いて、音源としての音楽と音場の評価に関する検討を行った。この評価手法がなぜ主観評価と高い対応関係を得たのかを検討する過程で、この手法が音源による音場の評価の違いについて追従していることが分かった。

これは主観評価の側面からの詳しい分析において明らかになったことである。まず、人の音場についての評価は、用いる音源 (音楽) によって傾向が違うということが分かった。

これは、例えば次の図に示すことができる。この図は類似度の主観評価で、部屋の各組み合わせについて、被験者全員の平均値をグラフ化したものである。部屋の組によっては、２種類の音楽で評価が大きく異なることが分かる。

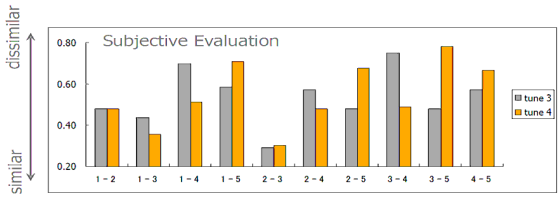


Fig. 3　用いる音楽による音場の評価の違い

そして、本研究での手法がこの傾向の違いに追従することが明らかになった。これが主観評価との高い対応関係という結果に結びついていることが考えられる。

これは従来の評価指標では成し得ないものである。なぜなら、従来は音場のみについて評価するだけなので、音源によって生じる評価の違いは構造的に考慮できない。

この結果は、音場を評価する際、音場単体ではなく、音源も考慮した上で評価を行う必要があることを示唆する。

4　考察

次に、この評価法において、どの部分が重要なのかを検討した。

4.1　音楽信号を扱うこと

インパルス応答を曲などに畳み込み、その結果を音として分析することが、本評価手法の特色であると考えられる。そして、それに

よって、聴覚に親和した処理が実現でき、音源による評価の違いを考慮できているのではないかと考えられる。

従来の音場評価ではインパルス応答のみを用いて、つまり音場の物理的情報のみを取り出して評価してきたが、本研究での評価が示唆するのは、音場単体の評価では充分ではなく、音源信号の特徴も考慮して評価することが必要なのではないか、ということである。

4.2 評価における重要項目

評価に大きく影響を与えるものとして、畳み込んだ音信号の音圧レベルの違いがあげられる。検討によって、評価結果の妥当性は、音信号間の音量の均等性に左右されることが明らかになった。そこで、評価の前段階に音圧を均等化する処理を加えると、音圧差の影響を排除することができ、主観との相関もわずかながら高くなることが分かった。

また、評価手法の他の部分についても重要性を調査したが、周波数領域でのフィルタの形状や、その数などについては、評価結果に与える影響は大きくないことが分かった。

5 応用例

5.1 多次元尺度構成法による布置

本研究での評価手法は、元々は2つの音場が異なる度合いを出力するのみであった。今回その出力を有効に活用すべく、多次元尺度構成法 (MDS) による各音場の類似関係の布置を求める処理を追加した。Figs. 4, 5 は結果として出力される布置の例を示すものである。

図は、5つの部屋の音場について、類似度から互いの類似関係を表したものである。すなわち、近いところに配置されている番号の組み合わせほど、似た音場と評価されているということである。この2つの図は、本研究での評価手法で評価した場合と、主観評価をもとに評価した場合のものである。回転を考えると2つの図において、大まかに見て似たような配置になっていると言えるだろう。よって、この手法は人の評価と大きな差のない評価をすることができているといえるのではないだろうか。

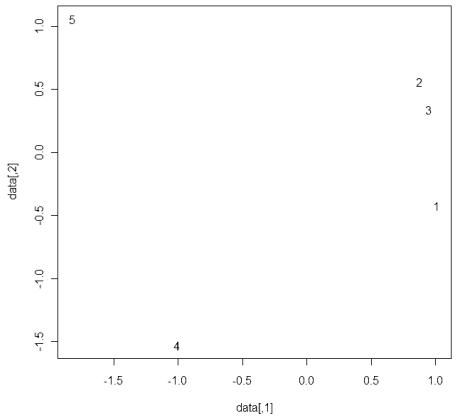


Fig. 4 MDS 布置の例 – 聴覚フィルタを用いた評価手法による評価

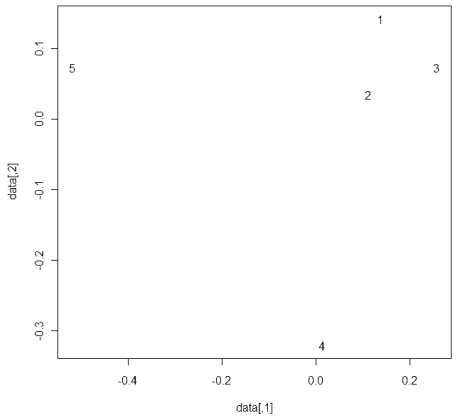


Fig. 5 MDS 布置の例 – 主観評価

5.2 ホール音場の評価

また、この評価法を応用し、ホール音場の全席評価を試みた。具体的には、ホールの全席で測定されたインパルス応答をもとに、それをすべて同一の音楽に畳み込み、基準となる席を指定して、その基準とどれくらい差があるかを席ごとに求め、その差の分布を表示するというものである。

その結果、ホール音場の左右対称性がはっきりと見られることや、評価の音圧依存性が大きいことなどが分かった。

これに関しては、さらに検討を深める予定である。

Fig. 6 は、ホール音場評価の例である。ホールの座席の見取り右側の濃い赤色の部分を基準として、他のすべての席でそれとどれくらい差があるかを求めたものである。

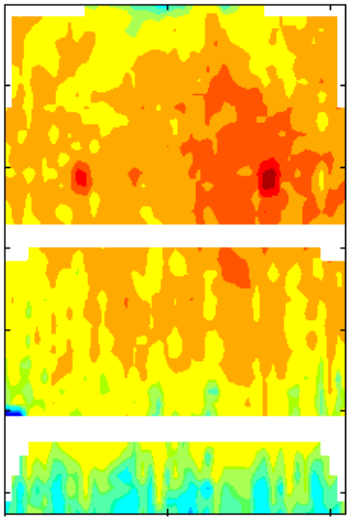


Fig. 6　ホール評価の例

色が薄くなるほど、差が大きい、つまり基準となる席と似ていないと評価されたことになる。全体的に見ると、席が離れるほど、また前方の列・後方の列になるほど、差が大きくなることが分かる。また、図の左側に色の濃い部分があり、これは左右対称性の影響が色濃く表れているものと解釈される。

6　まとめ

聴覚フィルターをもとにした新たな評価手法によって、まず、人の評価と似たような評価が可能であることを明らかにした。そして、その評価手法をもとに検討を加えることで、音源信号によって音場の評価は異なり、この手法による評価がそれに追従することなどが分かった。さらに、応用として、類似度評価から MDS によって類似関係を導き図示することや、ホールの全席評価に応用することができるという可能性を示した。

今後としては、音源による評価の違いや時間領域の評価についてさらに研究を深める予定である。音源を変化させ、音場の評価がどのように変わるかを調べたり、また、音源自体の特徴量なども調べていきたい。

参考文献

[1] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, M. H. Allerhand: Complex sounds and auditory images. Auditory Physiology and Perception: 9th International Symposium on Hearing, Oxford, Y. Cazals, L. Demany, K. Horner (eds.), Pergamon, pp.429-446. (1992)

[2] 鈴木正博, 尾本章, 中原雅孝 : 聴覚フィルターを用いた室内音場評価に関する研究, 日本音響学会秋季研究発表会講演論文集, 2-R-17 (2008)

反射音を可変とする音響壁面システムに関する研究

上川和久 尾本章 (九大芸工)

1　導入

1.1　研究背景

室内音響において、室の形状や使われている素材によって音響特性が決定される。例えば、コンサートホールは適度な響きのある空間に、スタジオはあまり響かないような空間にと、それぞれの目的に応じて建築音響設計が必要となる。そこで、あらゆる目的に応じた音響空間を一つの空間で実現することができれば実用的な空間が創れると考え、本研究のテーマである VRAWS が提案されている。

Fig. 1　photo of Variable Reflection Acoustic Wall System.

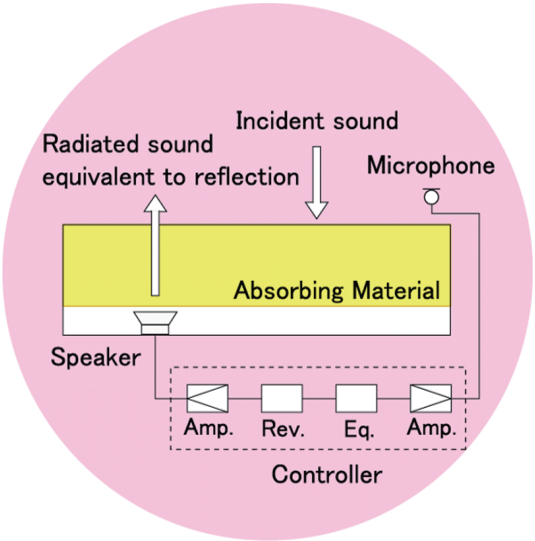


Fig. 2　Mechanism of VRAWS.

1.2　VRAWS について

VRAWS とは、『Variable Reflection Acoustic Wall System』の頭文字を取ったもので、“反射音を自由に変える”ことを目的とした、吸音～残響付加まで幅広く対応したハイブリッドな音響壁面システムである。仕組みとしては、壁面を模擬したこの VRAWS に何かしらの音が入射し、その一部は吸音材により吸収し、一方でユニットに設置したマイクロホンが音を検知し、所望の反射波をアンプやリバーブなどのエフェクター類で設定し、ユニットのスピーカーから放射するというものである。例えば、検知した音に 2 秒程度のリバーブをかけてある程度の音量で音を放射すればコンサートホールのような響きを持つ広がりのある音場になり、0.5 秒程度のリバーブをかけて大きめの音量で音を放射すれば反射音が大きく広がりのない音場になる。このように、残響時間やアンプの音量の設定次第で響きの無い音場から響きのある音場まで演出することが可能となり得るシステムである。現行の VRAWS 試作機の写真を Fig.1 に、仕組みの図を Fig.2 に示した。

1.3　先行研究 [1, 2]

VRAWS の提案から、プロトタイプユニットの作成、ユニットの放射特性・吸音特性の検討、反射波の周波数特性の変化・残響時間の可変性・模型実験による検討などが行われている。プロトタイプの作成は、初めは小さめの模型作りからはじまり、Fig.1 のように現在の大きさのものまで作成されている。ユニットの放射特性は、現在の VRAWS に様々な拡散体を装着させたことを想定して境界要素法によるシミュレーションを行ったり、その結果に基づいて実際に拡散体を装着させての実験が行われている。Fig.1 の VRAWS8 台のうちの 1 台に装着してある白いものが先行

A study of Variable Reflection Acoustic Wall System, by KAZUHISA Kamikawa and AKIRA, Omoto (Kyushu University)

研究において試行された 1 次元 PRD 拡散体である。吸音特性については、吸音材を装着した VRAWS 本体がどの程度音を吸収するのかについて検討されている。今後の課題としては、システムの独立化、実際に VRAWS で空間を囲っての主観評価実験などが挙げられている。

2　目的

2.1　研究の目的

この研究の究極的な目的は、あらゆる目的に応じた音響空間を一つの空間で実現することである。その中でも今現在は、普段生活しているような小さな空間でもコンサートホールのような豊かな響きを与える空間を創ることを目標にしている。この目標を達成した後は、パラメータの設定等によってスタジオやライブハウス、プレゼンテーションルームなど様々な空間を模擬して、様々な空間を演出することを新たな目標にしていきたい。

2.2　臨場感について

ここで言う臨場感とは一般的に映画館やホームシアターなど、実際その場に身をおいているかのような感じであったり、音に包まれている感じであると定義する。これまでは実際に存在するコンサートホールのインパルス応答などを用いて厳密に音場を再現する『高臨場感』が主流であった。例えば、空間上の点を制御するバイノーラルシステムやトランスオーラルシステム、空間上の領域を制御する 5.1ch(7.1ch) サラウンドシステムなどがこの高臨場感にあたる。最近ではユーザー一人ひとりが好みに応じて自由に制御することのできる、フレキシブルで従来のシステムではありえないような演出型の臨場感を創る『創臨場感』が提唱されている。本研究ではこの創臨場感のカテゴリーの属する演出型の臨場感を取り扱うことになる。

3　創臨場感を演出するために

3.1　空間を囲っての主観評価実験

今現在の VRAWS で空間を囲い、広がり感を中心にどのような効果が得られるかを 3 つの刺激音と VRAWS の設定 7 パターンで主観評価実験を行った。実験の風景を Fig.3 に示した。実験の概要としては、九州大学大橋キャンパス内の残響可変室において被験者の前方に 2ch の音源を置き、VRAWS を被験

Fig. 3　Experimental landscape.

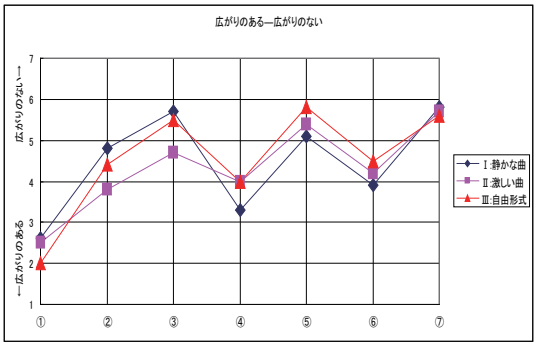


Fig. 4　Result of nature feeling.

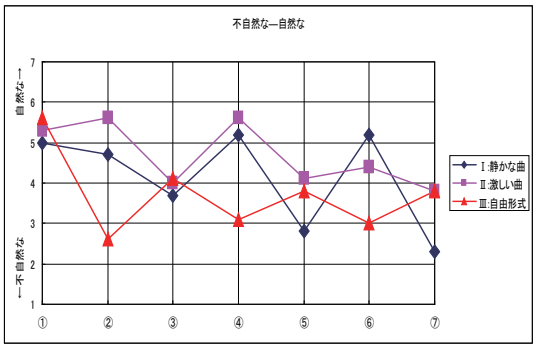


Fig. 5　Result of nature feeling.

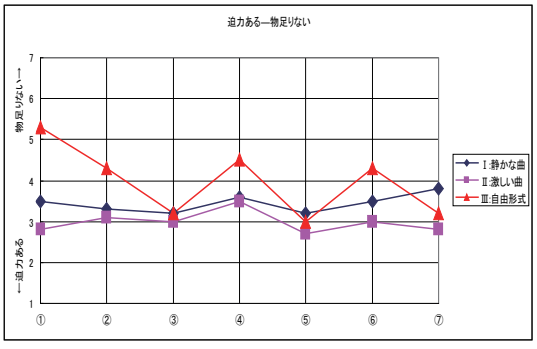


Fig. 6　Result of powerful feeling.

者の横に2台、後ろに2台の計4台を設置し、VRAWSの設定によってどのように印象が変化するかを調査した。用いた刺激音は、音源による大まかな違いを把握するために、1:静かな曲、2:激しい曲、3:被験者に自由に音を出してもらうものとした。今回の実験では15の刺激対を用いたが、その中でも注目すべき3つの刺激対、『広がりのある–広がりのない』・『不自然な–自然な』・『迫力のある–物足りない』についての結果をそれぞれFig.4,5,6に示した。

グラフ横軸のパターンは、1.VRAWS OFF、2.後ろのみ残響時間0.5秒付加、3.後ろのみ残響時間2.98秒付加、4.横のみ残響時間0.5秒付加、5.横のみ残響時間2.98秒付加、6.全て残響時間0.5秒付加、7.全て残響時間2.98秒付加である。

実験の結果、広がり感は全ての刺激において設定した残響時間との相関が見られた。特に静かな曲や自由に音を出してもらった場合に顕著な違いが見られた。また、自然さについては残響時間が長いほど不自然に感じるという結果が得られたが、不自然と感じながらもこのような空間で音楽を聴いてみたいという興味深い意見が得られた。迫力感に関しては、音楽を用いた場合は有意差が出るほどの顕著な違いは見られなかったが、音を自由に出してもらった場合は残響時間が長い方が迫力のある印象になることがわかった。これらのことから、VRAWSからの音の放射により音場や音場の印象に何かしらの影響を与えることがわかった。

3.2 VRAWS専用のリバーブの設計

システムとして確立するために、VRAWSに搭載するための専用のリバーブの設計を試みた。人工的なリバーブは異なる二つのアプローチによって得られる。一つは物理的アプローチであり、実際の部屋の正確なリバーブを人工的に再現しようとするものである。このような詳細さを得るために、通常、リバーブを含む信号は部屋のインパルス応答とソース信号を畳み込み演算することによって得られる。実際に存在するコンサートホールのインパルス応答などを元の音信号に畳み込み処理を施してその音場を再現するものがこれにあたる。 インパルス応答は部屋から直接記録するか、仮想的な部屋の幾何的モデルを用いることで得られる。後者のケースでは、部屋の幾何的な特徴（大きさや壁の材質など）がインパルス応答の係数を計算するために利用される。このアプローチはソースと聴き手の位置が与えられれば正確な表現が可能であるものの、リアルタイム・バーチャル・リアリティやゲームなどの用途においては柔軟性と効率の面で難がある。 例えば、3秒間のオーディオ信号と2秒間の部屋のインパルス応答（44.1kHzでサンプリング）を時間領域で畳み込み演算するには、おおよそ120億回の乗算と220,500回の加算が必要となる。等価な周波数領域についての畳み込み演算では220,500回の複素数の乗算が必要となる上、変換および逆変換作業のオーバーヘッドが加わる[3]。このように、主に高臨場感を演出する際に用いられるFIRを用いたリバーブは、再現度は非常に高いが演算処理が非常に重いために本研究で目的とするリバーブには適さない。

第二のアプローチは知覚的アプローチと呼ばれ、自然のリバーブと「知覚的に区別できない」人工的なリバーブ・アルゴリズムを生み出そうとするものである。 それらのアルゴリズムの目的は自然のリバーブの目立つ部分だけを再現することである。このアプローチは物理的なアプローチより一般にはるかに効率的で、理想的には生成されるアルゴリズムを完全にパラメータ化することができる可能性を持つ[3]。これらのことから、VRAWSに搭載するためのリバーブは知覚的アプローチのIIRを用いたリバーブを設計することにした。

IIRを用いた非常にシンプルなリバーブのアルゴリズムとして、コムフィルタを用いたリバーブが挙げられる。コムフィルタは減衰利得を持つフィードバックループと単独のディレイラインで構成される。この他にも、コムフィルタをいくつも並列につなぎ合わせたコムフィルタネットワーク、様々な周波数帯域に分けて処理するマルチレートアルゴリズムなど様々なアルゴリズムが存在する。

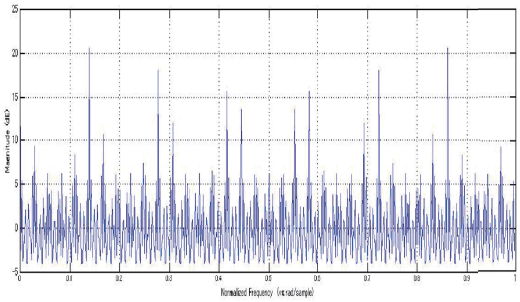


Fig. 7　Frequency characteristics of comb-filter.

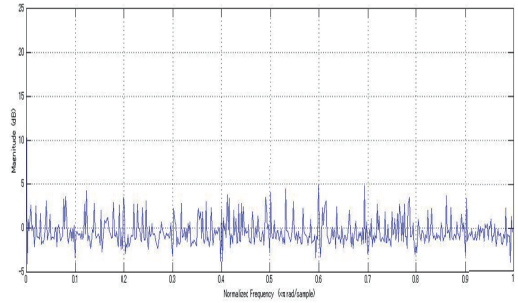


Fig. 8　Frequency characteristics of comb-filter network.

この中からいくつかのものを試し、今回はその中からコムフィルタとコムフィルタネットワークを用いたリバーブを紹介する。

リバーブ設計の流れとしては、まずMATLABでのシミュレーション、その後アルゴリズムが確定した後にDSPに書き込みリアルタイム処理での検証を行った。コムフィルタの数を増やせば増やすほど周波数特性はフラットになるが、DSPの性能上ディレイラインが12個程しか設定できなかったため、今回は12個でコムフィルタ単品との比較を行った。コムフィルタ単品の時の周波数特性をFig.7、コムフィルタネットワーク(12個並列時)の周波数特性をFig.8に示した。コムフィルタ単品のリバーブでは金属的な音が知覚され大変不快に感じるものであったが、コムフィルタをいくつか組み合わせたコムフィルタネットワークのリバーブではそれが改善され、コムフィルタの数を増やせば周波数特性がよりフラットになることがわかった。しかしながら12個並列につなげただけではまだ金属的な音が完全には打ち消されないため、ディレイラインを増やすための工夫や他のアルゴリズムを用いてより改善する必要がある。

4　おわりに

主観評価実験では、VRAWSの設定によって広がり感をはじめ、様々な印象が変化することがわかった。また、不自然と感じながらもこのような空間で音楽を聴いてみたいという興味深い結果から、VRAWSの存在意義が確たるものになる可能性が示唆された。
また、今回は４つのマイクで検知した音を１つのリバーブとハウリングキャンセラにより信号処理を行ったため、出力する際の音がMIXされた形になっている。このために音場がモノラル感があるなどの印象を与えてしまったものと考えられる。そのため今後は1台1台に独立したリバーブとハウリングキャンセラを搭載し、独立させた形ではモノラル感が改善されるか、広がり感はさらにどうなるかなどを検証するために主観評価実験を行いたい。

リバーブに関しては、コムフィルタネットワークでは周波数特性の抑制に限界があるため、今後はマルチレートアルゴリズムなどの別のアルゴリズムを用いてよりフラットな周波数特性を持ち、なおかつ迫力のあるリバーブの設計を試みたい。

また、残響時間やIACCなどによる物理指標も測定し、主観評価実験による心理的評価との相関や差異などを明らかにしていきたい。

参考文献

[1] Genta Yamauchi, et al.,“Variable reflection acoustic wall system by active sound radiation,”♯0028, Preprint AES 13th regional Convention, Tokyo2007, 2007.

[2] Y. Nagatomo et al.,“Variable reflection acoustic wall system by active sound radiation,”Acoust.sci. & Tech. 28 84-89, 2007.

[3] Jasmin Frenette,“Reducing artifical reverberation requirements using time-variant feedback delay networks”, Reserch of Master of Science, Music Engineering Technology, 2000.