5B−4　高域復元による音声の聞き易さ改善手法の性能評価

鵜木　恒有　　岩城　護　　木竜　徹
新潟大学大学院　自然科学研究科　人間支援科学専攻

1. はじめに

現在の電話音声は回線の効率的な使用の為に，音声符号化による圧縮や周波数帯域幅の制限がされて伝送されている．これにより電話音声は聴力が正常な者が音韻を正しく聞き取れる程度であり，特に高齢者や聴覚障害者にとって聞き取りにくいものとなっている．高齢化社会において電話音声を聞き易くする支援ツールが必要であると考えられる．そこで我々は，音声の低域スペクトルと高域スペクトルの相関が強いという仮定のもと，周波数帯域制限された音声の聞き易さ向上を目的とした音声改善手法を提案した[1]．そこで本稿では，より実際の電話音声に近い環境下におけるシステムの性能評価の為，様々な話者や音韻の組み合わせに対する性能を実験的に評価した．

2. 手法

2.1 概要

本研究での音声改善手法の概要を図1，設定値を表1に示す．高域復元は，LPF(バターワース，10次，fc=3.4[kHz])を通し電話音声を模擬した入力音声をフレーム分割し，各フレームの失われた高域部に予め用意したデータベースの中から最も適当だと思われる高域成分を決定し加えることによって実現する．予め用意するデータベースには，検索用インデックスとしてのLF成分，復元用データとしてのHF成分と位相成分をそれぞれ保存しておく．復元用データのHF成分と位相成分は，入力音声と検索用インデックスのLF成分の対数振幅包絡と位相よる比較から，入力音声に最も類似したLF成分の高域成分として決定され，声の大きさの指標となるゲインの調整と声の高さの指標となる基本周波数情報の調整を行い，入力音声に足し合わされる．以上の処理で高域復元された全フレームを合成することで高域成分を持った音声を得る．

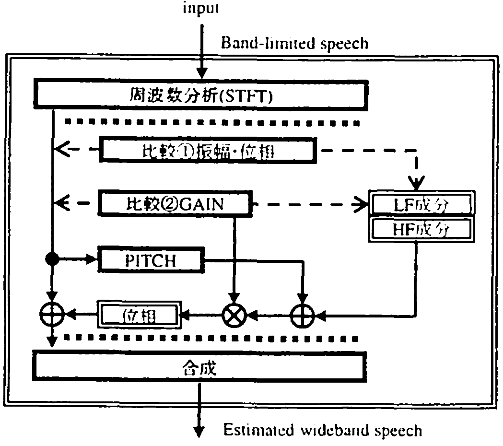

図1 システムの概要

2.2 音声特徴データベース

音声特徴データベースの概要を図2に示す．LF成分は，LPF後の音声をSTFTし，各フレームを保存する．HF成分は，原音声とLF成分をそれぞれSTFTした後，各フレームで引き算し，対数振幅包絡に変換して保存する．対数振幅包絡はケプストラムを用いて求め，リフタリングのサイズは2.0 [ms]である．また，対数振幅包絡に変換する前のHF成分から位相成分を算出し保存する．

2.3 HF成分の決定

入力音声とLF成分の比較によってHF成分を決定する．まず対数振幅包絡の比較によって，入力音声に最も類似した上位10個のLF成分を選出し，次に位相成分の比較によって先に選出された中から入力音声に

表1 システムの設定値

標本化周波数	20 [kHz]
量子化ビット数	16 [bit]
FFTの点数(*Nfft*)	1024 [pt]
シフト長	64 [pt]
リフタリング	2.0 [ms]
窓関数	ハニング

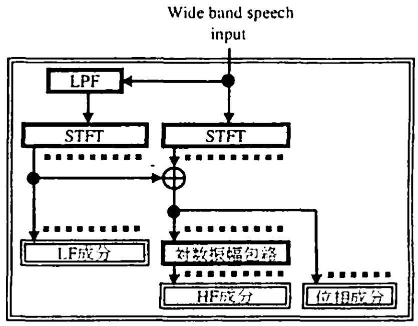

図2 音声特徴データベースの作成

最も類似した LF 成分を 1 つ選出する．この LF 成分と対となる HF 成分と位相成分を決定する．対数振幅包絡と位相の比較では共に式(1)で表される SD を用いた．ここで，S_1, S_2は比較する信号，FFT の点数を $Nfft$ として $N = Nfft/2+1$ である．

$$SD^2 = \frac{1}{N}\sum_{n=1}^{N}\left|S_1(n) - S_2(n)\right|^2 \qquad (1)$$

2.4 ゲイン調整

ゲイン調整は，平均パワーの調整量 ΔG を HF 成分に掛けることで実装した．ΔG を式(2)に示す．g_{in}, g_{LF} は，それぞれ入力音声と LF 成分の平均パワーである．

$$\Delta G = g_{in} - g_{LF} \qquad (2)$$

2.5 基本周波数情報の調整

基本周波数情報には，入力音声の微細構造を用いた．入力音声の微細構造をケプストラム分析によるスペクトル平坦化から算出し，高域に転写する．高域への転写は，0.5 - 3.0 [kHz]を 3.4 - 10 [kHz]に行う．

3. 実験

本稿における性能評価は，対数振幅包絡に対して式(1)で表される SD による客観評価実験を行った．使用音声は, 20 代男性 7 名(A～G)が発話した以下の 5 文(①～⑤)，計 35 音声である．

① おはようございます，きょうはさむいですね．
② はじめまして，よろしくおねがいします．
③ すみません，みちをおしえてください．
④ ここは，にいがただいがくこうがくぶです．
⑤ いま，おんせいのけんきゅうをしています．

電話音声に近い環境下での性能評価を行う為に，一つの入力音声に対するデータベースに含まれる音声の構築を，特定話者による構築(7 種類)と特定会話文による構築(5 種類)の計 12 種類で行った．入力音声を A-①としたときのデータベースの構築例を表 2 に，実験結果を図 3 に示す．得られた 420 個の出力音声（35 音声×12 種類）をデータベース構築方法毎に分けて解析を行った結果，平均 SD が特定話者で構築した場合と特定会話文で構築した場合で，それぞれ約 8.9 [dB]，約 8.8 [dB]で大きな違いは確認されなかった．特定話者で構築した場合で話者の違いによる改善効果に差が確認されたが，特定会話文で構築した場合では会話文による大きな違いは確認できなかった．特定話者で構築した場合，話者 E のとき約 10.2 [dB]であったのに対して話者 A のとき約 7.9 [dB]であり，SD による改善効果に約 2.3 [dB] の差が見られた．この差の原因としては，話者 E の音声の低域部と高域部の平均パワーの比率が他の話者のものと比べ，低域部が高いことが考えられる．

以上により，本稿の実験において，データベース内にある程度の音素が含まれていれば改善効果が期待できることが確認できた．また，データベースの構築に使用する話者によって改善効果が小さくなる場合があった．

4. まとめ

本稿では，一つの入力音声に対してデータベースの構築方法を 12 種類で行い，その性能を対数振幅包絡の SD による客観評価実験で行った．データベースを特定話者で構築した場合と特定会話文で構築した場合では大きな差は確認できなかったが，特定話者で構築した場合は話者間で音声改善効果に差があることが確認できた．今後の課題としては，話者による改善効果の差を小さくし，様々な話者に適応できるシステムの考案が必要であると考えられる．また，本稿では対数振幅包絡による SD の評価を行ったが，これは実際の聞こえとは異なることが考えられる為，高齢者や聴覚障害者にとって聞き易さの改善が得られるのかを聴取実験による主観評価で確認する必要がある．

参考文献

[1] 鵜木恒有，岩城護，木竜徹，"周波数帯域制限された音声の聞き易さ向上を目的とした音声改善手法の検討，" 信学技報，SP2005 - 37，2005.

表 2 データベースの構築例

特定話者による構築	特定会話文による構築
A-②,③,④,⑤	B,C,D,E,F,G-①
B-②,③,④,⑤	B,C,D,E,F,G-②
C-②,③,④,⑤	B,C,D,E,F,G-③
D-②,③,④,⑤	B,C,D,E,F,G-④
E-②,③,④,⑤	B,C,D,E,F,G-⑤
F-②,③,④,⑤	
G-②,③,④,⑤	

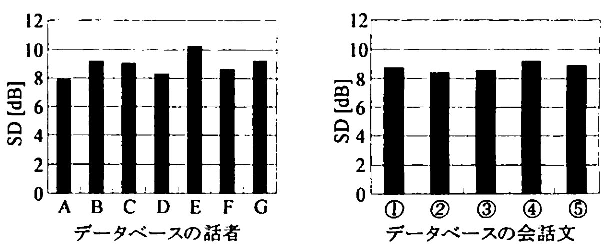

図 3 SD による客観評価実験結果