産業応用工学会論文誌, Vol. 6, No. 1, pp. 51-54 (Mar. 2018)
The Japanese Journal of the Institute of Industrial Applications Engineers
Online edition: ISSN 2187-5146, Print edition: ISSN 2189-373X, DOI: 10.12792/jjiiae.6.1.51

論　文

音声スペクトルの時間変化強調による明瞭性改善手法の提案

川原　竣介[a]　　平川　凜[a]　　中藤　良久[a,*]

Improvement Method of Speech Intelligibility by Enhancement of Spectral Variation

Shunsuke Kawahara[a,*], Rin Hirakawa[a], Yoshihisa Nakatoh[a]

(Received February 15, 2018; revised February 21, 2018; accepted February 23, 2018)

Abstract

It is hard to listen to an announcement in a noisy environments such as in a train while running. In this paper, to improve speech intelligibility in a noisy environment, we focused on the time variation of speech spectral representing spectral flux from the analysis result which clear speech has larger spectral flux than normal speech. Therefore, we propose speech enhancement method of spectral variation to improve the sound quality evaluation value of STOI. In the evaluation results of STOI score, enhanced speech by proposed method improved more than original speech in all the SNR conditions (SNR = -5, -10, -15 dB). Also, the significant difference in the dangerous rate of 5% was found in SNR = -10, -15 dB.

キーワード：音声強調，スペクトルフラックス，明瞭性，STOI
Keywords : speech enhancement, spectral flux, speech intelligibility, STOI.

1. はじめに

走行中の電車内や駅構内などでは，アナウンス音声が聞き取りづらくなることがある。その原因としては，周囲騒音や発話者の発話の仕方（発話スタイル）による影響などが考えられる。これまで，周囲騒音下での拡声音の明瞭度を改善する技術としては，MAEQ 法が報告されている[1-2]。一方，発話スタイルの影響については，明瞭音声は通常音声に比べて，スペクトルフラックス[3-5]が大きくなることがわかっている[6]。

本研究では，スペクトルフラックスを大きくすると雑音環境下における音声の明瞭性が改善すると仮定し，スペクトルフラックスを表す音声スペクトルの時間変化に着目し，音声スペクトルの時間変化強調方法を提案する。時間変化を強調する基準としては，主観的な音質評価指標である STOI [7-8]の音質評価値を向上させる方法を採用する。提案法に関して MAEQ 法と比較すると，MAEQ 法では周囲騒音のレベルに合わせて音声の補償量を計算し，音声強調を行うため，発話者の特性を考慮した音声強調はできない。しかし，提案法による音声強調では，音質評価値に着目しているため，雑音と発話者の特性を考慮した音声強調ができる。

* Corresponding author. E-mail: nakatoh@ecs.kyutech.ac.jp
a 九州工業大学
〒804-8550　福岡県北九州市戸畑区仙水町 1-1
Kyushu Institute of Technology.
1-1, Sensui-chou, Tobata-ku, Kitakyushui-shi, Fukuoka, Japan

2. 音声スペクトルの時間変化強調による明瞭性改善

2.1 音声スペクトルの時間変化強調方法

これまで，音声明瞭度を改善する技術として，在塚らが提案した周波数スペクトルの時間変化強調[9]や，柴田らが提案した動的特徴強調[10]等の音声の時間的な変化に着目した音声強調方法が提案されている[11-13]。

本研究では，在塚らが提案した周波数スペクトルの時間変化強調方法[9]を基にした上で，音声を帯域分割し，帯域毎に強調係数を設定し，スペクトルの時間変化を強調する方法を提案する。ここで，音声スペクトルの時間変化強調の式を以下に示す。

$$Y(t,f) = Y(t-1,f) + \alpha(j) * (X(t,f) - X(t-1,f)) \quad (1)$$

なお，式において，$X(t,f)$は強調前の音声のパワースペクトル，$Y(t,f)$は強調後の音声のパワースペクトル，$\alpha(j)$は強調係数，t はフレーム番号，f は周波数，j はサブバンド番号を表す。また，パワースペクトルは短時間フーリエ変換(STFT)から算出する。STFT の条件は，サンプリング周波数 16k Hz，フレーム幅 64 ms，フレームシフト 8 ms とし，窓関数はハニング窓を用いる。サブバンド毎のパワースペクトルの分割には，オクターブスケールバンドを用い，8 分割した。したがって，サブバンドは，

0~62.5 Hz, 62.5~125 Hz, 125~250 Hz, 250~500 Hz, 500~1k Hz, 1k~2k Hz, 2k~4k Hz, 4k~8k Hz とする。

2.2　強調効果の検証

提案法による音声スペクトルの時間変化強調の効果を確認するため，強調前の音声と強調後の音声の音声波形とスペクトルフラックスの変化の比較により強調効果の検証を行った。ここで，スペクトルフラックスの式を以下に示す。また，fs はサンプリング周波数を表す。

$$\mathrm{F} = \sqrt{\sum_{f=0}^{fs/2}(X(t,f) - X(t-1,f))^2} \qquad (2)$$

強調前の音声と強調後の音声の音声波形とスペクトルフラックスの変化を Fig.1 から Fig.4 に示す。使用した音声は，“アカハラ”と発話された単語音声である。強調後の音声については，500~4k Hz の強調係数を 2，それ以外の帯域の強調係数を 1 に設定した。また，強調前の音声と強調後の音声の音圧は 70 dB になるように正規化した。スペクトルフラックスは，2.1 節と同条件の STFT によりパワースペクトルを算出し，式(2)により求めた。

Fig.1，Fig.2 より，強調後の音声は強調前の音声に比べて，母音の”ア”の音節では音圧を正規化しているため，全体的に振幅が抑圧されていて，“カ”，“ハ”の子音を含む音節では振幅が大きい部分がより増幅されていることが確認できる。また，Fig.3，Fig.4 より，強調後の音声は強調前の音声に比べて，音声波形と同様に，母音の“ア”の音節ではスペクトルフラックスが抑圧されていて，“ハ”，“ラ”の子音を含む音節ではスペクトルフラックスが大きい部分でより増幅していることが確認できる。

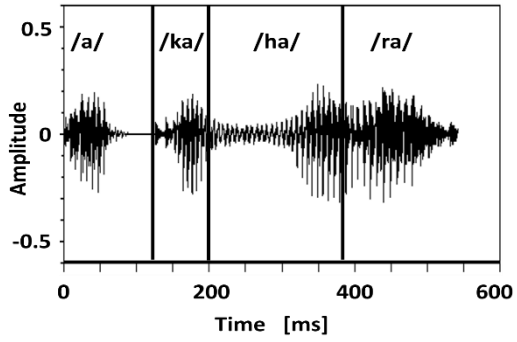


Fig. 1. Speech waveform before enhancement.

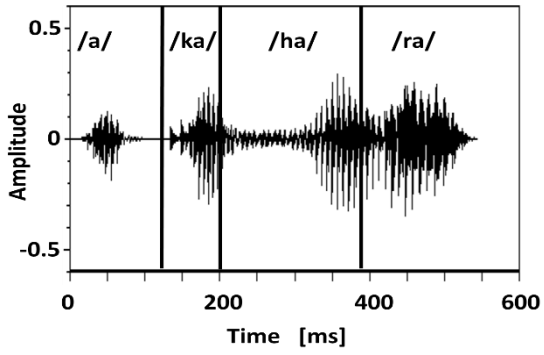


Fig. 2. Speech waveform after enhancement.

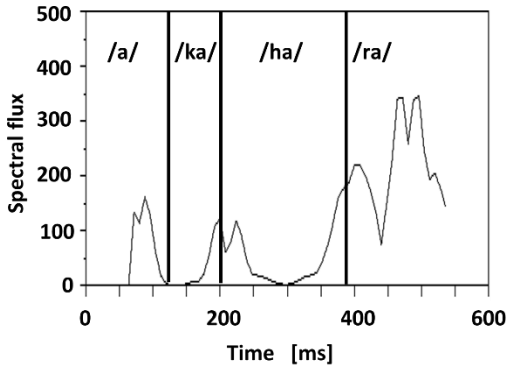


Fig. 3. Variation of spectral flux before enhancement .

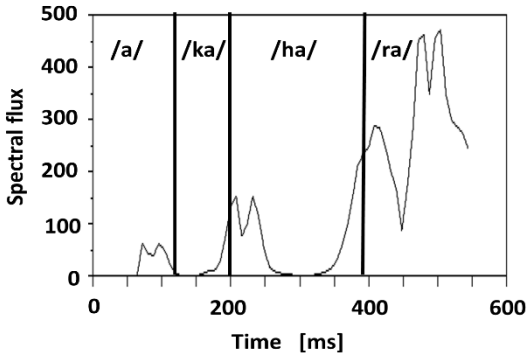


Fig. .4 Variation of spectral flux after enhancement.

2.3　ベイズ最適化による強調係数の設定方法

2.2 節において，提案法による音声スペクトルの時間変化強調の効果について確認することができた。しかし，式(1)において，α(j)を 1 より大きい値に設定することで帯域毎にスペクトルの時間変化を強調することができる反面，多次元関数となるため明瞭性改善に適した係数設定が困難となる。

そこで我々は，機械学習のハイパーパラメータ探索等に用いられているベイズ最適化[14]を用いて，STOI[7]の音質評価値が最大となるような強調係数の設定方法を提案する。ベイズ最適化とは，ブラックボックス関数の最大値または最小値となるパラメータを求める手法の一つである。また，STOI とは，聴取者の主観的な音質評価値を推定する客観評価法である。具体的には，クリーン音声と雑音環境下の音声を用いて，雑音環境下における音声の音質評価値を算出する。

次に，提案法における具体的な強調係数の設定方法について説明する。提案法における強調係数の設定方法を Fig.5 に示す。Fig.5 に示すように，原音声(Speech)と雑音(Noise)に対して，入力値(Input)が帯域毎の強調係数で，設定した強調係数に従って音声スペクトルの時間変化強調を施し，強調後の音声をクリーン音声，強調後の音声に雑音を重畳した音声を雑音環境下の音声として STOI により算出される音質評価値を出力値(Output)とする新たなブラックボックス関数を作成する。そして，作成し

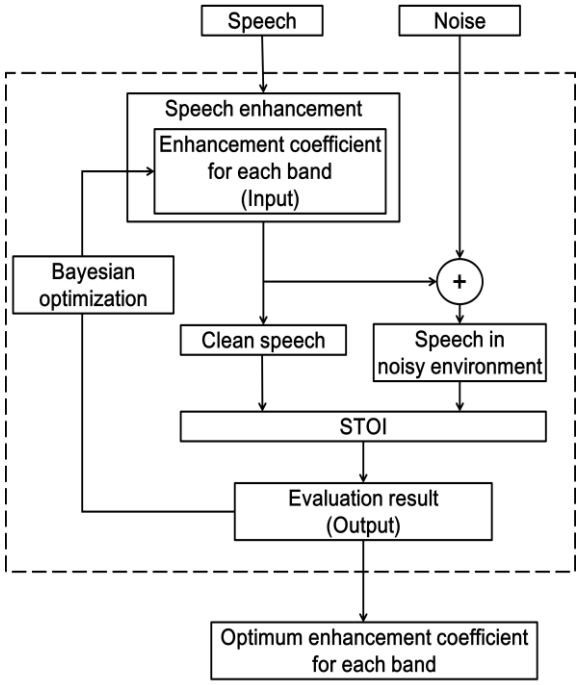


Fig. 5. Setting procedure of enhancement coefficient.

たブラックボックス関数についてベイズ最適化を用いて，STOI の音質評価値が最大となるような帯域毎の最適な強調係数を探索する。具体的には，ある強調係数に従って音声強調を施したときに算出される音質評価値の結果から，ベイズ最適化を用いて，これまでの結果から音質評価値が最大となるような強調係数を推定し、その強調係数に従って音声強調を施し，音質評価値を算出する。その過程を繰り返すことで，STOI の音質評価値が最大となるような帯域毎の最適な強調係数を探索し，設定することができる。

3. 明瞭性改善効果の検証

3.1 実験方法

提案法の雑音環境下における音声の明瞭性改善効果について，STOI の音質評価値を用いて評価した。評価対象の音声には，成人男性 5 名(21 歳~22 歳)が発話した親密度 1.0～2.5 の 4 モーラ単語 50 語[15]を用い，原音声と提案法により話者毎に最適化した強調係数を用いた強調音声（強調音声 1）と 5 名の最適化した強調係数の平均値を用いた強調音声（強調音声 2）を使用した。また，雑音には走行中の電車内騒音を使用した。強調係数設定のためのベイズ最適化は，SNR = -10 dB で，探索の試行回数を 200 回とした。また，音声と雑音の SNR = -5, -10, -15 dB になるように重畳し，音質評価値を算出した。

3.2 実験結果

5 名の話者(T1~T5)及び平均の強調係数の結果を Table 1 に示す。また，平均の音質評価値を Table 2 に，原音声の音質評価値と強調音声の音質評価値の間の有意差について，片側 t-検定による有意差検定の結果を Table 3 に示す。Table 1 より，T2 を除く全ての話者において，62.5~250 Hz の低域に比べて 250~4k Hz の中域～高域の強調係数の値が大きくなっていることが確認できる。Table 2 より，原音声に比べて，強調音声 1 と強調音声 2 どちらも全ての SNR 条件で音質評価値が向上した。Table 3 より，強調音声 1 と強調音声 2 どちらも SNR= -10, -15 dB において，有意差が危険率 5%で認められた。

Table 1. Results of enhancement coefficient.

(Hz)	T1	T2	T3	T4	T5	avg.
0~62.5	4.98	8.06	6.71	4.86	7.21	6.36
62.5~125	1.42	6.02	1.00	1.41	2.78	2.53
125-250	1.42	6.01	1.79	1.85	2.59	2.73
250~500	1.86	6.20	4.14	2.33	4.43	3.79
500~1k	3.68	4.27	3.21	3.22	5.62	4.00
1k~2k	4.95	4.22	4.23	4.07	6.71	4.83
2k~4k	4.51	4.96	11.8	4.64	6.77	6.53
4k~8k	1.18	4.61	1.31	3.26	2.84	2.64

Table 2. Evaluation results of STOI score (%).

SNR	原音声	強調音声 1	強調音声 2
-15dB	62.7±11.3	68.0±9.83	67.1±10.1
-10dB	73.0±9.71	75.9±8.29	75.3±8.62
-5dB	82.1±7.74	83.0±6.57	82.8±6.84

Table 3. Results of significant difference test.

SNR	強調音声 1	強調音声 2
-15dB	0.016	0.015
-10dB	0.015	0.037
-5dB	0.063	0.195

3.3 考察

Table 1 より，T2 を除く全ての話者において，62.5~250 Hz の低域に比べて 250~4k Hz の中域～高域の強調係数の値が大きくなっていることが確認できた。このことから，中域～高域のスペクトルの時間変化は，雑音環境下における音声の明瞭性への影響が大きいと考えられる。また，T2 の原音声は 500~8k Hz のスペクトルフラックスが 5 名の原音声の平均値の 2 倍以上大きかったため，その帯域の強調係数の値が小さくなったと考えられる。

Table 2 より，全ての SNR 条件において，原音声に比べて，強調音声 1 と強調音声 2 どちらも全ての SNR 条件で音質評価値が向上した。また，強調音声 1 の音質評価値は強調音声 2 よりも若干高い結果となったが，それらの差は小さかった。このことから，強調係数は話者毎に最適化した値を設定しなくても，最適化の平均値を用いることで改善効果が見込めることが示唆された。つまり，

電車のアナウンスシステムに提案法の処理を適用すると仮定したとき，発話者が既知の場合において，発話者に合わせて最適化した強調係数を用いて，音声強調することによる明瞭性改善効果が見込める。さらに，発話者が未知の場合においても，複数の発話者の最適化した強調係数の平均値を用いることで，音声強調することによる明瞭性改善効果が見込めて，発話者の人数を増やして強調係数の平均値を算出することで，より高い改善効果が得られることが期待される。

ところで，今回の実験では SNR = -10 dB で強調係数設定のためのベイズ最適化を行い，SNR = -5 dB においては有意差が認められなかった。そのため，SNR 条件に合わせた強調係数の設定や単語毎にベイズ最適化を行うなど音韻の特性を考慮した強調係数の設定が必要であると考えられる。

また，提案法による音声強調は，電車のアナウンス音声以外でも，ベイズ最適化に用いる雑音の種類を変えることで，電車内以外の雑音や残響のある環境下などにおける音声の明瞭性改善のために適用することができると考えられる。

4. おわりに

本研究では，雑音環境下における音声の明瞭性改善を目的とし，スペクトルフラックスを表す音声スペクトルの時間変化に着目し，主観的な音質評価指標である STOI の音質評価値を向上させるような音声スペクトルの時間変化強調方法を提案した。STOI による評価実験において，提案法による強調音声は原音声に比べて，全ての SNR 条件で音質評価値が向上し，SNR = -10, -15 dB において有意差が危険率 5%で認められた。今後は，SNR 条件や音韻の特性を考慮した強調係数の設定方法の検討や聴取者が高齢者での主観評価行う予定である。

文　献

(1) 村瀬，中村，飯田，“周囲騒音によるマスキングを考慮した音質制御方式”，日本音響学会講演論文集，pp.523-524，1997

(2) 高松，田中，中藤，“MAEQ 法の改良による拡声音の明瞭性改善効果の検討”，日本音響学会講演論文集，pp.837-838，2016

(3) Lie Lu et al., “Automatic mood detection and tracking of music audio signals”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, No. 1, pp.5-18, 2006

(4) 平賀，大石，武田，“主観評価に基づく楽曲間類似度算出モデル”，情報処理学会研究報告，pp.1-6，2009

(5) 吉田，“自然な音声合成実現に向けた音響的特徴の分析 “，日本工業大学研究報告，pp.129-132，2016

(6) 川原，中藤，“発話者の年齢や発話スタイルの異なる音声の音響的特徴の分析”，日本音響学会講演論文集，pp.351-352，2017

(7) C. Taal et al., “Short-Time Objective Intelligibility Measure for Time-Frequency Weighted Noisy Speech”, in Proc. ICASSP, pp.4214-4217, 2010

(8) 小泉 他,“聴感評点を向上させるための DNN 音源強調関数のブラックボックス最適化”, 日本音響学会講演論文集, pp.511-514, 2017

(9) 在塚, 禰寝,“周波数スペクトルの時間変化加工による音声強調”, 日本音響学会研究発表会講演論文集，pp.309-310，1995

(10) 柴田，坂野，板倉，“動的特徴に着目した音声分析合成音の明瞭性向上手法の提案 “，電子情報通信学会技術研究報告，pp.101-106，2010

(11) 荒井 他，“音声の定常部抑圧の残響に対する効果”，日本音響学会研究発表会講演論文集，PP.449-450，2001

(12) 安武，中島，“準実時間子音強調システム” 電子情報通信学会研究報告，pp.79-84，2005

(13) 小原，坂野，旭，“帯域ごとの動的特徴分析結果に基づいた音声の明瞭性向上手法の改良”，日本音響学会講演論文集，pp.273-274，2017

(14) Jacob R. Gardner et al., “Bayesian optimization with inequality constraints”, in Proc. ICML, pp.937-945, 2014

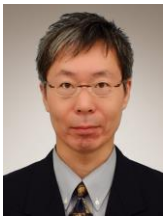
(15) 坂本 他，“親密度と音韻バランスを考慮した単語了解度試験用リストの構築”，日本音響学会誌，pp842-849，1998

川原　竣介

2016 年 3 月九州工業大学工学部電気電子工学科卒業。同年 4 月九州工業大学大学院に入学，現在に至る。音声情報処理の研究に従事。

平川　凜

2017 年 3 月九州工業大学工学部電気電子工学科卒業。同年 4 月九州工業大学大学院に入学，現在に至る。音声情報処理の研究に従事。



中藤　良久

1986 年 3 月信州大・工・電子卒．1991 年 3 月同大学院修士課程了。2007 年博士(工学)。1986～1989 年シャープ（株）勤務。1991～2010 年松下電器産業㈱(現パナソニック)勤務。2010 年九州工業大学大学院工学研究院・教授。音声認識，オーディオ符号化，補聴処理に関する研究開発に従事。産業技術総合研究所客員研究員。