加算性雑音に起因する音声の構造的表象の歪みに関する実験的検討*

◎村上隆夫, 朝川智, 峯松信明, 広瀬啓吉 (東大)

1 はじめに

音声には話者の声道形状の特性やマイクロフォンの特性などの非言語的特徴が不可避的に混入するが、近年、これらを表現する乗算性歪み・線形変換性歪みを原理的に保有しない音響的普遍構造が提案されている [1, 2]。これは構造音韻論の物理実装として、あるいは音声ゲシュタルトとして解釈される [1, 2]。音声に混入する非言語的特徴は加算性雑音・乗算性歪み・線形変換性歪みの 3 つに大別されるが、加算性雑音に対しては、音響的普遍構造はその形状が歪むと考えられる。しかし同時に、母音のスペクトル高域成分に多く含まれる話者性 [4] の情報を消失させる効果を持つことが予想される。本稿では、加算性雑音による構造の歪み、及びその際の話者性の消失についての定性的・定量的分析の結果を報告する。

2 音声の構造的表象 [1, 2]

音声における各事象が、ケプストラムによって構成される混合ガウス分布であると仮定する。この音声事象から任意の 2 事象の分布間距離を求め（距離行列）、音声を構造化することを考える。このとき、距離尺度としてバタチャリヤ距離の平方根を用いる。全てのケプストラム c に共通のアフィン変換 $Ac+b$ が施されても、バタチャリヤ距離は不変である。従って、構造も不変であり、これを音響的普遍構造と呼ぶ。A の乗算は線形変換性歪みに相当し、声道長・聴覚特性の差異が近似的にこれに含まれる。b は加算は乗算性歪みに相当し、マイクロフォンなどの伝送特性の差異がこの代表例である。声道形状の一部も近似的にこれに含まれる。A の乗算は構造の回転として、b の加算は構造のシフトとして観測される。

2 つの構造間差異として、距離行列をベクトルと見なしたときのユークリッド距離を用いる。これは、一方の構造を回転（A）・シフト（b）させた後の対応する点間距離の最小値を近似する尺度である [3]。

3 雑音による構造歪みに関する実験的検討

3.1 雑音による音声の構造的表象の歪み

加算性雑音が音響的普遍構造に与える影響について考えてみる。クリーンな音声、雑音、雑音下の音声のパワースペクトルをそれぞれ $|X(f)|^2$、$|S(f)|^2$、$|Y(f)|^2$ とし、

$$|Y(f)|^2 \approx |X(f)|^2 + |S(f)|^2 \qquad (1)$$

が成立すると仮定する。(1) 式は、対数パワースペクトル（$y(f) = \log|Y(f)|^2$）上では

$$y(f) \approx \log(\ \exp(x(f)) + \exp(n(f))\ ) \qquad (2)$$

と表される。従って、加算性雑音は、ケプストラムに対して非線形変換を施し、結果、音響的普遍構造はその形状が歪むものと予想される。Fig. 1 は男性話者の 5 母音を Ward 法によるボトムアップクラスタリングを用いて樹形図化したものである。このとき、分析条件は Table 1 のとおりで、MAP 推定を用いて音声事象分布を求めている。左はクリーンな音声の樹形図であり、右はこれに SNR=10[dB] の白色雑音を加えたものである。構造サイズは正規化してあるが、構造形状が雑音によって歪んでしまっている。特に/i/と/u/の距離が短くなっているが、これは/i/と/u/の第一フォルマントが互いに近傍にあり、他のフォルマントが雑音に埋もれてしまったためと考えられる。

Table 1 分析条件

サンプリング	16bit / 16kHz
窓	窓長 25msec、シフト長 10msec
パラメータ	FFTcep.（1〜12 次元）

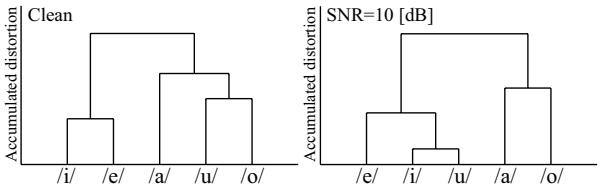

Fig. 1 男性話者の 5 母音の樹形図

3.2 雑音による話者性の消失に関する定性的分析

[4] は、母音のスペクトル包絡の約 2.2kHz 以上の高周波域成分には話者性が多く含まれていることを示している。[5] では、音声に LPF を施し、高周波域成分を抑制することで、日本語母音系列の認識性能を飛躍的に向上させた。しかし、ここでの目的は高周波域成分を音素間で揃えることであり、必ずしも下に揃える必要性はない。加算性雑音の重畳は、高周波域成分を上に揃える方法の一つである。Fig. 2

*Experimental study of distortion in the structural representation of speech caused by additive noise
By Takao Murakami, Satoshi Asakawa, Nobuaki Minematsu and Keikichi Hirose (University of Tokyo)

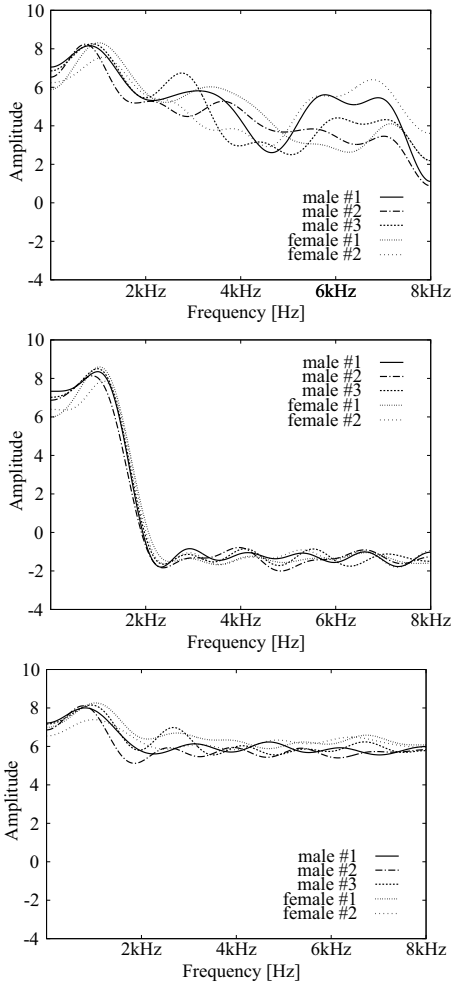

Fig. 2　5名話者の/a/のスペクトル包絡
(上：クリーン、中央：LPF、下：雑音)

は、5 名の話者が発声した/a/のスペクトル包絡である。上・中央・下の図は、それぞれクリーン音声・LPF（カットオフ周波数：2kHz）を施した音声・白色雑音（SNR=10[dB]）を重畳した音声に対応する。話者による違いが高周波域によく表れているが、加算性雑音を重畳することでそれが上に揃えられ、話者性の消失が効果的に行なわれていることが分かる。但し、Fig. 2 中央及び下の音声を聞いてみると、話者性は完全には消失されていないことが分かる。

3.3　雑音による話者性の消失に関する定量的分析

加算性雑音を重畳した母音はどれだけ話者性が消失されているのかを定量的に調べるために、雑音を付与した時の話者間構造差異・話者内構造差異の分散分析を行なった。用いた音声資料は、8 名話者（男性 4 名、女性 4 名）が 5 母音を 5 回発声したデータである。これに SNR $= \infty$（clean）, 20, 10, 0[dB] の白色雑音を重畳し、Table 1 に示す分析条件でケプストラムを求めた後、ML 推定を用いて分布化した。求めた音声事象分布から、各話者毎に 5 つの構造を抽出した。この際、全構造のサイズが等しくなるように正規化を施す場合と、施さない場合の 2 通りを試みた。抽出した構造から、話者間構造差異（${}_8C_2 \times 5^5 = 700$ 個）、及び話者内構造差異（$8 \times_5 C_2 = 80$ 個）を求め、これを分析対象とした。

Table 2　分析結果（構造サイズの正規化無し）

SNR[dB]	∞	20	10	0
話者間差異の平均	1.47	0.60	0.43	0.25
話者内差異の平均	0.99	0.36	0.25	0.17
構造サイズ	12.5	6.7	4.3	2.4

Table 3　分析結果（構造サイズの正規化有り）

SNR[dB]	∞	20	10	0
話者間差異の平均	1.17	0.81	0.85	0.95
話者内差異の平均	0.86	0.58	0.63	0.73
構造サイズ	12.5	12.5	12.5	12.5

分散分析の結果、危険率はどの場合においても $p < 0.001$ となった。これは、話者特有の発声の癖、方言差などの「構造の話者性」が消失し得ないことが原因と見られる。話者間差異の平均、話者内差異の平均を Table 2（正規化無し）、Table 3（正規化有り）に示す。Table 2 では、SNR の低下とともに構造サイズが小さくなり、その結果、話者間差異・話者内差異も減少している。しかし、構造サイズの正規化を施した Table 3 においても、雑音の重畳によって話者間差異・話者内差異が減少している。これは、雑音を重畳することで「音の話者性」の消失、及び話者内の発声の揺れ、即ち「構造揺らぎ」の抑制が行なわれている効果と見られる。但し、SNR の低下とともに話者間差異・話者内差異が増加している。これは低 SNR においては、音韻差異が不明瞭に（構造サイズが非常に小さく）なり、結果として構造形状が不安定になったことが原因と見られる。

4　まとめ

加算性雑音による母音構造の歪み、及びその際の話者性の消失に関する定性的・定量的分析を行なった。結果、加算性雑音を重畳することで、「音の話者性」の消失、及び話者内の「構造揺らぎ」の抑制が行なわれることが示された。音声の構造的表象を用いた加算性雑音音声認識については [6] を参照して戴きたい。

参考文献

[1] N.Minematsu, Proc. ICASSP, pp.889-892 (2005)
[2] 峯松他，信学技報，SP2005-12，pp.1-8 (2005)
[3] 峯松他，信学技報，SP2005-13，pp.9-12 (2005)
[4] T.Kitamura et al., JASJ(E) Vol.16, No.5 (1995)
[5] 村上他，信学技報，SP2005-14，pp.13-18 (2005)
[6] 村上他，音講論，1-P-2 (2005-9)