社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

雑音環境下における明瞭性に着目した 非可聴つぶやき強調処理における目標音声の評価

鶴田　さくら† 　田中　　宏† 　戸田　智基† 　Graham　Neubig† 　Sakriani　Sakti†
中村　　哲†

† 奈良先端科学技術大学院大学情報科学研究科　〒 630–0101 奈良県生駒市高山町 8916–5
E-mail: †{tsuruta.sakura.tl0,ko-t,tomoki,ssaktii,neubig,s-nakamura}@is.naist.jp

あらまし　非可聴つぶやき（Non-Audible Murmur: NAM）は，NAM マイクと呼ばれる専用の体表密着型マイクを用いて，体表から直接収録される体内伝導音声のひとつである．NAM は秘匿性の高い発話を可能とするため，サイレント音声通話への応用が期待されている．しかし，その音響特徴量は，通常の空気伝導音声のものとは大きく異なるため，明瞭性及び自然性が大きく劣化する．この問題に対処するため，統計的手法に基づき NAM から通常音声及びささやき声へと変換する NAM 強調法が提案されている．これまでに，遮音室のような静穏環境下での受聴評価において，NAM 強調法により NAM の品質を改善可能であること，また，通常音声よりもささやき声への変換の方が有効であることが報告されている．しかし，実環境下では，受聴時の環境は静穏環境下に限定されないため，同様の結果が得られるとは限らない．本研究では，NAM 強調技術の実環境への適用の第一歩として，受聴者が雑音環境下，発話者が静穏環境下にいる状況を想定し，明瞭性に着目して最適な変換目標音声の調査を行う．実験結果から，変換音声の有声化が明瞭性の向上に寄与することを示す．
キーワード　サイレント音声インターフェース，非可聴つぶやき，声質変換，雑音下での受聴，明瞭性

An Evaluation of Target Speech for Nonaudible Murmur Enhancement Focusing on Intelligibility under Noisy Environments

Sakura TSURUTA†, Kou TANAKA†, Tomoki TODA†, Graham NEUBIG†, Sakriani SAKTI†, and Satoshi NAKAMURA†

† Nara Institute of Science and Technology,　Tatayama-cho 8916–5, Ikoma, Nara, 630–0192 Japan
E-mail: †{tsuruta.sakura.tl0,ko-t,tomoki,ssaktii,neubig,s-nakamura}@is.naist.jp

Abstract　Nonaudible murmur (NAM) is a soft whispered voice recorded with a NAM microphone through body conduction. NAM is effective for silent speech communication as it allows a speaker to speak without making an audible sound. However, its intelligibility and naturalness are significantly degraded compared to natural speech owing to acoustic changes caused by body conduction. To address this issue, statistical voice conversion methods from NAM to normal speech (NAM-to-Speech) and to a whispered voice (NAM-to-Whisper) have been proposed. It has been reported that these NAM enhancement methods significantly improve speech quality and intelligibility of NAM, and that NAM-to-Whisper is more effective than NAM-to-Speech. However, it is still not obvious which method is more effective if a listener listens to the enhanced speech in noisy environments, a situation that often happens in silent speech communication. In this report, we assume a situation where NAM is uttered by a speaker in a quiet environment and conveyed to a listener in a noisy environment, and investigate what kinds of target speech is more effective for NAM enhancement. We also propose NAM enhancement methods for converting NAM to other types of target voiced speech. Experimental results show that the conversion process into voiced speech is more effective than that into unvoiced speech for generating more intelligible speech in noisy environments.
Key words　Silent speech interface, nonaudible murmur, voice conversion, listening in noisy conditions, intelligibility

1. まえがき

音声コミュニケーションは，私達の生活における基本的なコミュニケーション手段の 1 つであり，携帯電話やコミュニケーションソフトウェアなどの技術の発展によって，場所や時間を問わず，誰でも気軽に使用することが可能となった．しかしながら，発声を躊躇するような状況は未だ存在している．例えば，話し声が周囲の人の迷惑となってしまうような静穏環境下（図書館やオフィスなど）での会話をしたい場面や，大きな声を出さなければ相手に聞こえないような雑音のある人混みの中で，暗証番号のような秘匿性の高い情報を相手に伝えたい時などが挙げられる．

近年，周囲に声を漏らさずに発話を可能とするサイレント音声インターフェースが新しい音声コミュニケーション方法の 1 つとして注目を集めている [1]．サイレント音声インタフェースにおける入力情報の 1 つとして，周囲の人が聞き取りづらい発話を可能とする非可聴つぶやき（NonAudible Murmur: NAM）が提案されている [2]．NAM は，周囲の人が受聴困難なほど微小な音声信号であり，声帯振動を伴わずに発声される．NAM マイクと呼ばれる専用の体表密着型マイクを，図 1 に示すように耳介後方に圧着させることによって，体内伝導音声として収録することが可能である．NAM は，周囲に迷惑を掛けない秘匿性の高いコミュニケーションを可能とするという利点があるため，サイレント音声通話への応用が期待されている．しかし，口唇での放射特性の欠如や体内伝導によるローパス特性の影響により [3]，NAM の音響特徴量は通常の空気伝導音声と大きく異なる．そのため，自然性及び明瞭性が劣化するという欠点がある．

この問題に対処するため，統計的手法に基づく NAM 強調法が提案されている [4]．この手法は，統計的声質変換 [5] [6] の技術を用いて，NAM の音響特徴量を通常音声（NAM-to-Speech）又はささやき声（NAM-to-Whisper）などの空気伝導音声の音声特徴量へと変換するという枠組みである．文献 [4] の中で，NAM-to-Whisper に基づく強調音声は，NAM-to-Speech に基づく強調音声よりも高い明瞭性を有すると報告されている．その理由としては，1）NAM-to-Whisper では，NAM-to-Speech とは異なり，無声音である NAM からの F_0 パターン及び有声/無声情報予測といった本質的に困難な処理を回避できること，2) ささやき声は NAM と同じ無声音であるため，通常音声と比較すると，より NAM に近いスペクトル特徴量を持っているため，変換精度が高いこと，などが挙げられる．しかし，音声の受聴評価が遮音室のような静穏環境下でしか実施されておらず，NAM が実際に使用されると考えられる雑音のある実環境での NAM 強調法の効果が明らかになっていない．

本研究では，NAM 強調技術の実環境への適用を目指し，雑音環境下における NAM の明瞭性改善に最適な目標音声の評価を行う．なお，周囲に迷惑をかけずに発話可能である NAM の特性上，本技術の典型的な使用場面として，NAM の発話者は静穏環境下にいて，受聴者は雑音環境下にいる場合（例えば，図書館の外からかかってきた電話を図書館の中で受け取るなど）が想定される．そこで，本研究では，特に受聴側が外部雑音の影響を受けると想定して実験を行う．実験結果から，NAM 強調法において，有声音声への変換処理が，受聴時における外部雑音の影響を低減させ，雑音環境下で明瞭性向上に寄与することを示す．

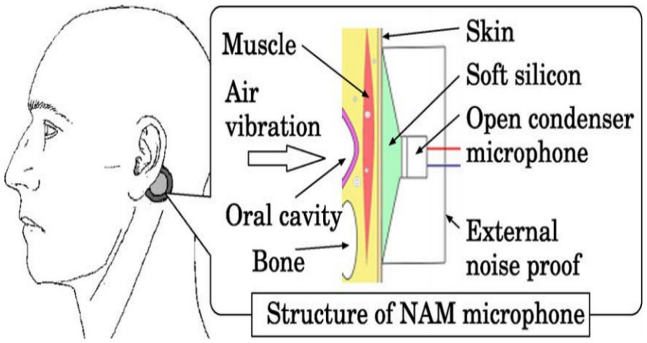

図 1　NAM マイクの構造と圧着位置

2. 統計的手法に基づく NAM 強調法

統計的手法に基づく NAM 強調法では，通常音声やささやき声といったより明瞭性の高い自然音声を目標音声として，NAM の音響特徴量を目標音声の音声特徴量へと変換することで，NAM の品質を改善する．

本手法は，学習処理と変換処理で構成される．学習処理では，NAM と目標音声の同一発話内容（パラレルデータ）を用いて，変換モデルを学習する．時間フレーム t において，前後 C フレームから計算される入力セグメント特徴量を $\mathbf{X}_t$ とし，出力静的・動的特徴量を $\mathbf{Y}_t = \left[\mathbf{y}_t^T, \Delta\mathbf{y}_t^T\right]^T$ とする．パラレルデータに対して動的時間伸縮を行い，対応付けを行った結合ベクトル $\left[\mathbf{X}_t^T, \mathbf{Y}_t^T\right]^T$ を用いて，次式に示すとおり，結合確率密度関数を混合正規分布モデル (Gaussian mixture model: GMM) でモデル化する [7]．

$$\mathbf{P}\left(\mathbf{X}_t, \mathbf{Y}_t | \boldsymbol{\lambda}^{(X,Y)}\right) = \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\left[\mathbf{X}_t^T, \mathbf{Y}_t^T\right]; \boldsymbol{\mu}_m^{(X,Y)}, \boldsymbol{\Sigma}_m^{(X,Y)}\right) \quad (1)$$

ここで，$\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は平均ベクトル $\boldsymbol{\mu}$ および共分散行列 $\boldsymbol{\Sigma}$ を持つ正規分布である．混合数 M の GMM のモデルパラメータセット $\boldsymbol{\lambda}^{(X,Y)}$ は，各分布 m の混合重み α_m，平均ベクトル $\boldsymbol{\mu}_m^{(X,Y)}$ および共分散行列 $\boldsymbol{\Sigma}_m^{(X,Y)}$ で構成される．変換処理では，最尤系列変換法 [6] により，NAM の音響特徴量系列を目標音声の音響特徴量系列へと変換し，音声波形を合成することで，強調音声を得る．

2.1 NAM から通常音声への変換

統計的手法に基づく NAM から通常音声への変換（NAM-to-Speech）[4] では，図 2 に示す通り，NAM のスペクトルセグメント特徴量（メルケプストラムセグメント）[8] を入力とし，

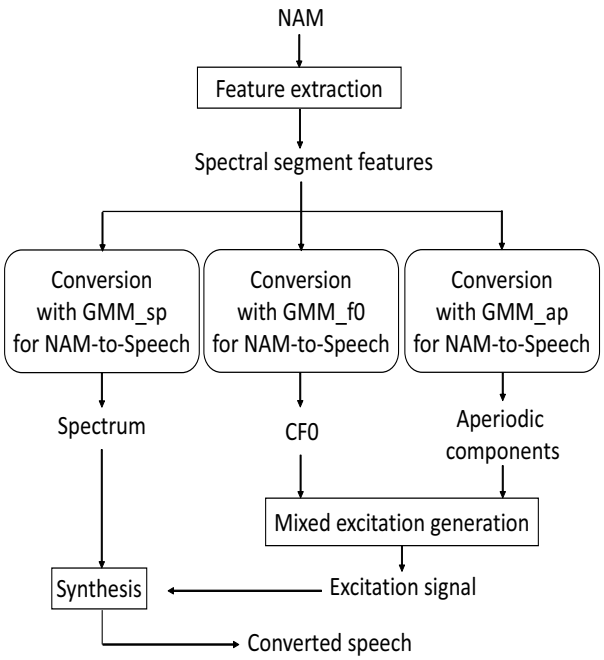

図 2　NAM-to-Speech の変換処理

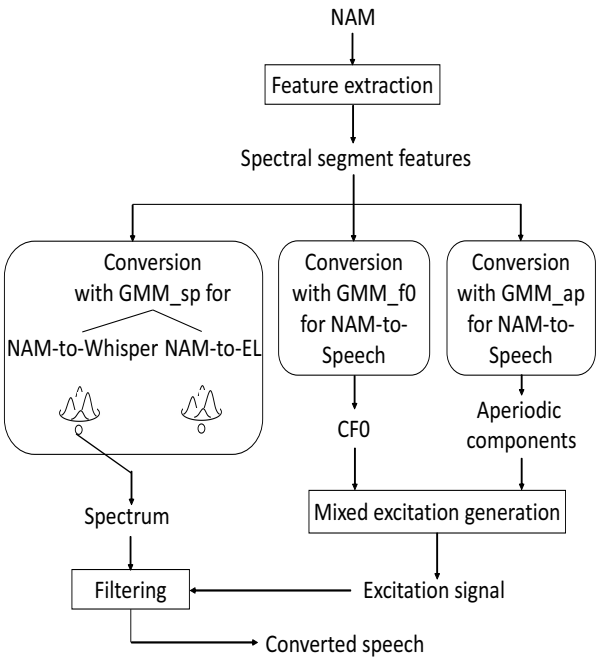

図 3　提案手法の変換処理

通常音声のスペクトル特徴量（メルケプストラム），有声無声情報を含む F_0，及び非周期成分 [9] の各々を出力とする計 3 つの GMM を学習し，変換時には個々の GMM を用いて各音響特徴量への変換を行う．変換後の F_0 および非周期成分を用いて混合励振源モデル [10] により音源信号を生成する．その後，音源信号に対して，変換後のスペクトル特徴量を畳み込むことで，強調音声を得る．

本手法により得られた強調音声は，NAM に比べて明瞭性が改善されており，通常音声と類似した音質を持つ．一方で，無声音である NAM のスペクトル特徴量から通常音声の有声/無声情報および F_0 パターンを推定することは本質的に困難である．結果，生成される強調音声は不自然な抑揚を持つ．

2.2　NAM からささやき声への変換

上記の問題を回避するため，NAM のスペクトルセグメント特徴量から，無声音であるささやき声への変換（NAM-to-Whisper) [4] が提案されている．ささやき声は，無声音であるが，NAM と比較して，馴染みのある空気伝導音声であるため，高い明瞭性及び自然性を有する．本手法では，NAM のスペクトルセグメント特徴量（メルケプストラムセグメント）を入力とし，ささやき声のスペクトル特徴量（メルケプストラム）を出力とする GMM を学習して，変換を行う．雑音源に対して，変換後のスペクトル特徴量を畳み込むことで，強調音声を得る．

本手法では，通常音声への変換とは異なり，有声/無声情報および F_0 パターン変換処理を回避可能である．また，無声音同士の変換であるため，スペクトル特徴量が通常音声よりも比較的類似しており，通常音声への変換と比較すると，より高いスペクトル特徴量変換精度が得られる．結果，NAM と比べて，強調音声の自然性および明瞭性は大幅に改善される．遮音室における実験的評価の結果から，通常音声への変換よりも高い性能が得られることが報告されている [4]．

3.　雑音環境下での受聴に適した変換目標音声

NAM は特別な発話様式であり，発話が周囲の迷惑となる図書館などの静穏環境下など，話者が NAM を使用し得る環境は比較的限られる．一方で，通話相手である受聴者の環境は限定されないため，雑音環境下で強調音声を受聴する状況が起こり得る．そのため，雑音環境下においても明瞭性の高い強調音声の実現が望まれる．なお，NAM を用いたサイレント音声通話では，相手が何を話しているかを理解出来るのが最も重要な要因であるため，本研究では，自然性よりも明瞭性や聞き取りやすさの改善に優先して取り組む．

本研究では，雑音環境下での受聴を想定して，明瞭性に着目した最適な目標音声の調査を行う．調査対象とする目標音声は，従来用いられていた通常音声およびささやき声に加え，有声化したささやき声および電気音声（EL: Electro laryngeal speech）[11] の 2 つも新たに検討する．

3.1　有声化ささやき声への変換

NAM-to-Whisper に基づく強調音声は，NAM-to-Speech に基づく強調音声よりも高いスペクトル変換精度を有することが文献 [4] の中で報告されている．一方で，雑音環境下において音声を受聴する際に，F_0 は一つの大きな手掛かりになると考えられるため，無声音であるささやき声に変換する NAM-to-Whisper に基づく強調音声は，雑音環境下における明瞭性が大幅に劣化する可能性がある．そこで，変換精度の高いささやき声のスペクトル特徴量と通常音声の音源特徴量を組み合わせることで，有声化ささやき声への変換を提案する．

有声化ささやき声への変換処理では，図 3 に示すように，

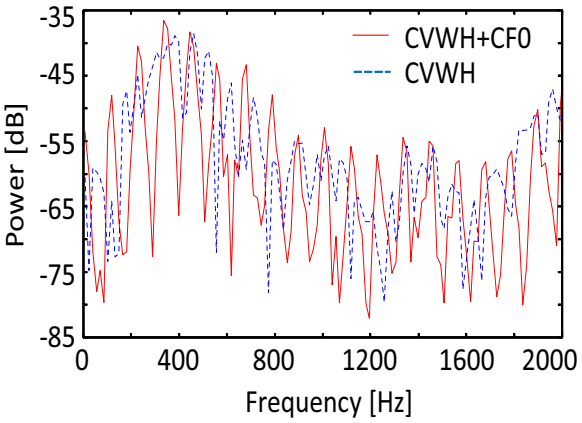

図 4　ささやき声への変換音声（CVWH）と有声化ささやき声への変換音声（CVWH+CF_0）のスペクトル

NAM のスペクトルセグメント特徴量から，ささやき声のスペクトル特徴量，通常音声の音源特徴量への変換を行う．音源特徴量として，連続 F_0 パターン（CF_0）[11] と非周期成分を用いる．これらの通常音声の音源特徴量を用いて混合励振源を生成し，変換後のささやき声スペクトル特徴量を畳み込むことで，変換有声化ささやき声（CVWH+CF_0）を生成する．

図 4 に，変換ささやき声と CVWH+CF_0 のスペクトルの一例を示す．なお，両変換音声の波形パワーは等しい．有声化スペクトルは調波構造を持つため，調波成分に対応する周波数成分のパワーが大きくなることが分かる．雑音環境下においては，有声化により，知覚に重要な周波数成分のパワーが相対的に大きくなるという効果が期待される．

3.2　電気音声への変換

CVWH+CF_0 は，無声音であるささやき声のスペクトルと有声音である通常音声の音源特徴量の組み合わせであるため，スペクトル特徴量と音源特徴量との間で有声無声情報の不一致が生じ，通常の音声生成過程では発声不可能な音声が生成される．この問題を回避するため，有声音であり，ささやき声への変換音声（CVWH）や CVWH+CF_0 と同様に有声無声情報を一切必要としない変換目標音声として，EL の使用を提案する．EL は，喉頭摘出者の代替発声器具の一つである電気式人工喉頭を用いて生成される音声であり，機械的に生成される有声音源信号を調音することで得られる有声音声である．EL は比較的高い明瞭性を有する一方で，自然な F_0 パターンの機械的な生成は本質的に困難であるため，自然性が大きく劣化する [12]．そのため，提案法では，NAM のスペクトルセグメント特徴量から，EL のスペクトル特徴量と，通常音声の音源特徴量である CF_0 及び非周期成分への変換を行うことで，より自然な F_0 パターンを持つ強調音声を生成する．これらの通常音声の音源特徴量を用いて混合励振源により音源波形を生成し，EL のスペクトル特徴量を畳み込むことで，変換音声（CVEL+CF_0）を生成する．なお，EL として，NAM の発話者と同一話者が電気式人工喉頭を用いて発声したものを用いる．

4.　実験的評価

各音響特徴量の変換精度に関する客観評価実験，および，聞き取りやすさ及び明瞭性に関する主観評価実験を行う．

4.1　実 験 条 件

男性話者 2 名と女性話者 1 名の通常音声（SP）とささやき声（WH），および，男性話者 1 名の EL を，各々空気伝導マイクを用いて収録する．また，同一話者の NAM を，NAM マイクと空気伝導マイクの両方を用いて同時に収録する．収録文は ATR 音素バランス文セット [13] 中の 50 文とする．また，親密度別単語了解度試験用音声データセット 2007 [14] 中の親密度 1 の単語 8 セット（計 160 単語）も収録する．客観評価実験および聞き取りやすさに関する主観評価実験では，学習データに ATR 音素バランス文 A セット中の 50 文中 40 文を用い，評価データに残りの 10 文を用いる．明瞭性に関する主観評価実験では，学習データに ATR 音素バランス文 A セット 50 文を用い，評価データに親密度別単語了解度音声データセット中の 160 単語を用いる．サンプリング周波数は 16 kHz とする．

NAM の音響特徴量として，FFT 分析による 0〜24 次のメルケプストラムセグメント特徴量（前後 4 フレーム相当）を用いる．分析フレームシフト長は 5 ms，通常音声及び EL のスペクトル分析には STRAIGHT 分析 [15] を用いる．スペクトル変換のために，1) 通常音声への変換用（GMM_SP），2) ささやき声への変換用（GMM_WH），3) 電気音声への変換用（GMM_EL）の計 3 つの GMM を学習する．また，音源特徴量変換のためにに，1) 通常音声の F_0 への変換用（GMM_F_0），2) 通常音声の CF_0 への変換用（GMM_CF_0），3) 通常音声の非周期成分への変換用（GMM_AP）の計 3 つの GMM を同様に学習する．GMM の混合数は 64（スペクトル変換用），32（F_0，CF_0 変換用），16（非周期成分変換用）とする．なお，文献 [4] とは異なり，NAM と通常音声においてフレーム間の対応付けを行う際には，空気伝導マイクで同期収録された NAM を用いる．強調音声として，従来の変換通常音声（CVSP）と変換ささやき声（CVWH），および，本報告で提案する CVWH+CF_0 と CVEL+CF_0 を合成する．

4.2　客観評価実験

EL を収録した男性話者 1 名に対して，各強調音声生成時における音響特徴量の推定精度を表 1，2 に示す．なお，メルケプストラム歪みの計算時には，0 次項を含めていない．表 1 に示す

表 1　スペクトル特徴量変換精度

	GMM_SP	GMM_WH	GMM_EL
Mel-cepstral distortion	4.2 dB	4.5 dB	5.4 dB

表 2　音源特徴量変換精度

	GMM_F_0	GMM_CF_0	GMM_AP
Aperiodic distortion	N/A	N/A	3.6 dB
U/V error rate	8.9 %	13.5 %	N/A
F_0 correlation coefficient	0.43	0.53	N/A

スペクトル特徴量変換精度の結果から，通常音声（GMM_SP）およびささやき声（GMM_WH）のスペクトル特徴量への変換精度と比較し，電気音声（GMM_EL）のスペクトル特徴量への変換精度が低いことが分かる．また，表 2 に示す音源特徴量変換精度から，連続 F_0 パターン（GMM_CF_0）を用いることで，有声／無声情報の変換精度は当然劣化するものの，F_0 パターンの相関係数は大きく改善することが分かる．

なお，表 1 において，通常音声のスペクトル特徴量への変換は，ささやき声のスペクトル特徴量への変換よりも高い精度が得られており，文献 [4] とは異なる傾向を示している．この原因を調べるために，もう 1 名の男性話者と女性話者 1 名に対するスペクトル特徴量変換精度を調査する．結果を表 3 および表 4 に示す．どちらの話者においても，ささやき声のスペクトル特徴量への変換（GMM_WH）の方が，通常音声のスペクトル特徴量への変換（GMM_SP）よりも精度が高いことが分かる．以上の結果から，ささやき声のスペクトル特徴量への変換は，必ずしも通常音声のスペクトル特徴量への変換よりも高い精度が得られるわけではなく，話者に依存することが分かる．

4.3 聞き取りやすさに関する主観評価実験

EL を収録した男性話者 1 名のデータを用いて，聞き取りやすさに関する主観評価実験を行う．評価は，音声の聞き取りやすさに関する 5 段階オピニオン評定により行う．評価対象音声は以下の 8 種類である．

- NAM
- WH: ささやき声
- EL: 電気音声
- SP: 通常音声
- CVSP: 変換通常音声
- CVWH: 変換ささやき声
- CVWH+CF_0: 変換ささやき声＋ CF_0
- CVEL+CF_0: 変換電気音声＋ CF_0

受聴環境として，雑音なしの静穏環境とオフィス雑音を SNR=0 dB 及び-10 dB で重畳した場合の雑音環境の計 3 種類を想定する．被験者は 12 名で，1 人あたり各環境，音声につき 10 サンプル，計 240 サンプルを受聴する．

図 5 に各環境下における聞き取りやすさに関する評価結果を示す．自然音声については，全ての環境において SP の聞き取りやすさが最も高く，NAM の聞き取りやすさは最も低い．また，静穏環境下においては EL よりも WH の方が聞き取りやすいのに対し，雑音環境である SNR=0 dB 及び-10 dB ではその差がほぼなくなる．この結果から，聞き取りやすさに関して，無声音は有声音よりも外部雑音の影響を受けやすいことが分かる．

表 3　スペクトル特徴量変換精度（男性話者）

	GMM_SP	GMM_WH
Mel-cepstral distortion	5.2 dB	5.0 dB

表 4　スペクトル特徴量変換精度（女性話者）

	GMM_SP	GMM_WH
Mel-cepstral distortion	5.7 dB	5.2 dB

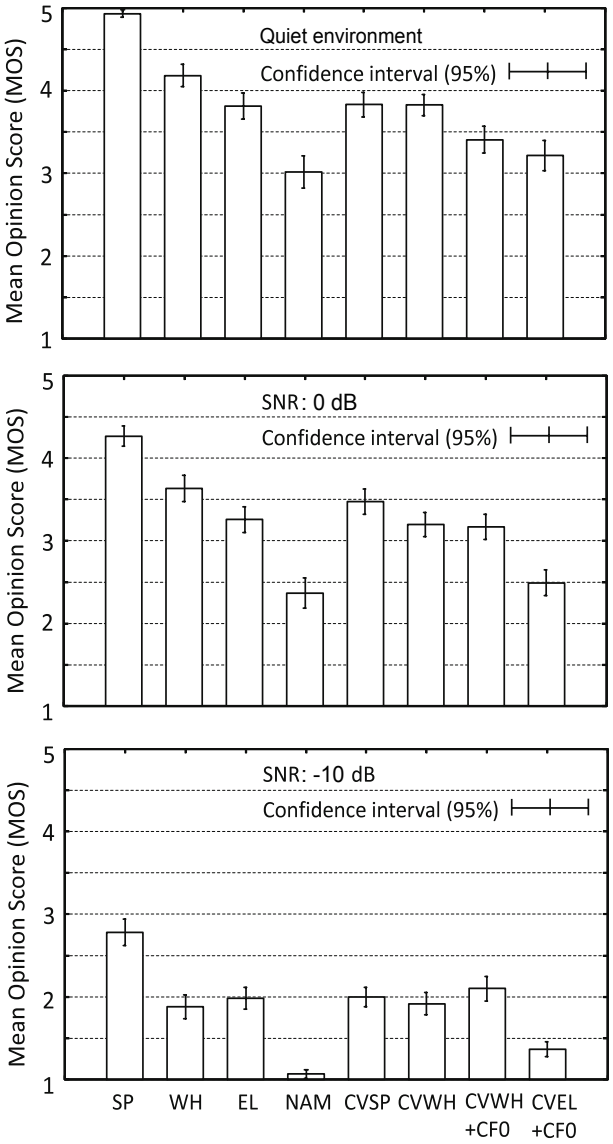

図 5　聞き取りやすさに関する主観評価実験結果（上図が静穏環境，中図が SNR ＝ 0 dB，下図が SNR ＝ -10 dB における聞き取りやすさに関する主観評価結果）

一方で，NAM と各種強調音声を比べると，全ての環境において NAM 強調法により NAM の聞き取りやすさを改善出来ていることが分かる．変換音声については，CVWH と CVWH+CF_0 を比べると，静穏環境下では CVWH の方が聞き取りやすいのに対し，SNR=0 dB ではその差がほぼなくなり，最も雑音の大きい SNR=-10 dB では結果が逆転する傾向にあることが分かる．このことから，ささやき声を有声化することで，外部雑音に対する頑健性を高めることが可能であると言える．一方で，CVEL+CF_0 は，全ての環境において他の強調音声と比べて聞き取りづらいことが分かる．これは，表 1 に示した通り，スペクトル変換精度が低いことが主な原因と考えられる．なお，CVSP は CVWH よりも同等か上回る値を示している．

表 5　書き取り試験結果

	Mora correct rate (Quiet environment)	Mora correct rate (SNR = 0 dB)
NAM	55.0%	42.5%
CVSP	67.8%	63.1%
CVWH	61.6%	55.0%
CVWH+CF_0	65.3%	62.5%

4.4　明瞭性に関する主観評価実験

聞き取りやすさに関する主観評価実験と同じ男性話者 1 名の音声データを用いて，明瞭性に関する書き取り試験を行う．使用する単語は，親密度別単語了解度試験用音声データセット 2007 中の親密度 1 の単語セットで，これらは全て 4 モーラで構成されている．評価音声は，変換音声から CVEL+CF_0 を除いた以下の 4 種類である．

- NAM
- CVSP: 変換通常音声
- CVWH: 変換ささやき声
- CVWH+CF_0: 変換ささやき声＋ CF_0

受聴環境は，雑音なしの静穏環境と，オフィス雑音を SN 比 0 dB で加えた雑音環境の計 2 種類とする．被験者は 4 名で，1 人あたり各環境，音声につき 20 サンプル，計 160 サンプルを受聴する．

表 3 に明瞭性に関する書き取り試験結果として，モーラ正解率を示す．まず，NAM と各種強調音声を比較すると，どちらの環境においても NAM 強調法により NAM の明瞭性を改善出来ることが分かる．CVWH は雑音環境下において明瞭性がかなり劣化しているのに対し，CVSP と CVWH+CF_0 は外部雑音による明瞭性の劣化は小さい．このことから，雑音環境下において，有声音声への変換は無声音声への変換よりも有効であることが分かる．また，CVWH と CVWH+CF_0 の比較から，静穏環境下においても，ささやき声を有声化することで，明瞭性を改善出来る傾向が見られる．

なお，文献 [4] の報告と異なり，CVSP が CVWH よりも明瞭性が高いという結果が見られる．この原因として，表 1 で示した通り，スペクトル変換精度が GMM_WH よりも GMM_SP の方が高いこと，文献 [4] では新聞記事の文を評価に用いているのに対し，本実験においては低親密度の単語を用いたこと，などが考えられる．今後，他の話者に対しても，同様の主観評価実験を行うことで，より汎用的な結果を得る必要がある．

5. おわりに

本報告では，NAM 強調技術の実環境への適用を目指し，雑音環境下における受聴を想定した際において，最適な変換目標音声の評価を行った．主観評価実験の結果から，通常音声や有声化されたささやき声のような有声音は，ささやき声のような無声音に比べて，聞き取りやすさ及び明瞭性の面において，外部雑音に対して頑健であることが分かった．今後は，より明瞭性の高い変換目標音声に対するさらなる調査を行うとともに，雑音環境下における通常音声に対する明瞭性改善技術を NAM 強調システムに導入する予定である．

謝辞　本研究の一部は，JSPS 科研費 22680060 および 24300073 の助成を受け実施したものである．

文　献

[1] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, “Silent speech interfaces,” *Speech Commun,* vol. 52, no. 4, pp. 270-287, 2010.

[2] Y. Nakajima, H. Kashioka, N. Cambell, and K. Shikano, “Non-audible murmur(NAM) recognition,” *IEICE Trans. Inf. Syst.,* vol. E89-D, no. 1, pp. 1-8, 2006.

[3] T. Hirahara, M .Otani, S. Shimizu, T. Toda, K .Nakamura, Y .Nakajima, and K. Shikano, “Silent-speech enhancement using body-conducted vocal-tract resonance signals,” *Speech Commun.,* vol. 52, no. 4, pp. 301-313, 2010.

[4] T. Toda, M. Nakagiri, K. Shikano, “Statistical voice conversion techniques for body-conducted unvoiced speech enhancement,” *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 20, no. 9, pp. 2505-2517, Sep. 2012.

[5] Y. Stylianou, O. Cappe, and E. Moulines, “Continuous probabilistic transform for voice conversion,” *IEEE Trans. Speech Audio Process.,* vol. 6, no. 2, pp. 131-142, Mar. 1998.

[6] T. Toda, A. W. Black, and K. Tokuda, “Voice conversion based on maximum likelihood estimation of spectral parameter trafectory,” *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 15, no. 8, pp. 2222-2235, Nov. 2007.

[7] A. Kain and M.W. Macon, “Spectral voice conversion for text-to-speech synthesis,” *in Proc. ICASSP,* Seattle, WA, pp. 285-288, May.1998

[8] T. Toda, A.W. Black, K. Tokuda, “Statistical mapping between articulatory movements and acoustic spectrum using a Gaussian mixture model,” *Speech Commun.,* vol. 50, no. 3, pp. 215-227, 2008.

[9] H. Kawahara, J. Estill, and O. Fujimura, “Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT,” in *Proc. MAVEBA,* Firenze, Italy, pp. 2266-2269, Sep. 2006.

[10] 大谷 大和，戸田 智基，猿渡 洋，鹿野 清宏，“STRAIGHT 混合励振源を用いた混合正規分布モデルに基づく最尤声質変換法，” 信学論，Vol. J91-D，No. 4，pp. 1082-1091，2008.

[11] K. Tanaka, T. Toda, G. Neubig, S. Sakti, S. Nakamura, “A Hybrid approach to electrolaryngeal speech enhancement based on noise reduction and statistical excitation generation,” *IEICE Transactions on Information and Systems,* Vol. E97-D, No. 6, pp. 1429-1437, Jun. 2014.

[12] H. Doi, T. Toda, K. Nakamura, H. Saruwatari, K. Shikano, “Alaryngeal speech enhancement based on one-to-many eigenvoice conversion.,” *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 22, no. 1, pp. 172-183, Jan. 2014.

[13] Y. Sagisaka, K. Takeda, M. Abe, S. katagiri, T. Umeda, and H. Kuwabata, “A large-scale Japanese speech database,” in*Proc. ICSLP90,* Kobe, Japan, Nov. 1990, pp.1089-1092.

[14] 近藤 公久，坂本 修 一，天野 成昭，鈴木 陽 一，“信号対雑音比調整による単語リスト間の単語了解度差補正：親密度別単語了解度試験用音声データセット (FW07) を用いた検証，” 日本音響学会誌，Vol. 69，no. 5，pp. 224-231，2013.

[15] H. Kawahara, I .Masuda-Katsuse, and A. de Cheveigne, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F_0 extraction: Possible role of a repetitive structure in sounds,” *Speech commun,* vol. 27, no. 3-4, pp. 187-207, 1999.