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話者照合

鎌田　敏明[†‡]　峯松　信明[‡]　長内　隆[†]　蒔苗　久則[†]　谷本　益巳[†]

†科学警察研究所　〒277−0882 千葉県柏市柏の葉 6−3−1
‡東京大学大学院新領域創成科学研究科　〒277−8561 千葉県柏市柏の葉 5−1−5
E-mail: †{kamada, osanai, makinae, tanimoto}@nrips.go.jp, ‡{kamada, mine}@gavo.t.u-tokyo.ac.jp

あらまし　法科学における電話音声を対象とした話者照合では，一般的な研究とは異なる雑音環境下における異同識別が求められることがある．異なる雑音環境とは SNR が 0dB 以下のような発声内容が聞き取れない劣悪な雑音環境のことであり，このような状況では第一に雑音を除去するための音声明瞭化処理が行われるが，明瞭化処理による個人性の変化に伴う話者照合の可否について明確な研究はこれまでほとんど行われていなかった．そこで本稿では，雑音環境下における音声に対して，音声明瞭化処理を伴う話者照合実験を行い，明瞭化処理と照合結果の関係について検討を行った．また電源雑音のような実在する雑音を使用した実験を行い，実環境下における話者照合精度について調べた．

キーワード　話者照合，テキスト依存型，音声明瞭化，電話音声，法科学

Speaker verification in noisy environment

T. KAMADA[†‡], N. MINEMATSU[‡], T. OSANAI[†], H. MAKINAE[†] and M. TANIMOTO[†]

† National Research Institute of Police Science　6−3−1 Kashiwanoha Kashiwa, Chiba, 277−0882 Japan
‡ Graduate School of Frontier Sciences, University of Tokyo　5−1−5 Kashiwanoha Kashiwa, Chiba, 277−8562 Japan
E-mail: †{kamada, osanai, makinae, tanimoto}@nrips.go.jp, ‡{kamada, mine}@gavo.t.u-tokyo.ac.jp

Abstract　In speaker verification for voice telephony on the forensic science, we might be requested forensic speaker identification in very noisy environment which is different from general research. In noisy environment, we do the clarification of speech at the first. However, previous study of speaker verification with clarification of speech was not sufficient. In this study, we experimented on speaker verification with clarification of speech in noisy environment, and we examined the relation of speech clarification and speaker verification results. Moreover, the experiment that used the existing noise like the power supply noise was conducted, and speaker verification accuracy in real environment was examined.

Keyword　speaker verification, text-dependent, clarification of speech, voice telephony, forensic science

1. はじめに

我々が行っている話者照合は，研究結果の求められる環境となる法科学の観点から，電話を通した音声を対象としている．これらの対象となる音声は，一般的に行われている雑音環境下における研究と比較すると非常に悪い状況が多く，発声内容がほとんど聞き取れない場合もある．このような状況においては，発声内容が聞き取れるための音声明瞭化処理を行っているが，これまで明瞭化処理による個人性の変化を伴う話者照合は十分行われていなかった．しかし実際にはこのような劣悪な雑音環境下における話者照合が要求される場合もある．

雑音環境下における音声認識や話者認識の一般的な研究では，雑音の特性をオフラインで調べることができず，リアルタイムな処理と結果が要求されることが多いが，我々が扱う法科学の分野においては，雑音の特性をオフラインで十分調べることができ，リアルタイム処理が要求されるような状況はほとんどない．そのため音声明瞭化処理についても，雑音の特性を詳しく調べた上で，最適なフィルタやそのパラメータを決定することができる．また未知資料は再収録不可能であり，雑音環境下で録音されることがあるが，比較する対照資料は後に収録されることが多いため，比較的雑音の少ないよい条件で録音されることというのも，法科学の分野における特徴の一つである．

そこで本稿では，劣悪な雑音環境下における音声に対して，法科学の用件を満たす前提条件で音声明瞭化処理を伴うテキスト依存型の話者照合実験を行い，明瞭化処理と照合結果の関係について調べた．対象雑音として狭帯域(周期性)雑音と広帯域雑音を取り上げ，雑音の特性の違いによる照合結果の比較を行った．更に狭帯域雑音の１つである実在する電源雑音を取り上

げ，電源雑音を含む音声に対する明瞭化処理を伴う照合実験を行い，実環境下における照合精度についての検討を行った．

2. 話者照合

2.1. 音声データベース

話者照合実験に使用した音声データベースは電話を介して発声した成人男性 3000 人程度の規模のデータベースを元にしている．この音声データベースから選んだ 300 人の音声資料を話者照合実験に使用した．表 1 に示した音声資料は 11.025 kHz のサンプリング周波数で A/D 変換されているが，電話音声であることから，実験に際しては 8 kHz にダウンダンプリングしたものを使用した．

2.2. DP マッチング

表 2 に示した分析を行い，2 つの音声資料における話者間距離の算出方法として，従来から利用されている DP マッチングを利用した．現在の音声認識や話者認識の研究では，メルケプストラムや MFCC のような特徴量が多く使われ，また VQ や GMM のような確率統計モデルを利用した分析手法が多く利用されている．我々は既にこれらの特徴量や分析手法による研究を行っているが，本稿では特徴量や分析手法による照合精度の向上を目的とせず，明瞭化処理と照合精度の関係を明らかにするため，従来からの研究で安定して利用されている手法を利用した．

2.3. 照合方法

テキスト依存型話者照合を行うため，同一の発声内容の音声資料に対する話者間距離を求めた．同一人の組合せとして同時期の音声資料は使用せず，1 時期目と 2 時期目及び 2 時期目と 3 時期目を比較して，1 人

表1 音声資料

話者数	300 人
発声時期差	3～4 ヶ月
発声時期	3 時期
1 時期の発声回数	3 回
発声内容	5 母音（ア,イ,ウ,エ,オ）， 6 単語（はい,車,電話,爆弾, 銀行,警察）
サンプリング周波数	8 kHz
量子化精度	16 bit

表2 分析条件

分析窓	ハミング窓
フレーム長	32 ms
フレーム周期	16 ms
高域強調	1 次の適応型
分析方法	LPC 分析
分析次数	12 次
特徴量	LPC ケプストラム係数

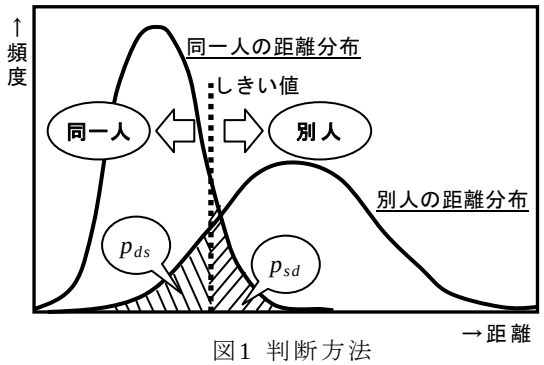

図1 判断方法

の話者間(話者内)距離分布を得た．同様に別人の組合せによる別人の話者間距離を得た．別人の組合せは非常に多くなるため,組合せについての間引きを行った．

図 1 に判断方法の模式図を示す．発声内容ごとに得られた同一人と別人の距離分布から等誤り率(p_{ds}=p_{ds})となるようにしきい値を求め，このしきい値を同一人かどうかの判断基準とした．判断基準から得られた正答率を照合実験における照合率とした．

3. 狭帯域雑音環境下の話者照合

3.1. 狭帯域雑音

法科学の分野における雑音環境として扱う雑音には狭帯域(周期性)雑音，広帯域雑音，定常衝撃(パルス性)雑音などが上げられるが,比較的多いのはブザーや警報機，電源雑音のような 1 つあるいは複数の狭帯域のスペクトル成分を持つ狭帯域雑音である．

狭帯域雑音を加法的に含む音声に対する明瞭化処理を行う場合，雑音除去フィルタの作成に関してパラメータの設定は試行錯誤的に行う必要があるが，明瞭化処理はオフラインで行うことが可能であり，また雑音の聴取や分析などから経験的に最適な処理手法(フィルタ)を選択することができる．

本稿では狭帯域雑音として，擬似的に作成したサイン波と実在する電源雑音を取り上げ，これらの雑音を含む音声に対する音声明瞭化処理と照合精度の関係を調べた．

3.2. 擬似雑音下での話者照合

3.2.1. 音声資料

実際に扱う狭帯域雑音は複数の狭帯域スペクトル成分を持つ雑音が多いが,雑音の特性である帯域(中心周波数と帯域幅)と明瞭化処理,話者照合の関係を明らかにすることが目的であるため，1 つの周波数成分を持ったサイン波を使用することにした．

音声信号 $s(t)$ の発話区間における RMS(Root Mean Square)を S_{RMS} , 雑音信号 $n(t)$ の RMS を N_{RMS} とすると，SNR [dB]となる雑音を含む音声 $x(t)$ は

$$x(t) = s(t) + \frac{S_{RMS}}{10^{SNR/20} N_{RMS}} n(t) \qquad (1)$$

で得られる．各発話において雑音が重畳された音声を求め，同一の発声内容で雑音成分の最大振幅が一定になるように規格化を行った．得られた音声資料を雑音環境下における話者が未知の音声資料とした．

3.2.2. 帯域除去フィルタ

サイン波が重畳された音声の明瞭化処理には帯域除去フィルタ(Band Elimination Filter:BEF)を利用した．FIR の設計にはカイザーフィルタなどが利用されることが多いが，本稿では汎用性を考慮して sox(Sound eXchange) によるカイザーフィルタを利用した．f [Hz]サイン波雑音に対して BEF の除去帯域幅を f_D [Hz] とした場合，BEF の帯域除去は $(f - f_D/2) \sim (f + f_D/2)$ となるようにした．それぞれのカットオフ周波数での利得は-6dB であり，また減衰率は雑音除去ができるように，適切にカイザー窓の窓幅や β を調整した．

3.2.3. 対照資料に処理を行わない話者照合

未知資料は 2kHz のサイン波雑音が重畳された音声とし，対照資料は雑音の重畳されていない音声とした．SNR は 10dB, 0dB, -10dB とした．未知資料の雑音除去を BEF で処理し，対照資料との話者照合実験を行った．照合実験の流れを図 2 に示す．BEF の除去帯域幅を可変とした照合結果に着目し，比較のために未知資料に対して BEF 処理を行わない条件(Noise)の照合を行った．また BEF 処理による照合率への影響を調べるために，未知資料に雑音が重畳されていない条件(Clear)についても実験を行った．本研究では明瞭化処理後の照合率を 100%に近づけるのが目的ではなく，あくまで雑音が無い条件(Clear)の照合精度まで改善させることが目的であるため，Clear における照合率が改善における目標値となる．実験における資料の照合条件を表 3 に示す．

表3 照合条件(1)

Noise	BEF 無し
200Hz~2kHz	BEF 処理，帯域幅は 200,400,600,1000,1400,2000[Hz]
Clear	SNR=∞

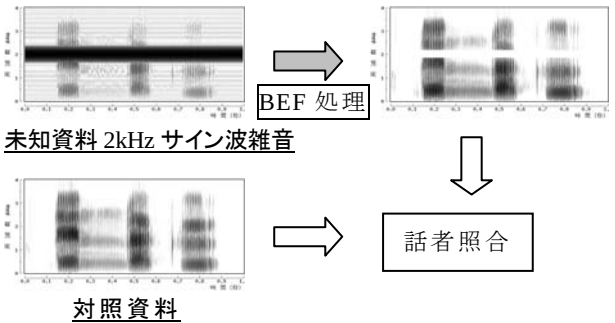

図2 話者照合の流れ(1)

3.2.4. 対照資料に処理を行わない照合結果

3 母音及び 3 単語における照合結果を図 4 に示す．BEF の帯域幅が同一である場合，SNR の違いによる照合率の差は見られない．この結果から少なくとも BEF による処理によって雑音除去可能であれば，SNR が -10dB のような悪い環境においても，雑音の少ない環境と同等の照合精度が得られることがわかった．また BEF の除去帯域幅 f_D に着目すると，f_D =200Hz であれば雑音のない環境と同等の照合精度が得られ，f_D の増加に伴い照合率は低下している．また f_D が 1kHz 以上の条件では BEF 無し(Noise)よりも照合率が低下していることから，雑音除去のための BEF が大きな帯域幅を必要とする場合は明瞭化処理を行わないほうが高い照合精度が得られるということがわかった．

3.2.5. 対照資料に同処理を行う話者照合(1)

未知資料のみに対して BEF 処理を行う場合，未知資料だけが狭帯域における特徴量(missing feature)が失われ，対照資料には存在しているために，帯域幅 f_D に関する頑健性が得られないことがわかった．そこでミッシングフィーチャー理論(Missing Feature Theory)により，対照資料に対して未知資料と同様の BEF 処理(ミッシングフィーチャーマスク)を行うことで，照合精度がどのように変化するのかについての実験を行った．照合実験の流れを図 3 に示す．SNR は 10dB, 0dB, -10dB とした．対照資料に対して BEF 処理を行う場合，未知資料と同じ雑音の事前重畳は行っていない．比較のための条件として，未知資料に BEF 処理を行わない条件(Noise)では，対照資料にも同 SNR となるように雑音を重畳した．その他の実験条件は 3.2.3.の実験と同様である．照合条件を表 4 に示す．

表4 照合条件(2)

Noise	BEF 無し，対照資料に同 SNR の Noise を重畳
200Hz~2kHz	BEF 処理，帯域幅は 200,400,600,1000,1400,2000[Hz]，対照資料にも同じ BEF 処理
Clear	SNR=∞

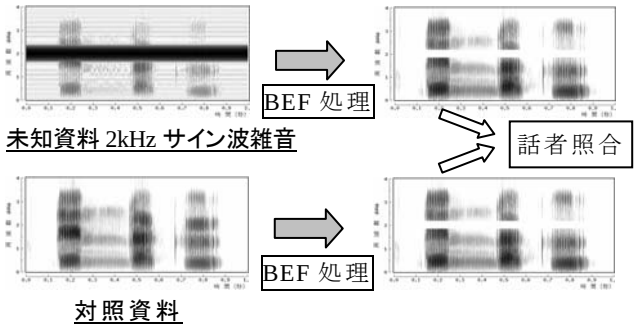

図3 話者照合の流れ(2)

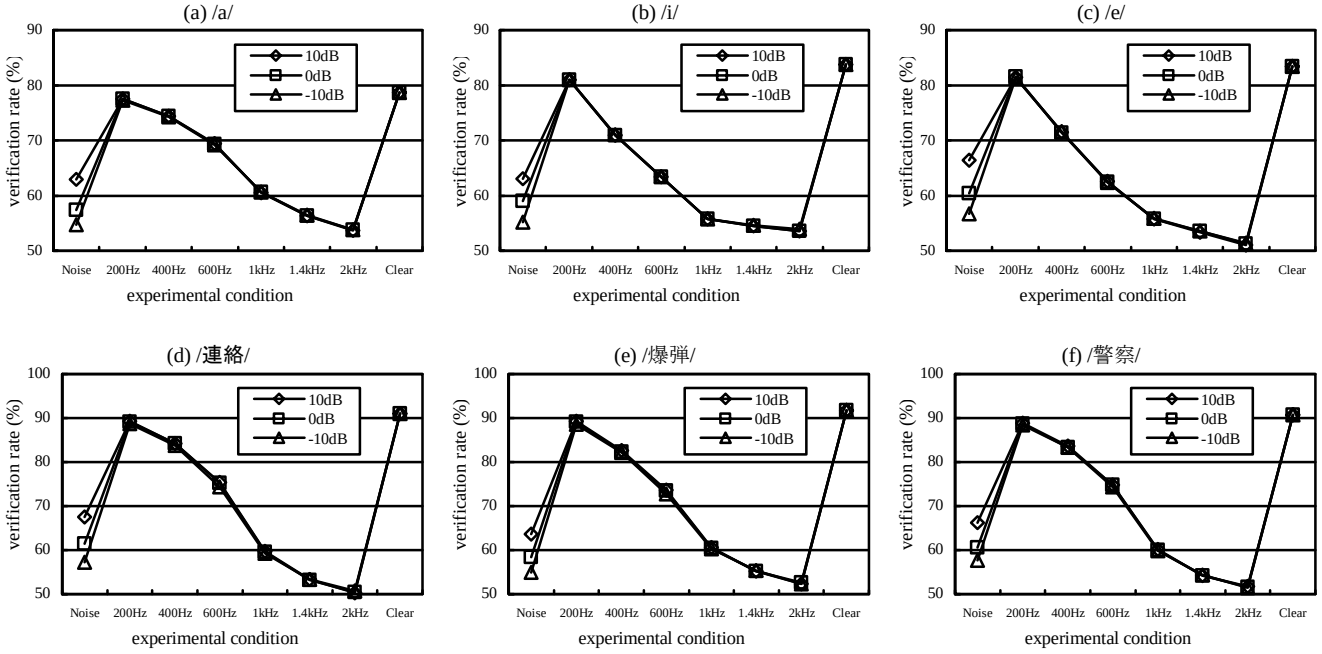

図4 対照資料に明瞭化処理を行わない話者照合結果

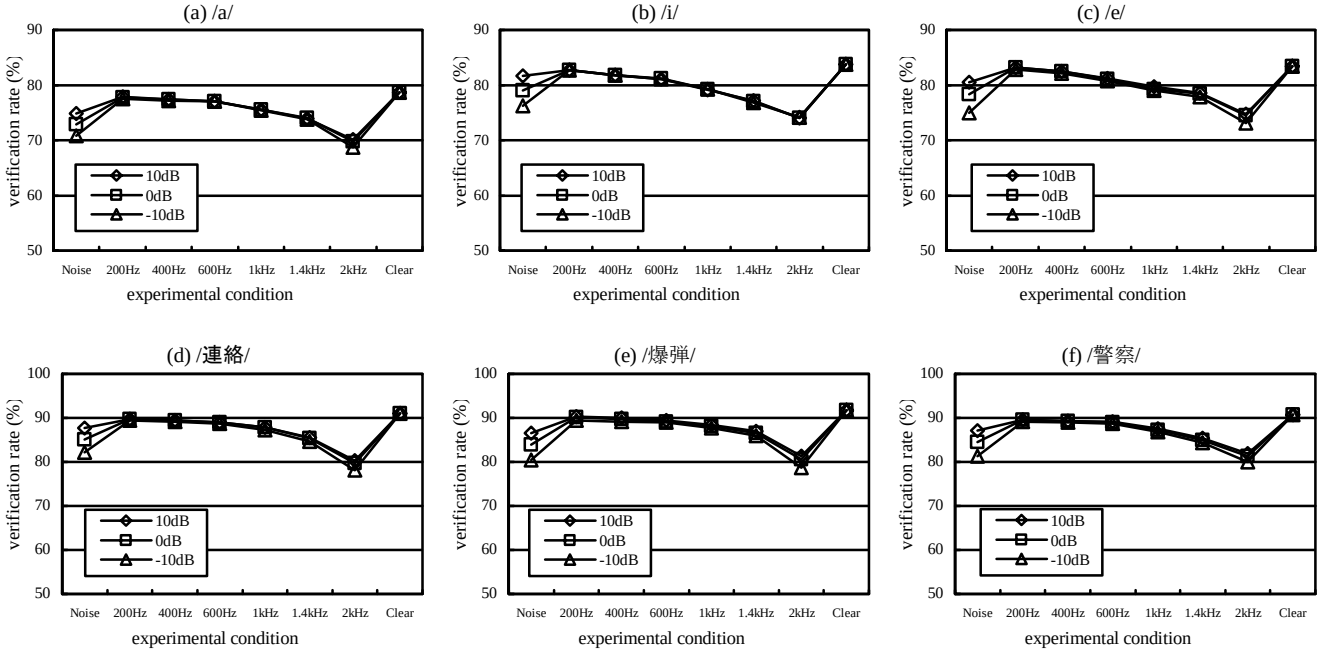

図5 対照資料に未知資料と同じ処理を行う話者照合結果(2kHz サイン波)

3.2.6. 対照資料に同処理を行う照合結果(1)

3 母音及び 3 単語における照合結果を図 5 に示す．図 4 の照合結果と比較すると，BEF 処理あり及び無し(Noise)いずれの条件においても照合率の改善が見られ，特に帯域幅 f_D が 1kHz の場合において大きく改善していることが確認された．これらの結果から，f_D が 600Hz 程度以下であれば，BEF 処理による照合精度への影響は非常に小さく，未知資料に雑音のない条件と同程度の照合率が得られることがわかった．

また未知資料に含まれる雑音の帯域幅が大きく，BEF の f_D が 2kHz 程度になる場合は，BEF の影響が照合精度に大きく現れるが，1kHz 以下であればミッシングフィーチャーマスクとして対照資料に同条件の雑音を付加するよりも，BEF による処理を行った方が高い照合率が得られ，音声明瞭化処理の効果が非常に高いということがわかった．また予備実験において，BEF の f_D が 400kHz 以下の条件で，BEF による雑音除去能力が低く雑音が完全に除去できない状況では，f_D が 600Hz～1kHz の BEF による雑音除去が成功している条件よりも照合精度が低くなることがわかった．これら

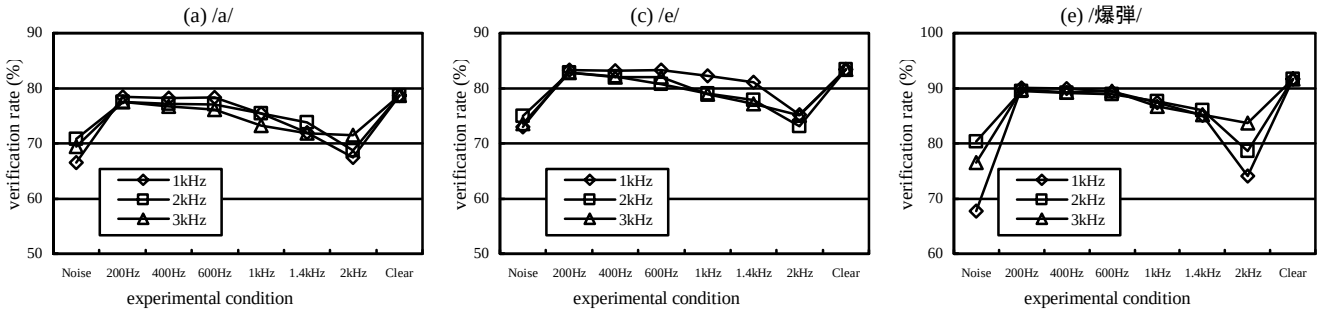

図6 対照資料に未知資料と同じ処理を行う話者照合結果(SNR=-10dB)

のことから，ミッシングフィーチャーマスクの条件を同一にすることが照合精度を改善するために最も効果的な手法であり，雑音除去能力が低い状況では，対照資料に対して雑音を重畳した後，BEF などの明瞭化処理を行うことが効果的であると考えられる.

3.2.7. 対照資料に同処理を行う話者照合(2)

BEF の帯域幅 f_D 及び f_D の存在する帯域と照合精度の関係を調べるために, 3.2.5.で行った話者照合について, SNR=-10dB の場合においてサイン波の周波数を可変とした話者照合実験を行った. 実験に使用したサイン波は 1kHz,2kHz,3kHz である.

3.2.8. 対照資料に同処理を行う照合結果(2)

2 母音及び 1 単語における照合結果を図 6 に示す. この結果から f_D が 600Hz 以下の条件では除去する帯域がどの周波数領域に存在していても目標とする照合精度まで改善できることがわかった. 一方 f_D が大きくなる場合は除去する帯域の存在する周波数領域に依存して，照合精度が変化することが確認された. f_D が 2kHz の条件を除けば, f_D が低い周波数領域よりも高い周波数領域に存在する場合，照合精度の低下が大きく現れていると思われる. これは低い周波数領域よりも高い周波数領域に個人性の特徴量が多く存在するということに符合していると考えられる.

3.3. 実環境雑音下における話者照合

3.3.1. 音声資料

3.2.では擬似的な雑音を取り上げたが，実環境下での明瞭化処理と照合精度の関係を調べるため，雑音として 50Hz の電源雑音を使用した. 3.2.1 と同様に SNR が 10dB,0dB,-10dB となるように重畳して雑音環境下の音声資料を作成した.

3.3.2. 周波数櫛型フィルタ

電源雑音は 50Hz の周波数構造を持つが，サイン波のような純音ではないため, 50Hz の高調波構造を併せ持った狭帯域雑音である. このような特性を持つ雑音は，音声明瞭化のためには BEF ではなく，周波数櫛型(comb)フィルタが有効である. 基本周波数とその高調波成分を除去する周波数フィルタを利用することで，明瞭化の効果は大きくなる. 櫛型フィルタの設計は比較的容易に可能である. 時刻 t における信号 $x(t)$ に対して周期 τ の基本波及び高調波成分を持つ雑音を軽減する櫛型フィルタは，処理後の信号を $y(t)$ とした場合

$$y(t) = x(t) - x(t-\tau) \qquad (2)$$

で実現できる. 電源雑音が重畳された未知資料に対して櫛型フィルタを利用した明瞭化処理を行った.

3.3.3. 実環境下における話者照合結果

電源雑音が重畳された未知資料と雑音のない対照資料に対して，表 5 に示す照合条件による明瞭化処理を伴う話者照合実験を行った. 照合結果を図 7 に示す. 櫛型フィルタを利用する場合 3.2.6.の実験結果と同様に，対照資料に対しても同じ櫛型フィルタによる処理を行った方が，照合精度の改善が大きい場合が多く見られる. SNR=-10dB の場合でも照合精度の低下は発声内容によって数%から 10%程度に抑えられており，単語による照合率もおおむね 80%を超えていることから, 狭帯域雑音を含む音声の明瞭化処理を伴う話者照合は実環境においても十分有効であることが示された.

4. 広帯域雑音環境下の話者照合

4.1. 音声資料及び音声明瞭化

ホワイトノイズを SNR=15dB で重畳した音声資料を作成し，広帯域雑音環境下の音声資料とした. 音声資料に対して SS (Spectral Subtraction)を利用した明瞭化処理を行い，明瞭化前後の音声資料を作成した.

4.2. 話者照合及び実験結果

表 6 に示した照合条件により話者照合実験を行った. 3 母音及び 3 単語の実験結果を図 8 に示す. 実験結果から SS 処理による照合精度の改善はほとんど見られないことがわかった. これはホワイトノイズの SS に

表5 照合条件(3)

	未知資料	対照資料(雑音無)
Noise	Filter 処理無し	同雑音重畳
Filter-F	Filter 処理有り	Filter 処理有り
Filter-C	Filter 処理有り	Filter 処理無し
Clear	SNR=∞	SNR=∞

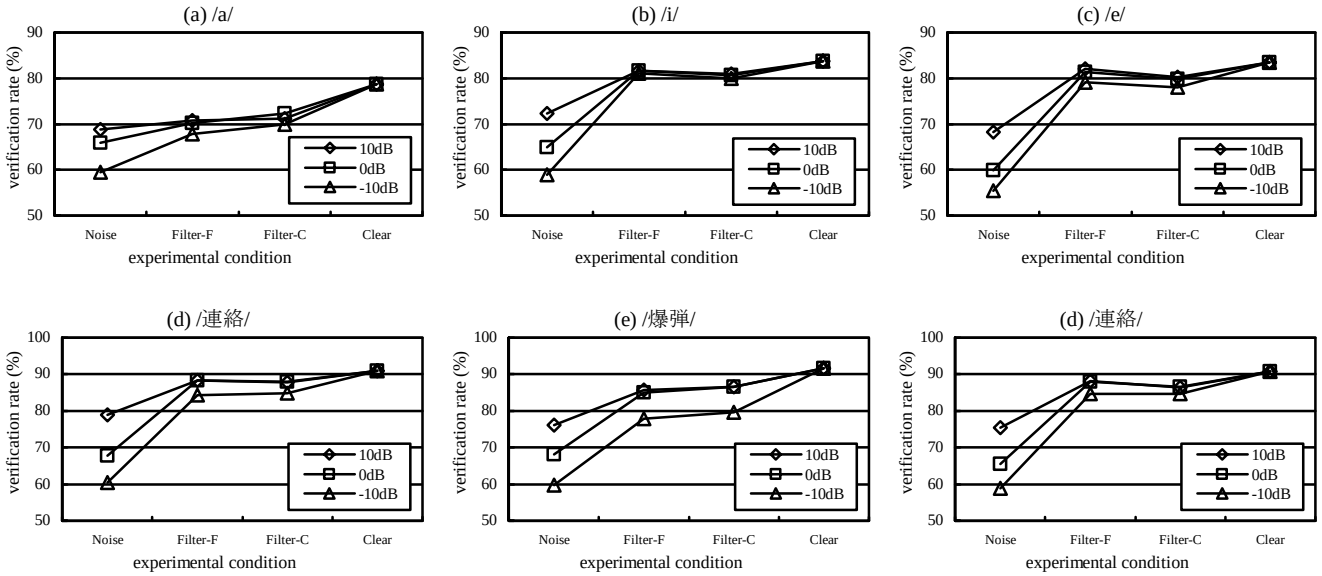

図7 実環境雑音下における話者照合結果

表6 照合条件(3)

	未知資料(SNR=15dB)	対照資料(雑音無)
条件 1	SS 処理無し	SS 処理無し
条件 2	SS 処理有り	SS 処理無し
条件 3	SS 処理無し	雑音重畳(SNR=10dB)
条件 4	SS 処理無し	雑音重畳(SNR=15dB)
条件 5	SS 処理無し	雑音重畳(SNR=20dB)

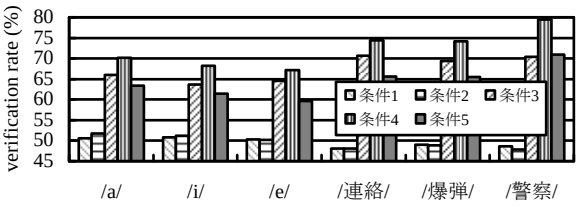

図8 広帯域雑音環境下における話者照合結果

よる明瞭化処理が困難であり，明瞭化の効果が低いことが原因であると考えられる.一方 SS 処理を行わず，対照資料に対して同じ雑音を重畳することにより，照合精度が大きく改善することがわかった．対照資料の雑音の重畳は未知資料と同じ SNR とすることで照合精度の改善が最大になることが確認された.

5. まとめ

法科学に分野における雑音環境下での話者照合実験を行い，音声明瞭化処理と照合精度の関係について調べた結果，狭帯域雑音に対する明瞭化処理が照合精度に及ぼす影響は小さく，明瞭化処理により照合精度が大きく改善できることがわかった．また電源雑音を使用した実験により，狭帯域雑音の明瞭化処理による照合精度の改善は，実環境下でも有効であることが確認された．広帯域雑音環境下では明瞭化処理の効果が低いため，対照資料に未知資料と同じ雑音を重畳する方法が有効であることがわかったが，狭帯域雑音では照合精度が改善された SNR が 0dB 以下の広帯域雑音環境下では，照合精度の改善が困難であることがわかっている．今後は研究結果を実務にどのように適応していくのかについて検討を行うとともに，照合精度を改善させるための効果的な明瞭化処理手法や,特徴量,照合手法などについて実験を行っていく予定である.

文　献

[1] 蒔苗久則，長内隆，鎌田敏明，谷本益巳，“周波数帯域を考慮した MFCC による話者特徴量の検討,”日本音響学会 2006 年秋季研究発表会講演論文集, pp.75-76, Sep.2006.

[2] 長内隆，尾関和彦，鎌田敏明，蒔苗久則，谷本益巳，“VQ によるテキスト独立型話者照合における特徴量変換,”日本音響学会 2006 年春季研究発表会講演論文集, pp.57-58, Mar.2006.

[3] Kun-Youl Park, Hyung Soon Kim, “Narrowband to wideband conversion of speech using GMM based transformation,” Proc. ICASSP, vol.3, pp.1843-1846, 2000.

[4] So, S., Paliwal, K.K., “Multi-Frame GMM-Based Block Quantization of Line Spectral Frequencies for Wideband Speech Coding,” Proc. ICASSP, vol.1, pp.121-124, 2005.

[5] Nishida, M., Kawahara, T., “Speaker model selection based on the Bayesian information criterion applied to unsupervised speaker indexing,”Speech and Audio Processing, IEEE Trans., vol.13, 4, pp.583-592, 2005.

[6] Shingo KUROIWA, Yoshiyuki UMEDA, Satoru TSUGE and Fuji REN, “Nonparametric Speaker Recognition Method Using Earth Mover's Distance.” IEICE Trans. on Information and Systems, E89-D(3), pp.1074-1081, 2006.

[7] 鎌田敏明，長内隆，蒔苗久則，谷本益巳，“劣悪な雑音環境下における話者照合,”日本音響学会 2005 年秋季研究発表会講演論文集，pp.130-131, Sep.2005.