

平成 29 年 5 月 23 日現在

機関番号：82505

研究種目：若手研究(B)

研究期間：2012～2016

課題番号：24710195

研究課題名(和文) 超低SNR法科学的音声資料での発話内容の理解を可能とする手法の開発と信頼性の評価

研究課題名(英文) Forensic intelligibility enhancement of recorded speech in high noise environments

研究代表者

蒔苗 久則 (Makinae, Hisanori)

科学警察研究所・法科学第四部・主任研究官

研究者番号：20415441

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：法科学的環境では、店内放送などにより完全に聴取が不可能なほどSNRが極めて低い信号が録音される機会が多いため、このような信号に有効な雑音抑圧手法を本研究では提案した。いくつかの提案手法では、法科学的な応用を考えると実時間処理が必ずしも必要とされないことから、雑音の音源を参照信号として処理に使用している。また、信号の表現に非負値行列因子分解(NMF)や正弦波モデルを全ての提案手法で用いた。抑圧実験の結果、いずれの手法でも一定の効果があることを確認した。また、こうした信号処理技術による手法に加え、音声認識技術の利用可能性についても予備的な検討を行った。

研究成果の概要(英文)：In a forensic situation, since speech samples are usually recorded in highly noisy environments, for example, in a store or a car where background music is played, the signal-to-noise ratio (SNR) of these samples is severely degraded. We propose speech enhancement methods effective to these samples in this study. In some proposed methods, source signals of the noise were used as reference signals, because real time processing is unnecessary for forensic purpose and the source signals used for background music is easily available. And non-negative matrix factorization (NMF) or sinusoidal modeling was used for signal representation. Experiment showed the effectiveness of the proposed methods to the severely degraded signals. In addition to these methods based on signal processing technique, preliminary examination was carried out to investigate the availability of speech recognition technology.

研究分野：法科学

キーワード：信頼性工学 裁判科学

1. 研究開始当初の背景

店舗や街頭に網羅的に設置された防犯カメラ、ICレコーダに代表される小型で安価な録音機材、携帯電話やスマートフォンの普及などにより、過去と比較できないほど音声録音される機会が増加している。これにともない、コンビニ強盗や脅迫といった犯罪についても、その模様が記録される機会が増加している。記録された犯人の発話内容は犯罪の全容解明のための重要な手がかりであり、事件によっては唯一の手がかりとなる。しかしながら、店内放送などの著しい背景雑音により聴取が完全に不可能となる例が少なくなく、発話内容が理解出来なければ、その価値は著しく損なわれる。

犯罪捜査といった法科学的な場面では、得られた音声信号を人間が事後的に聴取することが可能であるため、0dB 程度の信号対雑音比 (SNR) であれば、記録された発話内容の理解は比較的容易である。そのため、法科学的な応用を想定すると、SNR が 0dB よりもはるかに低く、人間による聴取が困難なほど劣悪な音声信号を対象とする必要がある。こうした際、実環境で機能する音声信号処理システムの要素技術として精力的に研究が進められている雑音抑圧技術や音声強調技術の利用が考えられるものの、人間による聴取が完全に不可能なほど聴取性が低下した音声信号を対象とした技術開発の試みは乏しいのが現状である。

2. 研究の目的

本研究では、社会の安全維持に密接に関わる法科学的な応用を前提に、著しい背景雑音により聴取が不可能となるほど SNR が低下した音声から、その発話内容の理解を可能とする手法の開発を試みる。

3. 研究の方法

音楽などの背景雑音は音声に加算的に重畳されることが多いため、雑音の特性が高精度に推定できれば、減算処理による雑音抑圧が有効である。また法科学的な応用の場面では、事後的な処理が主となるため、雑音源に相当する楽曲 CD などを処理に利用可能であり、効果的な抑圧が期待できる。そこで本研究では、雑音の音源を参照信号として処理に利用し、信号の表現に非負値行列因子分解 (NMF) や正弦波モデルを使用した手法を提案した。また、正弦波モデルを用いた手法を拡張し、参照信号を必要としない手法の開発も行った。

これらの信号処理による手法の開発に加え、音声認識技術の利用可能性についての予備的な検討も行った。雑音環境下では、システム設計時の想定と利用時の環境とのミスマッチによる認識性能低下が避けられない。そのため、既存技術では、ミスマッチを低減する適応処理などの利用が試みられている。一方、法科学的な応用の場面では、実時間処

理は必須では無く、事後的な処理が可能であるため、実際の利用環境に応じたシステム調整が可能である。そこで、本研究では、利用環境に応じた音響モデルの再構築による認識性能の改善を試みた。

4. 研究成果

(1) NMF を用いた手法

店内放送として用いられた楽曲が雑音となり聴取性が著しく低下した観測信号を想定し、楽曲の音源を参照信号として利用する雑音抑圧手法の開発を行った。このような場合、既存技術では適応処理やスペクトル減算法 (SS 法) の利用が考えられるが、図 1 に示したように、楽曲の特性が急激に変化する時刻で抑圧性能の低下が生じる場合がある。本研究では、このような抑圧性能の低下を避けるため、NMF を参照信号の表現に利用した雑音抑圧手法を提案した。

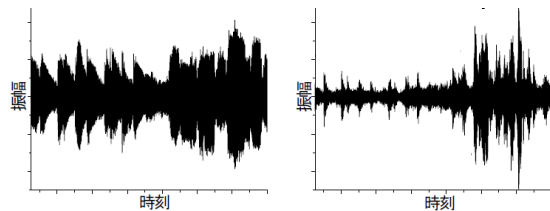


図 1 抑圧処理前(左)と抑圧処理後(右)

NMF は要素が非負である行列 S を、同じく要素が非負である 2 つの行列 W と H の積で近似する手法である。つまり S として与えられたデータを、これを構成する要素に分解することが可能であり、顔画像認識や楽曲の自動採譜などに利用されている。本研究では、NMF を参照信号の振幅スペクトログラム R の表現に使用し、SS 法と同様の減算による雑音抑圧を行った。NMF を用いた信号表現では信号の周波数特性と時間特性が分離して表現されており、これに注目することによって、SS 法や適応処理では生じた雑音の周波数特性が急激に変化する時刻での抑圧性能の低下の低減を図る。

参照信号のスペクトログラム R を、行を時刻、列を周波数に対応させた F 行 T 列の行列とみなして NMF を行うと、 R は行列積 WH によって近似表現される。 W と H は、 R と WH 間の距離 $J_1 = \|R - WH\|^2$ などを目的関数とした最小化問題により算出される。こうして得られた W と H は、 F 次元のベクトルである $w_i (1 \leq i \leq K)$ と、 T 次元のベクトルである h_i を用いて、 $W = (w_1, w_2, \dots, w_K)$ や $H = (h_1, h_2, \dots, h_K)^T$ (上付きの T は、行列の転置を意味する。) の様に表現される。つまり、 R の基底スペクトルと見なせる周波数成分を w_i で表現し、その時間特性が h_i で表現されている。このように、信号を構成する周波数特性と時間特性が分離されて表現されることが、NMF を用いた信号表現の利点である。

NMFにより得られた W と H を使用した雑音抑圧は、SS法と同様の減算処理、 $0-WH$ (0 : 観測信号のスペクトログラム) によって行う事ができる。本研究ではさらに、楽曲の急激な周波数特性の変化に対応させるため、時間特性 H に事後的な修正を行った H' を使用した。 H の事後的な修正は、最大値フィルターを h_i に適用することにより行った。これにより、 h_i で表現された時間特性の立ち上がりが早まるのと同時に、立ち下がりに遅れが生じ、急激な特性変化に追従出来ないために生じた減算不足に起因する副次的な雑音の発生が抑制される。

以上の処理を行った例を図2に示す。図1と比較して、雑音特性が急激に変化する際に生じていた雑音が減少した事に加え、定常的な白色雑音状の雑音も減少した事が分かる。

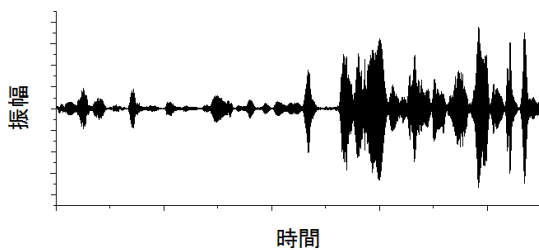


図2 NMFを用いた抑圧処理結果

次に、やはり NMF を使用し、法科学の場面では広く雑音となるインパルス性雑音を対象とした雑音抑圧手法の開発を行った。

法科学の場面では、録音機が自動車内のダッシュボードに置かれた場合のように、エンジンや走行に伴う様々な振動や、録音機に加わった衝撃などにより、インパルス性の雑音が混入するケースがある。インパルス性雑音が混入すると、聴取者の注意が向かうために著しく聴取性が損なわれるのに加え、広い周波数領域に雑音のエネルギーが存在するため、周波数フィルターを用いた比較的単純な方法では、雑音の抑圧は困難である。そこで本研究では、NMF を利用したインパルス性雑音の抑圧を提案した。

提案手法では、先と同様に、観測信号のスペクトログラム 0 を行列とみなし、NMFを行う。ただし、 W と H については、音声を表現する部分行列 (W_1 と H_1) と雑音を表現する部分行列 (W_2 と H_2) に分割して推定を行う。つまり、 $W=[W_1, W_2]=[w_{11}, w_{12}, \dots, w_{1p}, w_{1p+1}, \dots, w_{1p+q}]$ と W を表現し H についても同様に $H=[H_1, H_2]^T=[h_{11}, h_{12}, \dots, h_{1p}, h_{1p+1}, \dots, h_{1p+q}]^T$ と表現する。目的関数 J には、観測信号 0 から音声と雑音のスペクトログラムを $W_1H_1^T$ や $W_2H_2^T$ として分離するため、距離項 J_1 だけでなく、音声と雑音の周波数特性の違いを評価する $\text{trace}\{W_1H_1^T(W_2H_2^T)^T\}$ などの制約条件項 J_2 を課す。こうして構成した目的関数 $J = J_1 + J_2$ (λ : 調整パラメータ) の最小化は W_1, H_1, H_2 についてのみ行う。雑音の周波数成分を表現する W_2 については、音声为重畳されておら

ず雑音のみが録音された時間区間での信号を利用し、別途、目的関数として距離項 J_1 のみを含む NMF により推定する。そして、処理結果は、 $W_1H_1^T$ や観測信号から雑音成分を減算した $S-W_2H_2^T$ とした結果として得られる。図3示したインパルス性雑音の抑圧例から分かるように、提案手法の有効性が確認出来る。

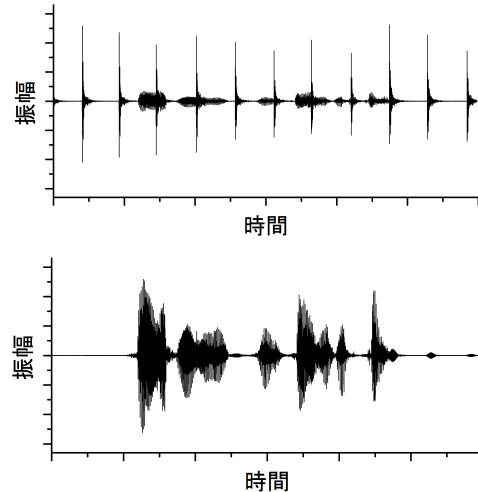


図3 インパルス性雑音の抑圧
上：抑圧前、下：抑圧後

(2) 正弦波モデルを用いた手法

先述の(1)で参照信号の表現に用いた NMF にかわり正弦波モデルを利用した手法の開発を行った。正弦波モデルは、様々な振幅と周波数の正弦波の和として信号をモデル表現する手法である。正弦波モデルを信号表現に利用すれば、NMF を用いた方法と同様の雑音抑圧が可能であり、実際にその効果を確認した。

しかし、これらの手法では参照信号として楽曲音源などを別途入手する必要があり、その適用範囲には自ずと限界がある。そこで次に、正弦波モデルを使用し、参照信号を利用しない手法の開発を行った。

この手法では、本研究が対象とする観測信号の SNR の低さに注目している。観測信号に含まれる所望の音声信号の振幅は極めて小さい一方で、雑音の振幅は極めて大きい。そのため、雑音に由来すると考えられる大振幅成分のみを除去すれば、聴取性の向上が期待できる。提案手法では最初に、観測信号を正弦波モデルにより様々な周波数と振幅を示す正弦波の和として表現する。次に、観測信号を表現する正弦波のうち、振幅の大きなものを雑音由来と見なし、閾値処理により除去する。この結果、振幅の小さな正弦波の和に対応する所望の音声信号が得られる。以上の処理を行った例を図4に示す。ここでは雑音として、入店時のチャイム音を想定し、計算機上で音声と重畳して作成した信号を対象として抑圧処理を行っている。

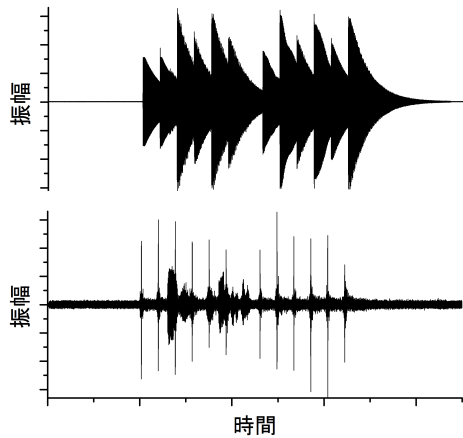


図4 正弦波モデルによる雑音抑圧
上：抑圧前、下：抑圧後

図4示すように、雑音の特性が変化する時刻では、インパルス性の雑音が生じているものの、大部分の雑音が抑圧されている事が分かる。次に、抑圧対象とした信号のSNRと、抑圧性能の関係を図5に示す。ここでは、抑圧性能の指標としてケプストラム距離 D_r を用いている。図5が示すように、SNRの低下にともない抑圧性能は低下するものの、0dBから-30dB程度まではケプストラム距離の変化は乏しく、安定した抑圧性能を示す事が分かる。

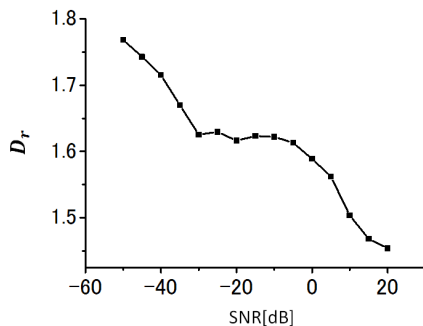


図5 SNRと抑圧性能の関係

(3)音声認識技術の利用

ここまで行ってきた信号処理技術を応用した手法に加え、音声認識技術の法科学環境下での利用可能性について予備的な検討を行った。HMMを用いた標準的な音声認識技術を用いて、音響モデルの構築や認識実験を行った。これらに使用した当所で構築した音声データベースには、クリーン音声に加え、同時録音された携帯電話音声と骨導音声も収録されている。クリーン音声を用いて構築した音響モデルを使用し、携帯電話音声や骨導音声を用いた認識実験を行うと、クリーン音声を用いた場合と比べ、認識性能が低下した。次に、携帯電話音声や骨導音声を用いて音響モデルを再構築し、2~4モーラからなる単語を用いた認識実験を行うと、クリーン音声から構築した音響モデルを用いた際の認識性能と比べ、音素認識率で20%程度の認識性

能の改善を確認した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計2件)

長内 隆、蒔苗 久則、網野 加苗、音と法科学、日本音響学会誌、査読有、vol.72、2016、74-80

蒔苗 久則、鎌田 敏明、長内 隆、法科学分野における話者認識のための大規模音声データベースの構築、査読有、vol.62(1-2)、2013、53-57

〔学会発表〕(計11件)

蒔苗 久則、網野 加苗、鎌田 敏明、長内 隆、正弦波モデルを用いたブラインド雑音抑圧、本法科学技術学会 第22回学術集会、2016-11-10 - 2016-11-11、中野サンプラザ(東京都中野区)

蒔苗 久則、鎌田 敏明、網野 加苗、長内 隆、非定常雑音の抑圧性能の評価に関する研究、日本法科学技術学会、2015-11-12 - 2015-11-13、柏の葉カンファレンスセンター(千葉県柏市)

蒔苗 久則、網野 加苗、鎌田 敏明、長内 隆、正弦波モデルを用いた非定常雑音の抑圧、日本法科学技術学会第20回学術集会、2014-11-13 - 2014-11-14、ホテルフロラシオン青山(東京都港区)

Kanae Amino、Hisanori Makinae、Tatsuya Kitamura、Nasality in Speech and Its Contribution to Speaker Individuality、Interspeech 2014、2014-09-14 - 2014-09-18、Singapore

Kanae Amino、Hisanori Makinae、Tatsuya Kitamura Nasality in Oral Sounds? - Perception and Analysis of Oro-Nasal Signals-、日本音響学会 2014 年春季研究発表会、2014-03-10 - 2014-03-12、日本大学(東京都千代田区)

網野 加苗、蒔苗 久則、鎌田 敏明、長内 隆、同一話者内における母音の無声化の再現性、日本音響学会 2014 年春季研究発表会、2014-03-10 - 2014-03-12、日本大学(東京都千代田区)

蒔苗 久則、網野 加苗、鎌田 敏明、長内 隆、非負値行列因子分解を用いたインパルス製雑音の抑圧、日本法科学技術学会 第19回学術集会、2013-11-14 - 2013-11-15、ホテルフロラシオン青山(東京都港区)

四宮 康治、蒔苗 久則、網野 加苗、鎌田 敏明、長内 隆、伊藤 仁、フォルマント周波数を用いた話者照合法の統計的評価、日本法科学技術学会 第 19 回学術集会、2013-11-14 - 2013-11-15、ホテルフロラシオン青山（東京都港区）

網野 加苗、蒔苗 久則、鎌田 敏明、長内 隆、本人および両親の生育地が母音の無声化に与える影響、日本音響学会 2013 年秋季研究発表会、2013-09-25 - 2013-09-27、豊橋技術科学大学（愛知県豊橋市）

網野 加苗、蒔苗 久則、鎌田 敏明、長内 隆、母音の無声化頻度と話者の出身地に関する考察、日本音響学会 2012 年春季研究発表会、2013-03-13 - 2013-03-15、東京工科大学（東京都八王子市）

蒔苗 久則、鎌田 敏明、網野 加苗、長内 隆、非負値行列因子分解を用いた非定常雑音の明瞭化、日本法科学技術学会 第 18 回学術集会、2012-11-15 - 2012-11-16、ホテルフロラシオン青山（東京都）

〔図書〕

なし

〔産業財産権〕

なし

〔その他〕

なし

6. 研究組織

(1) 研究代表者

蒔苗 久則 (MAKINAE, Hisanori)

科学警察研究所、法科学第四部、主任研究官

研究者番号：20415441

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

佐藤 正学 (SATO, Masamichi)