特　集「センシングウェブ」

音声に含まれるプライバシ情報の保護

Privacy Protection of Speech Signals

中川　聖一　豊橋技術科学大学情報工学系
Seiichi Nakagawa　Department of Information and Computer Sciences, Toyohashi University of Technology.
nakagawa@slp.ics.tut.ac.jp, http://www.slp.ics.tut.ac.jp/~nakagawa/

山本　一公　（同　上）
Kazumasa Yamamoto
kyama@slp.ics.tut.ac.jp, http://www.slp.ics.tut.ac.jp/~kyama/

土屋　雅稔　豊橋技術科学大学情報メディア基盤センター
Masatoshi Tsuchiya　Information and Media Center, Toyohashi University of Technology.
tsuchiya@imc.tut.ac.jp, http://www.imc.tut.ac.jp/~tsuchiya/

Keywords: speech information, privacy, voice conversion, speech removal, proper noun extraction.

1. はじめに

センシングウェブで扱うセンサ情報の中には，マイクロホンセンサによる情報，すなわち音情報が含まれる．センシングウェブでは，ウェブ上を流れるセンサデータに対してプライバシ処理が施されていることが前提となっているが，音情報の中でも，音声情報はプライバシ情報の塊であるため，プライバシ処理が非常に重要になってくる．

現在，インターネットを経由してSkypeなどによりボイスチャットを楽しむユーザが増えてきている．しかし，不特定の相手としゃべる際に，自分の声をそのまま相手に聞かせたくないとして，ボイスチェンジャを経由して音声を発信するユーザも少なくないようである．これは，自分の声そのものがプライバシ情報であるという認識があるためであろう．また，マイクロソフトでは，オンラインゲームのように同時多数のボイスチャットが行われる場合や生放送において，不適切な発言を自動で除去するためのシステムとして，リアルタイム放送禁止用語除去システム [Danieli 08] が開発されている．これは直接プライバシ情報を扱うわけではないが，センシングウェブにおいては重要な技術であると考えられる．

本解説では，音情報の中でも音声情報について，どのようなプライバシ情報が含まれているか，また，それらをどのように処理することでプライバシ情報が保護できるかについて述べる．

なお，ここで扱うプライバシ情報とは“個人に属する情報”のことであり，「スピーチプライバシ」（「会話の秘話性・漏えい性」，「他人の声が聞き取れない（邪魔にならない）音環境」）[佐藤 08] と呼ばれるものとは異なることをあらかじめ断っておく．

2. 音情報・言語情報に含まれるプライバシ情報

マイクロホンで収録される音は，プライバシ情報を扱ううえでは次の二つに大別できる．

- 背景音：騒音（自動車走行音，ファンノイズなど），音楽など
- 音声：人声雑音（複数人の声の集合で，誰が何をしゃべっているか個々には識別できないもの；ヒューマンスピーチライク雑音 [小林 95]），人の声（誰が何をしゃべっているか識別可能なもの）

背景音には，どういう場所で話しているかというプライバシ情報は含まれるが，話者そのもののプライバシ情報は含まれないので，一般にプライバシ処理を行わずにそのまま用いることができる*1．主にプライバシを考慮すべきなのは音声である．

音声も，複数人の音声が重なり合うと，雑音のような性質をもつ [小林 95]．人間には“カクテルパーティ効果”と呼ばれる能力があり，背景雑音の中から目的とする音声だけに注目して聞き取り理解することが可能であるが，多人数（3人以上）の声を同時に聞いてそのすべてについて言語的な意味まで理解するのは難しい．そのため，人の声であっても誰が何をしゃべっているか個々に識別できないものについては，背景雑音と同等に扱ってよい．そのため，ここでは“音声”を“誰が何をしゃべっているか識別可能な人の声”として扱う．

音声により表現され，伝達される情報は

(1) 言語的情報（linguistic information）
(2) パラ言語的情報（para-linguistic information）

*1 音楽をネットワークに流すことについては著作権的な問題が生じると考えられるが，ここではその問題は扱わない．

(3) 非言語的情報 (non-linguistic information)
の3種類に大別できる [藤崎 94]. パラ言語的情報は, 音声に込められた意図, 発話者の態度など, 言語的情報以外で発声者が意識的に制御できる情報 (主に韻律的特徴) を指す. 非言語的情報は声質, 性別などの発話者の個性や感情などの話者が意識的に制御しない情報である. すなわち, 非言語的情報はそのすべてがプライバシ情報となり得る.

言語情報に含まれるプライバシ情報は, 発話内容に含まれるプライバシ情報と, 発話スタイルに含まれるプライバシ情報の2種類に大別することができる. 発話内容には, 発話者本人のプライバシ情報と会話中で言及されている人についてのプライバシ情報が含まれる. そのプライバシ情報は, 以下のような情報すべてである.

- 個人情報 (氏名, 住民基本台帳番号など)
- 身体特徴 (年齢, 身長, 性別など)
- 居住地 (住所, 電話番号など)
- 社会的地位 (職業, 勤務地, 会社名, 学校名など)
- その他 (成績, 収入など)

なお, 上記のような情報であっても, 広く一般に知られている情報 (例えば, 有名野球選手の身長) はプライバシ情報とはいわない. しかし, その情報が知られている範囲を推測することは難しいため, 安全のためには, 上記のような情報をすべてプライバシ情報として扱うのが望ましい. 発話スタイルに含まれるプライバシ情報とは, 類義語の選択, 機能語の選択, フィラーの挿入頻度などにおける地域的・個人的な癖である. 典型的な例としては, ある地方の方言に特有の文末表現などが使用されていると, 発話者の出身地を推測することが可能である.

次章以降は, これらの個々の要素に対してどのようにプライバシ保護を行えばよいか述べていく.

3. 音声を除去することによるプライバシ保護

言語的情報を手掛りとするプライバシ保護技術は, 音声認識技術を必要とするため, その精度が音声認識精度に依存する. しかし, 現在の音声認識システムは実環境下で十分に機能するとはいえず, また一般に計算コストも高いため, センサノードに必要なリアルタイム動作が難しい. 本章では, 音声信号そのものをプライバシ情報とみなしてこれを丸ごと除去する音声除去システムについて述べる.

3・1 一般的な音声除去・雑音除去手法

"音声除去" の研究は多くは行われておらず, あまり一般的ではないため, まず, 一般的な雑音除去手法について述べる. その後, 雑音除去手法で音声と雑音を逆に考えることで, 音声除去手法として用いる場合について述べる.

雑音除去の最も一般的な手法としてあげられるのは, スペクトルサブトラクション (Spectral Subtraction: SS) 法である [Boll 79]. 音声信号を $s(t)$, 背景雑音信号を $n(t)$ とすると, 観測される信号 $x(t)$ は

$$x(t) = s(t) + n(t) \tag{1}$$

となる. 音声信号のスペクトルを $S(\omega)$, 背景雑音信号のスペクトルを $N(\omega)$ とすれば, 観測される信号のスペクトル $X(\omega)$ は, 式 (1) をフーリエ変換することで

$$X(\omega) = S(\omega) + N(\omega) \tag{2}$$

となる. ここで背景雑音信号は定常であると仮定し, あらかじめ音声が含まれていない背景雑音から背景雑音のパワースペクトル $|\tilde{N}(\omega)|^2$ を推定しておく. 音声信号と背景雑音信号が互いに無相関であるとすれば, 式 (2) より,

$$|\tilde{S}(\omega)|^2 = |X(\omega)|^2 - |\tilde{N}(\omega)|^2 \tag{3}$$

とすることで, 音声スペクトルを推定することができる.

もう一つの代表的な手法として Wiener フィルタがあげられる [西山 01]. Wiener フィルタは, 信号や雑音を確率過程とみなして, 観測信号を入力とし, 所望信号 (ここでは音声信号) との平均二乗誤差を最小にする最適推定値を出力するフィルタである. Wiener フィルタの周波数領域表現を $H(\omega)$ とすると, 音声の推定スペクトルは

$$\hat{S}(\omega) = H(\omega) X(\omega) \tag{4}$$

となる. この場合, Wiener フィルタの伝達特性は

$$H(\omega) = \frac{|S(\omega)|^2}{|S(\omega)|^2 + |N(\omega)|^2} \tag{5}$$

で与えられる. 音声信号と背景雑音が無相関であると仮定すると, 式 (3) から

$$H(\omega) = \frac{|X(\omega)|^2 - |\tilde{N}(\omega)|^2}{|X(\omega)|^2} \tag{6}$$

となり, この形で広く用いられている.

上記の手法は, 単一マイクロホン (1チャネル) の手法であるが, 複数マイクロホンを用いることを前提とした手法もある. 代表的な手法として, マイクロホンアレーによるビームフォーミングがあげられる [Flanagan 85, Griffiths 82] また, 聴覚情景解析 (Auditory Scene Analysis) [Bregman 90] をもととして, 独立成分分析 (Independent Component Analysis: ICA) に基づく音源分離 (Blind Signal Separation: BSS) も最近活発に研究が行われている [猿渡 07].

次に音声除去の手法について述べる. 基本的には, 雑音除去手法において雑音と音声を入れ換えて考えれば, 既存手法を音声除去手法として用いることができるが, 音声は非定常であるため, SS 法のように雑音の定常性を仮定している手法をそのまま音声除去手法に用いることはできない. そこで我々は, VQ コードブックマッピングによるスペクトル置換を用いた音声除去システムを提案した [山本 08]. 図1に手法の概略図を示す. 本手法では, あらかじめ背景雑音を重畳した音声と, それに対応するクリーン音声 (背景雑音がない音声) のスペクトルをペアとしたものを特徴量として VQ コードブック

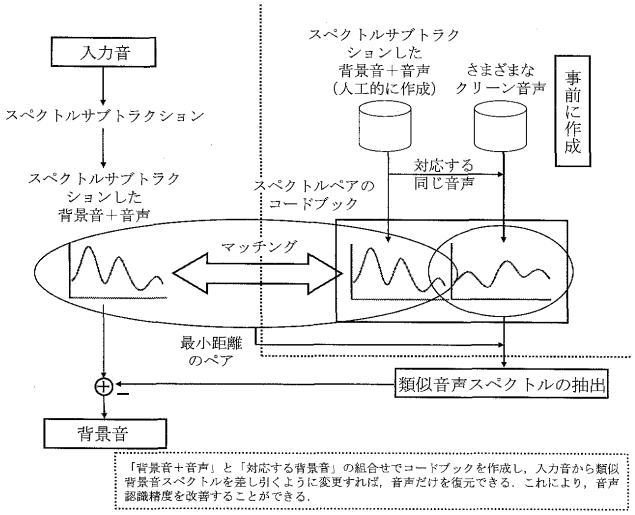

図1 コードブックマッピングによる音声除去システムのブロック

を作成しておき，入力音声（背景音＋音声）とVQコードブックの背景雑音を重畳した音声部分のスペクトルで距離を計算した上で，最も距離が近いVQコードからクリーン音声部分のスペクトルを取り出し，そのスペクトルを入力音声スペクトルから差し引くことで，背景音のみを復元する．

精度良く音声除去を行うためには，この手法を統計的な手法に拡張する必要がある．クリーンな音声コーパスと雑音が重畳した音声コーパスを並列に用いることで雑音を除去する統計的な手法として，SPLICE[Droppo 01]がある．また，単一マイクロホンでの音源分離手法として，周波数スペクトルを帯域ごとに混合ガウス分布モデル（Gaussian Mixture Model: GMM）の分散で重み付け分配することで信号を分離する手法[Benaroya 06]が提案されている．複数マイクロホンを用いる場合には，音源方向に死角を形成することでその音源から到来する信号を抑制する死角型ビームフォーマーも使われる[Araki 01].

3・2 求められる音声除去技術

センシングウェブにおいてセンサノードとして動作するためには，その動作速度にリアルタイム性（一定遅延時間内での動作）が求められる．そのため，動作が複雑で計算量が多いアルゴリズムや，発話区間を検出してその全体の特徴量を用いるようなアルゴリズムの適用は難しい．また，マイクロホンから入力される音声は複数話者同時発話であることも予想され，2人程度の同時発話であれば，これらが言語的に識別可能であることから，このような音声を雑踏（人声雑音）と区別したうえで除去することが必要になる．

4. 声質変換によるプライバシ保護

声質は，誰がしゃべっているかという情報を含んでいるため，プライバシ情報となる．本章では，声質を変換することによりプライバシを保護することについて述べる．

4・1 一般的な声質変換手法

一般に声質は，声道形に起因する音声のスペクトルピーク位置やピークの鋭さ，ホルマント周波数と，音源に起因するスペクトルの傾きやピッチ，アクセントなどに現れるとされ，これらが個人性とされる[Childers 85, 談 94].

従来，これらのパラメータを個別にターゲット話者に近づけることで声質変換が行われてきた[Childers 85]. テレビ放送などでプライバシ保護のために声質を変換することがあるが，その場合は主にピッチの変更が行われる*2単純なピッチの上下では，元の音声を復元される可能性があるため，複数のピッチ周波数に変換した後それらを合成することで，元の音声が復元できないようにする場合が多い．

最近は，スペクトル変換による声質変換が主流となっている．まず，マッピングコードブックによる方法が提案された[Arslan 97]. その後，GMMを用いた統計的手法に拡張され[Stylianou 98]，さらに固有声（Eigen Voice: EV）を用いた手法に発展している[戸田 06]. 固有声ベクトル変換は，各固有ベクトルの大きさを制御することにより，音声の感情変換にも用いられる[橘 07]. ここでは，EVを用いて任意の話者を1人の話者の声質$\{\boldsymbol{Y}_t\}$に変換する手順について簡単に述べる（図2参照）.

（1）入力話者と出力話者のすべてのパラレルデータを用いて，不特定入力話者GMM $\boldsymbol{\lambda}(0)$を学習する.

$$\boldsymbol{\lambda}^{(0)}=\arg\max_{\lambda}\prod_{s=1}^{S}\prod_{t=1}^{T_s}p(\boldsymbol{X}_t^{(s)},\ \boldsymbol{Y}_t|\boldsymbol{\lambda}) \tag{7}$$

$$p(\boldsymbol{X}_t,\boldsymbol{Y}_t|\boldsymbol{\lambda})=\sum_{i=1}^{M}\alpha_i\mathcal{N}(\boldsymbol{X}_t,\boldsymbol{Y}_t;\boldsymbol{\mu}_i^{(X,Y)},\ \Sigma_i^{X,Y}) \tag{8}$$

$$\mu_i^{(X,Y)}=\begin{bmatrix}\mu_i^X\\ \mu_i^Y\end{bmatrix} \tag{9}$$

$$\Sigma_i^{(X,Y)}=\begin{bmatrix}\Sigma_i^{(XX)} & \Sigma_i^{(XY)}\\ \Sigma_i^{(YX)} & \Sigma_i^{(YY)}\end{bmatrix} \tag{10}$$

ここで，$\boldsymbol{Y}_t$は出力話者のフレームtにおける特徴ベクトル，$\boldsymbol{X}_t^{(s)}$はs番目の事前学習用入力話者のフレームtにおける特徴ベクトル，Sは事前学習用入力話者数，T_sは各データのフレーム数，MはGMMの混合数である．$\boldsymbol{X}_t^{(s)}$と$\boldsymbol{Y}_t$はDPマッチングによって，あらかじめフレームごとに対応づけておく.

（2）各事前学習用入力話者sそれぞれに対するパラレルデータを用いて，$\boldsymbol{\lambda}^{(0)}$の入力平均ベクトル$\boldsymbol{\mu}_i^{(X)}$のみを更新することで，入力話者依存GMM $\boldsymbol{\lambda}^{(s)}$を学習する.

（3）学習された$\boldsymbol{\lambda}^{(s)}$の入力平均ベクトル$\boldsymbol{\mu}_i^{(X)}(s)$を接続することで，各入力話者に対してスーパベクトル$\boldsymbol{SV}^{(s)}=[\boldsymbol{\mu}_1^{(X)}(s)^T,\cdots,\boldsymbol{\mu}_M^{(X)}(s)^T]^T$を構成する．全入力

*2 製品としては，携帯電話用音声変換機NATEC VC102（http://www.natec-j.com/）などがある.

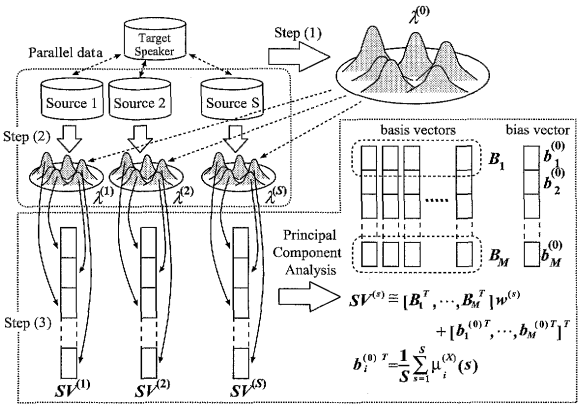

図2 固有声による声質変換システムのモデル学習

話者のスーパベクトルに対して主成分分析を行うことで，バイアスベクトル $\boldsymbol{b}_i^{(0)}$ および固有ベクトルによる行列 $\boldsymbol{B}_i$ を決定する．図中の $\boldsymbol{w}^{(s)}$ は s 番目の入力話者に対する $J(<S)$ 個の主成分である．得られた $\boldsymbol{b}_i^{(0)}$，$\boldsymbol{B}_i$ および $\boldsymbol{\lambda}^{(0)}$ により，EV-GMM $\lambda^{(EV)}$ が構成される．

(4) EV-GMM を用いて声質変換を行う．まず，任意の入力話者の特徴ベクトル $\boldsymbol{X}^{(\mathrm{input})}$ に対して，最適な重み $\hat{\boldsymbol{w}}$ を

$$\hat{\boldsymbol{w}}=\arg\max_{\boldsymbol{w}} p(X^{(\mathrm{input})}|\lambda^{(EV)}) \qquad (11)$$

により最尤推定する．次いで，$\boldsymbol{X}^{(\mathrm{input})}$ に対する変換静的特徴量 $\hat{\boldsymbol{Y}}$ を，$\hat{\boldsymbol{w}}$ と $\boldsymbol{\lambda}^{(EV)}$ を用いて求める．

(5) 得られた特徴量時系列に対して，適当な音源を構成し，MLSA フィルタリングを行うことで音声を合成する．

4・2 求められる声質変換技術

音声除去技術と同様に，センシングウェブにおいてセンサノードとして動作するためには，その動作速度にリアルタイム性が求められる．現在の高品質な声質変換技術は，計算量が多いため，これをリアルタイムで動作するように簡略化する必要がある．複数人の話者が会話している状況の場合，多対一変換によってすべての音声を同じ声質に変換すると，何人で会話しているかという情報が消えてしまうことになるため，センサ情報としては，会話人数に応じてターゲット話者の人数を変更する（1対1，2対2，3対3など）必要がある．これには，声質変換技術とともに，その場に話者が何人いるのか，いつ話者が交代しているのかを検出する技術が同時に必要になる．また，背景雑音中の音声は背景音を変形しないで音声だけを変換する必要がある．

5. 言語情報に対するプライバシ保護

音声に含まれるプライバシ情報から，言語情報として含まれる部分に対して処理を行うには，最初に，音声認識を行って音声をテキストに変換し，得られたテキストを対象として各種の処理を行う必要がある．本節では，音声をテキストに変換する技術，発話内容に含まれるプライバシ情報に対する保護手法，発話スタイルに含まれるプライバシ情報に対する保護手法について，順次述べる．

5・1 音声からテキストへの変換

音声をテキストに変換するために必要なのが，音声認識技術である．近年の音声認識技術の発展はめざましいものがあり，静かな環境で接話マイクロホンで丁寧にしゃべった場合には，90%以上の単語認識率が得られている．しかしながら，現時点では実環境下で十分な認識性能を得ることはできていない．

実環境下での音声認識を考える場合に問題となるのが背景雑音である．実環境では，接話マイクロホンは使用できないため，遠隔発話（マイクロホンから離れて発話した音声）を認識することになり，背景雑音の影響がより大きくなる（SNR の低下を招く）．また，マイクロホンの設置場所によって，残響の影響が出ることがあり，これに対しても何らかの対処を行うことが必要になる．頑健な音声認識については，3・1節で述べた雑音除去技術が有効に働くが，センサノードとしての動作を考えた場合には，やはりリアルタイムでの動作が要求される．

5・2 発話内容に含まれるプライバシ情報の保護

発話内容には，2章で述べたとおり，多種類のプライバシ情報が含まれている．これらの多くは人名や地名などの固有表現であるが，通常の固有表現には含まれない表現も多い．これらを取り除くことが，プライバシ情報の保護のためには必要である．

プライバシ情報を取り除く単位として，2通りの単位が考えられる．第1は，プライバシ情報を含む発話を検出し，そのような発話を取り除く方法である．固有表現には，未知語が特に多く出現するため，誤認識が起きやすい．そのため，[辻村 06] は，固有表現の周辺の表現を素性として考慮して，固有表現そのものではなく，固有表現が含まれる発話を検出する方法を提案している．この方法を応用すると，プライバシ情報を含む発話を検出し，そのような発話を取り除くことができる．ただし，単位が粗いため，プライバシ情報を含まない部分も多く取り除いてしまう欠点がある．

第2は，プライバシ情報を含む部分のみを取り除く方法である．音声認識用言語モデルをクラス言語モデルとし，プライバシ情報（固有表現など）からなるクラスを設定しておくと，認識結果からプライバシ情報クラスに含まれる部分を取り除くことによって，プライバシ情報の保護が実現できる．そのためには，プライバシ情報クラスの辞書に，プライバシ情報（固有表現など）を単語として登録しておく必要がある．しかし，センシングウェブの適用対象は今現在行われている通常の会話であ

形態素素性 MF		類似形態素素性 SF		文字種素性 CF	チャンクラベル
表層形	品詞	表層形	品詞		
今日	名詞-副詞可能	今日	名詞-副詞可能	⟨1, 0, 0, 0, 0, 0⟩	0
の	助詞-連体化	の	助詞-連体化	⟨0, 1, 0, 0, 0, 0⟩	0
石狩	名詞-固有名詞	**関東**	名詞-固有名詞	⟨1, 0, 0, 0, 0, 0⟩	B-LOCATION
平野	名詞-一般	平野	名詞-一般	⟨1, 0, 0, 0, 0, 0⟩	I-LOCATION
は	助詞-係助詞	は	助詞-係助詞	⟨0, 1, 0, 0, 0, 0⟩	0
晴れ	名詞-一般	晴れ	名詞-一般	⟨1, 1, 0, 0, 0, 0⟩	0

図3 学習データの例

り，そのような会話には，当然，新規のプライバシ情報（例：新たに出生した子の名前）が頻出する．事前にすべての新規プライバシ情報を辞書登録しておくことは不可能なため，ウェブなどから新規プライバシ情報を絶えず自動収集し，プライバシ情報クラス辞書に登録しなければならない．

音声認識処理とプライバシ情報の抽出が独立であると仮定すると，プライバシ情報の抽出は，音声認識結果として得られたテキストに対するチャンキングとして定式化できる．固有表現を対象とするチャンキングには，統計的機械学習に基づく手法 [Isozaki 02, Sekine 98, 内元 00, 山田 02] が有効である．しかし，プライバシ情報の抽出を統計的機械学習に基づくチャンキングとして定式化する場合，新規プライバシ情報に対して十分な量のラベルありデータが必要であるが，それを用意することが事実上不可能であるという問題がある．

この問題に対応するには，大量のラベルなしデータと少量のラベルありデータを併用する半教師あり学習が有効である．[Miller 04] は，ラベルなしデータからクラス言語モデルを作成し，ラベルありデータにクラスタリング結果を素性として追加して学習を行う方法を提案している．[Ando 05] は，ラベルなしデータから単語の予測問題を学習し，その予測問題を解いた予測値を素性として追加して学習を行う方法を提案している．[Suzuki 08] は，Conditional Random Field のポテンシャル関数と学習方法を拡張し，ラベルありデータとラベルなしデータを直接に併用する手法を提案している．著者らは，ラベルなしデータから求めた類似語を素性として追加して学習を行う方法を提案している [土屋 08]．図3は，少量のラベルありデータ中では非頻出（出現頻度が5未満）の単語「石狩」に対して，ラベルありデータ中に頻出（出現頻度が5以上）する単語の集合から，単語「石狩」に最も類似している単語「関東」を選び，素性として加えた学習データの例である．類似度には，大量のラベルなしデータから求めた周辺ベクトルの cosine 距離を用いている．ラベルあり新聞記事（10日分）とラベルなし新聞記事（6年分）を併用して実験したところ，ラベルあり新聞記事に出現しない個人情報で最も重要な未知の人名に対する検出性能（F 値）が 0.69 から 0.79 に向上した．

5・3 発話スタイルに含まれるプライバシ情報の保護

発話スタイルには，類義語の選択，機能語の選択，フィラーの挿入頻度などにおける地域的・個人的なくせという形で，プライバシ情報が含まれている．そのため，類義語や機能語の言換えを行い，発話スタイルの個人性を取り除いて，プライバシ情報を保護することが必要である．

発話スタイルの変換としては，話し言葉テキストと書き言葉テキストの変換が広く研究されている．[鍛治 04] は，待遇表現に着目して話し言葉コーパスおよび書き言葉コーパスをウェブから自動収集し，コーパスに基づいて，書き言葉特有の用言を話し言葉でも用いられる用言に言い換える手法を提案している．[下岡 04, 秋田 05] は，統計的機械翻訳の枠組みに基づいて，発話スタイルを変換する手法を提案している．この手法では，同一の内容に対する発話であるが，発話スタイルが異なるテキストからなるパラレルコーパスから変換モデルを学習し，その変換モデルを用いて話し言葉テキストを書き言葉テキストに整形したり，その逆の処理を行う．これらの方法により，話し言葉テキストを書き言葉テキストに変換すると，発話スタイルの個人性はかなり取り除かれる．

個人性が現れる表現を含まない制限言語 [Hujisen 98] を設計し，任意のテキストを，その制限言語内の表現に言い換えることによって，プライバシ情報を保護するという方法も考えられる．[Mitamura 01] は，機械翻訳の精度を向上することを目的として，機械翻訳の容易な表現からなる制限言語を設計し，制限言語に含まれない表現を自動的に検出・言い換える方法を提案している．

6. むすび

本解説では，音声に含まれるプライバシ情報の分類と，そのプライバシ情報の保護手法について，音声除去，声質変換，固有表現除去，発話スタイル変換の観点から述べた．要求される技術の多くは確立された技術ではないため，今後も新たな技術開発が必要となる．

謝辞

本研究は，文部科学省の科学技術振興調整費（科学技術連携施策群の効果的・効率的な推進）による「センサ情報の社会利用のためのコンテンツ化」の一環として実施したものである．

◇ 参考文献 ◇

[秋田 05] 秋田祐哉，河原達也：統計的機械翻訳の枠組みに基づく言語モデルの話し言葉スタイルへの変換，情処学研報，No. 2005-SLP-127, pp. 109-114（2005）

[Ando 05] Ando, R. K. and Zhang, T.: A high-performance semi-supervised learning method for text chunking, *Proc. ACL' 05*, pp. 1-9（2005）

[Araki 01] Araki, S., Makino, S., Mukai, R. and Saruwatari, H.: Equivalence between frequency domain blind source separation and fequency domain adaptive null beamformers,

Proc. EUROSPEECH2001, pp. 2595-2598 (2001)
[Arslan 97] Arslan, L. M. and Talkin, D.: Voice conversion by code-book mapping of line spectral frequencies and excitation spectrum, *Proc. EUROSPEECH' 97*, pp. 1347-1350 (1997)
[Benaroya 06] Benaroya, L., Bimbot, F. and Gribonval, R.: Audio source separation with a single sensor, *IEEE Trans. Audio, Speech, and Lang. Process.*, Vol. 14, No. 1, pp. 191-199 (2006)
[Boll 79] Boll, S. F.: Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. on Acoust., Speech, and Signal Process.*, Vol. ASSP-27, No. 2, pp. 113-120 (1979)
[Bregman 90] Bregman, A. S.: *Auditory Scene Analysis*, MIT Press (1990)
[Childers 85] Childers, D. G., Yegnanarayana, B. and Wu, K.: Voice conversion: Factors responsible for quality, *Proc. ICASSP' 85*, pp. 748-751 (1985)
[Danieli 08] Danieli, D. V.: Automatic censorship of audio data for broadcast, United States Patent PN/7,437,290 (2008)
[Droppo 01] Droppo, J., Deng, L. and Acero, A.: Evaluation of the SPLICE algorithmon on the Aurora2 Database, *Proc. EUROSPEECH2001*, pp. 217-220 (2001)
[Flanagan 85] Flanagan, J. L., Johnston, J. D., Zahn, R. and Elko, G. W.: Computer-steered microphone arrays for sound transduction in large rooms, *J. Acoustical Society of America*, Vol. 78, No. 5, pp. 1508-1518 (1985)
[藤崎 94] 藤崎博也:音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出, 信学技報, No. HC94-37, pp. 1-8 (1994)
[Griffiths 82] Griffiths, L. J. and Jim, C. W.: An alternative approach to linearly constrained adaptive beamforming, *IEEE Trans. Anntenas Propagation*, Vol. 30, No. 11, pp. 27-34 (1982)
[Hujisen 98] Hujisen, W.-O.: Controlled language: An introduction, *Proc. 2nd Int. Workshop on Controlled Language Applications* (*CLAW*), pp. 1-15 (1998)
[Isozaki 02] Isozaki, H. and Kazawa, H.: Efficient support vector classifiers for named entity recognition, *Proc. 19th Int. Conf. on Computational Linguistics*, pp. 1-7, Morristown, NJ, USA, Association for Computational Linguistics (2002)
[鍛治 04] 鍛治伸裕, 岡本雅史, 黒橋禎夫: WWW を用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え, 自然言語処理, Vol. 11, No. 9, pp. 19-37 (2004)
[小林 95] 小林大祐, 梶田将司, 武田一哉, 板倉文忠: ヒューマンスピーチライク雑音における音声的特徴の分析, 信学技報, No. SP95-105, pp. 85-92 (1995)
[Miller 04] Miller, S., Guinness, J. and Zamanian, A.: Name tagging with word clusters and discriminative training, *Proc. HLT-NAACL 2004*, pp. 337-342 (2004)
[Mitamura 01] Mitamura, T. and Nyberg, E.: Automatic rewriting for controlled language translation, automatic paraphrasing: Theories and applications, *Proc. NLPRS2001 Workshop*, pp. 1-12 (2001)
[西山 01] 西山 清: 最適フィルタリング, 培風館 (2001)
[猿渡 07] 猿渡 洋: 独立成分分析による音源分離技術, 電学誌, Vol. 127, No. 7, pp. 413-416 (2007)
[佐藤 08] 佐藤 洋, 清水 寧: スピーチプライバシー研究の歴史と近年の動向, 日本音響学会誌, Vol. 64, No. 8, pp. 475-480 (2008)
[Sekine 98] Sekine, S., Grishman, R. and Shinnou, H.: A decision tree method for finding and classifying names in Japanese texts, *Proc. 6th Workshop on Very Large Corpora 1998*, pp. 171-178 (1998)
[下岡 04] 下岡和也, 南條浩輝, 河原達也: 講演の書き起こしに対する統計的手法を用いた文体の整形, 自然言語処理, Vol. 11, No. 2, pp. 67-83 (2004)
[Stylianou 98] Stylianou, Y., Cappé, O. and Moulines, E.: Continuous probabilistic transform for voice conversion, *IEEE Trans. on Speech and Audio Process.*, Vol. 6, No. 2, pp. 131-142 (1998)
[Suzuki 08] Suzuki, J. and Isozaki, H.: Semi-supervised sequential labeling and segmentation using Giga-word scale unlabeled data, *Proc. ACL' 08-HLT*, pp. 665-673 (2008)
[橘 07] 橘 誠, 小林隆夫: 平均声モデルを用いる合成音声の話者性とスタイルの同時多様化の検討, 信学技報, No. SP2007-87, pp. 7-12 (2007)
[談 94] 談 旋, 松下貴光, 丁 文, 粕谷英樹: 話者認識に寄与する音源及び声道特徴の音響的性質, 信学技報, No. SP94-49, pp. 9-16 (1994)
[戸田 06] 戸田智基, 大谷大和, 鹿野清宏: 固有声に基づく声質変換法, 信学技報, No. SP2006-40, pp. 31-36 (2006)
[土屋 08] 土屋雅稔, 肥田新也, 中川聖一: 非頻出語に対して頑健な日本語固有表現の抽出, 情処学研報, No. 2008.NL.46, pp. 1-6 (2008)
[辻村 06] 辻村裕史, 秋葉友良: 音声文書を対象とした質問応答のための固有表現検出法の検討, 日本音響学会 2006 年秋季研究発表会講演論文集, pp. 143.144 (2006)
[内元 00] 内元清貴, 馬 青, 村田真樹, 小作浩美, 内山将夫, 井佐原均: 最大エントロピーモデルと書き換え規則に基づく固有表現抽出, 自然言語処理, Vol. 7, No. 2, pp. 63-90 (2000)
[山田 02] 山田寛康, 工藤 拓, 松本裕治: Support Vector Machine を用いた日本語固有表現抽出, 情処学論, Vol. 43, No. 1, pp. 44-53 (2002)
[山本 08] 山本一公, 土屋雅稔, 中川聖一: センサネットワークにおける音情報の扱いとそのプライバシー保護に関する検討, 日本音響学会 2008 年秋季研究発表会講演論文集, pp. 215-218 (2008)

2009 年 1 月 13 日 受理

著 者 紹 介

中川 聖一 (正会員)

1976 年京都大学大学院工学研究科博士課程修了. 同年, 京都大学情報工学科助手. 1980 年豊橋技術科学大学情報工学系講師. 1990 年教授. 1985 ～ 86 年カーネギメロン大学客員研究員. 音声情報処理, 自然言語処理, 人工知能の研究に従事. 工学博士. 1977 年電子通信学会論文賞, 1988 年 IETE 最優秀論文賞, 2001 年電子情報通信学会論文賞, 各受賞. 電子情報通信学会フェロー. 情報処理学会フェロー. 著書「確率モデルによる音声認識」(電子情報通信学会),「情報処理の基礎と応用」(近代科学社),「パターン情報処理」(丸善) など.

山本 一公

1995 年豊橋技術科学大学工学部卒業. 2000 年同大学院工学研究科博士後期課程電子・情報工学専攻修了. 博士 (工学). 2000 年信州大学工学部助手. 2007 年より豊橋技術科学大学情報工学系助教. 音声情報処理 (主に音声認識) に関する研究に従事. 日本音響学会, 電子情報通信学会, 情報処理学会各会員.

土屋 雅稔

1998 年京都大学工学部卒業. 2004 年同大学院情報学研究科知能情報学専攻博士課程単位認定退学. 博士 (情報学). 2004 年豊橋技術科学大学情報処理センター助手. 2007 年より同大学情報メディア基盤センター助教. 自然言語処理に関する研究に従事. 情報処理学会, 言語処理学会各会員.