

音声・音響分析技術と パブリックソリューションへの応用

越仲 孝文 宝珠山 治 大西 祥史 磯谷 亮輔 谷 真宏

要旨

ビッグデータ時代の到来とともに、実世界の膨大な情報を高速かつ安価に処理して有益な情報を得ようとするデータ分析はますます注目を集めています。実世界情報の重要な一部分である音声・音響情報を扱う技術もまた、データ分析の文脈で重要度をいっそう増しており、社会の安全・安心・効率・公平を支えるパブリックソリューションへの早期適用が期待されています。本稿では、音声・音響というメディアから意味のある情報を抽出する先端技術を、そのパブリックソリューションへの応用に触れつつ紹介します。



音声認識／話者認識／感情認識／雑音除去／音響イベント検知／ノイズサプレッサ／
ビームフォーミング／ビッグデータ

1. はじめに

人の声を扱う音声処理、音一般を扱う音響処理は、画像・映像処理と並び多くの社会課題の解決に貢献し得る有望なメディア処理技術です。ビッグデータ時代を迎え、膨大な量のマルチメディア情報がクラウドで高速かつ安価に処理可能となった昨今、これらの技術は実世界の大規模データから価値のある情報を獲得するデータ分析の手段としていっそうの注目を集めています。

音声・音響情報は、物理的には空気などの媒体の振動として伝わる波に過ぎません。しかしそこには、人間にとって意味のある情報が内在します。例えば音声であれば、言語によってさまざまに表現される言葉、話し手の性別や年齢からその時々感情まで、さまざまな情報が含まれます。音声以外の一般の音であっても、例えば小鳥のさえずりや木々を通る風の音などから、今そこにあるモノや起こっているコトが分かります。コンピュータの世界では、これらの情報は物理的な波というデータを分析することで抽出できます。

本稿では、音声・音響というメディアから意味のある情報を抽出するデータ分析の観点で音声・音響処理をとらえ、これを実現する先端技術を、そのパブリックソリューションへの応用に触れつつ紹介します。

2. 音声分析

2.1 テキストベース分析

人の声すなわち音声はさまざまな情報を運びます。なかでも最も豊富かつ重要な情報は、テキスト（文字）で表現される「言葉」でしょう。音声からテキストを抽出する音声認識の始まりは1950年代にさかのぼります¹⁾。現状の技術水準はまだまだ人間の聴取能力には遠く及ばないというのが学術界の共通認識ですが、一方で産業界では、スマートフォン（スマホ）のユーザーインターフェースとして近年急速に広まり、一般消費者には大変身近なものとなっています。

NECも音声認識の研究開発に早くから取り組んできました。近年は「人对機械から人对人へ」というビジョンのもと、単語や定型文をコンピュータに入力する一手段だけでなく、人と人の会話を理解するコンピュータの実現を目指し、研究開発と製品化に取り組んでいます（図1）。

その1つの実現形態が、議事録作成支援ソリューション「VoiceGraphy」です²⁾。本ソリューションは、議会、官公庁や企業の会議などでの人と人の会話を自動的に書き起こし、議事録作成に掛かる時間とコストを大幅に削減するものです。弊社の技術は、裁判所での尋問など従来音声認識が困難とされた場面でも実用化されています³⁾。



図1 音声認識は人对機械から人対人へ

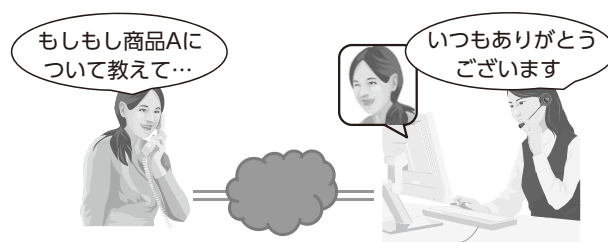


図2 音声から個人を特定する話者認識技術

2.2 非テキストベース分析

音声には、テキスト（文字）で表現される「言葉」以外にも、さまざまな情報が含まれています。ここでは特に、近年実用段階に入ってきた話者認識と感情認識について紹介します。

話者認識は、音声から個人を特定する技術です。音声は指紋、顔、静脈などに続く新しく手軽なバイオメトリクス（生体情報）として、産業応用の進展が期待されています。話者認識の応用は多岐にわたります。身代金要求の電話から誘拐犯を特定するなどの犯罪捜査や、公共の場での事件や事故の監視など、社会の安全・安心に関わるソリューションは最も有望視されるどころです。また、電話による銀行取引（テレホンバンキング）での本人確認やコールセンターでのお客様確認による対応品質向上など、民生用途への広がりも期待されます（図2）。

弊社では、かねてからこの技術の研究開発と製品化に取り組んでおり、長年の音声研究を通して培った高速かつ高精度な音声照合方式による大規模データベース検索が官公庁で実用化されています。

次に、感情認識に関する技術開発とコールセンター応用について紹介します。コールセンターは、企業にとってお客様の生の声を聴く重要な接点です。また、現在では多くの政令指定都市や公共機関がコールセンターを設置しており、市民の声を把握する場としても注目されています。

弊社では、コールセンターの音声に埋もれた不満や要望を効率的に把握する会話分析技術の研究開発を行っています。不満コール検知システム（図3）は、お客様の怒り感情を認識することで、不満コールを絞り込みます。更にオペレータの謝罪表現を認識します。オペレータは、「申し訳ありませんがご住所をお願いします」といった形式的な意味でも謝罪表現を用いるため、表層的な言葉だけでなく、深い感情のこもった謝罪のみを選別します。これにより、お客様の怒りとオペレータの深い謝罪が共に現れる先頭付近に集

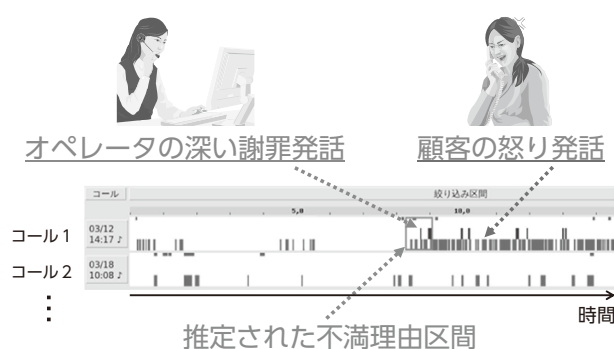


図3 不満コール検知システム

中する、不満の理由を含む会話区間（不満理由区間）の特定を可能としました。

更に、コールセンターなどの顧客接点での応用を念頭に、お客様の製品知識の有無を判別する技術を開発しました⁴⁾。これにより、お客様層の分析や相手の水準に合った応対といった応用が可能となります。このように感情認識の研究対象は近年広がりを見せており、将来は性格、体格、心身疾患などが音声から分かるようになるかもしれません。

2.3 雑音除去

テキストベースの分析、非テキストベースの分析のいずれにおいても、音声から情報を抽出する際、音声に重畳する雑音が課題となります。スマホで音声入力する際も、まわりがうるさいと誤認識が発生することがよくあります。

弊社は古くからこの課題に取り組み、2マイクを用いた雑音除去技術により工場・倉庫などの騒音下でも使用可能な音声認識装置「VoiceDo」を製品化しています。この技術を更に発展させ、「2マイク雑音抑圧」と「モデルベース音声強調」を組み合わせた新たな技術を開発しました（図4）。

本技術は、最新の2マイク雑音除去に加えて、静かな環境で想定される音声のモデル（知識）を用いた1マイクの音声強

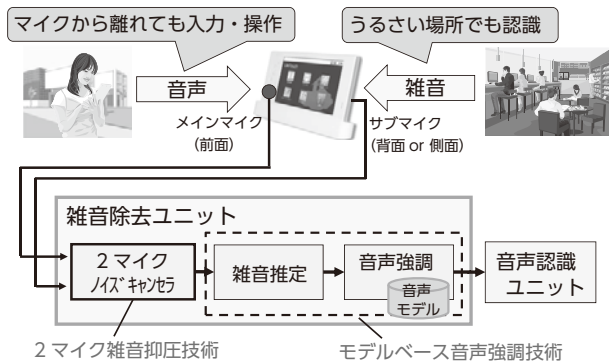


図4 音声認識向け雑音除去技術

調⁵⁾を適用することにより、雑音除去による音声の歪みを抑え、よりうるさい環境でマイクから離れて話した音声でも認識できるようにしました。本技術をカーナビ音声入力に適用し、従来に比べ5倍雑音が多い環境での音声認識を可能にしました。現在、公共エリアなどさまざまな環境での種々のテキスト・非テキスト分析への適用を目指し研究開発を進めています。

3. 音響分析

3.1 音響イベントの検知

実世界には、音声以外にもさまざまな音が存在します。このような一般の音をコンピュータで分析して「いつ」「どこで」「何が起こった」ととらえる試みが始まっています。

弊社は、一般の音から有用な情報を抽出する音響分析の研究開発に取り組んでおり、その一環として特定の事象（イベント）に関係する音を検知する音響イベント検知技術を開発しました。本技術は、例えば公共エリアに設置されたマイクでとらえた環境音から悲鳴、ガラスの割れる音などの異常な音をリアルタイムに検知、通知することができ、事件や事故の早期発見、早期解決に役立ちます（図5）。開発した技術では、検知したい目的音だけでなく、検知の邪魔になる他の音もシステムに学習させ、目的音とそれ以外の音の差異に着目することで、実世界の種々の音の中から目的音を正確に検知しています。さまざまな音が混在する公共エリアでの数か月間の実証実験で有効性を立証しました⁶⁾。

3.2 遠隔音の抽出

音響イベントの検知では、入力される音の明瞭さが検知性能を左右します。一般の音を扱う場合、音声と違ってマイ

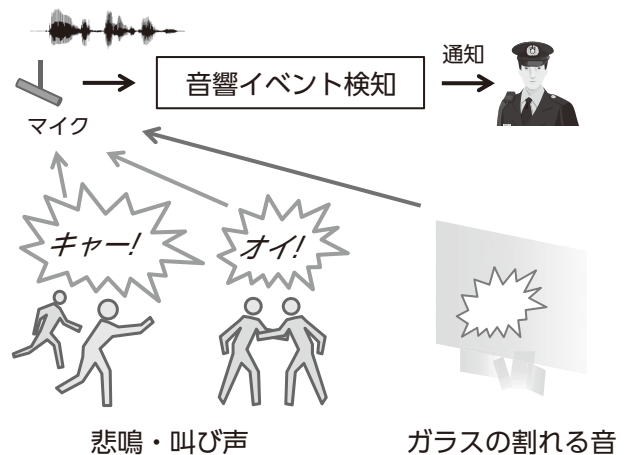


図5 音響イベント検知技術のユースケース

クの近くで音が発生するわけではなく、周囲の雑音や反射による残響などの外乱が混入します。これらの外乱を取り除き検知性能を向上すべく、ノイズサプレッサやビームフォーミングの研究開発を進めています。

ノイズサプレッサは名前のとおり雑音を抑圧する技術で、目的音と雑音の性質の違いを見分け、雑音だけを低減して欲しい音を抽出します。携帯電話による通話での周囲騒音抑圧などで豊富な実績があり、これまで難しいと言われてきた風音の抑圧にも最近成功しました⁷⁾。

ビームフォーミングは、複数のマイクを用いて特定の方向から到来する音だけを分離・抽出する技術です⁸⁾。明瞭な音が獲得できるだけでなく、音の到来方向が「どこで」を示す有用な情報として音響分析に寄与します。例えば街頭でガラスが割れた音がした場合、歩道から聞こえれば、家宅侵入や喧嘩の可能性がありますが、一方、車道から聞こえれば、事故や車上荒らしの可能性が高くなります。

本節で例示したような安全・安心に関わるソリューションでは、音響分析は映像分析と組み合わせることではしばしば相乗効果を発揮します。各種メディアを組み合わせたマルチモーダル分析についても目下検討を進めています。

4. おわりに

本稿では、音声・音響分析技術とそのパブリックソリューション領域への応用について俯瞰しました。音声・音響分析技術が今後いっそう、パブリックソリューションに活用され、各方面で社会課題の解決に貢献することを期待します。

参考文献

- 1) 古井貞照：コンピュータによる音声認識のこれまでと今後の展望，日本音響学会研究発表会講演論文集（春季），2011.3
- 2) 寺西博人：音声認識技術活用 議事録作成支援ソリューション「VoiceGraphy」2013年度グッドデザイン賞受賞，情報処理学会研究報告（音声言語情報処理），2014.10
- 3) 越仲孝文ほか：裁判員裁判向け音声認識システム，NEC 技報，Vol.63, No.1, 2010.2
- 4) 北出祐，大西祥史：電話問合せにおける発信者の知識量判別，日本音響学会研究発表会講演論文集（秋季），2014.9
- 5) 荒川隆行，辻川剛範：Model-Based Wiener Filterによる雑音下音声認識，NEC 技報，Vol.63, No.1, 2010.2
- 6) Paul Wang ほか：「MAG1C」における大規模メディア解析及び共有デジタルサイネージ機能，NEC 技報，Vol.67, No.1, 2014.11
- 7) M. Kato, & A. K. Sugiyama: A wind-noise suppressor based on wind-onset detection and spectral gain modification, Int' l Workshop on Acoustic Signal Enhancement, Sep., 2014.
- 8) O. Hoshuyama, A. K. Sugiyama: Robust adaptive beamforming," Book chapter of "Microphone Arrays", Editors: M. Brandstein & D. Ward, Springer, 2001.

執筆者プロフィール

越仲 孝文

情報・メディアプロセッシング研究所
主幹研究員
電子情報通信学会、日本音響学会
各会員
工学博士

宝珠山 治

情報・メディアプロセッシング研究所
主任研究員
電子情報通信学会、IEEE 各シニア
会員、日本音響学会、日本VR学会
各会員
工学博士

大西 祥史

情報・メディアプロセッシング研究所
主任研究員
日本音響学会会員
理学博士

磯谷 亮輔

情報・メディアプロセッシング研究所
主任研究員
電子情報通信学会、日本音響学会
各会員

谷 真宏

NEC Asia Pacific Pte. Ltd.
NEC Laboratories Singapore
主任

関連 URL

議事録作成支援ソリューション VoiceGraphy

<http://jpn.nec.com/voicegraphy/>

音声認識応答ソリューション VoiceDo

<http://jpn.nec.com/voicedo/>

NEC、自動車走行中の雑音が多い環境下においても、音声入力
で快適なカーナビ操作を実現する雑音除去技術を開発

http://jpn.nec.com/press/201404/20140418_01.html

NEC技報のご案内

NEC技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご覧ください。

NEC技報WEBサイトはこちら

NEC技報(日本語)

NEC Technical Journal(英語)

Vol.67 No.1 社会の安全・安心を支えるパブリックソリューション特集

社会の安全・安心を支えるパブリックソリューション特集によせて
NECが目指すパブリックソリューションの全体像
NECのパブリックセーフティへの取り組み

◆ 特集論文

効率・公平な暮らし

マイナンバー制度で実現される新しいサービス
ワールドカップを支えた「NECのスタジアム・ソリューション」
魅力あふれるフライトインフォメーションシステムの実現
駅の新サービス実現を加速するSDNソリューション
マルチデバイス対応テレビ電話通訳の通訳クラウドサービス
カラーユニバーサルデザインを採用した使いやすいスマートフォン向けネットバンキングサービス
安全・安心を実現する世界一の顔認証技術
顔認証製品と社会ソリューションでの活用

安全・安心な暮らし

ICTを活用したヘルスケアへの取り組み
組織間の安全な情報共有を実現する「MAG1C」の情報ガバナンスソリューション
「MAG1C」における大規模メディア解析及び共有デジタルサイネージ機能
シンガポールにおけるより安全な都市「セーフター・シティ」の構築
アルゼンチン ティグレ市の未来を守るビデオ解析ソリューション
群衆行動解析技術を用いた混雑推定システム
音声・音響分析技術とパブリックソリューションへの応用
昼夜を問わず24時間監視を実現する高感度カメラ
人命救助を支援するイメージソリューション
Emergency Mobile Radio Network based on Software-Defined Radio

重要インフラの安全・安心

新幹線の安全・安定輸送を支える情報制御監視システム
水資源の有効利用をICTで実現するスマートウォーターマネジメント技術の研究開発
センサとICTを融合させた漏水監視サービス
沿海域の重要施設へ接近する不審対象を監視する港湾監視システム
インバリアント解析技術(SIAT)を用いたプラント故障予兆監視システム
赤外線カメラの画像処理技術と応用例
高度化するサイバー攻撃への取り組み「サイバーセキュリティ・ファクトリー」

社会の安全・安心を支える先端技術

国家基盤を支える指紋認証の高速高精度化技術
次世代放送を支える超高精細映像圧縮技術とリアルタイム4K映像圧縮装置

◆ NEC Information

NEWS

NEC「衛星インテグレーションセンター」の稼働を開始
陸上自衛隊の活動を支援する「浄水セット・逆浸透2型」の開発



Vol.67 No.1
(2014年11月)

特集TOP