裁判員裁判向け音声認識システム

越仲 孝文・江森 正・大西 祥史

要　旨

2009年に始まった裁判員制度による裁判では、裁判に不慣れな裁判員の負担を軽減し円滑に裁判を進行することを狙い、音声認識技術を活用したシステムが導入されています。このシステムで用いられている音声認識技術は、2006年度から3年間に渡って最高裁判所の協力の下で研究開発した技術です。本稿では、裁判員裁判向け音声認識システムについてその機能を簡単に紹介するとともに、裁判の音声を直接認識できるようになった音声認識技術について、マルチチャネル音声検出技術を中心に３年間の研究開発成果について説明します。

キーワード

●裁判員制度 ●裁判音声認識 ●最高裁判所 ●マルチチャネル音声検出 ●評議

1. はじめに

2009年8月の東京地方裁判所における初の裁判員裁判以来、日本全国の地方裁判所において裁判員裁判が実施されています。この裁判員裁判では、裁判官と裁判員による評議などにおいて利用することを目的として、裁判員裁判向けの音声認識システムが導入されています。

この裁判員裁判向け音声認識システムで用いられている音声認識技術は、2006年度当初に最高裁判所が技術の研究開発業者を公募されて弊社が受注し、以来3年間に渡って最高裁判所の協力の下で研究開発した技術です。

本稿では、現在裁判員裁判で活用されている音声認識システムについてその機能を簡単にご紹介するとともに、裁判の音声を直接認識できるようになった音声認識技術について、性能向上の要であるマルチチャネル音声検出技術を中心に、3年間の研究開発成果について説明します。

2. 裁判員裁判向け音声認識システム

まず、システムの用途と機能についてご説明します。

2.1 システムによる裁判員裁判の支援

本システムが支援する裁判の場面は、証人や被告人に対し尋問などが行われる公判と、その後裁判官と裁判員のみで判決内容を決定する評議の2場面になります。裁判に不慣れな裁判員の方は、数時間を超える公判の際の発言を記憶しておき評議の際に正確に思い出すことが困難であろうことから、記憶喚起の必要が生じた場合に備えて公判の音声・映像を記録しておき、その検索・頭出しに音声認識技術を活用しています。

2.2 システムの機能

具体的には、公判の際の裁判官・裁判員、検察官、弁護人、証人・被告人の発言を音声として、証人・被告人の表情を映像として記憶しておきます。また音声・映像の記録と同時に入力音声に対し音声認識を実行し、音声認識結果をその音声の時刻情報とともに記録しておきます。評議の場面で公判を振り返る際には、その振り返りたいシーンの前後で使われたキーワードを入力し検索すると、そのキーワードを音声認識

図1　公判の場面での音声認識画面

図2　評議の場面での検索・再生画面

結果の中で検索し、見つかったキーワードの時刻情報を取り出し、その時刻から音声及び映像を再生することで、容易に音声・映像の頭出し再生が行えるようになっています。

キーワードが公判の複数個所で見つかる場合もありますが、検索条件に発言者や時刻範囲情報などを追加して絞り込むことも可能となっており、また検索された複数個所についてその前後の発言が音声認識結果で確認できるため、適切な発言を容易に選択しそこから音声・映像が頭出し再生できるようになっています。

図1 が公判の場面でリアルタイムに音声認識を実行する画面で、 **図2** が評議の場面での検索・再生の画面となります。

3. 裁判環境に対する音声認識技術

裁判で交わされる直接の会話を認識するためには、従来の音声認識技術を導入するだけでは不十分で、法廷環境や公判運営を考慮した音声認識技術を新規に研究開発する必要がありました。NECでは入力となるマイク機器から出力となる音声認識結果の可読性向上まで、技術実用化に向け研究開発に取り組みました。

とりわけ、発話と発話の重なりがしばしば発生するような会話音声に対して、その認識性能を著しく向上させるマルチチャネル音声検出技術を、認識性能向上の中心技術として実現しましたので、この技術を中心に研究開発した技術について説明します。

3.1 マルチチャネル音声検出技術

裁判環境では、1つの法廷に複数のマイクがあり、主に証人もしくは被告人に対するほかの3者（裁判官・裁判員、検察官、弁護人）からの質問によって裁判が進行します。この中で、ある話者の発話の終端と次の別の話者の発話の始端が重なるという現象がしばしば見られます。

このような会話のマイク音声を1チャネルにミキシングして音声認識した場合、複数の人の発言をつなげて1人の1回の発話として認識してしまうので、認識性能が著しく低下します。このような認識率低下が生じることがあらかじめ予想できたので、私たちは裁判で発言する人の主な4種類の役割ごとに1つの音声チャネルを割り当て、4チャネル同時に並行して音声認識する方式を採用しました。

しかし、このように4チャネル同時並行に認識しようとすると、裁判官の発言が証人のマイクにも回り込んで入ってしまい、証人チャネルの音声認識結果に不要な認識結果文字列が生じてしまうという問題が生じます。もちろん複数のチャネルで音声検出を排他的に起動することでこの問題を回避する方法もありますが、それでは発話終端と次の発話始端とが重なった2人の音声を並列で音声認識処理する意義が薄れますので、音声検出結果の音声認識処理として「2人が同時に発声している場合には2つの処理を並行動作させ、1人の発声が2つ以上のマイクに回り込んでいる場合には1つの処理のみを動作させる」ことを実現するマルチチャネル音声検出技術を新規に開発しました。

この技術を実現するに至った着眼点は、1人の音声が回り込んで2つ以上のマイクに入った場合には、元の音が同じ音声ですので各マイクの音を周波数軸上で比較するとマイクから遠い音が近い音に包含される形状となりますが、2人が同時に発声している場合には、それら2つの音は別の声質の音声で、かつ発声した音（あいうえお…）もほぼ間違いなく別音となるので、おのおののマイクに入った音の波形で比較すると、波形が包含関係とならずどこかで交差することになり、その交差関係を抽出すると2人の同時発話が検出可能となる、というものです。

この着眼点に基づいた処理を高速な処理として実現するため、おのおののマイクに入った音の比較を一定幅の周波数帯（サブバンド）に区切り、そのサブバンドごとの音のパワー（音量）を比較するというのがマルチチャネル音声検出技術

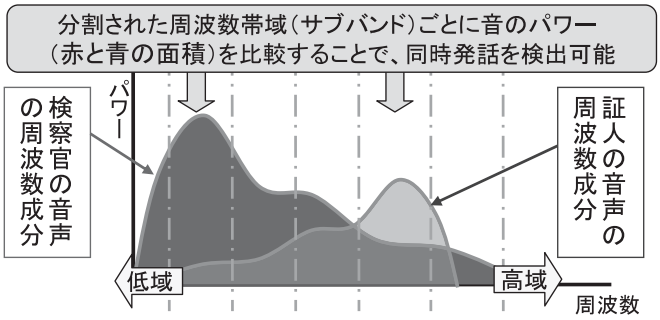

図3　マルチチャネル音声検出技術

です。図3がその検出の概念図となっています。

3.2 その他の新規開発技術

裁判音声認識技術の性能を向上させ実用化するために、入力から出力まで全般にわたる技術開発が必要となりました。これらすべての詳細については本稿では省略しますが、業界的に実用化例があまり見かけられない技術開発成果として、以下が挙げられます[1, 2]。

(1) 音声入力系

- Bluetoothワイヤレスピンマイクの採用
- 24bitワイドレンジ音声の採用

(2) 音声認識アルゴリズム

- リアルタイム教師なし話者適応技術
- 高速声道長正規化技術

(3) 音響/言語モデル開発

- 役割別音響/言語モデル
- 関西語への言語モデル変換技術

(4) 可読性向上

- 句読点自動挿入、フィラー除去

上記技術をすべて研究開発し裁判音声認識システムへ統合することにより、裁判員裁判において実用可能な音声認識システムを構築することができました。

4. おわりに

実裁判の音声データを対象とした評価では、2006年度の開始時に認識率60%を割っていた音声認識エンジンに対し、前述したような新規に技術開発した技術・モデルを投入しました。裁判音声・映像の検索・頭出し再生機能として活用する際には、おおよそ8割程度の認識率であれば実用に耐え得ると考えられるのですが、2008年後半には一定程度の認識率を実現できたことが、裁判員裁判での音声認識システム実用化のポイントであったと考えています。

今後の方向性としては、裁判音声認識で開発した技術を一般企業の会議へ展開することが考えられます。一般会議への音声認識技術の導入を考えた際には、認識性能をある程度以上に高く維持するために、部屋の中に複数個のマイクを設置することが考えられます。そこで複数の人が会話するような環境では、本稿で説明したマルチチャネル音声検出技術の利用が不可欠になると思われます。本システムの研究開発で実現したマルチチャネル音声検出などの技術を搭載した音声認識エンジンを活用し、音声認識活用事業を各種の会議へ展開していく予定です。

5. 謝辞

本裁判員裁判向け音声認識システムの3年間の研究開発期間の全般にわたり、最高裁判所事務総局総務局の方々には研究開発及びシステムの設計・開発・検証・運用のすべてに対しご指導・ご鞭撻を賜りました。この場を借りて深く感謝を申し上げたいと存じます。

参考文献

1) T. Koshinaka et al. ; " Online speaker clustering using incremental learning of an ergodic hidden Markov model" ,Proc. of ICASSP, 2009.
2) T. Emori et al. ; " Vocal tract length normalization using rapid max - imum-likelihood estimation for speech recognition" ,Proc. of Euro - speech 2001.

執筆者プロフィール

越仲 孝文
共通基盤ソフトウェア研究所
電子情報通信学会
日本音響学会各会員

江森 正
NEC情報システムズ
先端技術ソリューション事業部
日本音響学会会員

大西 祥史
共通基盤ソフトウェア研究所
日本音響学会
日本物理学会各会員