

ICASSP2007 報告

戸田 智基[†] 篠崎 隆宏[‡] 秋田 祐哉[‡]

[†]奈良先端科学技術大学院大学 情報科学研究科
〒630-0192 奈良県生駒市高山町 8916-5
[‡]京都大学 学術情報メディアセンター
〒606-8501 京都市左京区吉田二本松町

あらまし 本稿では、2007年4月に米国ハワイ州ホノルルにて開催された国際会議 IEEE ICASSP2007 の報告を行う。

ICASSP2007 Report

Tomoki TODA[†] Takahiro SHINOZAKI[‡] Yuya AKITA[‡]

[†]Graduate School of Information Science, Nara Institute of Science and Technology,
8916-5, Takayama-cho, Ikoma-shi, Nara 630-0192, Japan,

[‡]Academic Center for Computing and Media Studies, Kyoto University,
Sakyo-ku, Kyoto 606-8501, Japan

Abstract This paper reports IEEE International Conference on Acoustics, Speech and Signal Processing held at Honolulu, Hawai'i, USA, in April 2007.

1 はじめに

本稿では、2007年4月15日から20日にかけてアメリカ合衆国ハワイ州ホノルルにて開催されたICASSP2007の報告を行う。各著者の担当分野は次の通りである。

- 戸田 音声合成・符号化など、学生コンテスト
- 篠崎 音声認識（音響モデル、識別学習、ロボストネスなど）
- 秋田 音声認識（言語モデル、システム）・音声言語処理

2 音声合成・音声符号化

音声合成及び音声符号化関連では、2つのオーラルセッションと3つのポスターセッション及び1つのスペシャルセッションにて、計45件（音声合成：27件、音声符号化：18件）の論文発表が行われた。オーラルセッション会場は、十分な聴講スペースが設けられるだけの広さを持っていた。ポスターに関しては1セッションあたり発表件数が10本以内に抑えられており、議論するのに十分なスペースが確保されていた。全ての発表会場が同一フロアにあり、非常に移動しやすい配置となっていた。

ここ数年の傾向といえるが、多様性に富む音声の合成技術が注目を集めている印象を受けた。特に、統計的手法に基づく音声合成・変換技術は盛んに研究されており、Statistical Parametric Speech Synthesis と題したスペシャルセッションが行われた。音声認識の分野と比較すると研究者数は少ないものの、広い会場に見合う数の聴講者が参加していた事から、本研究分野の注目度の高さが十分に伺えた。オーラルセッションにおいて、20代から30代の若い研究者が積極的に質問する場面が多く見られ、今後この分野のさらなる発展が大いに期待できると感じた。

近年、注目を集めているHMMに基づく音声合成技術[1]に関しては、取り組む研究者の数がここ数年で大幅に上昇していると感じられた。上述のスペシャルセッションにおいては、名高い音声符号化の研究者の方々も多数聴講された

と伺った。統計処理のみでなく音声信号処理も重要な役割を担う技術であるため、今後符号化の知識に富んだ研究者の方々の参入により、さらなる発展がもたらされるのではないかと予想される。

今回のICASSPに限らず近年の国際会議でよく感じる事であるが、中国からの研究発表は目を見張るものがある。例えば、音声合成に関連する論文に絞って見ると、一般セッションにて発表された論文21件の国別の内訳は、

7件：中国(1)

3件：日本(3), アメリカ(2)

2件：ベルギー

1件：台湾, インド, フランス, スペイン,
ギリシャ, フィンランド

であった。ここで、括弧内の数字はスペシャルセッションにて発表された論文6件の内訳である。明らかに、中国からの発表件数が多いことが分かる。なお、単に発表件数が多いだけではなく、その内容も理論的に優れているものが多い。今回の会議で一際目を引いた論文として、1) HMMに基づく音声合成において、HMMの全学習過程をパラメータ生成誤差最小化基準で行うもの[2]や、2) HMM尤度に基づくフレーム単位素片選択において、音素単位素片の併用を導入することで高速化を図るもの[3]の2つが挙げられるが、どちらも中国からの発表である。

(戸田智基)

3 音響モデル

音響モデルに関連した発表は100件ほどであった。その内容は主に、学習アルゴリズム、話者適応、頑健性、特徴量抽出、隠れマルコフモデルによらない又はそれを拡張した統計的モデル化方法、マルチモーダルなどの複数特徴量の利用に関するもの、システムコンビネーションなどであった。図1に、これら話題を大まかに分類して示す。

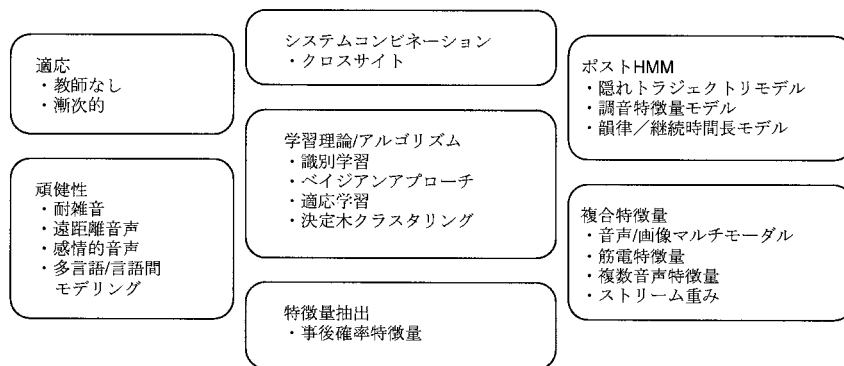


図 1: 音響モデルに関する話題

会議では特に、最近のトレンドである識別学習を中心とした学習アルゴリズムに関する話題が、多く注目を集めていた。識別学習関連の発表内容としては、定式化のバリエーションの比較検討や、新しい定式化法の提案などであった。話者適応では教師なし適応や逐次適応などに関する発表が行なわれた。頑健性に関する研究では、雑音の重畳や話者の感情変化が声質に与える影響、離れた位置からの音声の入力、言語をまたいでの音声の利用や認識などが話題となっていた。特徴量抽出では、事後確率特徴量が普及しつつある状況が見られた。従来の隠れマルコフモデルによらない音声のモデル化手法としては、隠れトラジェクトリモデル、ベイジアンネットを用いた調音特徴量モデル、韻律や音素継続時間長を組み込んだモデルなどに関する発表があった。調音特徴量モデルについては、ジョンズホプキンス大での夏季ワークショップの報告に比較的多くの人が集まっていたようである。システムコンビネーションやマルチモーダルは、話題的には一時期に比べるとやや落ち着いてきたような印象を受けた。

(篠崎隆宏)

4 言語モデル・システム

言語モデルのセッションは口頭発表およびポスター発表の計 2 件で、このうち言語モデルそのものに関する発表は 10 件程度であった。発表

内容はモデル化手法と適応手法に大別できるが、新たなモデル化の観点から興味深い発表として混合ガウス分布に基づくモデル [4] が挙げられる。一方適応手法としては、潜在ディリクレ分析 (LDA) などの潜在的な (話題の) モデルを用いる手法が一般化しており、今回の会議においても報告があった。ただし言語モデルの話題は比較的少なく、一定の聴衆を集めてはいたものの盛り上がりは今ひとつに見受けられた。

音声認識システム (LVCSR/ASR) 関連では、口頭セッションおよびポスターセッション (4 件) の計 5 セッションが設けられた。毎回恒例ながら、現在欧米で進行中の大型プロジェクトに関するシステムの発表があり、GALE に関連するアラビア語や中国語のシステム、AMI や TC-STAR に関するシステムが紹介された。また複数の認識システムの出力を統合する手法の提案も複数見られた。研究機関が共同で進める現在の研究トレンドもあり、複数システムの結合や出力の補正といった研究に関心が集まっているようである。なお、ポスターセッションのうち 1 件は “Resource Constrained” すなわち携帯電話などの計算資源に制約のある音声認識を対象としたもので、デモなども見られた。

(秋田祐哉)

5 音声言語処理

音声言語理解に関しては口頭発表およびポスター発表の計2セッションで、先述の言語モデルセッションで扱われていた分を含めるとおよそ20件の発表があった。主要な話題として、文境界や品詞タグなどのアノテーション、また対話行為の推定手法についてそれぞれ複数の発表が見られた。音声対話関連では、対話管理の学習手法への関心の高まりを反映して、部分観測マルコフ決定過程 (POMDP) に基づく対話管理手法 [5] が多くの聴衆を集めていたのが印象的であった。

音声翻訳については、ポスターセッションに加えて音声認識と機械翻訳の統合に関するスペシャルセッションが設けられており、計12件の発表があった。これらに聴衆が多数集まっていたことから、音声翻訳に対する音声コミュニティの関心の高さがうかがわれる。ただし、特にスペシャルセッションでは音声認識システムと機械翻訳システムの個別の説明に多くの時間が割かれ、統合に関する考察や議論、質疑は低調であった。GALEやTC-STARなどの関連プロジェクトが推進中ということもあり、今後の発展が期待される。

(秋田祐哉)

6 Student Paper Contest Awards

ICASSPでは、優れた論文を発表した学生に対して賞を授与する制度があり、受賞者の選定システムは毎年異なっている。今年は、各分野のTechnical Committeeから選出されたファイナリスト計54名の中から最優秀論文発表者を決定するために、1日をかけて審査会が行われた。ファイナリストは研究分野に基づいて3つのグループに分けられ、自身の論文に関して10分程度のプレゼンテーションと5分程度の質疑応答を行った。1グループにつき3名の審査委員により、論文の内容及びプレゼンテーション能力などが総合的に評価された。最終的に各グループから3名ないし4名が厳選され、その中からFirst Prizeとして1名、Second Prizeとして

2名、Third Prizeとして3名、及びHonorable Mentionと称してさらに4名が選出された。音声言語の分野からはファイナリストとして9名が選ばれ、その中の1名がThird Prizeを受賞した [6]。

(戸田智基)

参考文献

- [1] K. Tokuda, H. Zen, J. Yamagishi, T. Masuko, S. Sako, A.W. Black, and T. Nose. The HMM-based speech synthesis system (HTS). <http://hts.sp.nitech.ac.jp/>.
- [2] Y.-J. Wu, R.-H. Wang, and F. Soong. Full HMM training for minimum generation error in synthesis. *Proc. ICASSP*, vol. 4, pp. 517–520, Hawaii, USA, Apr. 2007.
- [3] Z.-H. Ling and R.-H. Wang. HMM-based hierarchical unit selection combining Kullback-Leibler divergence with likelihood criterion. *Proc. ICASSP*, vol. 4, pp. 1245–1248, Hawaii, USA, Apr. 2007.
- [4] M. Afify, O. Siohan and R. Sarikaya. Gaussian mixture language models for speech recognition. *Proc. ICASSP*, vol. 4, pp. 29–32, Hawaii, USA, Apr. 2007.
- [5] S. Young, J. Schatzmann, K. Weilhammer and H. Ye. The hidden information state approach to dialog management. *Proc. ICASSP*, vol. 4, pp. 149–152, Hawaii, USA, Apr. 2007.
- [6] <http://www.icassp2007.org/SPCWinners.asp>.