

論文 / 著書情報  
Article / Book Information

|                   |                                                                                                                                                                                                |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 題目(和文)            | DNN音声合成の多様化と音声対話システムへの応用に関する研究                                                                                                                                                                 |
| Title(English)    | A Study on DNN-Based Speech Synthesis with Diverse Voices and Its Application to Spoken Dialogue Systems                                                                                       |
| 著者(和文)            | 北条伸克                                                                                                                                                                                           |
| Author(English)   | Nobukatsu Hojo                                                                                                                                                                                 |
| 出典(和文)            | 学位:博士(工学),<br>学位授与機関:東京工業大学,<br>報告番号:甲第11435号,<br>授与年月日:2020年3月26日,<br>学位の種別:課程博士,<br>審査員:小林 隆夫,山口 雅浩,奥村 学,杉野 暢彦,篠崎 隆宏                                                                         |
| Citation(English) | Degree:Doctor (Engineering),<br>Conferring organization: Tokyo Institute of Technology,<br>Report number:甲第11435号,<br>Conferred date:2020/3/26,<br>Degree Type:Course doctor,<br>Examiner:,,,, |
| 学位種別(和文)          | 博士論文                                                                                                                                                                                           |
| Category(English) | Doctoral Thesis                                                                                                                                                                                |
| 種別(和文)            | 論文要旨                                                                                                                                                                                           |
| Type(English)     | Summary                                                                                                                                                                                        |

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

|                                            |              |          |                                                                        |
|--------------------------------------------|--------------|----------|------------------------------------------------------------------------|
| 系・コース：<br>Department of, Graduate major in | 情報通信<br>情報通信 | 系<br>コース | 申請学位(専攻分野)： 博士 (工学)<br>Academic Degree Requested Doctor of Engineering |
| 学生氏名：<br>Student's Name                    | 北条 伸克        |          | 指導教員(主)：<br>Academic Supervisor(main) 小林 隆夫 教授                         |
|                                            |              |          | 指導教員(副)：<br>Academic Supervisor(sub)                                   |

要旨(和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

本論文は、音声対話システムで使用されるテキスト音声合成 (text-to-speech; TTS) に関する検討を行う。本研究の目的は、「合成音声から、システムの発話意図がいかに関に自然に感じられるか」という尺度に基づき、合成音声の品質を改善することである。本研究では、この尺度を「発話内行為自然性 (illocutionary act naturalness; IAN)」と呼ぶ。

第 1 章では、本論文の背景および研究目的について述べる。

第 2 章では、まず、多様な音声を合成するため、DNN 音声合成による複数話者モデルを検討する。多様な話者や発話スタイルの音声を合成するためには、全体のデータコストを考慮した場合、一つの話者や発話スタイルの学習データが少量であっても高品質な音声を合成可能な技術が必要である。このため、本章では、まず、複数の話者の音声コーパスを使用し、合成音声の品質を改善するための検討を行う。本章の提案法は、DNN 音声合成の枠組みにおいて、話者の情報をベクトル (話者コード) により表現し、入力情報として使用する。本論文では、客観評価実験および主観評価実験により、提案法を従来法と比較する。客観評価実験により、提案法は、従来の話者依存 DNN および DNN による複数話者モデリングの従来法 (隠れ層共有モデル) に比べ、特徴量の生成誤差を改善することが明らかとなった。また、主観評価実験により、提案法は、従来の話者依存 DNN および HMM に基づくモデル適応手法を上回る自然性を示すことが明らかとなった。提案法による自然性の改善は、特に、目標話者の学習データが少量の場合に顕著であることが明らかとなった。

第 3 章では、発話意図を表現する音声合成システムを構築し、発話内行為自然性に関する評価を行う。本研究では、対話システムの発話意図として、雑談対話システム等で使用される対話行為タグの使用を検討する。本章では、まず、対話行為タグが付与された音声コーパスを構築する。続いて、構築した音声コーパスに基づき、音声合成システムを構築する。本章では、第 2 章の提案法における話者コードを対話行為コードに置き換えたモデルを提案する。評価実験では、従来の HMM に基づく手法 (スタイル混合モデル手法およびモデル適応手法) と提案法を比較する。客観評価実験では、大域的韻律 (発話の平均 F0 や平均継続長) および文末音調 (発話末尾 1 音節の F0) の再現性を評価する。主観評価実験では、発話内行為自然性を評価する。客観評価実験により、スタイル混合モデルに比べ、提案法は、発話意図の大域的な韻律的特徴の再現性を改善することが明らかとなった。一方、モデル適応手法に比べ、提案法は、文末音調の再現性を改善することが明らかとなった。主観評価実験では、提案法は、二つの従来法よりも高い発話内行為自然性を示すことが明らかとなった。客観評価実験および主観評価実験の結果を比較すると、大域的な韻律的特徴および文末音調の特徴の再現性が向上す

る場合に、発語内行為自然性が改善される傾向が示された。これらの再現性が、発語内行為自然性の向上に寄与することが示された。最後に、対話行為を考慮しない従来法と提案法を比較すると、提案法により発語内行為自然性が改善されることが示された。この結果から、対話行為を考慮することで、発語内行為自然性が改善されることが示された。一方、以上の主観評価実験では、評価コスト等の理由により、評価に使用可能な発話の数に制約があった。このため、様々な状況の発話を評価し、提案法の効果や課題を詳細に分析することが困難であるという課題があった。

第4章では、音声合成のために対話行為を使用する手法の効果や課題を詳細に分析するため、発話の文末音調ラベルを推定する問題に焦点を当て、評価結果を分析する。文末音調ラベルは、客観評価により十分な妥当性の評価が可能であると考えられるため、交差検定により音声コーパス中の全ての発話を評価に使用することができる。このため、様々な状況の発話について、対話行為を考慮する音声合成の効果や課題を分析することができる。本章では、まず、対話行為を考慮する文末音調ラベルの推定手法を検討し、評価結果を分析する。検討のため、第3章で構築した音声コーパスに対し、文末音調ラベルを付与し、文末音調ラベル推定用のコーパスを構築する。提案法では、従来法で使用される、発話文章の形態素および品詞などの特徴量に加え、対話行為タグを文末音調ラベルの推定のために使用する。客観評価により、提案法では、推定ラベルと正解ラベルの一致係数 (Cohen の  $\kappa$ ) が 0.12 改善されることを示した。これにより、提案法により対話行為を考慮する有用性が示された。また、実験結果の分析により、提案法による対話行為の考慮は、主に3種類の表現 (「自問納得」「上昇イントネーションによる疑問表現」「自問/フィラー」) のために有用であることが明らかとなった。また、提案法により文末音調ラベルが正しく推定されなかった発話を分析した。分析の結果、推定結果を改善するためには、「話者の感情を表明する/しない自己開示」「驚き・意外性を表明する/しない繰り返し」等、より詳細な意図を対話行為タグにより表現する有用性が示された。将来的に、音声対話システムにおいてこれらの対話行為タグを使用し、対話制御および音声合成のために使用することで、合成音声により自然にシステムの発話意図を表現することが可能となると考えられる。

第5章では、本論文の結論および今後の課題について述べる。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1 copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ (T2R2) にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).