

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), 学位授与機関:東京工業大学, 報告番号:甲第11435号, 授与年月日:2020年3月26日, 学位の種別:課程博士, 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), Conferring organization: Tokyo Institute of Technology, Report number:甲第11435号, Conferred date:2020/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第		号	学位申請者氏名		北条 伸克	
論文審査 審査員		氏 名		職 名		氏 名	職 名
	主査	小林 隆夫		教 授		篠崎 隆宏	准教授
	審査員	山口 雅浩		教 授	審査員		
		奥村 学		教 授			
		杉野 暢彦		教 授			

論文審査の要旨（2000 字程度）

本論文は「A Study on DNN-Based Speech Synthesis with Diverse Voices and Its Application to Spoken Dialogue Systems (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)」と題し、英文5章からなっている。

第1章「Introduction (序論)」では、最近のテキスト音声合成 (text-to-speech, TTS) 技術を概観した後、TTS を音声対話システムに使用する場合、合成音声の品質を評価する際に「システムの発話意図がいかに自然に感じられるか」という尺度が重要であることを述べている。本論文ではこの評価尺度を発語内行為自然性 (illocutionary act naturalness) と呼び、この評価尺度の観点から合成音声の品質を向上させることが研究目的であると述べている。

第2章「DNN-based speech synthesis using speaker codes (話者コードを用いた DNN 音声合成)」では、まず多様な話者の音声合成可能にするため、深層ニューラルネットワーク (DNN) に基づいた音声合成の枠組みを用いた新たなモデリング手法を提案している。提案法では、話者の情報をベクトル (話者コード) により表現し、DNN の入力情報として使用する。客観評価実験により、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、提案法は特徴量の生成誤差が改善されることを示し、さらに主観評価実験により、提案法は従来の話者依存 DNN および隠れマルコフモデル (HMM) に基づくモデル適応手法を上回る自然性を示すこと、自然性の改善は目標話者の学習データが少量の場合に顕著であることを明らかにしている。

第3章「DNN-based speech synthesis using dialogue-act information and its evaluation with respect to illocutionary act naturalness (対話行為情報を用いた DNN 音声合成と発話行為自然性に関する評価)」では、発話意図を表現する TTS システムを構築し、発語内行為自然性に関する評価を行っている。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ている対話行為タグを利用することを考え、話者コードの代わりに対話行為コードを用いてモデルを学習することを提案している。提案手法の有効性を示すために、対話行為タグが付与された音声コーパスを作成し、これを用いて音声合成システムを構築している。HMM に基づく従来手法 (スタイル混合モデル手法及びモデル適応手法) と比較した結果、客観評価では提案法はスタイル混合モデルに比べ、発話意図の大域的な韻律的特徴 (発話の平均基本周波数や平均継続長) の再現性を改善することを明らかにしている。また主観評価では、発語内行為自然性において、提案法はスタイル混合モデルに比べ、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を明らかにすると共に、モデル適応手法に比べて文末音調の再現性を改善すること、さらに従来法よりも高い発語内行為自然性を示すことを示している。

第4章「Sentence-final tone labels estimation using dialogue-act information (対話行為情報を用いた文末音調ラベル推定)」では、音声合成に対話行為情報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、評価結果を分析している。ここでは、まず対話行為を考慮する文末音調ラベルの推定手法の検討及び評価結果を分析するため、第3章で構築した音声コーパスに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している、そして、従来法で使用されている発話文章の形態素および品詞などの特徴量に加え、対話行為情報を使用する文末音調ラベル推定手法を提案している。客観評価により、提案法により推定ラベルと正解ラベルの一致係数 (Cohen の κ) が 0.12 改善されることを示し、対話行為を考慮する有用性を示している。また、実験結果の分析により、提案法による対話行為情報の考慮は、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を明らかにしている。

第5章「Conclusion and future work (結論と今後の課題)」では、本論文で得られた成果をまとめ、今後の課題を述べている。

以上を要するに、本論文では発語内行為自然性の観点から音声合成システムの品質を向上させることを目的として、話者コード情報や対話行為情報を利用する音声合成システムを提案し、有効性を示したものであり、工学上ならびに工業上貢献するところが大きい。よって本論文は博士 (工学) の学位論文として価値あるものと認められる。

注意：「論文審査の要旨及び審査員」は、東工大リサーチポジトリ (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。