様式３

論　文　内　容　の　要　旨

氏　名　（　呉　　伯　文　）	
論文題名	Co-speech Gesture Generation using Deep Generative Models (深層生成モデルを用いた音声に伴うジェスチャの生成)

論文内容の要旨

Co-speech gestures, spontaneous hand movements accompanying speech, complement and reinforce spoken messages. Their prominence in human interactions extends to human-like robots and avatars, particularly those with arm-like features. Integrating co-speech gestures in these agents is vital for enhancing the naturalness and effectiveness of human-robot or human-avatar communication.

This research aims to develop a co-speech gesture generation system for humanoid robots and avatars based on speech inputs, addressing the following key challenges: generating natural motions, expressing intrinsic character, ensuring applicability, and being coherent with context. The core of this thesis is the adoption of deep generative models for gesture generation, chosen for their advanced capability in capturing complex patterns and generating diverse outcomes compared to deterministic deep learning methods. Specifically, the thesis approaches the goal by making the following contributions: To enhance naturalness, where prior models showed limitations due to deterministic approaches, this thesis introduces a conditional generative adversarial network-based probabilistic generation method. This model improves gesture naturalness, synchronizes better with speech, and produces a wider range of plausible gestures. In addressing the intrinsic character, the thesis tackles the challenge of generating gestures that express specific personality traits, a task hindered by the lack of personality-labeled training data in existing methods. A novel approach involving automatic labelling and a Wasserstein generative adversarial network-based model is proposed, facilitating the generation of gestures that align with speech and convey personality nuances like extroversion or introversion. Regarding the applicability, especially in generating gestures within physical constraints like confined spaces, traditional methods often result in unnatural, truncated movements. This thesis proposes a new diffusion model combined with a sampling algorithm, aiming to ensure the generation of natural and coherent gestures within such constraints.

In summary, this thesis introduces methods addressing key challenges in co-speech gesture generation through deep generative models. These approaches significantly enhance gesture naturalness, effectively capture intrinsic character traits, and ensure practical applicability under physical constraints. The systems developed in this thesis show considerable promise in improving interactions between humans and robots or avatars.

様式７

論文審査の結果の要旨及び担当者

氏　名　（　呉　伯　文　）		
	（職）	氏　名
論文審査担当者	主 査　教 授	石黒　浩
	副 査　教 授	飯國　洋二
	副 査　教 授	佐藤　宏介

論文審査の結果の要旨

人の発話に伴う自発的な手の動き（ジェスチャ）は、発話内容を補完する機能があり、円滑なコミュニケーションを行う上で重要である。音声対話システムでは、発話に伴うジェスチャによる効果を望めないが、人間型のロボットやアバターが人間と自然に，円滑にコミュニケーションを行うためには，このような発話に伴うジェスチャを生成することが望まれる。

本研究は、人間型ロボットおよびアバターの発話に伴うジェスチャを、発話情報から自動的に生成するシステムを開発することを目指しており、自然な動きの生成、動きによる個性の表現、適用性の向上、文脈に一貫性のある動きの生成などの重要な課題に取り組んでいる。これらの課題を解決するために本論文で提案した手法の核心部分は、複雑なパターンを捉え、決定論的な深層学習方法と比較して多様な結果を生成する能力がある深層生成モデルを採用したことである。具体的には、自然な動作生成においては、決定論的手法を用いる従来モデルと異なり、本研究では条件付き敵対的生成ネットワークに基づく確率的生成方法を導入している。確率的生成手法により、ジェスチャの自然さや発話との同期性が改善され、ある発話に対してより多様なジェスチャを生成することができる。個性特性を表現する動き生成の課題においては、既存手法では個性をラベル付けされた学習訓練データが不足しているという問題があることに対して、データへの自動ラベリングとWasserstein敵対的生成ネットワークを用いた手法提案している。これにより、発話と一致し、かつ外向的、あるいは内向的な特性を伝えるジェスチャを生成することが可能となった。最後に、本論文は手法の拡張性についても取り組んでいる。新しい拡散モデルとサンプリングアルゴリズムによって、手が動かせる空間が限られるというような物理的制約内でも、自然で、かつ表現内容が制約によらず一貫しているジェスチャを生成する手法を提案している。

まとめると本論文は、深層生成モデルに基づいて発話に伴うジェスチャ生成の主要な課題に取り組んでいる。本論文のアプローチによって、生成されたジェスチャの自然さが改善され、個性という付加的な情報をも表出でき、物理的制約があっても適用できる実用的な動作生成手法を実現している。この研究で開発されたシステムは、人間とロボットやアバターとのインタラクションにおいて重要な役割を果たすことが期待される。ジェスチャ生成における主要な課題を独自のアイデアで解決し、実用的なシステムを構築したこの論文は博士（工学）の学位論文として価値のあるものと認める。