

論 文 内 容 の 要 旨

博士論文題目 Statistical electrolaryngeal speech production toward voice restoration

(邦題：音声回復に向けた統計的電気音声生成)

※ 論文題目が外国語の場合はワープロ等を用いること。また、その邦文を論文題目の下に () で記入すること。

氏 名 田中 宏

要 旨

Laryngectomees are people with disabilities as result of an operation to remove their larynges including the vocal folds for several reasons such as an injury and laryngeal cancer. An electrolarynx is a medical device to help them to produce quite intelligible speech, electrolaryngeal (EL) speech, by mechanically generating artificial excitation signals, instead of vocal fold vibrations. Unfortunately, there are mainly three disadvantages: 1) its sound is characterized as mechanical and robotic because of the fundamental frequency (F0) pattern of the monotonic excitation signals, 2) the excitation signals are absolutely emitted outside as noise because of EL speech production mechanism, and 3) one hand is occupied.

This research aims to provide laryngectomees more smooth speech communication and deals with speaking aid system for EL speech enhancement to recover losing information such as F0 patterns by using techniques of speech processing. One is A) a speaking aid system capable of modifications at the acoustic level. In this dissertation, we propose a hybrid approach using a noise reduction method for enhancing spectral parameters and statistical voice conversion for predicting excitation parameters. Through experimental evaluations, we demonstrate that the proposed hybrid approach effectively address not only the issue (1) but also (2) by achieving to improve the naturalness of EL speech while preserving the intelligibility of EL speech. However, this approach may not suitable for face-to-face conversation, because the enhanced EL speech presented from a loudness speaker. Therefore, we propose another method B) which is a speaking aid system capable of

modifications at the physiological level which allows laryngectomees to directly produce the enhanced EL speech from their mouths. In this dissertation, to make it possible to control without conscious operation, we propose a control method of the F0 patterns of the electrolarynx by applying a statistical F0 pattern prediction method to the system (B). To address the latency issues, we also propose two methods of F0 pattern modeling. Through experimental evaluations, we demonstrate that the proposed control method incorporating the proposed F0 pattern modeling effectively addresses the issue (1) by achieving to improve the naturalness of EL speech.

To recover more natural F0 patterns, this dissertation further proposes a new statistical F0 pattern prediction method, which is applicable to the above-mentioned approaches, considering speech production mechanism. Since the F0 patterns are physically constrained by the actual control mechanism of the thyroid cartilage, we incorporate a generative model of F0 pattern into the conventional statistical model to predict F0 patterns of natural speech. This approach is noteworthy in that it allows to generate an F0 patterns that is both statistically likely and physically natural. From experimental evaluations, it is demonstrated that the proposed systems successfully address the issue (1) by improving the prediction accuracy of F0 patterns.

氏名	田中 宏
----	------

(論文審査結果の要旨)

本論文では、喉頭摘出者のための健常者と同等の音声コミュニケーションの実現に向けて、以下の3つの要素技術について論じた。

1つ目は、音響学的レベルでの強調を可能とする発声障害者支援システムに関する研究である。提案法は、喉頭摘出者の声道は健在であることに着目し、単純な信号処理技術と複雑な統計的手法を組み合わせることで、電気音声の高い明瞭性を劣化させることなく自然性を大幅に改善することを示した。

2つ目は、生理学的レベルでの強調を可能とする発声障害者支援システムに関する研究である。提案法は、リアルタイム統計的音源予測を用いて電気式人工喉頭を直接制御することで、話者による手動での制御を必要としない、かつ、電気音声の自然性を大幅に改善することを示した。

3つ目は、上記の両アプローチの主要技術である統計的F0パターン予測法に関する研究である。提案法は、F0パターンの物理的な生成過程に着目し、F0パターン生成過程の確率モデルを従来の統計的F0パターン予測法へ組み込むことで、統計的に尤度の高い、かつ、物理的に自然なF0パターンの復元を可能とし、電気音声の自然性を改善することを示した。

これらの成果は、従来技術では本質的に解決困難であった問題に対する解決策を示しており、これらの結果は1編の学術論文、10編の査読付き国際会議論文、22件の国内会議として発表していることから、研究業績として非常に高く評価できる。以上、本博士論文の審査を行い、本論文は、博士論文(工学)として十分な価値があるものと判断した。