氏　　名	LI, Yongwei			
学位の種類	博士(情報科学)			
学位記番号	博情第 405 号			
学位授与年月日	平成 30 年 12 月 21 日			
論文題目	A study on effects of glottal source and vocal tract cues on perception of emotional vowels			
論文審査委員	主査	赤木　正人	北陸先端科学技術大学院大学	教授
		党　　建武	同	教授
		鵜木　祐史	同	教授
		河原　英紀	和歌山大学	名誉教授
		北村　達也	甲南大学	教授

論文の内容の要旨

In human-human communication, speech is the most direct way for communication. Speech contains a lot of information of speaker, such as emotion, gender, age, native and level of education. Emotions play a vital role in speech understanding. By using appropriate emotions, the same textual information can be used to convey different meanings. Emotions in speech can be used not only to express intentions, but also to understand intended information by combining potential emotions, voice information, and other linguistic factors.

Although acoustic features of emotional speech have been investigated, it is still difficult to model emotions by using only these acoustic features. Many studies have shown that the speech production organs features, such as glottal waveforms and vocal tract shapes are important. However, the properties of glottal source and vocal tract to expressive emotions via acoustic features have not been investigated deeply yet. Thus, this study focuses on investigating the effects of glottal source and vocal tract cues on emotional speech perception, particularly for vowel, since vowel plays more important role in emotional speech.

The research aims to investigate the effects of glottal source and vocal tract cues on perception of emotional vowels, especially after removing known effects of dominant prosody features (e.g., pitch, intensity, and duration). Thus, (1) an analysis-by-synthesis method is firstly developed for estimating glottal source waveform and vocal tract shape of emotional vowel. (2) the glottal source waveform and vocal tract shape are estimated from Japanese vowel /a/, and the spectral tilt and character of vocal tract shape are consistent with previous results. (3) F_0/pitch (fundamental frequency), intensity (E_e-related), duration of source related features (prosody features), spectral tilt of glottal source waveform, and F_1 (first formant frequency) are discussed, which in a controlled way of modifying the estimated glottal source waveform and vocal tract shape and utilized for establishing an analysis-by-synthesis method for resynthesizing the emotional vowels. Then, Japanese natives with normal hearing participate in the evaluation of perceptual rating emotions in the valence and arousal space. The results show that the glottal source information plays an essential role in perception of

emotions in vowels, whereas the vocal tract information contributed to the valence and arousal perception after neutralizing the F_0, intensity, and duration cues effects.

This study investigates emotional vowels from the point of view of speech production. The results contribute to further understanding the emotional speech production mechanism, also can enlighten many emotional speech fields, such as emotional speech recognition, synthesis and conversion. Moreover, an accurate estimation method of glottal source waveform and vocal tract shape is proposed for vowels in this study. It can be used in many speech signal processing fields, for example, speech analysis, speech synthesis, voice pathology detection, speaker recognition, and speech recognition.

Keywords: Emotional vowel production, emotional vowel perception, glottal source waveform, vocal tract, ARX-LF model, valence and arousal

論文審査の結果の要旨

本論文は，Source-Filter モデルを用いた音声からの声帯音源波形と声道形状の同時推定とその感情音声分析への応用に関する研究報告である。

音声には豊かな情報が含まれている。人 - 人の音声コミュニケーションでは，言語情報だけではなく，音声に含まれるその他の情報（感情など）も送受され，コミュニケーションを円滑にしている。しかし，現有の音声システムでは，未だに言語情報の送受のみが中心であり，感情等の情報を自在に扱えているとは言い難い。この原因の一端は，音声に含まれる感情の生成及び知覚を一体として議論できておらず，何が感情等の生成・知覚に関係しているか完全には把握できていないことによる。

本研究では，他の研究で行われているように音響特徴と感情の関係のみを議論するのではなく，音声生成→音響特徴→音声中の感情知覚という枠組みで，総合的に音声に含まれる感情情報を議論している。具体的には，(1) 音声生成における Source-Filter モデルの一種である ARX-LF モデルを用いて，音声波形から ARX フィルタと LF 声帯音源の同時逆推定を行い，声帯音源波形と声道形状を抽出した。そして，(2) この手法を感情（Joy, Angry, Sad, Neutral）を含む母音に適用し，感情の違いによる声帯音源波形と声道形状の変化を明らかにした。(3) さらにこの手法を応用して，ARX-LF モデルのパラメータ値を変更しながら，声帯音源と声道形状がどのように感情知覚に影響するのかを系統的に調査し，感情知覚にかかわる音響特徴とその音響特徴を生成する固有の声道形状および声帯音源波形について考察した。この結果，母音という限られた範囲ではあるが，生成系のどのような変化が音声中の感情情報を生起し，これがどのように聴取者に知覚されるかという一連の流れが説明できることとなった。

現在，音声分野では，表現豊かな音声の分析・再合成に関する研究が各所で盛んにおこなわれている。本論文の成果は，感情音声のみにとどまらず表現豊かな音声全般の新たな

分析および表現法の可能性を提示したことであり，今後もこの分野での使用および発展が大きく期待できるものである。

以上のように，本研究は音声生成→音響特徴→音声知覚という拡大された枠組みのもとで，感情音声生成と知覚を議論するための手法を実現し，感情知覚にかかわる主要な特徴を特定した研究であり，学術的に貢献するところが大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。