



HOKKAIDO UNIVERSITY

Title	Analysis on Acoustical and Perceptual Characteristics of Whispered Speech and Whisper-to-Normal Speech Conversion [an abstract of dissertation and a summary of dissertation review]
Author(s)	今野, 英明
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12482号
Issue Date	2016-12-26
Doc URL	https://hdl.handle.net/2115/64439
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Hideaki_Konno_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 今野 英明

学 位 論 文 題 名

Analysis on Acoustical and Perceptual Characteristics of Whispered Speech and Whisper-to-Normal Speech Conversion

（ささやき声の音響的・知覚的特徴の分析と通常音声への変換）

近年、喉頭摘出手術により声帯を失った人が、電気咽頭等の医療機器を頼らずに、自らの声でコミュニケーションできるようにすることなどを目的として、ささやき声を通常の音声に変換する研究が行われている。本論文は、通常音声との比較分析を通じて日本語のささやき声の特徴を明らかにし、その知見を活かして、ささやき声から変換した通常音声（変換音声）の韻律を改善する手法を提案するものである。

音声を持つ特徴には音韻性と韻律がある。音韻性は音素を特定する音色であり、韻律は音素より長い時間区間において観察されるアクセントやイントネーション、リズム等である。音声の言語情報の多くを担うのは音韻性であり、韻律は言語情報に加えて感情等の非言語的な情報の伝達にも重要な役割を果たす。

音声は発声時に声帯振動を伴うか否かによって有声音と無声音に大別され、声帯振動の有無は子音の音韻性に係わる。一方の韻律には声の高さ（ピッチ）が強く関与する。通常音声においては、ピッチを決める主たる音響的特徴は声帯の振動数に起因する基本周波数（F0）である。したがって、声帯振動は通常音声の音韻性と韻律の双方に対して重要な役割を担う。

ささやき声は、大きな声を出ることができない状況下等において、日常的に使われる音声の一様式である。ささやき声は声帯の間を通る呼気の乱流を音源として生成され、声帯振動を伴わずに発声される。そのため、ささやき声は常に無声音であり F0 も存在しない。にもかかわらず、ささやき声でも通常音声と同様に音韻性や韻律の伝達が可能である。本研究は数理解析的な手法でこの現象の解明を試みたものである。

さて、音韻性に関わる主要な音響的特徴は母音のホルマント周波数に代表される音声のスペクトル包絡すなわち声道特性である。ささやき声の母音にも、通常音声と同様にホルマント周波数があるものの、同じ母音であっても通常音声のホルマント周波数とは異なることが、種々の言語に関して報告されている。その原因について、音声の生成面からの説明はなされているものの、両者のホルマント周波数の違いがささやき声の音韻性にどのように関わるかは明らかではない。一方、韻律に大きく影響するピッチは高低の印象を与える音の知覚的属性の一つである。通常音声とは異なり、ささやき声には F0 が存在しないため、それに代わる特徴がピッチを生じさせると考えるのは妥当である。しかしながら、ささやき声のピッチに関係する音響的特徴は十分に解明されておらず、ささやき声のピッチを定量的に表現する方法も確立していない。そこで、本論文では、ささやき声の音韻性とピッチに関する音響分析と知覚実験を行った上で、ささやき声を通常音声に変換する手法を検討した。

ささやき声から通常音声への変換では F0 の時間軌跡（F0 パターン）の生成が不可欠である。従来の研究では、ホルマント周波数やスペクトル包絡の補正に重きが置かれ、F0 パターンの生成に関

しては研究が不十分であった。近年、自然な変換音声のための F0 パターンの生成法が開発されているものの、変換元のささやき声を持つ韻律を再現する観点からの検討は行われていない。そこで本研究では、ささやき声を持つピッチを定量的に予測し、それに基づく F0 パターンを生成することにより、変換元のささやき声の韻律を変換音声に保持させることを目的とした。

本研究の貢献は以下の三つである。1) 日本語のささやき声の母音に関して、通常発声の母音とのホルマント周波数の違いを音韻性に着目して定量的に示した。2) 単一の周波数によってピッチを表現できるささやき声の収録方法を考案し、ささやき声の母音のピッチとスペクトル形状との関係を定量的に分析した。3) 前項の分析結果からささやき声のピッチの推定式を作成し、ささやき声の単語をピッチアクセントを保持して通常音声に変換する方法を提案した。本論文の概要は以下のとおりである。

第 1 章では、本研究の背景や目的を述べた。

第 2 章では、ささやき声の音韻性に関する従来研究を紹介した後、日本語のささやき声の音響分析と聴取実験について述べた。ささやき声の母音に関しては、通常母音において低い値をもつホルマント周波数ほど大きな割合で高い周波数に移動していることを明らかにした。また、合成音を用いた聴取実験により、ささやき声と通常音声の母音では、ホルマント周波数が同じであっても、音源の違いによって音韻性に変化が生じる場合があることを示し、このことにピッチが関連している可能性を述べた。さらに、日本語の 110 個の単音節に対して音韻同定実験を行い、ささやき声の音韻の正答率は通常音声より 20% 低下することを示すとともに、ささやき声における有声・無声子音の発声や知覚には、通常音声の場合とは異なり、スペクトルの低域・高域成分のバランスが大きく関与する可能性を示した。

第 3 章では、ささやき声のピッチに関する従来研究を紹介した後、ピッチの異なるささやき声の母音を分析し、ピッチの変化に伴ってホルマント周波数やスペクトルのエネルギー分布が変化することを確認した。また、メルフィルタバンク分析の結果に主成分分析と重回帰分析を適用して、話者が意図したささやき声のピッチを単一の周波数として予測する重回帰モデルの構築法について述べた。

第 4 章では、ささやき声から通常音声への変換に関する従来研究を紹介した後、第 3 章で得たささやき声のピッチを F0 として用いた通常音声への変換手法について述べた。この手法を用いて、アクセント位置によって意味の異なる日本語単語を変換し、アクセントの正答率が、F0 一定の変換音声に比して 55.5% から 72.0% に改善されることを示した。

第 5 章では、本研究全般に関する考察を行い、この研究の将来展望について述べた。第 6 章で結論を述べた。