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_review.pdf, 審査の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　　氏名　今野　英明

審査担当者　主　査　教　授　工藤 峰一
副　査　教　授　今井 英幸
副　査　教　授　杉本 雅則

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）

本研究はささやき声に関する研究であり、通常の声とは違った原理で声の高さを生成するささやき声の機構解明を数理解析的に行うとともに、ささやき声を通常音声へと変換する方式の高品位化を図ったものである。

研究背景はつぎのようなものである。喉頭摘出手術などにより声帯を失った人の多くは電気咽頭等の医療機器に頼って声を出すほかない。しかし、その声は本人のそれとは大きく異なる。一方、声帯を失ったとしても、呼気により、ささやくように話すことができる場合がある。そのような場合、ささやいた声を通常の音声に変換することができれば、より満足度の高い意思伝達が可能となる。また、健常な人でも、携帯電話の利用時など、公共の前では通常の音量で話すことが憚られることもあり、そのような場合でもささやき音声の明瞭度を向上させることの効果は大きい。

本研究は、申請者が 20 数年来続けてきた、ささやき声の発声上および聴覚上の特性解析結果の集大成であり、声帯振動のないささやき声においていかにして「聴覚上の声の高さ」が知覚されるのかについて、従来の知見を踏まえつつ解析を重ねた成果である。また、応用として、ささやき声で失われやすい韻律 (アクセントやイントネーション, リズム等) を、推定した声の高さを利用して復元する試みも紹介している。

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、声の高さ (ピッチ) は韻律に強く関与する。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。一方、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。本研究ではこの機構を明らかにすべく統計解析を行っている。

従来研究は、様々な聴覚実験を通して、ささやき声独自のピッチ調節方法を定性的傾向として述べたものが殆んどであった。例えば、同じ母音において通常音声とささやき音声をスペクトルで比較した場合、スペクトル包絡形状に差があることなどがたびたび報告されていた。また、その原因に関しても音声の生成面からの解釈はある程度行われていた。しかし、そのような相違や関連を定量的に明らかにしたのは本研究がはじめてである。これが実現できたのは、本来存在しない F0 値を、巧妙な聴覚実験を考案して、発声・聴覚的な F0 値として取得する方法を確立したためであ

る。実際には、人工的に生成した単一周波数の純音に対して、その音と「同じ高さにあると感じるささやき母音」を発声してもらうようにした。これにより、音声波形およびそのスペクトル波形とその波形が意味する「音の高さ」の両方が入手可能となり、それらの関係を関数として表現できるようになった。さらに、この実験方式により男女複数人のささやき母音を採取して、音声波形からの F0(ピッチ) 推定を、主成分分析や線形回帰を通して行った。その結果、波形から F0 を得る具体的な推定方式を示している。また、話者や性別、母音の種類に関する推定方式の依存性も検証している。

次に、この F0 推定法を利用して、ささやき声を通常音声に変換する手法を提案している。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ため、アクセントなどを適切に再現することが難しかった。本研究では、F0 の値を波形から逐次推定することで、滑らかに変わる音の高さを適切に変換音に持たせることができるようになった。アクセント位置によって意味の異なる日本語単語を用いて聴取実験を行い、アクセントの正答率を、F0 一定の変換音声に比して 55.5% から 72.0% に改善している。

本研究の貢献は以下の三つである。1) 日本語のささやき母音において、これまで訐れなかった F0 値を疑似的に得る巧妙なデータ採取方式を確立した。2) その方式によって採取したデータを基に、統計解析技法により、ささやき音声波形から F0 推定が可能であることを示した。3) 前項の F0 推定方式を利用してささやき声を通常音声に変換する方法を提案し、アクセントを回復できることを実験的に確認した。

これを要するに、著者は、ささやき音声における声の高さの生成および知覚過程において波形とピッチの関係を定量評価とするとともに、波形からのピッチ推定法を確立することにより、ささやき声から通常音声への高品位な変換法を提案した。この成果は音声研究に貢献するところ大なるものがある。よって著者は、北海道大学博士 (情報科学) の学位を授与される資格あるものと認める。