【助成 38－25】

表現力豊かな音声の声質を決める声門流波形と声道伝達関数特徴の解析

代表研究者　富山県立大学工学部　教授　MOKHTARI Parham
共同研究者　富山県立大学工学部　講師　森川大輔
共同研究者　富山県立大学工学研究科　大学院生　河内太一
共同研究者　富山県立大学工学研究科　大学院生　松本幸輝

〔研究の概要〕

This research aimed towards automatic analysis and synthesis of speech that is rich in expressivity. In line with the classical source–filter model of speech production, we sought improvements in the performance of algorithms designed to separate the influence of the glottal flow and the vocal-tract on the speech signal. We improved a benchmark algorithm for estimation of the glottal flow from the speech signal, and we quantified the relationship between physiological parameters of Titze's vocal fold model and acoustic parameters describing the shape of the simulated glottal flow.

〔研究経過および成果〕

Aiming toward physiologically relevant analysis, modelling and synthesis of expressive human voices, our research prioritizes automatic estimation of the laryngeal source (glottal flow) from the acoustic speech signal, and its synthesis with physiological models and computer simulation methods.

In the mid-term report we summarized our research results on automatic formant estimation [1], glottal flow estimation [2][3], and a new database of recorded voice qualities [4]. Here below, we additionally summarize our subsequent research results obtained since the mid-term report.

Iterative adaptive inverse filtering (IAIF) is a benchmark method for estimating the glottal flow from the speech signal, and iterative optimal preemphasis (IOP) has also been proposed as an improvement of that algorithm. However, the conventional method typically analyses a frame of speech spanning 30–50 ms, which includes several glottal cycles; it therefore suffers from inaccuracy due to the inclusion of the glottal open phase, during which there is acoustic influence of the sub-glottal tract (the

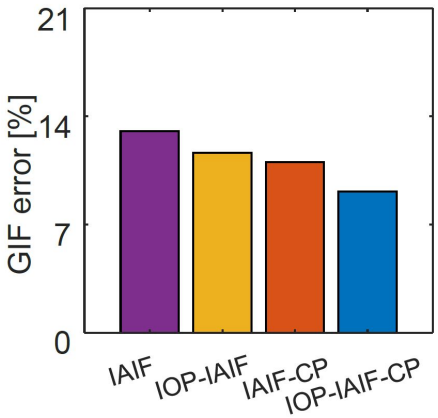

Fig.1　Glottal flow estimation error decreased with closed-phase (CP) analysis [5]

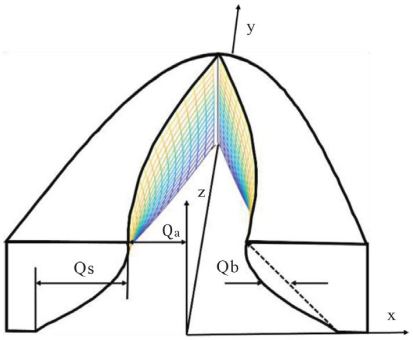

Fig.2　Titze's vocal-fold model [6]

trachea) and the glottis itself. We therefore proposed to combine the benchmark IAIF algorithm with closed-phase (CP) analysis, so that the vocal-tract characteristics and therefore the glottal flow signal too, could be potentially derived more accurately. As shown in Fig.1, evaluation using the OpenGlot dataset revealed overall lower errors for the proposed methods involving CP analysis, and the lowest overall error was obtained for the IOP-IAIF-CP method [5].

Another ongoing project in our laboratory is physiologically-based modelling and synthesis of various voice qualities, aiming at expressive speech synthesis. For this purpose, we investigated the characteristics of a vocal-fold simulation model originally proposed by Titze, to assess whether it is capable of producing various voice qualities. As shown in Fig.2, Titze's model is represented by 3 parameters Qa, Qs, and Qb that describe the pre-phonatory posture of the vocal folds, plus one additional parameter Qp that controls the phase difference between the upper and lower folds during dynamic oscillation. We used this model with a wide range of parameter settings to synthesize numerous glottal flow waveforms, then extracted 3 acoustic parameters describing the shape of the glottal flow pulse (OQ, NAQ, and openNAQ). As a result, we were able to quantify the relationship between the vocal-fold model (physiological) parameters and the glottal flow (acoustic) parameters [6].

〔発表論文〕

1. 河口竜馬、モクタリパーハム、森川大輔(**2021.6**)「ラーモニック減算法を用いたフォルマント周波数の自動推定法の検討」、情報処理学会研究報告、Vol.2021-SLP-137 No.22、5 pages.
2. 平野雄大、モクタリパーハム、森川大輔(**2021.9**)「プリエンファシスの適用回数が IAIF に基づく声門流推定の精度に与える影響」、日本音響学会講演論文集、939-940.
3. 平野雄大、モクタリパーハム、森川大輔(**2021.9**)「音声のスペクトル傾斜モデリングが IAIF に基づく声門流推定の精度に与える影響」、日本音響学会音声研究会資料 Vol.1 No.2 SP-2021-26、135-137.
4. Mokhtari P., Morikawa D.(**2022.3**)"Introducing a Japanese multi-talker database of laryngeal voice qualities,"日本音響学会講演論文集.
5. 河内太一、モクタリパーハム、森川大輔(**2023.9**)「声門の閉鎖期の音声信号による IAIF に基づく声門流推定」、日本音響学会講演論文集、1115-1116.
6. 松本幸輝、モクタリパーハム、森川大輔(**2023.11**)「Titze の声帯モデルのパラメータが声門流波形の音響特徴に与える影響の検討」、日本音響学会音声研究会資料 Vol.3 No.5 SP-2023-42、9-14.