5Q-06

高品質音声符号化のための基本周波数量子化の知覚的影響

宮下玄太† 森勢将雅‡
山梨大学大学院医工農学総合教育部† 山梨大学大学院総合研究部‡

1. はじめに

本研究ではサンプリング周波数が 40 kHz 以上のフルバンド音声の，品質劣化の伴わない符号化に取り組んだ検討を実施している．前報[1][2]ではそれぞれ，フルバンド音声の周波数軸方向，時間軸方向の符号化について検討を行った．本稿では，フルバンド音声の音声分析を行った際に得られる，音声パラメータの量子化について検討を行う．特に，基本周波数 (F0)について検証を行った．音声分析合成において，分析時に複数の音声パラメータが生成される．本稿では，この分析パラメータの表現に必要なデータ量の削減を，量子化によって行うことを目的とする．量子化では，データの表現に必要なビット数を削減する．しかし，量子化誤差によって元となる分析合成音から音質の劣化が生じる．この誤差が人間の聴覚に影響を与えない範囲を，主観評価実験によって測定する．本稿では，複数の音声パラメータの中で，基本周波数についての量子化を，主観評価実験を用いて検証する．

2. 実験内容

実験用のボコーダには，高品質音声合成系 WORLD[3] (D4C edition[4]) を使用した．WORLD による音声分析合成の概要を図 1 に示す．実験は，まず，フルバンド音声の音源を用意し，それらをボコーダによってパラメータ化する．次に，得られた F0 の量子化を行う．最後に，量子化 F0 と，量子化しないスペクトル包絡，非周期性指標を用いて音声合成を行う．合成時の F0 について，量子化ビット数の異なる分析合成音声 5 種の聴取比較を行う．

2.1 F0 量子化

F0 の量子化手順について述べる．まず，F0 の上限値と下限値を設定する．本稿では，全ての人間発話の条件を網羅するために，上限値は 4096 Hz，下限値は 1 Hz とした．次に，F0 を対数化し，対数軸上の下限値から上限値の間を任意のビット数で等間隔に分割する．最後に，指数化してパラメータに戻す．

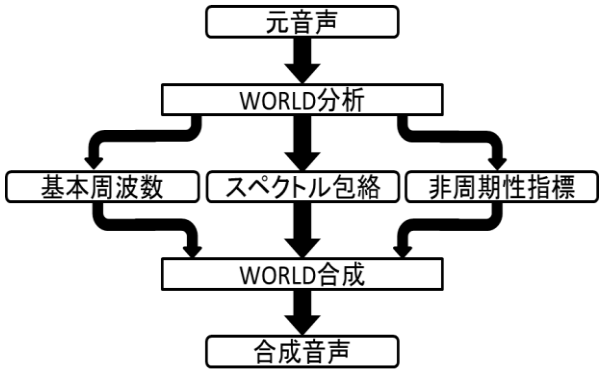

図 1 WORLD による音声分析合成の概要.

量子化を行った後のパラメータを用いて合成した分析合成音を，符号化音声と記述する．本稿では，F0 のみを量子化し符号化音声の作成を行った．

3. 実験条件

表 1 に主観評価の条件を示す．実験の評価方法として MUSHRA 法 (Method for the subjective assessment of intermediate quality levels of coding systems)を用いた．MUSHRA 法とは，提示された音声の品質を評価する方法の 1 つで，被験者は GUI を用いて 0 から 100 の尺度で音声刺激を採点する．この手法は，一般的に音質の違いの評価に使用されていて，MOS 評価よりも差の検出力が高い手法と言われている．

実験は，暗騒音の A-weighted SPL が 18 dB の防音室を使用し，正常な聴力を有する 10 人が評価に参加した．音声刺激はヘッドフォン (SENNHEISER HD650)を用いて与えた．

主観評価に用いた音声刺激は，2 人の男性と 2 人の女性による各 5 発話の，計 20 発話である．サンプリング周波数は 48 kHz，量子化ビット 16 bit である．発話内容は日本語による子音を含む 4 モーラ単語であり，親密度別単語了解度試験用音声データセット 2007(FW07)を使用した[5].

被験者は，同時に提示される 5 種類の音声刺激を順に評価する．5 個の内訳は，WORLD を用いて分析合成された，F0 の量子化ビット数が異な

Perceptual influence of fundamental frequency quantizion for encoding of full-band speech
†Genta MIYASHITA, Integrated Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi
‡Masanori MORISE, Integrated Graduate School, University of Yamanashi

Copyright ©2018 Information Processing Society of Japan.
All Rights Reserved.

表 1 実験条件.

評価手法	
比較方法	MUSHRA 法
被験者	20 代男性 10 名
実験環境	防音室 (A-weighted SPL: 18 dB)
オーディオ インターフェース	Roland QUAD-CAPTURE
ヘッドフォン	SENNHEISER HD650
使用音声	親密度別単語了解度試験用音声データセット 2007(FW07)[5]
評価用音声	
発話者	4 人(男女各 2 名)
A/D 変換	48 kHz/16 bit
音源数	全 20 音声(各発話者 5 音声)
音声種	4 モーラ単語
比較条件	
比較用音声	分析合成音 5 種
量子化パラメータ	基本周波数 (F0)
パラメータ上限値	4096 Hz
パラメータ下限値	1 Hz
量子化ビット数	64, 10, 9, 8, 7 bit

る 4 個の符号化音声と，1 個の非符号化音声(64 bit)となっている．量子化を行うパラメータは F0 であり，上限値は 4096 Hz，下限値は 1 Hz とした．5 個の音声刺激のパラメータの量子化ビット数は 7，8，9，10，64 bit とする．これらの値は，予備実験にて決定した．

4. 実験結果・考察

実験結果を図 2 に示す．縦軸は MUSHRA 法による評点を，横軸は F0 の量子化ビット数を示す．誤差棒は 95%信頼区間を表す．複数の比較を行うため，有意差検定には Benhamini-Hochberg 法[6]に基づく 2 段階線形上昇手順[7]を用いて p 値の補正を行った．

補正 p 値の検証を行った結果，9, 10, 64 bit 間はそれぞれに有意差がなく，7-8, 8-9 bit 間には補正 p 値 0.001 以下の有意差が検出された．これらの結果から，9 bit 表現の F0 を用いた符号化音声は，64 bit 表現の分析合成音から音質の有意差がないことが確認された．本稿の実験条件では，全ての人間発話を含むために F0 の上限値をかなり高く設定している．そのため，音声を通常発話に限定し，上限値と下限値を適切に設定すれば，8 bit 以下に削減することも可能であると考えられる．例えば，本稿では下限値を 1 Hz に設定したが，人間の可聴域は 20 Hz 以上であり，過剰にデータ表現の領域を占めている．

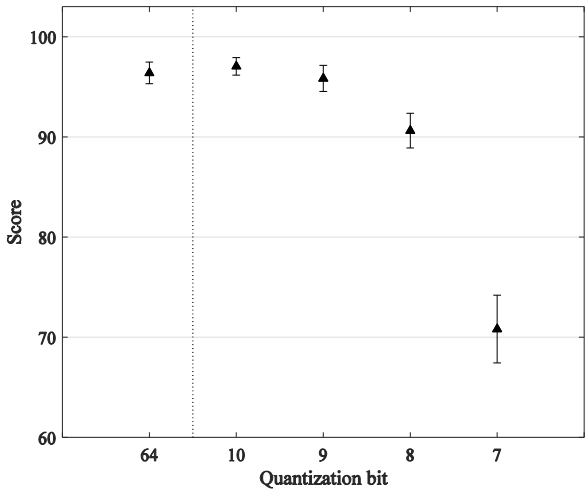

図 2 MUSHRA 法による符号化音声の評価結果.

5. おわりに

本稿では，高品質音声の分析合成における，F0 の量子化について検討した．量子化パラメータを用いて音声を合成し，その音質劣化と量子化ビット数の関係を，主観評価実験によって検証した．その結果，F0 は 9 bit あれば十分な音質を保つことができることが確認された．

今後の発展としては，基本周波数以外の音声パラメータの量子化と，その主観評価を行う計画である．これらの結果から，品質劣化の生じない，フルバンド音声を表現可能なデータ転送レートを明らかにする．

謝辞 本研究は，JSPS 科研費 JP15H02726, JP16H05899, JP16K12511, JP16H01734 の支援を受けて実施された．

参考文献

[1] 宮下玄太他，日本音響学会秋季研究発表会，pp.259-260, 2017.
[2] 宮下玄太他，情報処理学会音楽情報科学研究会,vol. 2017-MUS-115, no. 23, pp. 1-6, 2017.
[3] M. Morise et al., IEICE transactions on information and systems, vol.E99-D, no.7, pp.1877–1884, 2016.
[4] M. Morise, Speech Communication, vol.84,pp.57–65, 2016.
[5] 近藤公久他, 電子情報通信学会技術報告, vol. 107, no. 432, pp. 43-48 , 2008.
[6] Y. Benjamini et al., J.R.Statist. Soc. Series B, vol.57, no.1, pp.289–300, 1995.
[7] Y. Benjamini et al., Biometrika, vol.93, no.3, pp.491–507, 2006.

Copyright ©2018 Information Processing Society of Japan.
All Rights Reserved.