

# 物理乱数音源とフォルマントを用いた口笛音合成

伏木田 勝信  
KF 研究所

## 1. まえがき

先に、声道の共鳴特性を表すフォルマントと乱流音源を用いた口笛の合成方式について提案を行った[1]。しかしながら口笛の自然性や個人性の音質近似に関してはまだ改善の余地があると考えられた。その原因としては、声道の共鳴特性を全極モデルで近似を行ったためそれ以外の零点等の特性が反映されていなかったことがあげられる。今回、音源波形として全極近似の近似誤差として与えられる予測残差波形を物理乱流音源として用いる方式について検討し音質が改善されることを確かめたので報告する。

## 2. 口笛の生成システム

図-1 に示すように口笛の発声システムは、声道内に作られた狭めによって生ずる乱流音源と、共鳴管としての口腔の共鳴フィルタとから構成される。舌と歯茎による狭めの形成とすぼめた唇との間での短い共鳴口腔を形成し肺から供給される強い呼気あるいは吸気により乱流が生じ音源となって口笛音が生成される。

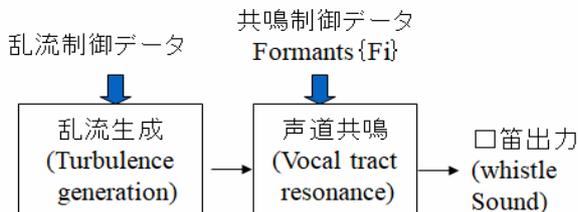


図 - 1. 口笛の発声システム

## 3. 共鳴パラメータ (フォルマント) 抽出

口笛音のフォルマント抽出は文献[1]と同様のMSF(Multi-Step-Focusing)-ADIF(Autocorrelation-Domain-Inverse-Filtering)方式[2],[3],[4]を用いた。

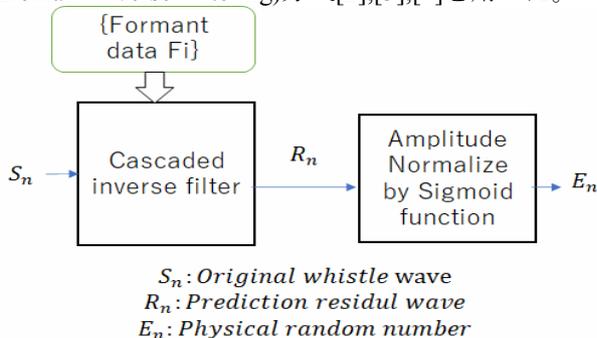


図 - 2. 予測残差波形による物理乱数の算出方式

フォルマント抽出の際に得られる予測残差情報は原音の周波数スペクトルとの誤差を表している。このフォルマント (全極型フィルタ) の予測残差情報 (prediction residual) は音声の品質改善に寄与することが報告されている[5]。このため予測残差情報は口笛の音質改善にも有効であると考えられる。

## 4. 物理乱数の算出

### 4.1 予測残差波形の算出

予測残差波形には原周波数スペクトルと全極近似スペクトルの誤差情報が含まれている。予測残差波形は物理乱数の一種と見ることができる。擬似乱数にみられる乱数の限界を克服できる可能性がある。

原音の口笛音サンプルは男性1名が呼気により発声した1オクターブの音域の口笛音である。サンプリング周波数は22.05kHz、フォルマント数 (NOF: Number of Formants) の変化範囲は5から9を用いた。フォルマントの量子化テーブルとしては140Hz~11kHzの範囲を対数スケールで2048レベルに非線形に量子化したものを用いた。この際、バンド幅Bは尖鋭度Qが一定 ( $Q=F/B=10$ ) として設定した。

予測残差波形  $R_n$  は図 - 2 に示されるようなフォルマントに対応する2次のLPC (Linear predictive coefficient)の縦列逆フィルタで求めた。また、 $E_n$  は振幅の正規化としてシグモイド (Sigmoid) 関数により非線形符号化を行ったものを用いた。ここでは、 $E_n$  を物理乱数音源と呼ぶ。

### 4.2 物理乱数の特性の評価、振幅値の分布

予測残差波形の振幅分布を図 - 3 に示す。振幅分布は一樣分布ではなく、ガウス (Gauss) 分布に近いものであった。

### 4.3 モンテカルロ法による $\pi$ と $\epsilon$ の近似度と近似速度の評価実験

乱数の物理、AI 応用としてボルツマンマシンやモンテカルロ法 (Monte Carlo method) などが知られている[6]。ここでは予測残差波形を物理乱数として用いてモンテカルロ法による円周率 ( $\pi$ ) と自然対数の底 ( $\epsilon$ ) の近似値を求める実験を行った。表 - 1 に各 2000 個の乱数を用いた実験結果例を示す。前期物理乱数では一樣乱数に比較して収束速度の向上が見られた。なお、 $\epsilon$  は [0 - 1.0] での乱数の和が1を超える個数の期待値として求められた。この結果例は、使用する物理乱数の数が一樣乱数に比べて少なく済み効率的に近似できることを示唆している。この特性は高速の最適化が要求される場合の乱数として有効であると考えられる。例えば、前記の多段フォーカシング

Whistle synthesis using formant and predictive residual as physical random number  
Katsunobu Fushikida, KF Laboratory

(MSF) 法により大規模で複雑な構造を持つ系に対して高精度なフォルマントを求める場合には、各段の候補の生成の際に物理乱数等を用いて多様な候補を生成する手段が必要であると考えられる。

表-1. モンテカルロ法による  $\pi$  と  $\varepsilon$  の近似実験結果例 ( $\pi$  and  $\varepsilon$  approximation by Monte Carlo method)

$\pi$ の近似値比較	近似値	真値	誤差
擬似乱数	3.182	3.141593	0.040407
予測残差波形	3.13	3.141593	0.011593
$\varepsilon$ の近似値比較	近似値	真値	誤差
擬似乱数	2.753443	2.718282	0.035161
予測残差波形	2.714674	2.718282	0.003608

上の表は乱数の数として各 2000 個を用いた結果である

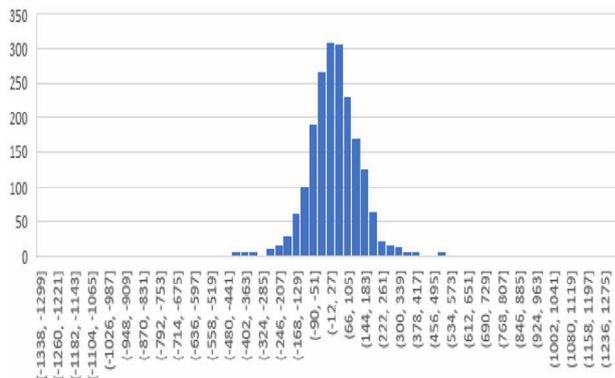


図 - 3. 口笛の正規化予測残差波形  $E_n$  の振幅分布 (ヒストグラム) 例

### 5. 合成システム

口笛音合成における物理乱数音源の有効性を確かめるために、図 - 4 に示されるような一様乱数音源による合成音との音質比較実験を行った。

口笛の乱流音源のスペクトルは、口腔共鳴の影響を受け類似のスペクトル構造を持つと仮定し、口腔のフォルマント周波数によって制御される共鳴フィルタによりモデル化した[1]。図 - 5 に予測残差波形による物理乱数音源  $E_n$  を用いた口笛の合成モデルを示す。なお、フォルマントの量子化テーブルは分析の際に用いたものと同様である。

### 6. 合成実験と結果

呼気の口笛の分析の結果得られたフォルマント周波数を用いて口笛の合成実験を行った[7]。一様乱数を用いた合成システム (図 - 4) で作成された合成音と物理乱数を用いた合成システム (図 - 5) で作成された合成音を対比較で試聴したところ後者が原音の音質に近いことが分かった。

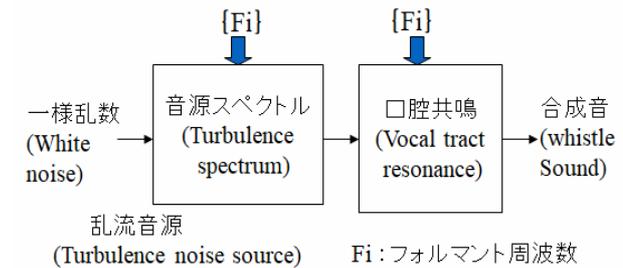


図 - 4. 一様乱数を用いた口笛の合成モデル  
擬似乱数の生成はC言語の関数 r(n)を用いた

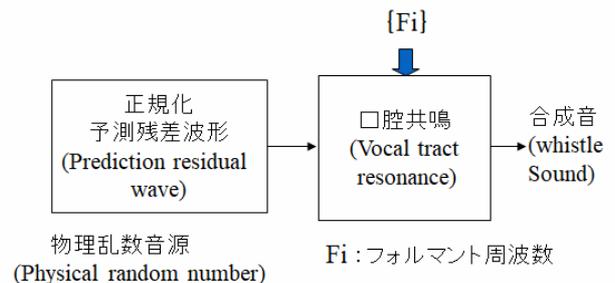


図 - 5. 物理乱数音源による口笛の合成モデル  
物理乱数としては  $E_n$  を用いた。

## 7. あとがき

口笛音の生成モデルとして全極モデルと物理乱数としての予測残差を組み合わせた方式について検討した。この物理乱数はボルツマン分布に近い分布を持ちモンテカルロ法による近似実験においては一様乱数より誤差が小さく近似速度が速いことを示した。また、予測残差による物理乱数音源を用いた口笛音は、全極モデルと一様乱数により合成した場合に比べ周波数スペクトルの近似が向上することを指摘し、一様乱数を用いた音源による合成よりも原音に近い音質の口笛音を実現出来ることを確かめた。

参考文献

- [1] 伏木田, 「フォルマントパラメータを用いた口笛の分析と合成」, 情報処理学会第 70 回全国大会講演論文集, 2-59-2-60, 2008.
- [2] 伏木田, 「DPを用いたフォルマント数の最適選択方式」 電子情報通信学会総合大会, D-14-25,, 2006-3
- [3] 伏木田, 「自己相関領域で逆フィルタリングを用いたホルマントの多段推定方式」 日本音響学会, 音声研究会資料, S81-41, pp. 323-329, 1981-10.
- [4] Fushikida, "A formant extraction method using autocorrelation domain inverse filtering and focusing method", Proc. IEEE ICASSP, E2.8:2260-2263, 1988.
- [5] Mitome and Fushikida "A speech synthesis device using formant and residual information" IEEE Transaction on Consumer Electronics, vol.29, Aug. 1983.
- [6] 磯、伏木田「ニューロコンピュータ」第 4 章、ボルツマンマシン、技術評論社、1989.
- [7] <http://www.5b.biglobe.ne.jp/~hfyf/fk/Laboratory/Laboratory.htm>