

フォルマントパラメータを用いた口笛の分析と合成

伏木田 勝信

無所属

1. まえがき

さまざまな発声の口笛音のフォルマント抽出実験を自己相関領域でのインバースフィルタリングを用い、フォルマント数を可変としたAbS法により行い比較検討を行った。また、分析によって得られたフォルマント周波数値等により制御される共鳴回路と乱流音源を用いて口笛の合成実験を行った。分析・合成実験により、フォルマント周波数と音階周波数との関係、澄んだ音の生成と制御方法等の検討を行った結果について述べる。

2. 口笛の発声法

図-1 に示すように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強い呼気あるいは吸気により口笛を発声した。

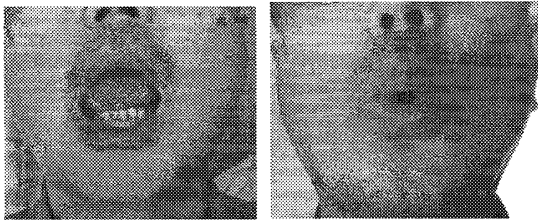


図-1 口笛を鳴らす際の舌の位置 (左) と唇の形状 (右)

発声の仕方により澄んだ音質や擦れた音質になる。ここでは、澄んだ音は擦れた音に比べて透明度が高いと呼ぶ。

3. 分析方式

MSF(Multi-Step-Focusing)-ADIF(Autocorrelation-Domain-Inverse-Filtering) とダイナミックプログラミング (DP) を用いたフォルマント数 (NOF: Number of Formants) 可変の AbS (Analysis-by-Synthesis) 的フォルマント抽出方式[1],[2],[3],[5]により分析を行った。

この方式では、2 次の ADIF は、入力、出力の自己相関値をそれぞれ、 r , v とするとフォルマントに対応する 2 次の線形予測係数 $\{\alpha_1, \alpha_2\}$ より次の(1)式により効率よく算出される。[2],[3],[5]

$$r_i = (1 + \alpha_1^2 + \alpha_2^2)v_i - (\alpha_1 - \alpha_1\alpha_2)(v_{i-1} + v_{i+1}) - \alpha_2(v_{i-2} + v_{i+2}) \quad (1)$$

ここで、 i は時間遅れを表す。

最適な NOF 値及びフォルマント周波数は、評価関数 G として予測残差パワー (縦列 2 次 ADIF の出力) の RMS 値 ($g = \text{SQRT}(r_0)$) の積算値を用いて(2)式を用いて DP の計算を行い求めた。[1]

$$G_{i,j} = g_{i,j} + \min\{w_1 \times G_{i-1,j-1}, w_0 \times G_{i-1,j}, w_1 \times G_{i-1,j+1}\} \quad (2)$$

ここで、 i, j はそれぞれ NOF 候補、分析フレーム番号を、 $w_0=1, w_1>1$ は NOF の連続性に対する重み係数を表す。

4. 分析実験と結果

4.1 実験条件

口笛音サンプルは男性 1 名が発声した呼気および吸気による約 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 22.05kHz、NOF の変化範囲は 5 から 9 までを用いた。

フォルマントの量子化テーブルとしては 140Hz~11kHz の範囲を 2048 レベルに dF/F (F はフォルマント周波数) が一定となるよう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。

4.1 実験結果

表-1 に呼気により発声された口笛のフォルマント分析結果例を示す。NOF は音階周波数の増加とともに減少している。これは、高い音を出すためには口腔長を短くすることに対応していると考えられる。また、F1 と F2 は口腔の共鳴と乱流音源の特性とが重なりあつて現れていると考えられる。

澄んだ口笛音の場合、抽出されたフォルマント周波数 F1, F2 の平均値は平均律の音階周波数とほぼ一致しており、ピッチ周波数の抽出を別途行わなくても音階周波数を抽出することができることを示している。

F1 と F2 は重なっている場合もあるが分離している場合もある。ここでは F1 と F2 との分離度 (d) を次の(3)式によって定義する。

$$d = (F_2 - F_1) / F_1 \quad (3)$$

呼気、擦れ気味の呼気、吸気、吸気の場合の分離度 d はそれぞれ 0.01, 0.26, 0.23 であった。この分離度の逆数 ($1/d$) は透明度を与えるものと解釈できる。

吸気による口笛の場合は、呼気のように明らかに NOF の変化は見られず NOF=7 でほぼ一定であつた。

た。吸気の場合、音程の制御において口腔長変化以外に、呼気流速の影響、零点の影響などが推測されるが詳細の解明は今後の課題である。

5. 合成方式

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ーとから構成される。乱流音源は、口腔と共鳴したフォルマント周波数に近い回転周波数を持つ多様な渦音[4]の発生などによる乱流が音源となっていると考えられる。

音源波形としてサンプリング周期の精度のパルスを用いる方式は、正確な音階に必要な精度が高い周波数では得られないという難点を持つため、ここでは、白色雑音を入力とし、フォルマント(極)周波数 F_1 と F_2 の平均値を音階周波数として口笛音の合成を行う方式とした。

また、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ーによりモデル化した。図-2に乱流音源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一のものを用いた。

6. 合成実験と結果

呼気的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。

6.1 3オクターブ合成実験

3オクターブ(590-4324Hz)の音域で $(F_1+F_2)/2$ として平均律音階の理論値を用いることにより、乱流音源による上記口笛合成モデルで合成音を作成し試聴した。その結果、正確な音程を持つ合成音が得られることが確かめられた。

6.2 インパルス音源と乱流音源の音質比較実験

インパルス音源と乱流音源の口笛音の合成を行い音質比較を行った結果では、後者のほうが自然性に優れていた。混合音源は中間的な音質であった。

6.3 透明度の制御実験

透明度の制御方法としては、バンド幅(あるいはQ)の制御、共鳴フィルターの多重接続の制御などが考えられるが予備的な実験によってもこれらの方法の有効性が確かめられた。また、 F_1 と F_2 との分離度 d を澄んだ口笛および擦れた口笛の分析で得られた分離度に基づいて設定し、生成された合成音を試聴して透明度の比較実験を行った。

この結果、透明度の変化が聴覚的にも認められ透明度の制御が F_1 と F_2 分離度によっても有効に制御できることがわかった。

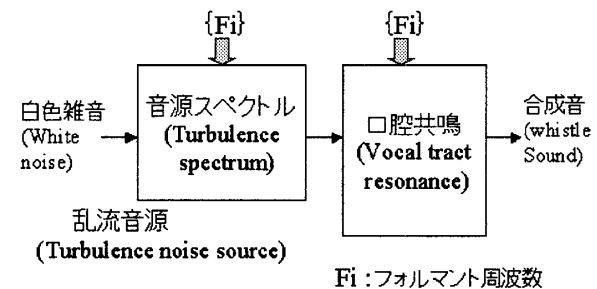


図-2: 乱流音源による口笛の合成モデル

7. あとがき

フォルマント数可変の Abs 的フォルマント抽出方式により口笛音の分析を行い、フォルマント周波数から音階周波数が求められること、また、乱流音源による口笛の合成モデルにより透明度の高い口笛音が得られ、フォルマント周波数 F_1, F_2 の分離度などを制御することにより透明度の制御が可能であることを分析・合成実験により確かめた。

参考文献

- [1] 伏木田, 電子情報通信学会総合大会, D-14-25., 2006-3
- [2] 伏木田, 日本音響学会, 音声研究会資料, S81-41, pp. 323-329, 1981-10.
- [3] Fushikida, Proc. IEEE ICASSP, E2.8:2260-2263, 1988.
- [4] 神部, ながれ 20, 流れと音の物理, pp. 174-186, 2001
- [5] <http://www.5b.biglobe.ne.jp/~hfyf/fk/Laboratory/Laboratory.htm>

表-1: 口笛(呼気)のフォルマント周波数分析結果例

Formant analysis of whistle (Expiration)/do-re-mi-fa-so-ra-si-do/											(Hz)	
	NOF	F1	F2	F3	F4	F5	F6	F7	F8	F9	(F1+F2)/2	平均律
do	9	1081	1081	2892	4297	5567	6830	8102	9293	10414	1081	1081
re	8	1205	1205	3283	4791	6238	7654	9056	10361		1205	1213
mi	8	1327	1327	3038	4741	6303	7753	8939	10361		1327	1362
fa	7	1419	1449	3207	5424	7137	8644	10441			1434	1443
so	7	1574	1649	3689	5509	7193	8688	10254			1612	1620
ra	7	1773	1819	3142	5424	7137	8688	10307			1796	1818
si	5	2018	2028	5382	7673	9940					2023	2041
do	5	2164	2164	5355	7673	9889					2164	2162

* NOF: Optimum number of formants