

減算合成による母音の合成と評価

2015SC014 後藤峻汰

指導教員：大石泰章

1 はじめに

どんなに複雑な周期波形であっても大小さまざまな正弦波の重ね合わせによって合成できることを保証するのが「重ね合わせの原理」であり、これにもとづいて音を作り出すテクニックが「加算合成」である。足し算の発想で音を作り出す加算合成は音響合成の最も基本的なものとして位置づけられている。しかし、加算合成による音作りは、重ね合わせる正弦波の数によって音質が左右されるが、正弦波の数が増えるとコントロールしなければならないパラメータの数もそれだけ増えて音作りが困難になる。

これに対して、容易に扱える音響合成のテクニックが「減算合成」である。あらかじめ多数の周波数成分を含んだ波形を用意し、こうした「原音」からフィルタを使って不要な周波数成分を削り取って音を作り出す。減算合成は、帯域通過フィルタの遮断周波数をコントロールするだけでよいため音作りの仕組みとして簡単である [1]。

本研究では、減算合成を用いて母音を生成する。具体的には、文献 [1] のプログラムを参考にし、パラメータ等を操作することでさまざまな母音を合成し評価する。

2 音声合成のメカニズム

人間が音声を生成する際に重要な役割を担っているのが「声帯」と「声道」という2つの音声器官である。肺から押し出された呼気は声帯を周期的に振動させ、多数の倍音を含む原音を作り出す。こうした原音が口腔や鼻腔を通過すると、その形状にしたがって周波数特性が変化し、音声は生成される。口腔や鼻腔はフィルタとして働き、あごを上下に開いたり、舌を前後に動かしたりすると、それにとまってフィルタの周波数特性は変化することになる。音声の通り道であることからこうした音声器官をまとめて声道と呼んでいる [1]。

図 1 に示すように、声道フィルタは帯域通過フィルタ（以下、BPF）を組み合わせたものでモデル化できる。それぞれのスペクトルのピークはフォルマントと呼ばれ、周波数の低いものから順番に「第1フォルマント（F1）」、「第2フォルマント（F2）」、「第3フォルマント（F3）」、「第4フォルマント（F4）」と名付けられる。

フォルマントをコントロールすると、原音はすべて同じでもパラメータしだいで異なる母音を生成することができる。

3 合成実験

文献 [1] に基づいて、音声を合成する。すなわち、原音としてすべての倍音が同じ大きさになっているパルス列を用い、合成したい音声のフォルマントに対応した BPF で

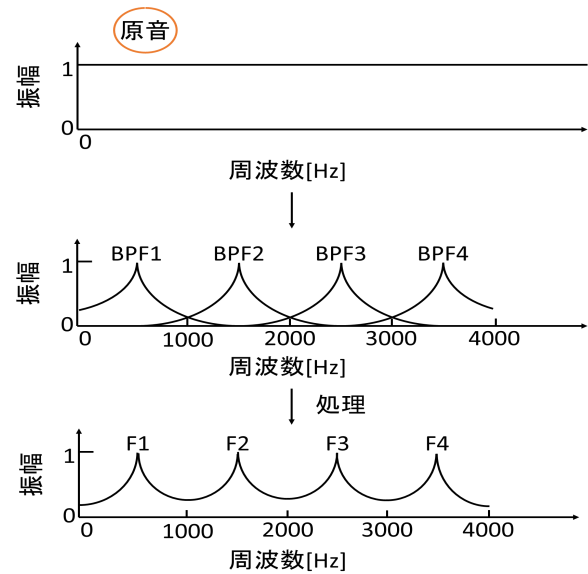


図 1 減算合成の概略図

それぞれフィルタリングしたのち、それらの結果を足し合わせる。

3.1 フォルマント周波数の調整

母音「ウ」、「エ」、「オ」の周波数特性が似ていることに着目し、フォルマントをコントロールすることで通常の母音以外の音声を合成できるかどうか検証する。母音「ウ」、「エ」、「オ」の周波数特性をそれぞれ図 2、図 3、図 4 に示す。それぞれの周波数特性について、F3、F4 はすべて等しい。特に「エ」と「オ」に関しては F1 も等しい。

まず F2 以外のすべてのフォルマントが等しい「エ」と「オ」について考え、F2 を 800[Hz] から 1900[Hz] まで 100[Hz] 刻みで推移させる。F2 を 800[Hz] から 1200[Hz] まで推移させると、「オ」の音声にだんだんと「ウ」の音声は混じっていくような音声を聞き取ることができた。理由としては、F2 が 1200[Hz] になった際には F1 以外のすべてのフォルマントが「ウ」と「オ」で一致するからであると考えられる。F2 を 1200[Hz] から 1900[Hz] まで推移させると、「オ」や「ウ」といった音声の特徴が薄まり「エ」が強調されていくように聞き取れた。こうした結果から、フォルマントをコントロールすることによってある母音とその他の母音の間の音声を作り出せることが確認できた。このことを応用して合成した音声に個性を作り出すことも可能であると考えられる。

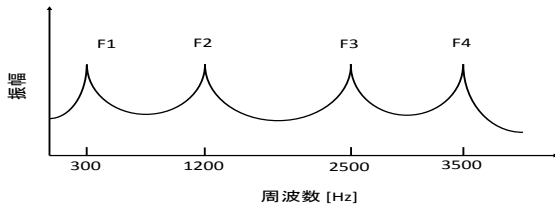


図2 音声の周波数特性：「ウ」

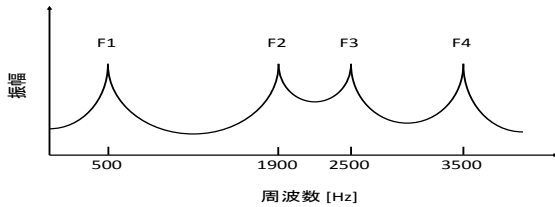


図3 音声の周波数特性：「エ」

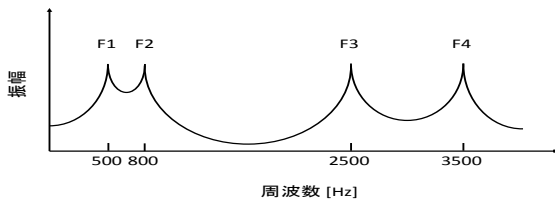


図4 音声の周波数特性：「オ」

3.2 フォルマントの数の増減

またフォルマントの数を増減させてどのような音を作り出せるか検証した。母音「ウ」に関してまずはF1のみ設定し、フォルマントを1つにして音を作り出したところ、音声というほどのものではなくただ単に機械的な音になった。次にF1とF2を設定し、フォルマントを2つにして音を作り出した。1つのときよりも音がこもったようになり豊かさが増したように思えたが、機械的な音に変わりはなかった。最後にF1、F2、F3を設定し、フォルマントを3つにして音を作り出したところ、フォルマントが1つや2つのときよりも音が母音に近づき音声ともとれるものに変わった。母音「エ」や「オ」の場合も同様であった。この結果から、人間の音声に近づけるためにはフォルマントの数は3つ以下では不十分であり、4つ以上が適当であることがわかった。

3.3 フォルマントの振幅の調整

これまでの実験では、フォルマントの振幅はすべて一定の値にしていた。しかし、文献 [2] の図 3 によると、それぞれのフォルマントの振幅は必ずしも一定の値をとるとは限らない。そこで、文献 [2] の図 3 に基づいて、「ア」と「イ」について F1 から F5 の振幅を変化させ、聞き取れる音の違いを確認した (図 5)。

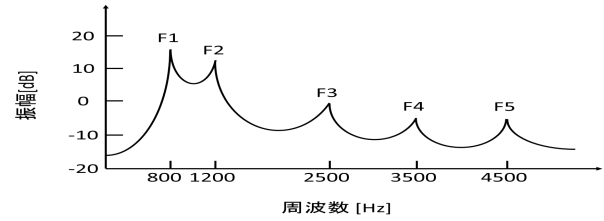
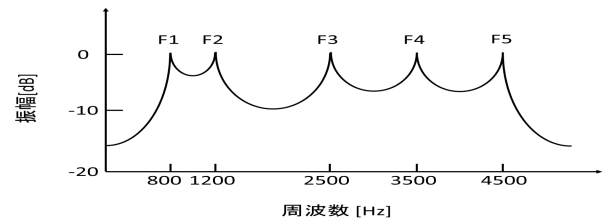


図5 母音「ア」のフォルマントの振幅変化

振幅を変化させる前とさせた後では音の大きさ等に多少の変化を感じ取れたものの、これによって大きく人間の音声に近づいたとは言い難い。この結果から、振幅の変化による音声の変化はほとんどないことが確認できた。

4 おわりに

本研究では、母音を合成するプログラムを実装し、フォルマントをコントロールすることで母音やそれ以外の音声を合成して評価した。結果としては、減算合成だけでは十分に人間の音声に近づけることはできなかった。今後の課題は、加算合成による実装や実験、評価等があげられる。

母音以外にも子音であったり、ゲームの BGM や効果音に利用されている電子音、アナログシンセサイザによる楽器音などの音響を合成できるプログラムが存在する。また、マイクから入力された音を加工すると同時にスピーカーから出力するといったリアルタイム処理のサウンドプログラミングなど音響合成の分野は幅広い。そういったプログラムの実装や改良も課題としてあげられる。

参考文献

- [1] 青木直史：『サウンドプログラミング入門—音響合成の基本とC言語による実装』。技術評論社、東京、2013。
- [2] 廣谷定男：「母音のフォルマント分析—過程と仮定を知る」。日本音響学会誌，70 卷 (2014) 10 号，pp. 538-544。