

デジタル・パターン・プレイバック*

荒井隆行, 安啓一, 後藤崇公 (上智大・理工)

1 はじめに

パターン・プレイバック (Pattern Playback) はサウンドスペクトログラムから音声信号を復元する機械として、Haskins 研究所の Cooper らによって 1940 年代終わり頃に作られ[1]、その後の音声研究の飛躍的な発展に大きく寄与した[2-4]。特に、この機械で可聴化することによって、スペクトログラムに反映される情報の中でもどのような音響的キューが音声知覚に重要であるかを確認することができる。場合によってはその音響的キューを単純化して表現し、また系統的に変化させながら合成することも可能となる。これによって、閉鎖子音の知覚において後続する母音の第 2 フォルマントの軌跡が重要であることを説明するローカス理論など多くの実験が行われた[5]。ところで、現代版パターン・プレイバックはデジタル信号処理技術を用いて実現することが可能で、またその教育的応用の価値は高いと推測される。そこで、本研究ではデジタル・パターン・プレイバックを実現した。

2 原理

オリジナルのパターン・プレイバックは、Fig. 1 に示すような仕組みで構成される[1]。光源とトーン・ホイールによって 120 Hz の整数倍の光の正弦波が作られ、その倍音構造がスペクトログラムによって振幅変調 (AM) される。スペクトログラムはベルトの上に乗せられ一定の速さで送られることで、AM 信号が出力される。

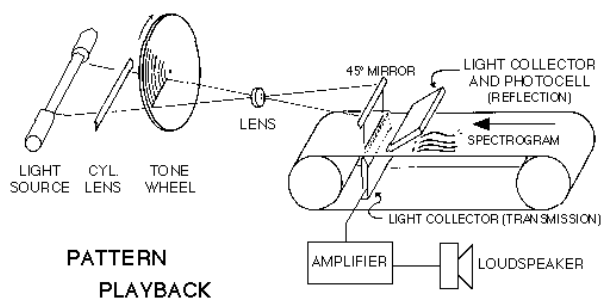


Fig. 1 Pattern playback [1].

そのアナログ版パターン・プレイバックは、近年のデジタル処理により容易にデジタル化され得る。現に、Nye らは“Digital Pattern Playback”に関するレポートを Haskins 研究所から出している[6]。本稿では、デジタル・パターン・プレイバックのアルゴリズムの可能性を再検討し、その実現を試みた。

第 1 のアルゴリズムとして、Fig. 2 に示すようにデジタル的に発生させた正弦波による倍音構造を駆動源とし、それをスペクトログラムによって AM 変調する手法が考えられる (AM 法)。これは、オリジナルの原理をそのままデジタル的に実現したものであり、音源フィルタ理論的な考えに基づいている。

このアルゴリズムでは、各倍音成分の振幅をスペクトログラム上で対応する領域の濃淡によって変調する。入力する正弦信号群の基本周波数を変えることによって、ピッチ変化 (イントネーションなど) を付加することが可能になる。また、雑音駆動に置き換えることも可能である。

なお、スペクトログラムから位相を含めて音声信号を復元する試みも行われているが[7]、教育的応用を考えた場合、フォルマント遷移の重要性などに焦点を当て、なるべくシンブルに復元したいなどの理由もあり、位相の復元は行わないこととした。

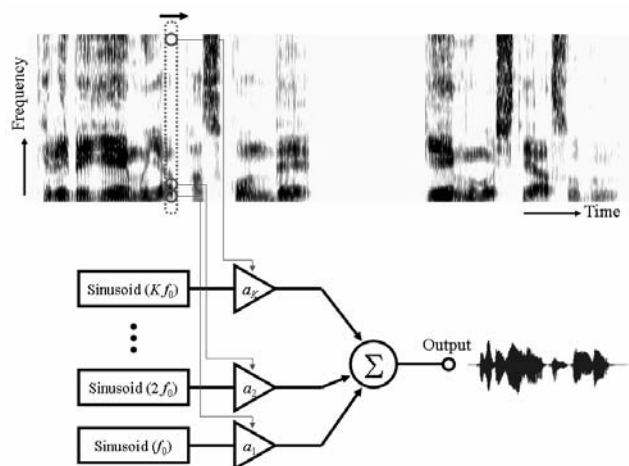


Fig. 2 Block diagram for the Digital Pattern Playback (AM-based algorithm).

* Digital Pattern Playback by ARAI, Takayuki, YASU, Keiichi, and GOTO, Takahito (Sophia Univ.).

第2のアルゴリズムは、Fig. 3のようにFFT（高速フーリエ変換）を用いる方法が考えられる（FFT法）。スペクトログラムの濃淡からフレームごとにスペクトル包絡を求め、それをIFFTすることによってその時点での声道のインパルス応答が近似される。フレーム周期（シフト幅）をピッチ周期に合わせることで、容易に倍音構造を達成することができるようになる。その際、フレーム周期を3~13ms程度にすることによって妥当な音声信号を実現することができる。なお、時間軸上において各フレームの配置を制御することによって、基本周波数を変化させることも原理的には可能となるが、ここでは簡単のためにフレームシフト幅は固定とした。

フレームに関する各種パラメータは、様々な値を取り得る。例えば標本化周波数（ F_s ）は8/16/48kHz、フレーム長（ N ）は256点や512点、フレーム周期は10ms（この場合、基本周波数は100Hz）などである。これらは、組み合わせによってはoverlap-addをする必要がなくなる。音声スペクトルの表現は0-8kHzの帯域に対して40点程度あれば問題なく明瞭な音声が発元される。

3 スペクトログラムの可聴化

3.1 オフライン版

スペクトログラムを画像ファイルとしてMatlab上で読み込み音声信号を合成した結果、AM法・FFT法の両手法によって明瞭な音声を得られた。また、webカメラから画像を読み込んだ直後に合成することも可能であることが確認された。

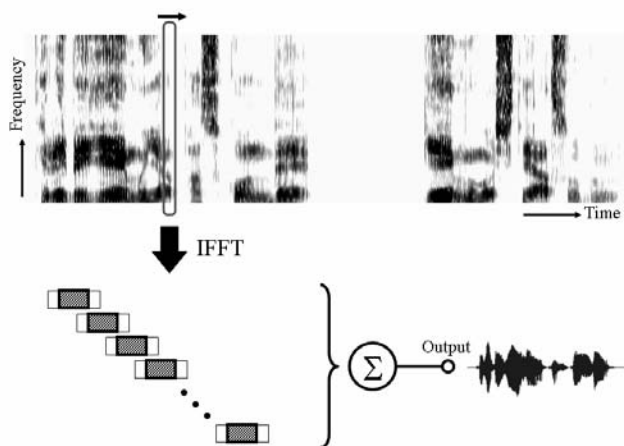


Fig. 3 Block diagram for the Digital Pattern Playback (FFT-based algorithm).

3.2 実時間処理に関する検討

実時間処理を考えた場合、画像取り込みと1ピッチ周期の音声波形の計算ならびに出力を1フレームとして1つのループに出来れば、FFT法のほうが都合がよい。カメラから画像を逐次読み込み、その都度IFFTした結果を音声信号として出力させる手法などが考えられる。本処理をVideo Portをペリフェラルとして持っているTexas Instruments社のDSP、DM642を用いて検討を行った結果、動作周波数720MHzという高性能を生かすことで処理の実時間処理化が可能であると考えられた。

4 おわりに

現代版パターン・プレイバックを、デジタル信号処理技術を用いて実現した。ピッチが一定の倍音構造を用いて/s/などの音が出ないように思うが、まばらな小さい斑点を用いることでそのような音を実現される[4]。このデジタル・パターン・プレイバックは教育的応用の意義が大きいと考えられることから、多くの教育現場で活用していきたい。

謝辞

本研究の一部は日本学術振興会の科研費（A-2, 16203041ならびにC-2, 17500603）から助成を得た。

参考文献

- [1] <http://www.haskins.yale.edu/haskins/MISC/PP/pp.html>
- [2] Cooper *et al.*, "The interconversion of audible and visible patterns as a basis for research in the perception of speech," *PNAS*, 37, 318-325, 1951.
- [3] Cooper *et al.*, "Some experiments on the perception of synthetic speech sounds," *J. Acoust. Soc. Am.*, 24, 597-606, 1952.
- [4] Borst, "The use of spectrograms for speech analysis and synthesis," *J. Audio Eng. Soc.*, 4, 14-23, 1956.
- [5] 荒井他, 音声の音響分析, 海文堂, 1996.
- [6] Nye *et al.*, "A digital pattern playback for the analysis and manipulation of speech signals," *Haskins Lab. Status Report on Speech Research*, SR-44, 95-107, 1975.
- [7] Slaney, "Pattern playback from 1950 to 1995," *Proc. IEEE Systems, Man and Cybernetics Conf.*, 1995.