ディジタル・パターン・プレイバックを用いた音響学ならびに信号処理工学への教育的応用*

○荒井隆行（上智大・理工）

1　はじめに

サウンドスペクトログラムを音声信号に変換する技術はパターン・プレイバック（pattern playback, 以後 PP と略す）と呼ばれ，1940 年代終わりにアメリカ Haskins 研究所の Cooper らによって開発されてから，その後の音声研究に大きく貢献した[1-3]. 現在は計算機やアルゴリズムの著しい発展によって, PP そのものが音声研究に用いられることはほとんどなくなったが，その教育的価値の高さから荒井らはディジタル版の PP である「ディジタル・パターン・プレイバック（digital pattern playback, 以後 DPP と略す)」を開発し，教育的応用を進めている[4-6].

本稿では, その DPP について以下の２つの点について試みた結果を報告する：

１）正弦波加算法によるアルゴリズムの簡素化,

２）基本周波数（f0）を変化させる試み.

2　正弦波加算法によるDPP

最初に開発された DPP のアルゴリズムにはAM法とFFT法の2つが存在する[4,5]. AM 法は PP が実現する光学式の合成法と基本的な考え方は同じであり，倍音成分の 1 つ 1 つをキャリア信号と見なしてスペクトログラムから得られる濃淡情報でそのキャリア信号を振幅変調（amplitude modulation, AM）するものである．一方，FFT 法はスペクトログラムを数十 ms 程度のフレームに分割し，そこから得られる濃淡情報を短時間スペクトルと見なして FFT（高速フーリエ変換, fast Fourier transform）による逆フーリエ変換を実施後，overlap add（OLA）するものである．ところで，逆フーリエ変換はFFTを使わなくても実現でき，もともとの PP が倍音構造を前提としていることから，正弦波を加算することによってフーリエ合成するほうが，アルゴリズムもシンプルに実現される．そこで考えられたのが，図 1 に示すような正弦波加算法による DPP アルゴリズムである．この図を見ると分かるように，図の上部にあるスペクトログラムのうちある時刻に注目し（図中で縦長の長方形), その時点における濃淡情報を振幅スペクトルと見なしてそれを a_i という係数においている(図の左下)．係数 a_i は倍音構造を持つ正弦波群のそれぞれの振幅として乗算され，それらを加算することによって時間波形を得る．その時間波形は周期信号になるが，これはその時点における声道インパルス応答が繰り返されたものと見なすことができる．この時間波形を有限長で切り出すために窓関数を掛け，注目する時刻（フレーム）を順次左から右にずらす（フレームシフト）ことにより得られる時間波形を OLA することで，最終的な出力信号が求められる．

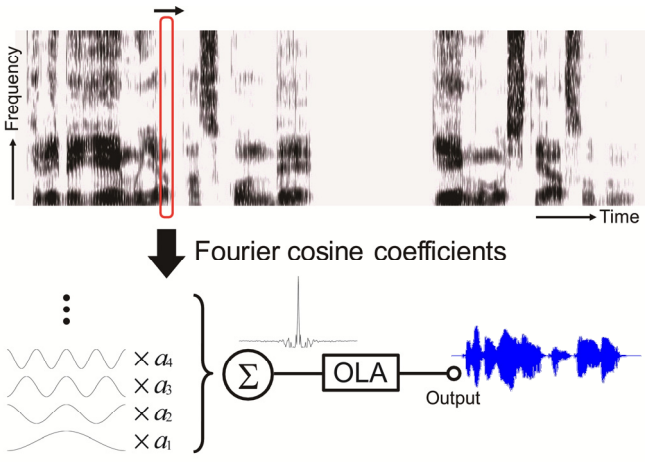

Fig. 1. Schematic representation of DPP algorithm based on the additive synthesis of sinusoidal harmonics. The overlap-add (OLA) technique was only applied when we used a non-constant pitch contour.

* Applying the digital pattern playback to education in acoustics and signal processing engineering, by ARAI, Takayuki (Sophia University).

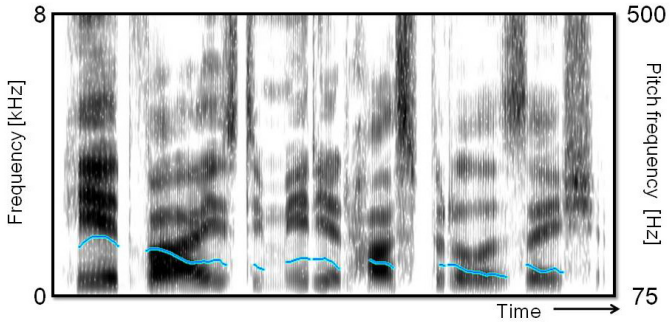

Fig. 2. Input image of sample sentence in the spectrographic representation with overlaid pitch contour.

より厳密に言えば，正弦波群を加算することによって得られる周期信号は，その正弦波群の基本周波数の逆数（基本周期）ごとにインパルスが並んだインパルス列に声道インパルス応答がたたみ込まれた信号ということになり，一般にはインパルス応答の長さが基本周期より長くなることもあり得る．しかし，正弦波群の位相をすべて 0 としフーリエ cos 合成を行うことによって，上記の周期信号は各周期の両端で振幅が十分小さくなる結果，方形窓で切り出すことが可能となる．正弦波群の基本周波数を例えば後の例のように 100 Hz と低く設定することによって基本周期が 10 ms と長くなると，さらに両端での振幅は小さくなる．そして，フレームシフトの幅を基本周期と同じ（上の例の場合，10 ms）とすれば，OLA ではなく波形を単にフレームごとに接続することで，最終的に基本周波数が 100 Hz の音声信号が得られる．

3 f0 可変式DPP

オリジナルの PP は f0 を変化させず，一定である．それを踏襲し，また教育的にもシンプルに実現するために DPP では f0 を固定にしている．一方，過去にはいくつもの研究がスペクトログラムからオリジナルの f0 曲線を推定し，抑揚のある音声信号を復元する試みが行われている（例えば[7]）．DPP においても，比較的シンプルに f0 を変化させる試みを行った．その際，正弦波加算法において正弦波群に含まれる倍音成分の基本周波数は固定とし，フレームのシフト幅を所望の f0 曲線から得られる f0 周波数の逆数から求めた．つまり，ある時刻における声道インパルス応答がフーリエ合成によって得られ，そのようなインパルス応答が声帯振動に合わせてたたみ込まれるという考え方である．

3.1 スペクトログラムに重ねて描かれたf0曲線からのDPP

我々は以前から，紙に印刷されたスペクトログラムをカメラで画像として取り込むことで行う DPP のデモンストレーションが，PC 上だけで完結する DPP よりも効果的であることを実証している[5]．そこで，紙の上にスペクトログラムと f0 曲線が重ねて描かれた印刷物から，カメラを経由して画像を取り込むことで DPP を行うことを試みた．幸い，多くの音声分析ソフトウェアにおいて，両者を重ねて描画するモードが存在するため，それを印刷して使用することを考えた．Fig. 2 に音声分析ソフトウェアの Praat [8]を用いて描画したスペクトログラムならびに f0 曲線を示す．この図で音声サンプルは，TIMIT コーパス[9]からの“He advised immediate hospitalization” という男性の発話である．

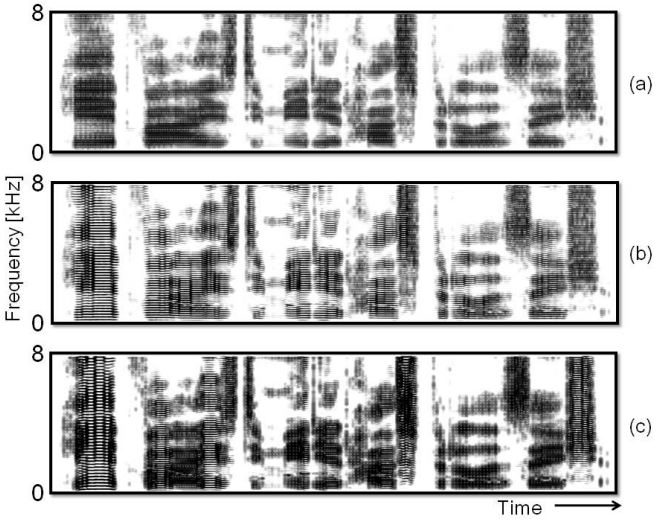

Fig. 3. Narrow-band spectrographic representations of synthesized signals; (a) the original DPP, (b) DPP with the measured pitch contour, and (c) DPP with the estimated pitch contour.

```
Fs = 16000;              % 標本化周波数（Hz）
f0 = 100;                % 基本周波数（Hz）
T = Fs/f0;               % 基本周期（サンプル）
f_pixel = 80;            % 縦ピクセル数
t_pixel = 240;           % 横ピクセル数
S = imread('samples.jpg');    % 画像読み込み
output = zeros(T*t_pixel,1);  % 出力用初期化
for frame=1:t_pixel           % フレーム用ループ
    ns = T*(frame-1)+1;       % 開始サンプル
    ne = ns+T-1;              % 終了サンプル
    for k = 1:f_pixel     % フーリエ合成用ループ
        A = 255-double(S(f_pixel-k+1,frame));
                              % フーリエ係数
        output(ns:ne) =       % cos 関数の加算
            output(ns:ne)
            +A*cos(2*pi*k*[-T/2:T/2-1]/T)';
    end
end
soundsc(output, 16000)        % 音声出力
```

Fig. 4.　Matlab script for additive synthesis of sinusoidal harmonics for DPP.

Fig. 2 では，スペクトログラムは白黒の濃淡で 0 から 8 kHz の範囲で描画されているのに対し，f0 曲線は 75 から 100 Hz の範囲でカラー（シアン）で描画されている．両者は異なる色で描画されている結果，色情報からそれらを分離し，f0 曲線を反映させた DPP 出力を得ることが可能である．結果として得られた復元音声信号の狭帯域スペクトログラムを Fig. 3 に示す. Fig. 3 の(a)は f0 曲線を反映させていない DPP 出力であるのに対し, (b)は f0 曲線を反映させたものである．

3.2 インテンシティーからのf0 の推定

Saikachi らは，電気喉頭音声の f0 を自動的に推定することを目的として，f0 曲線を音声信号の瞬時インテンシティーから求めることを試みている[10]．この手法によって求められた f0 曲線を用いて DPP した結果を Fig. 3 (c)に示す．結果として得られる音声信号のイントネーションは, 3.1 節のものに比べるとそこまで自然ではないものの, f0 が一定のものに比べると自然性が向上することが確認された．

4　教育現場での応用例

4.1 概要

上智大学理工学部情報理工学科の 2 年次生を対象にして行われている実験（情報理工学実験Ⅰ）にて，実際に DPP が導入されている．この実験は全体で 14 週あるが，1 週につき 2 コマ連続（3 時間）を使い，2 週で 1 テーマをカバーする．DPP は「信号処理」というの章の一部（2 週目後半）に登場するが，この章ではまずフーリエ級数展開，フーリエ合成，正弦波と音（音階と等比数列を含む）などを Matlab による演習形式で学ぶ．学生は教員とティーチングアシスタントの指導のもと, 自分で Matlab スクリプトを入力しながら実行結果を確認する. DPP については, 教員による説明の後, 教員と一緒に Matlab スクリプトを自分で試しながら進めるが，示されるスクリプトにはアルゴリズム上のバグわざと 2 か所, 埋め込まれている．学生は教員のヒントを手掛かりにデバッグをしながら正しい出力を得る．

4.2 学生実験で用いるアルゴリズム

シンプルなアルゴリズムで DPP を実現するため，2 節で説明した正弦波加算法を採用し，f0 は一定とした．Fig. 4 に，実際に実験で用いられている Matlab スクリプトを示す．実験では各学生に対して入力画像をファイルで与えるものとし，画像は 240 x 80 ピクセルの JPEG 形式とした．画像ファイルは，ある音声信号から音声分析ソフトウェアなどを用いて事前にこの大きさに作成した．このような低い解像度を用いた理由は，やはりアルゴリズムをシンプルにするためである．標本化周波数を 16 kHz とし f0 を 100 Hz とした場合，縦を 80 ピクセルにすることによって，スペクトログラムの縦の周波数範囲である 0～8000 Hz において，1 ピクセルがちょうど 1 つの倍音成分の周波数位置に対応する．また，フレームシフトを 10 ms として横方向の 1 ピクセルをちょうどフレームシフトに合わ

Table 1　Number of students who understood the algorithm of the DPP vs. who got interested in the DPP.

		興味を持てましたか？		
		持てず	まあ	とても
理解できましたか？	できず	4	7	3
	まあ	2	35	43
	とても		1	4

せれば，240 ピクセルで 2.4 s となり，標準的な「短い文」にちょうど良い長さとなる．

上述のように縦 1 ピクセルが一つ一つの倍音成分に対応しているので，縦方向の各ピクセルに対応した濃淡値を *a* 係数と見なして cos 関数に乗算している．また，やはり上述のように横 1 ピクセルがフレームシフト幅に対応しているため，横方向の各ピクセル毎に時間波形（160 サンプル）を計算し，それを順次接続している．なお，正弦波群として cos 関数を用いていることによって，加算後の時間波形は 10 ms の周期ごとに振幅はかなり小さくなる．そのため，そこで波形を切って次のフレームの時間波形と接続しても不連続は十分無視できる程度である．

4.3 評価

実験の後，受講した全学生 112 名に対してアンケートを実施した．約 53%の受講生が同時期に「ディジタル信号処理」の講義を履修中であった．アンケートの結果の一部を Table 1 に示す．この表を見ると分かるように，約 86%の学生がまあまあ理解できた，あるいはとても理解できたと回答した．一方，約 94%の学生がまあまあ興味を持てた，あるいはとても興味を持てたと回答した．必ずしも理解が十分でない場合でも，強い興味を示す傾向があることが分かった．

5　おわりに

著者らによって提案している DPP について，さらなる教育的応用を目的にアルゴリズムの簡素化や簡単に f0 を変化させる手法について提案した．実際に教育現場で応用した結果，一定の教育効果を確認した．

謝辞

f0 可変式 DPP の開発にあたり，アドバイスをいただきました ATR-Promotions の正木信夫さんに感謝いたします．内容の一部は日本学術振興会の科学研究費補助金（21500841），及び文部科学省私立大学学術研究高度化推進事業上智大学オープン・リサーチ・センター「人間情報科学研究プロジェクト」の助成を得た．

参考文献

[1] F. S. Cooper, A. M. Liberman and J. M. Borst, “The interconversion of audible and visible patterns as a basis for research in the perception of speech,” *PNAS*, 37, 318-325, 1951.

[2] F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst and L. J. Gerstman, “Some experiments on the perception of synthetic speech sounds,” *J. Acoust. Soc. Am.*, 24(6), 597-606, 1952.

[3] J. M. Borst, “The use of spectrograms for speech analysis and synthesis,” *J. Audio Eng. Soc.*, 4, 14-23, 1956.

[4] 荒井隆行, 安啓一, 後藤崇公, “ディジタル・パターン・プレイバック,” 音講論, 429-430, 2005.9.

[5] T. Arai, K. Yasu and T. Goto, “Digital pattern playback: Converting spectrograms to sound for educational purposes,” *Acoust. Sci. Tech.*, 27(6), pp. 393-395, 2006.

[6] 英語版 Wikipedia "Pattern Playback" http://en.wikipedia.org/wiki/Pattern_playback

[7] M. Slaney, “Pattern playback from 1950 to 1995,” *Proc. IEEE Int’l Conf. Systems, Man and Cybernetics Conf.*, 4, 3519-3524, 1995.

[8] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glot International*, 5(9/10), 341-345, 2001.

[9] V. Zue, S. Seneff and J. Glass, “Speech database development at MIT: TIMIT and beyond,” *Speech Communication*, 9(4), 351-356, 1990.

[10] Y. Saikachi, K. N. Stevens, R. Hillman, "Development and perceptual evaluation of amplitude based F0 control in electrolarynx speech," *Journal of Speech, Language, and Hearing Research*, 2009.