（招待講演）パターン・プレイバックから スペクトログラム・ボコーダへ*

○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フ（sound spectrograph; 以下，スペクトログラフ）は1940年代にBell研究所でPotterらによって開発された[1]．第2次世界大戦においてスクランブルされた音声を解析するのにも用いられたとの話もあるが[2]，分析結果として表わされるサウンドスペクトログラム（sound spectrogram; 以下，スペクトログラム）は，時間と共に変化する周波数特性が濃淡として表わされる．そしてこのスペクトログラム上に現れる音声信号の時間ー周波数パターンを見ることによって様々な分析を行うことが可能となり，その後の音声研究において音響分析には欠かせない存在となった．

初期のスペクトログラフはアナログ式のものであり，その基本的な考え方はフィルタバンクによる周波数分析である．周波数軸上に配置された複数のバンドパスフィルタ（band-pass filter; 以下，BPF）によって音声信号を分析し，その出力レベルに応じて濃淡表示される．ヘテロダイン方式と呼ばれるものでは，固定のBPFを1つ用意し，音声信号に少しずつ異なる周波数で変調を掛けながら対象とする周波数帯域を連続的変化させて分析を行う[3,4]．いずれにおいても，BPFの帯域幅が狭ければ狭帯域分析，広ければ広帯域分析と呼ばれる．

ところで，スペクトログラフによる音声の分析が盛んに始まった1940年代後半，スペクトログラムの濃淡表示から音声を合成する試みがHaskins研究所のCooperらによって進められた[1]（Fig. 1）．その結果，作られたのがパターン・プレイバック（pattern playback; 以下，PP）であり，その後の音声研究の飛躍的な発展に大きく寄与した[5-7]．このPPを利用する最大の利点は，スペクトログラム上に現れるパターンの中で，どの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点である．それを確かめるためには，例えばその音響的キューだけを抜き出してパターンを描いてみる，あるいはそのパターンを変化させてみることが有用である．そのためにスペクトログラムを単純化して表現し，系統的に変化させながらPPによって合成することにより，その変化がどのように知覚上の違いを生み出すかを研究できるようになった．その結果，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[4]など多くの実験が行われた．

一方，音声の合成に関しては，1930年代後半にBell研究所のDudleyによってVoderが開発された[1,8]. 1939-1940年に開催された世界博覧会にこのVoderが出展され有名となったが，合わせてDudleyは音声をより少ない情報量で伝送することを試みた．それがボコーダ（Vocoder）の始まりであり，初期のものは音声信号を複数の周波数帯域に分割し，各チャンネルの時間包絡情報に加え，雑音音源やパルス音源を組み合わせた音声合成器を電子回路にて実現した（チャンネル・ボコーダ）[9].

以上のように初期の分析・合成はアナログ式であったが，近代になってディジタル信号処理が発展すると，様々な処理がプロセッサやコンピュータ上で実現されるようになった. ボコーダについてもその後，様々なものが提案され，それらは例えば位相ボコーダ，ケプストラム・ボコーダ，LPCボコーダなど多岐にわたる[10]. そして，1990年代に入り，STRAIGHTによるボコーダによって高精細な音声分析・合成が可能となった[11].

そしてPPについても，近年になってディジタル処理によって容易に実現できるよ

* From pattern playback to spectrogram vocoder, by ARAI, Takayuki (Sophia University).

うになった[12]．実際，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[13]．そしてその後，著者らによってディジタル・パターン・プレイバック（digital pattern playback; 以下，DPP）が見直され，教育目的に応用してきている[14-16]．著者らが開発した最初の DPP では，PP をそのままディジタル版にすることが目的であったため，音源の基本周波数（f0）は一定であった．しかも，アルゴリズムはオリジナルの PP を模擬するため，固定の f0 に対する倍音を発生させ，その倍音に対応する正弦信号をスペクトログラムの濃淡で振幅変調することで合成を行った．なお，アナログ版 PP では，倍音に対応する正弦信号の発生から振幅変調までを光学的に実現しており，その原理はチャンネル・ボコーダと同じである．

しかし，ディジタル信号処理の観点において DPP で f0 を時間と共に変化させることは容易に実現される上，アルゴリズムも他のボコーダのように時間フレームに対応したスペクトル分析に基づくものなど，多岐に実現される．このことを受け，著者らは簡単なフーリエ合成を応用し，簡易的な f0 可変式の DPP を使って教育目的に応用している[17-19]．

ところで f0 を可変にした DPP については，音声をスペクトログラムと音源情報に分離して表現し，再び再合成する分析合成系とみなすことができる．すなわち，これはいわば「スペクトログラム・ボコーダ（以下，SGV）」である（Fig. 2）．そして，スペクトログラムを経由することにより，もともと PP の利点であった「注目する音響的キューを抽出し，必要に応じて系統的に変化させて音声を出力する」という点を最大限に活用できる．その観点から，本稿ではこの SGV について，DPP と合わせて教育的及び研究的側面の両面から考えてみる．

2　音声知覚における音響的キュー

2.1 接近音の例

PP を中心とする合成手法によって議論されてきた音声知覚における古典的な例として，まず接近音の例に触れる．接近音では声道における形状が母音のように定常的ではなく，時間と共に変化する．しかし，閉鎖音や摩擦音のように声道内において閉鎖や強い狭めが作られることはなく，いわばある母音から別の母音に遷移するのに近い．そして，またその変化する速度は接近音ごとに最適値は異なるものの，一般に二重母音よりも速い．その結果，スペクトログラム上では時間と共に変化するフォルマントのパターンの違いとして接近音が表わされる．

そこで，フォルマント遷移を音声知覚上，重要な音響的キューとして捉え，PP によってどのようなフォルマント遷移が最もその子音らしいかを合成音声によって確認することが可能となる．実際，DPP による合成音声を聞くと，各パターンに従ってその違いを確認することができる．

2.2 閉鎖音の例

接近音と同様に，閉鎖音では声道における形状が時間と共に変化する．しかし，閉鎖音では口腔が完全に閉鎖され，口腔内圧が高まった後，解放される際に一気に調音器官が動くため，結果として描かれるフォルマント遷移は接近音の場合よりも速い．

調音位置の異なる有声閉鎖音の/b/, /d/, /g/に対し，フォルマント遷移のパターンについても実際，DPP による合成音声を聞くと，それぞれのパターンに従ってその違いを確認することが可能である．

3　SGVの教育応用例

3.1 音響音声学デモンストレーション

前節では，音響的キューの例としてフォルマント遷移パターンの違いによって，接近音や閉鎖音の調音位置の違いなどが異なって知覚される古典的な例を見た．しかし，一方でこのような実験を再現したり実演したりする機会が思うほどないのではないかと，著者は長年感じていた．そこで，著者が中心となって「音響音声学デモンストレーション（Acoustic-Phonetics Demonstrations, APD）」を作り始めた[20]．この APD では DPP や SGV，フォルマント合成器などで作られる刺激音や連続体などによるデモも複数含まれており，web 上で公開されている[21]．今後，さらなるデモの追加やインタラクティブなものも増やす計画である．（現

在，DPP に関してはユーザが準備した画像ファイルを音声ファイルに変換するページなども公開している）．

3.2 博物館での応用例

日本音響学会では 2007 年から音響教育調査研究委員会が中心となって，夏休みの期間中の 2 日間，国立科学博物館にて「サイエンススクエア」という催しに音に関するブースを出展している[22]．2011 年からは音声のブースも加わり，子どもたちの声を分析してスペクトログラムを見せたり，声道模型の実演などと合わせ声の仕組みに関する体験型の学習の場を提供している．その一環として，来場者の声を録音し，スペクトログラムを印刷後，DPP を使って再合成することも行っている．印刷されたスペクトログラムをお土産として持ち帰ることができることも相まって，人気の展示の一つになっている[22]．2012 年からは f0 曲線を赤い針金などで実現し，印刷されたスペクトログラム上に乗せて一緒にコンピュータに取り込んだり，事前に準備されたメロディーと組み合わせてイントネーション付きの音声，あるいは歌声を合成することも試みており，子どもたちの反応も上々である[22]．

一方，ソニー・エクスプローラサイエンスにおいても，企画展として 2008 年に DPP の展示を行った．展示の前半では声道模型の体験コーナーや，自分の声をスペクトログラム分析するコーナーもあり，その上で，スペクトログラムから音声を DPP を使って合成する．それも，日本語のモーラに対応させて，時間幅は短く，周波数方向に長いスペクトログラムの断片をプラスティック製の板に貼り付け，そのような板を複数モーラ単位で並べ，長いスペクトログラムを作って DPP にて合成するというものである．来場者はパズル感覚で音声合成を体験でき，この展示も人気を博していた[17]．

3.3 信号処理教育応用

2009 年からは，上智大学理工学部情報理工学科の 2 年生を対象に，情報理工学実験Ⅰという実験科目において，SGV を扱っている．特に最初の数年はスペクトログラムは提供したものを使い，f0 も一定にしていた．しかし，2013 年からは自分の声を実際に録音し，その録音音声からスペクトログラムを計算させ，f0 も自分でメロディーを作って組み合わせることで歌声を合成するという内容に改めた．アルゴリズムの中でフーリエ合成を行っている様子も解説しながらの実習のため，信号処理教育も兼ねており，また最終的な音声が合成された際の達成感が得られる点からも，学生から高い評価を得ている[18]．

4 おわりに

本稿では SGV の教育的応用の例を複数示した．しかし，研究上の意義も忘れてはいけない．もともと，PP しか音声合成の手段がなかった時代，PP の出現によって音声研究が飛躍的に進歩した．その後，ディジタル信号処理によって多数のアルゴリズムやソフトウェアが出回るようになった．しかし，今一度，スペクトログラムなどの時間・周波数表現を経由する有用性をあえて強調したい．f0 曲線を含め，そこには音韻情報と韻律情報など，音声の主要な情報が凝縮されている．そして，それらを視覚的に捉えることにより，把握しづらかった音響的キューを判断し，実際にそのキューを操作することで再合成音によって確認することが可能となる．この一連の作業は，今の時代の音声研究においても重要であると考える．そして，その分析・合成には STRAIGHT を用いる「ハイブリッド型」にすれば，高品質な SGV が実現される．

謝辞

河原英紀先生，聴覚研究会幹事団の皆様，上智大学荒井研メンバーに御礼申し上げます．内容の一部は日本学術振興会の科学研究費（24501063）の助成を得た．

参考文献

[1] G. J. Borden, K. S. Harris and L. J. Raphael, *Speech Science Primer*, 4th ed., Lippincott Williams & Wilkins Inc., 2003.（廣瀬肇訳, 新 ことばの科学入門, 医学書院, 2005.）

[2] 城生佰太郎, 福盛貴弘, 斎藤純男編, 音声学基本事典, 勉誠出版, 2011.

[3] W. Koenig, H. K. Dunn and L. Y. Lacey, “The sound spectrograph,” *J. Acoust. Soc. Am.*, 18(1),

19-49, 1946.

[4] R. D. Kent and C. Read, *Acoustic Analysis of Speech*, 2nd ed., (Singular, San Diego, CA, 2001).（荒井隆行, 菅原勉監訳, 音声の音響分析, 海文堂, 1996.）

[5] 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.

[6] 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.

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

[8] B. Gold and N. Morgan, *Speech and Audio Signal Processing*, John Wiley & Sons, 2000.

[9] H. Dudley, "Remaking speech," *J. Acoust. Soc. Am.*, 11(2), 169-177, 1939.

[10] 古井貞熙, ディジタル音声処理, 東海大学出版会, 1985.

[11] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigné, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," *Speech Communication*, 27(3-4), 187-207, 1999.

[12] M. Slaney, "Pattern playback from 1950 to 1995," *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics Conf.*, 4, 3519-3524, 1995.

[13] P. W. Nye, L. J. Reiss, F. S. Cooper, R. M. McGuire, P. Mermelstein and T. Montlick, "A digital pattern playback for the analysis and manipulation of speech signals," *Haskins Lab. Status Report on Speech Research*, SR-44, 95-107, 1975.

[14] 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.

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

[16] Wikipedia (http://en.wikipedia.org/wiki/Pattern_playback).

[17] T. Arai, "Digital pattern playback for education in digital signal processing and speech science," *Proc. of the IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2769-2772, Kyoto, 2012.

[18] 荒井隆行, "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," 音講論, 1467-1470, 2012.3.

[19] 荒井隆行, "メロディーを楽音と歌声で合成しながら学ぶ," 音講論, 1617-1620, 2014.9.

[20] T. Arai, "Learning acoustic phonetics by listening, seeing, and touching," *Proc. of Meetings on Acoustics*, Vol. 19, 025017, pp. 1-9, 2013.

[21] http://www.splab.net/APD/

[22] 網野加苗, 荒井隆行, 佐藤史明, 中村健太郎, 西村明, 横山栄, "国立科学博物館「夏休みサイエンススクエア」への出展," 日本音響学会誌, 70(5), 296-298, 2014.

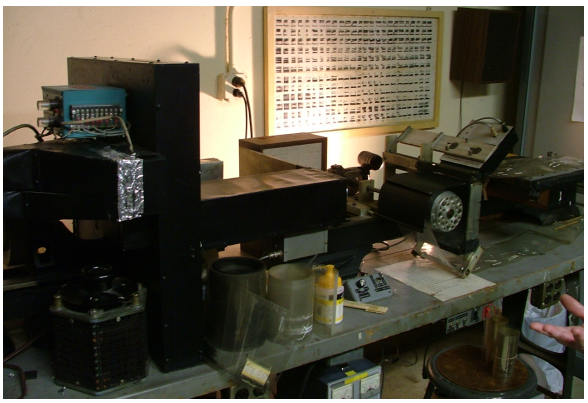

Fig. 1: Pattern Playback (Haskins Lab.).

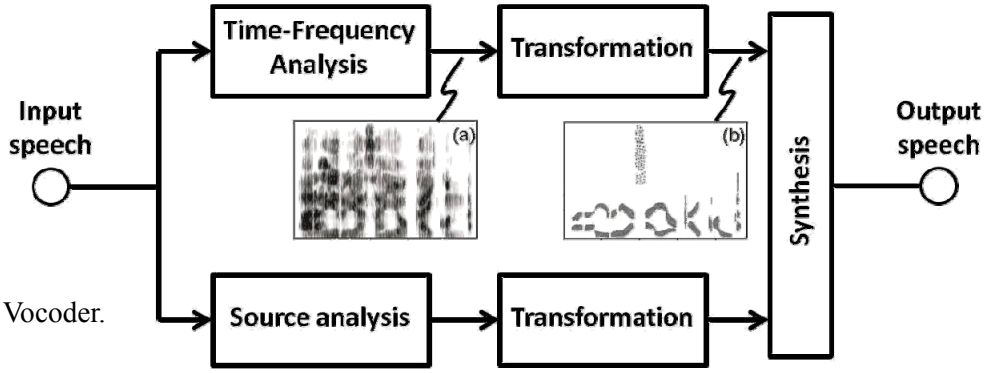

Fig. 2: Spectrogram Vocoder.