

# velvet noise とその変種の聴覚心理・生理研究への応用可能性について

河原 英紀<sup>†</sup> 津崎 実<sup>††</sup> 坂野 秀樹<sup>†††</sup> 森勢 将雅<sup>††††</sup> 松井 淑恵<sup>†††††</sup>  
入野 俊夫<sup>†</sup>

<sup>†</sup> 和歌山大学 〒640-8510 和歌山市栄谷 930

<sup>††</sup> 京都市立芸術大学 〒610-1197 京都市西京区大枝沓掛町 13-6

<sup>†††</sup> 名城大学 〒468-8502 名古屋市天白区塩釜口一丁目 501 番地

<sup>††††</sup> 山梨大学 〒400-8510 甲府市武田 4-4-37

<sup>†††††</sup> 豊橋技術科学大学 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: †{kawahara,irino}@sys.wakayama-u.ac.jp, ††minoru.tsuzaki@kcua.ac.jp, †††banno@meijo-u.ac.jp,  
††††mmorise@yamanashi.ac.jp, †††††tmatsui@cs.tut.ac.jp

あらまし velvet noise は、90～85%が零となり、1 と -1 がランダムに出現する疎な離散信号である。同じパワーの白色ガウス雑音よりも滑らかな印象を与えることから、velvet noise と命名されている。この velvet noise の構成法を周波数領域で余弦級数との畳み込みと併用することで、時間的に局在しかつ平坦なパワースペクトルを有する雑音様の信号を生成することができる。これらを生成する際に用いる乱数の seed の設定と、配置や分割数などの生成パラメータを設定することにより、雑音から周期信号、周期的パルス列からランダム位相の周期信号の間を連続的に埋める信号を生成することができる。この信号は、聴覚における処理と知覚を調べるための新しい手段を提供する。

キーワード ランダム信号、周期信号、位相、群遅延、ピッチ

## Possible application of velvet noise and its variant in psychology and physiology of hearing

Hideki KAWAHARA<sup>†</sup>, Minoru TSUZAKI<sup>††</sup>, Hideki BANNO<sup>†††</sup>, Morise MASANORI<sup>††††</sup>,  
Toshie MATSUI<sup>†††††</sup>, and Toshio IRINO<sup>†</sup>

<sup>†</sup> Wakayama University, 930 Sakae-dani, Wakayama, Wakayama, 640-8510 Japan

<sup>††</sup> Kyoto City University of Arts, 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto 610-1197 Japan

<sup>†††</sup> Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502 Japan

<sup>††††</sup> University of Yamanashi, 4-4-37, Takeda, Kofu, Yamanashi, 400-8510 Japan

<sup>†††††</sup> Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, 441-8580 Japan

E-mail: †{kawahara,irino}@sys.wakayama-u.ac.jp, ††minoru.tsuzaki@kcua.ac.jp, †††banno@meijo-u.ac.jp,  
††††mmorise@yamanashi.ac.jp, †††††tmatsui@cs.tut.ac.jp

**Abstract** We propose a set of test signals for psychology and physiology of hearing. They are the Velvet noise and its variant. The Velvet noise is a discrete random sequence of 1 or -1, which are sparsely (5 to 10% of available time slots) allocated on the discrete time axis. The name “velvet noise” was given because it sounds smoother than white Gaussian noise. A variant of the velvet noise is generated on the discrete frequency axis and using convolution with a cosine series to assure the flat power spectrum with temporally localized power distribution. Setting random seed conditions and generation parameters such as allocation period and the number of division enable to generate signals which span random to periodic, periodic pulse to random phase periodic signal seamlessly. These signals provide a unique tool for investigating psychology and physiology of hearing.

**Key words** random signal, periodic signal, phase, group delay, pitch

## 1. はじめに

全く違うように見える波形を有する音でも、同じ音として聴こえることがある。様々な年齢、性別、体型の人が発声した同じ（例えば）単語の波形も、一致することは無いにもかかわらず、同じ言葉に聴こえる。2016年の秋に突然発表された WaveNet と一連の研究 [1–3] は、技術を一挙に十年のオーダーで進め、柔軟で高品質な音声合成・変換を実用レベルのものとした。しかし、それらは人間の音声コミュニケーションの本質的な理解につながる「聴覚は何を聴いているのか」という素朴かつ根本的な疑問に、答えを与えるものではない。<sup>(注1)</sup>

ここでは、パルスと雑音の知覚と信号の物理的性質との関連を調べるための新しい可能性について紹介する。ここで用いる信号は、velvet noise と呼ばれており、音響効果の分野で発明されたものである [5]。著者の一人は、この velvet noise の発想を周波数領域で借用することにより、新たに有用な信号を作成できることを示した [6]。velvet noise と、こうして作られた velvet noise の変種を用いることで、上記の疑問の理解に貢献できる可能性がある。また、それらの理解は、合成音声の品質に大きく影響する「バズ感」や有声/無声判定誤りに伴う耳障りな劣化を体系的に制御することにもつながる。

## 2. Violet noise と変種

velvet noise は、自然な残響を付与するための方法として提案された信号であり、ランダムに 1 と -1 の値をとるパルスを、離散時間軸上に疎に配置して作られる。標本化周波数が 44,100 Hz の場合、一秒間のパルス数が 2,000 個を越えると、正規乱数から作られる白色雑音よりも滑らかに感じられると報告されている [5]。この velvet noise に関連する幾つかの変種も作られ、生成法のパラメタと知覚との関係が調べられている [7]。

最初に提案された velvet noise を (OVN: Original Velvet Noise) と記す。OVN の  $m$  番目の信号の位置を表す離散時刻  $k_{\text{ovn}}(m)$  は、次式で表される。

$$k_{\text{ovn}}(m) = \lceil mT_d + r_1(m)(T_d - 1) \rceil, \quad (1)$$

ここで  $\lceil \cdot \rceil$  は、値を最も近い整数とする丸め関数を表し、 $T_d$  は、パルスの平均時間間隔を表す。また、 $r_1(m)$  は、0 と 1 の間に分布する一様分布の系列を表す。時刻  $n$  におけるパルスの値  $s_{\text{ovn}}(n)$  は、次式で求められる。

$$s_{\text{ovn}}(n) = \begin{cases} 2\lceil r_2(m) \rceil - 1 & n = k_{\text{ovn}}(m) \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

ここで  $r_2(m)$  は、0 と 1 の間に分布する一様分布のもう一つの系列を表す。

### 2.1 Velvet noise を利用した allpass filter の設計

velvet noise を直交行列を用いて離散 Fourier 変換すると、離散周波数軸上の複素数の実部と虚部の前半は、それぞれ、独立な正規乱数とみなすことができる [8]。したがって、時不変の帯域通過フィルタを通過した velvet noise の実部と虚部は、それぞれが帯域通過フィルタの利得により整形された雑音となる。これを応用すると、周波数軸上で velvet noise と同様に位相特性を局所的に操作する関数を配置することで、振幅特性が平坦で、位相特性が（滑らかに）ランダムに変化する信号を作ることができる。

このように、振幅周波数特性が平坦で位相特性が変化する信号をインパルス応答とするフィルタを all-pass filter と呼ぶ [9]。しかし、極と零の対を用いる方法 [9] や、群遅延を直接操作する方法 [10] では、最近提案された対数領域パルスモデル [11] と同様に、インパルス応答が目的とする範囲外に拡がるという問題があった。前の段落で説明した方法で設計した信号は、インパルス応答の局在性と平坦な周波数特性を両立させることができる。周波数領域で設計する velvet noise の変種なので、以下ではこの信号を周波数領域 velvet noise (FVN: Frequency domain Velvet Noise) と呼ぶことにする。

#### 2.1.1 余弦級数を用いた設計: FVN

まず、Nuttall の方法 [12] を用いて設計される余弦級数  $w_p(k, B_k)$  を、離散周波数  $k$  の関数として用意する。

$$w_p(k, B_k) = \sum_{m=0}^M a(m) \cos\left(\frac{\pi km}{B_k}\right), \quad (3)$$

ここで  $B_k$  は離散周波数軸上での  $w_p(k, B_k)$  の定義域の 1/2 を表し、 $M$  は余弦級数の次数を表す。この定義域 ( $-B_k \leq k \leq B_k$ ) の外では、 $w_p(k, B_k) = 0$  とする。なお、 $B_w = B_k/M/2$  を、設定帯域幅と呼ぶことにする。

この余弦級数を位相を操作する単位として使い、離散周波数軸上での余弦級数の中心の位置  $k_{\text{fvn}}(m)$  を、velvet noise と同様の手順で配置する。なお、添字  $\text{fvn}$  は、FVN に関するものであることを表す。

$$k_{\text{fvn}}(m) = \lceil mF_d + r_1(m)(F_d - 1) \rceil, \quad (4)$$

ここで  $F_d$  は、周波数間隔の平均を表す。また、 $r_1(m)$  は、(0, 1) に分布する一様乱数の系列を表す。なお、離散 Fourier 変換を用いるため、 $0 < k_{\text{fvn}}(m) < K(f_s/2)$  とする。ここで  $K(\cdot)$  は、周波数を離散周波数に変換する離散値をとる関数である。

$\mathbb{K}$  を、このようにして配置した  $k_{\text{fvn}}(m)$  を要素とする集合とする。実信号の Fourier 変換の位相の性質である奇対称を保証するため、以下のようにして FVN の位相を設定する。

$$\varphi_{\text{fvn}}(k) = \sum_{k_c \in \mathbb{K}} \varphi_{\max}(w_p(k - k_c, B_k) - w_p(k + k_c, B_k)). \quad (5)$$

この位相を有する複素指数関数を離散逆 Fourier 変換することにより、目的とする FVN 信号  $h_{\text{fvn}}(n)$  を得る。

$$h_{\text{fvn}}(n) = \frac{1}{K} \sum_{k=0}^{K-1} \exp\left(\frac{2kn\pi j}{KN} + j\varphi_{\text{fvn}}(k)\right), \quad (6)$$

$N$  は離散 Fourier 変換の周期を表し、虚数単位を  $j$  とした。

一連の数値実験に基づいて、余弦級数として Nuttall による 4 項のもの (MATLAB の関数 `nuttallwin`) を用いることとした。この場合、 $\varphi_{\max} = \pi/2.5$ 、 $F_d = B_w/5$  とすることで、 $B_w$  のみの調整により、目的とする持続時間を有するインパルス応答を求めることができる。

#### 2.1.2 極と零点による all-pass filter を用いた設計: CFVN

極と零の対を用いた all-pass filter は、応答が無限に続くものの、設計に必要なパラメタが極だけなので、使いやすい [9]。ここでは、この all-pass filter を用いて、もう一つの FVN を構成する。余弦級数を用いたものと異なり、こうして構成された応答は因果律を満たすため、CFVN (Causal FVN) と呼ぶことにする。

一次の all-pass filter  $H(z)$  は、 $z$  変換を用いて次式で表される。

(注1)：なお、この研究を行った DeepMind をかかえる Alphabet は、同時に、真剣にこの疑問に取り組んでいるグループを傘下の Google に有している [4]。

$$H(z) = \frac{z_c^* - z^{-1}}{1 - z_c z^{-1}}, \quad (7)$$

ここで  $z_c$  は、極を表す（一般には）複素数であり、 $z_c^*$  は、その共役複素数を表す。極が実数の場合は、そのまま使うことができるが、複素数の場合には、共役複素数である  $z_c^*$  を極とするものと組み合わせて、2次の IIR フィルタとして用いる。なお  $z_c$  は、極の周波数  $f_c$  と帯域幅  $b_w$  を用いて、次のように表される。

$$z_c = \exp(-b_w \pi + 2\pi f_c j). \quad (8)$$

この周波数  $f_c$  に対応する離散周波数  $k_{f_{vn}}(m)$  は、FVN になって式 (4) に基づいて設定することができる。また、 $b_w$  は、必要とする等価矩形長に応じて設定することになる。

ただし、極と零を配置する方法では、velvet noise 構成の次のステップである極性のランダム化を行うことができない。また、極の個数  $K$  に応じて  $2\pi K$  だけ位相が遅れるため、応答波形の重心は  $2K$  サンプル後の離散時刻になる。これらの影響は、 $f_c$  を配置する各帯域の幅をランダムに変化させることで軽減することができる。

### 2.1.3 周波数毎の等価矩形信号長の設定

ここまでで、設定帯域幅あるいは極の帯域幅を操作することにより、等価矩形信号長を制御できることを見てきた。これらの帯域幅を周波数の関数として変化させることにより、等価矩形信号長を周波数毎に設定することができる。ここでは FVN を例として説明する。CFVN も、同じ手続きで設定が可能である。

設計の目的である周波数  $f$  の関数としての等価矩形信号長を  $g(f)$  とする。ここでは、 $g(f) > 0$  の条件を満たす正值関数を対象とする。FVN の位相  $\varphi_{f_{vn}}$  は、定数の  $B_w$  に対して生成されている。この際に用いられる周波数を  $\nu$  とする。こうして生成された位相  $\varphi_{f_{vn}}(\nu)$  を、周波数を  $a$  倍することで圧縮して  $\varphi_{f_{vn}}(a\nu)$  とすると見かけ上の設定帯域幅が  $B_w/a$  となり、等価矩形信号長が  $a$  倍になる。言い替えると、 $B_w$  が定数になる周波数  $\nu$  を目的とする周波数の関数としたときの導関数は、 $g(f)$  に比例する。したがって、 $g(f)$  を積分して定数倍することによって目的とする周波数  $f$  の関数として  $\nu(f)$  を求めることができる。

$$\nu(f) = \alpha \int_0^f g(u) du, \quad (9)$$

ここで校正用の係数  $\alpha$  は、拘束条件  $\nu(fs/2) = fs/2$  から求められる。

FVN の位相の生成に用いる設定帯域幅  $B_{worg}$  は、 $g(f)$  が最大となる周波数  $f_{max}$  における  $\nu(f)$  の導関数とその周波数における設定帯域幅  $B_{wtgt}(\nu(f_{max}))$  から、次式を用いて求められる。

$$B_{worg} = B_{wtgt}(\nu(f_{max})) \left. \frac{d\nu(f)}{df} \right|_{f=f_{max}}, \quad (10)$$

こうして求められた FVN の位相  $\varphi_{f_{vn}}(\nu)$  を、目的とする周波数  $f$  から FVN 生成用の周波数軸への変換関数  $\nu(f)$  を用いることで、周波数領域での等価矩形信号長が設計された FVN の変種の位相  $\varphi_{mod}(f)$  を求めることができる。具体的には、次式となる。

$$\varphi_{mod}(f) = \varphi_{f_{vn}}(\nu(f)). \quad (11)$$

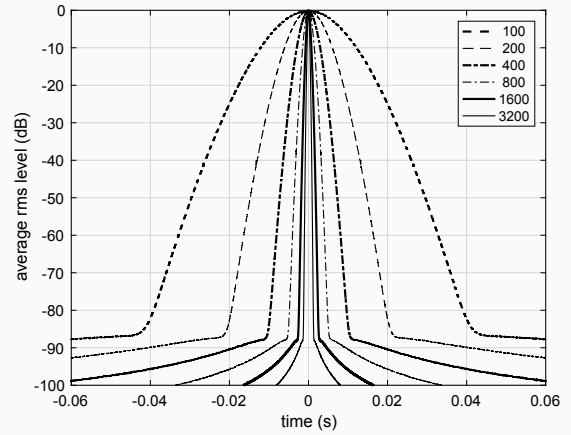


図 1 Average RMS (root mean squared) values of FVN shapes.

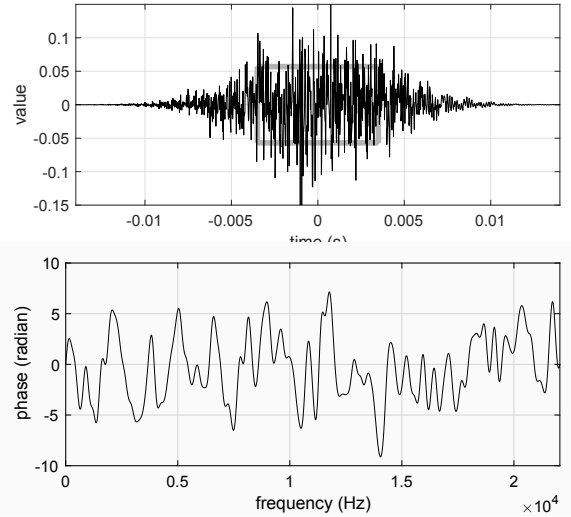


図 2 Example of the generated FVN waveform with  $B_w = 200$  (upper plot) and the corresponding phase (lower plot).

ここまでの定式化では、 $g(f) > 0$  のみを条件としてきた。ただし  $g(f)$  に不連続や急激な変化があると、変換された  $\varphi_{mod}(f)$  にも急速に変化する成分が含まれ、求められた all-pass filter の応答が時間方向に広がる。実用上は、 $g(f)$  には滑らかな関数を用いることが適切である。

## 2.2 数値例

ここでは FVN と CFVN について、具体例を幾つか示す。まず、FVN について、設定帯域幅  $B_w$  による変化を示す。

### 2.2.1 設定帯域幅が定数の場合：FVN

図 1 に、凡例に示す設定帯域幅のそれぞれについて 5,000 個の FVN を発生させた結果をまとめたものを示す。縦軸は波形の自乗平均値 (RMS) を表す。これらの数値実験の結果を利用することにより、目標とする FVN の等価矩形長  $T_{ERL}$  (単位: s) から、次式を用いて生成に用いる設定帯域幅  $B_w$  (単位: Hz) を定めることができる。

$$B_w = \frac{1.4}{T_{ERL}}. \quad (12)$$

図 2 に、 $B_w = 200$  Hz として生成される位相特性  $\varphi_{f_{vn}}(k)$  と波形の例を示す。この位相特性を用いることにより、図 2 の下に示される FVN が生成される。なお、図中には波形とそれに対応する等価矩形長で同じエネルギーを有する矩形を示す。

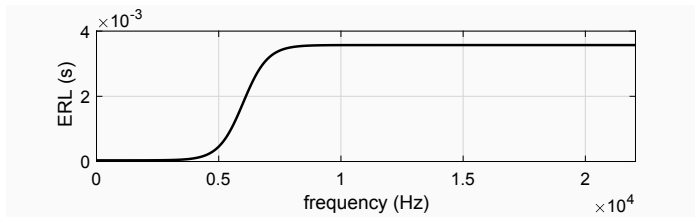


図 3 Designed ERL frequency characteristics.

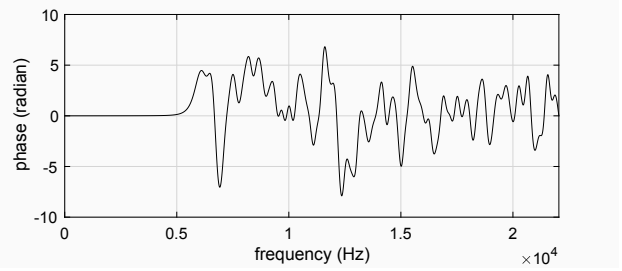


図 5 Example of a designed phase.

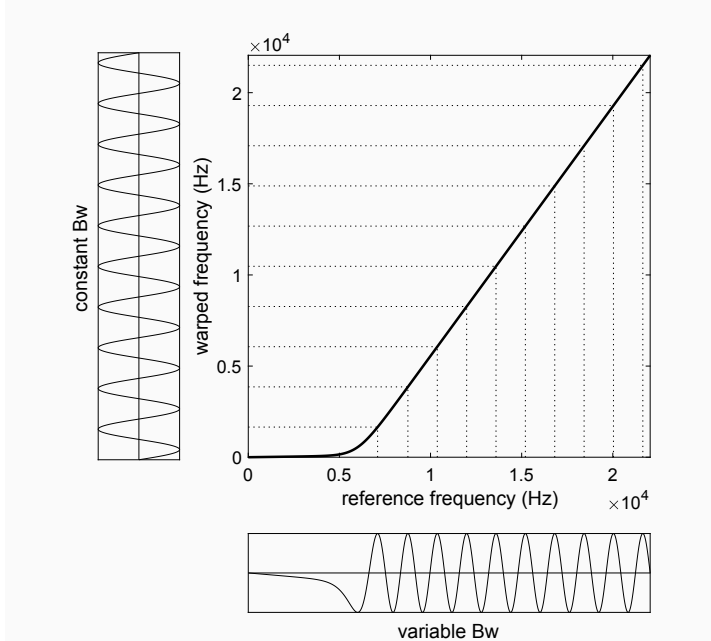


図 4 Designed frequency axis mapping between original and the warped one. A generated phase function  $\varphi_{\text{FVN}}(\nu)$  shown on the left using a constant  $B_w$  is converted to the modified phase function  $\varphi_{\text{mod}}(f)$  shown on the bottom, which corresponds to the frequency dependent  $B_w$ .

### 2.2.2 設定帯域幅が周波数の関数の場合：FVN

ここでは等価矩形長が図 3 に示す周波数特性を有する場合を例として提案した方法の動作を説明する。この特性は、変曲点の周波数が 6,000 Hz、1 Hz あたりの傾斜が 1/500 である sigmoid 関数を重みとして、高い周波数端で  $B_{\text{wtgt}}$  が 200 Hz、低い周波数端で  $B_{\text{wtgt}}$  が 20,000 Hz となるようにした。

図 4 に、図 3 に設定した等価矩形長に基づいて求めた元の周波数から変形した周波数への変換関数を示す。ここでは、併せて、定数の  $B_w$  を用いて生成した位相特性  $\varphi_{\text{FVN}}(\nu)$  の例を変換関数の左側に、変換関数を用いて求められた周波数依存性を持つ位相特性  $\varphi_{\text{mod}}(f)$  を変換関数の下側に配置した。また、変換関数の動作を見易くするために、位相特性のピークの対応関係を点線で示した。

この変換関数を用いて求められた位相特性の例を図 5 に示す。低い周波数領域で等価矩形長が小さくなるように設計しているため、その領域での位相特性の変化がゆっくりとしていることが分かる。

図 6 に、周波数特性を有する等価矩形長を用いて生成された FVN と、定数の等価矩形長を用いて生成された FVN の例を示す。この例では、定数の等価矩形長として、周波数特性を有する等価矩形長の Nyquist 周波数での値を用いた。原点から離れた位置では、原点からの距離に応じて両者の波形の振幅が同じような傾向で減少していることが分かる。

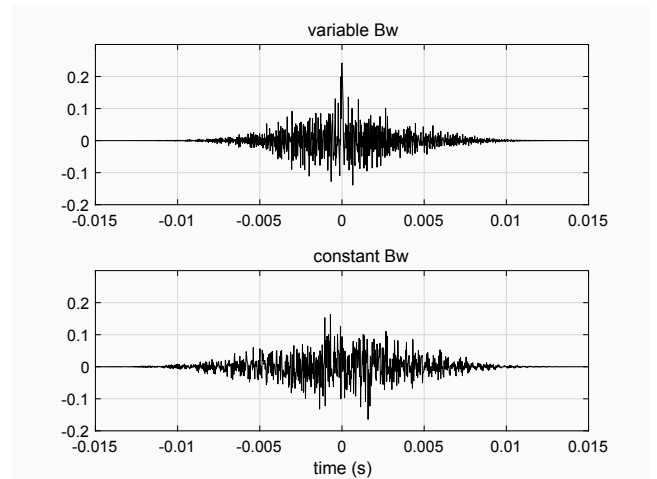


図 6 Example of the generated FVN. The upper plot shows the FVN with the frequency shape design. The lower plot shows the FVN with a constant  $B_w$ .

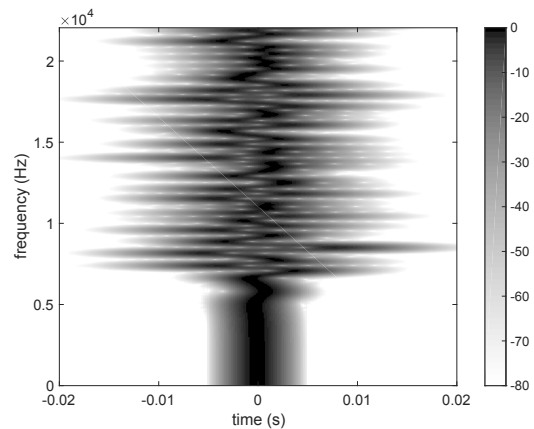


図 7 Spectrogram of the example FVN using 10 ms Nuttall window.

生成された信号を定性的に理解するためにスペクトログラムを求めた例を示す。図 7 に、10 ms の Nuttall による 4 項の余弦級数窓を用いて 0.1 ms 毎に求めたスペクトログラムを示す。この表示では周波数分解能が高いため、設定した位相特性に対応する群遅延を周波数の関数として把握することができる。

音声合成への応用では、周波数帯域ごとに信号の持続時間（あるいは等価矩形長）を指定したいことがある。標準化周波数を 16 kHz とした統計的音声合成システムの例では、[0-1], [1-2], [2-4], [4-6], [6-8] kHz の 5 個の帯域に分割している [13]。

図 8 に、上記の帯域ごとに等価矩形長を設定して生成した信号のスペクトログラムを示す。帯域の境界周波数において、応

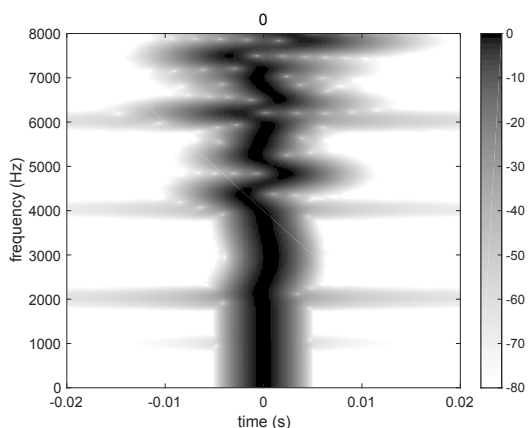


図 8 Spectrogram of the band-wise ERL defined FVN example using 10 ms Nuttall window

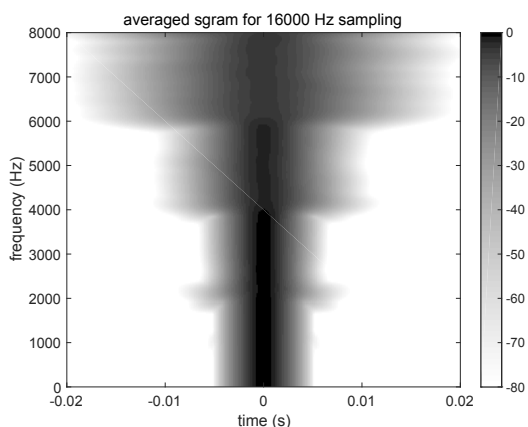


図 9 Averaged spectrogram of the band-wise ERL defined FVN using 10 ms Nuttall window

答が時間方向に広がっていることが分かる。

図 9 は、底辺が 400 Hz の raised cosine を使って  $g(f)$  を平滑化した場合の結果を示す。ここでは、結果を解釈しやすくするために、スペクトログラムを 1,000 回加算平均している。境界周波数での広がり、平滑化によってほぼ消失している。

### 2.2.3 帯域幅が同一の場合：CFVN

CFVN については、帯域幅が同一の all-pass filter を用いた構成についてのみ、数値例を紹介する。標本化周波数は、44,100 Hz とする。図 10 に、44 個の 2 次 の all-pass filter を配置した例を示す。極を配置する各帯域の幅を、0.3 と 1 の間に一様分布する乱数を用いて設定した。極の帯域幅は凡例に示すように、200, 400, 800, 1600, 3200 Hz とした。図中の○印は、時刻 0 での応答の値を示す。平均振幅のピークは、標本化周期の 2K 倍に相当する 2 ms 付近にある。ピークから離れる際の傾斜の絶対値は、帯域幅に比例している。これらの性質は、目的とする応答を設計する際に利用できる。

図 11 に帯域幅を 800 Hz とした場合に生成された信号と、対応する群遅延の例を示す。

## 3. OVN の応用

OVN は、パルス数が多い場合には、正規分布する乱数で発生させた白色雑音と同じような雑音と感じられる。両者を注意して比較すると、OVN の方が変動が少なく感じられ滑らかに

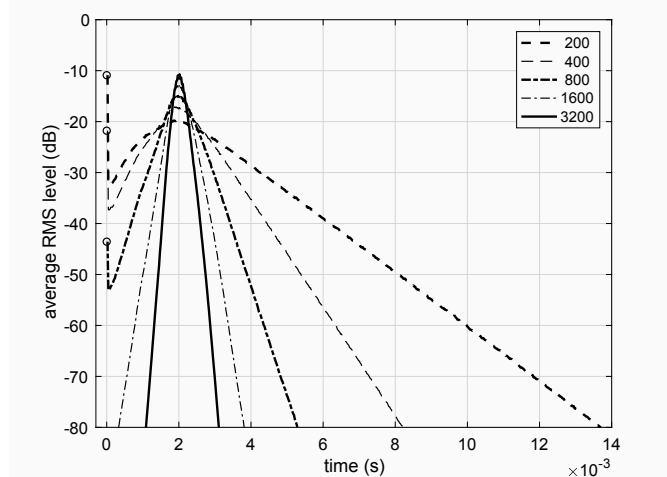


図 10 Average RMS (root mean squared) values of CFVN shapes. Legend shows bandwidths.

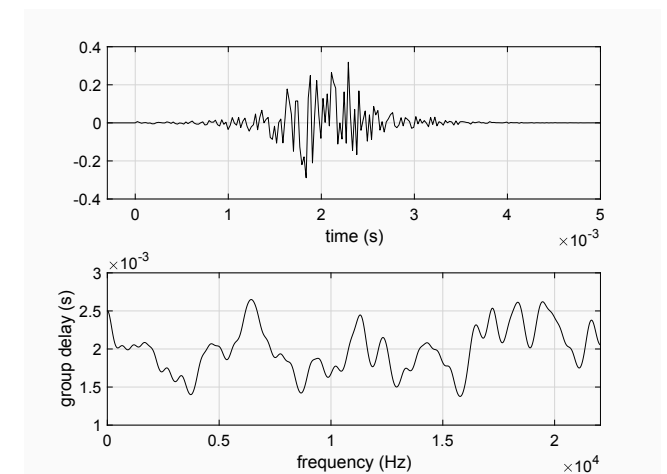


図 11 Examples of the impulse response and group-delay.

聴こえる。それでも、さらに注意すると、局所的に音色が変化するパターンを感じることもある。パルス数を少なくしていくと、そのような局所的な音色は聴こえなくなり、数えることはできないものの、ランダムに刺激される感じが生ずる。

ある長さで切り出したガウス雑音を繰り返すことにより、ピッチと、ビート (motorboating) と、繰返しが知覚される境界が 19 Hz, 4 Hz, 1 Hz にあると報告されている [14]。また、この繰返しの知覚は 0.5 Hz 以下になると、急速に困難になると報告されているこれは、聴覚における (波形に近い) 低レベルでの記憶バッファの長さが 1 秒程度であることを意味していると解釈して良いのであろうか？

自分を被験者として実験してみると、雑音の揺らぎで生ずる特有の音色のパターンに注目して、(意識的に／無意識に) 数えようとしていることが自覚される。OVN を用いて、系統的に手がかりを少なくして行った場合、ビートと繰返しの知覚には、同様な周波数に同様な境界はあるのであろうか？どのような結果になるにせよ、聴覚における低次の情報表現と処理機構を理解するための有用な手がかりが得られると考えられる。

## 4. FVN、CFVN の応用

聴覚は、同じ統計的パラメタから生成される二つの独立な正規雑音や OVN を区別できない。このことは、聴覚は位相の情

報の全てを利用している訳ではないことを示唆する。しかし、聴覚は、条件が整えば各調波の相対位相が異なる調波複合音を区別することができる [15,16]。我々の聴覚がどのような条件でどのような位相の違いを知覚しているか、それは何故なのかについて、見通しの良い理解はまだ得られていない。FVN は、時不変の all-pass filter として、信号の位相構造を局所的に少ない副作用で操作する手段を提供する。また、固定した FVN と毎回ランダムに生成される FVN の位相を線形補間することにより、ランダム信号から周期信号にいたる刺激音を、シームレスに合成することができる。これらの性質から、FVN はピッチ知覚の研究に広く用いられている IRN(Iterated Rippled Noise) [17] を補完する信号として利用できる可能性がある。

基底膜を伝播する信号の遅延を補償するチャープ信号を用いることにより、聴性脳幹反応が最大化される [18] にも関わらず、知覚される信号の凝集度 (compactness) は、単発のクリック (パルス) が最大であり、遅延を補償するチャープ信号を時間反転させた信号の方が元の信号よりも高いとの報告がある [19]。CFVN は、時間軸を反転させたものと対して用いることにより、この問題を含め、聴覚における信号処理と知覚の関係を、より深く調べるための有用なツールとなる可能性がある。

音声合成において問題となる「バズ感」の解消への応用を想定して、群遅延操作の知覚への影響を組織的に検討した報告がある [20]。FVN は時間方向の振幅分布の局在性とスペクトルの平坦性が両立している信号であり、これらについて、より統制された組織的検討を可能にする。ここでも、CFVN とその時間反転したものの対は、有用なツールとなる可能性がある。

簡単な予備検討では、一個の FVN を時不変のフィルタとして STRAIGHT による合成音声処理するだけで「バズ感」を大きく軽減できる可能性が示されている。また、駆動音源として用いる場合、帯域ごとの等価矩形長を設定したランダムな FVN を利用する通常の方法に加え、一個の固定した FVN と毎回ランダムに生成される FVN を周波数特性のある重みを用いて混合する方法も、実用的な方法として利用できそうである。

## 5. おわりに

ここでは、ランダムに 1 と -1 の値をとるパルスを、離散時間軸上に疎に配置して作られる velvet noise OVN と、周波数軸上に配置して作られる変種 FVN、因果律を満たす CFVN を提案した。これらの信号は、聴覚における処理や情報表現を研究するためのツールとして有用である可能性があることに加え、音声合成の分野でも興味深い応用の可能性がある。今後は、これらの可能性を追求して行く予定である。なお、関連する MATLAB による実装を、筆頭著者のサイトで公開している。

### 謝辞

本報告とその発表は、科研費挑戦的萌芽研究 16K12464 と、基盤 (B)15H02726 の支援を受けている。

文 献

- [1] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “WaveNet: A generative model for raw audio,” arXiv preprint arXiv:1609.03499, pp.1–15, 2016.
- [2] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, “Neural Discrete Representation Learning,” ArXiv e-prints, Nov. 2017.
- [3] A. van den Oord, Y. Li, I. Babuschkin, K. Simonyan, O. Vinyals, K. Kavukcuoglu, G. van den Driessche, E. Lockhart, L.C. Cobo, F. Stimberg, N. Casagrande, D. Grewe, S. Noury, S. Dieleman, E. Elsen, N. Kalchbrenner, H. Zen, A. Graves, H. King, T. Walters, D. Belov, and D. Hassabis,

- “Parallel WaveNet: Fast High-Fidelity Speech Synthesis,” ArXiv e-prints, Nov. 2017.
- [4] R.F. Lyon, Human and Machine Hearing: Extracting Meaning from Sound, Cambridge University Press, 2017.
- [5] H. Järveläinen and M. Karjalainen, “Reverberation modeling using velvet noise,” AES 30th International Conference, Saariselkä, Finland Audio Engineering Society., pp.15–17 2007.
- [6] H. Kawahara, “Application of the velvet noise and its variant for synthetic speech and singing,” IPSJ SIGMUS, vol.2018, no.118, , feb 2018.
- [7] V. Välimäki, H.M. Lehtonen, and M. Takanen, “A perceptual study on velvet noise and its variants at different pulse densities,” IEEE Transactions on Audio, Speech, and Language Processing, vol.21, no.7, pp.1481–1488, July 2013.
- [8] R.H. Lyon, “Statistics of combined sine waves,” The Journal of the Acoustical Society of America, vol.48, no.1B, pp.145–149, 1970.
- [9] A.V. Oppenheim and R.W. Schaffer, Discrete-time signal processing: Pearson new International Edition, Pearson Higher Ed., 2013.
- [10] H. Kawahara, I. Masuda-Katsuse, and A. deCheveigne, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction,” Speech Communication, vol.27, no.3–4, pp.187–207, 1999.
- [11] G. Degottex, P. Lanchantin, and M. Gales, “A log domain pulse model for parametric speech synthesis,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.26, no.1, pp.57–70, Jan. 2018.
- [12] A.H. Nuttall, “Some windows with very good sidelobe behavior,” IEEE Trans. Audio Speech and Signal Processing, vol.29, no.1, pp.84–91, 1981.
- [13] H. Zen, T. Toda, M. Nakamura, and K. Tokuda, “Details of the Nitech HMM-based speech synthesis system for the Blizzard Challenge 2005,” IEICE transactions on information and systems, vol.90, no.1, pp.325–333, 2007.
- [14] N. Guttman and B. Julesz, “Lower limits of auditory periodicity analysis,” The Journal of the Acoustical Society of America, vol.35, no.4, pp.610–610, 1963.
- [15] R. Plomp and H. Steeneken, “Effect of phase on the timbre of complex tones,” The Journal of the Acoustical Society of America, vol.46, no.2B, pp.409–421, 1969.
- [16] R.D. Patterson, “A pulse ribbon model of monaural phase perception,” The Journal of the Acoustical Society of America, vol.82, no.5, pp.1560–1586, 1987.
- [17] W.A. Yost, R. Patterson, and S. Sheft, “A time domain description for the pitch strength of iterated rippled noise,” The Journal of the Acoustical Society of America, vol.99, no.2, pp.1066–1078, 1996.
- [18] T. Dau, O. Wegner, V. Mellert, and B. Kollmeier, “Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion,” The Journal of the Acoustical Society of America, vol.107, no.3, pp.1530–1540, 2000.
- [19] S. Uppenkamp, S. Fobel, and R.D. Patterson, “The effects of temporal asymmetry on the detection and perception of short chirps,” Hearing research, vol.158, no.1-2, pp.71–83, 2001.
- [20] H. Banno, K. Takeda, and F. Itakura, “The effect of group delay spectrum on timbre,” Acoustical Science and Technology, vol.23, no.1, pp.1–9, 2002.