2019 年 12 月 14 日，日本音響学会聴覚研究会資料，Vol. 49，No. 8，H-2019-108
Proc. Auditory Res. Meeting, The Acoustical Society of Japan, Vol. 49, No. 8，H-2019-108，December 14, 2019

Strange aftereffect caused by periodic allocation of a frequency domain variant of velvet noise

Hideki KAWAHARA†

† Wakayama University
930 Sakaedani, Wakayama, Wakayama, 640-8510 Japan

Abstract We report a strange aftereffect. We found the effect accidentally while testing a new test signal for acoustic measurements. The test signal uses frequency domain variants of velvet noise (FVN) as the building block. One instance of FVN is an impulse response of an all-pass filter. The gain of the filter does not depend on frequency. A procedure inspired by the original velvet noise allocates a phase modification function on the frequency axis using two sets of random number sequences. Exposure to periodic signals made from a frozen velvet noise caused the strange aftereffect. The exposure made everyday sounds sounded like altered by using the "flanger" effector. We discuss the implication of this aftereffect in hearing research. We also discuss a possible application to altered auditory feedback research.

Key words periodic signal, all-pass filter, group delay, aftereffect, timbre perception, auditory evoked potential

1. Introduction

We found a strange aftereffect accidentally while testing a new test signal for acoustic measurements [1]. The test signal consists of a unit signal which has a randomly fluctuating group delay, and the unit periodically repeats in a regular interval on the time axis. While hearing the sound about a minute, everyday sounds sounded like modified by using "flanger," for example, for several seconds. It reminded us of contradictory reports on "compactness" perception [2, 3]. This report introduces how to make the signal and discusses how to test the phenomenon.

An acoustic impulse propagates on the basilar membrane of the cochlea and stimulates auditory nerve fibers connected to the inner hair cells at different timing because of the propagation delay. Consequently, the auditory evoked potential responds to this propagation delay and has temporally spread shape. A chirp signal which compensates for this propagation delay stimulates auditory nerve fibers at the same time. Consequently, the auditory evoked potential shows a sharper response [2]. However, the perceived "compactness" of the chirp with compensating propagation delay is weaker than the chirp with enhancing propagation delay. Furthermore, an acoustic impulse without any modification of the propagation delay provides the highest "compactness." [3].

We expect that test signals made from FVN can provide clues to solve this apparent contradiction. This article first introduces the FVN definition and its specific features. Second, we describe the strange aftereffect and how we found them. Based on these, we speculate a plausible mechanism underlying the aftereffect. Finally, we propose several FVN-based test signals and discuss how to use them to investigate detailed temporal information processing in our auditory system.

2. Velvet noise and FVN

The original velvet noise is a randomly allocated binary-valued (± 1) pulses [4, 5]. This sparse signal sounds like noise and provides impression smoother than Gaussian white noise. This impression is the reason for the naming "velvet noise."

The FVN uses a similar procedure for manipulating phase in the frequency domain. Because of this similarity of the procedure, we name it as a "frequency domain variant of velvet noise" (FVN) [6].

2.1 FVN: generation procedure

FVN generation uses two random number sequences $r_1[n]$ and $r_2[m]$ $(n, m \in \mathbb{Z})$, which are sampled from the uniform distribution in $(0, 1)$. The first random sequence $r_1[n]$ defines the position of the phase manipulation function $w_p(f)$ in each subdivided frequency segment, where n represents the index of the segment. The segment length in the frequency domain is F_d, and the lower boundary of the segment spans from 0 to $f_s/2$, where $f_s/2$ is the sampling frequency of the discrete-time signal. The following equation defines the n-th location $f_c[n]$ of the center of the phase manipulation function $w_p(f)$.

$$f_c[n] = (n - 1 + r_1[n])F_d. \quad (1)$$

The second random sequence $r_2[n]$ defines the coefficient of phase manipulation $c_\varphi[n]$ of the n-th segment using the

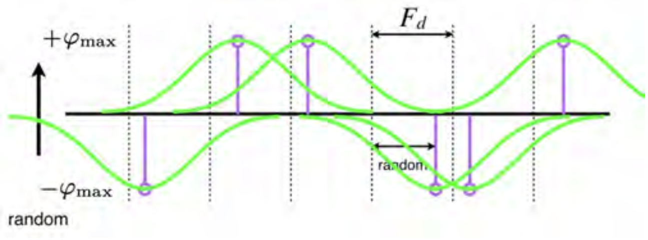

Fig. 1 Schematic diagram of the allocation of phase manipulation function $w_p(f - f_c[n])$ on the frequency axis. Each violet vertical line represents the position of $f_c[n]$.

maximum magnitude of the phase manipulation $\varphi_{\max}$. The following equation defines the coefficient of the n-th phase manipulation $c_\varphi[n]$.

$$c_\varphi[n] = (2\|r_2[n]\| - 1)\varphi_{\max}, \tag{2}$$

where $\|a\|$ represents the nearest integer to a real number a.

Figure 1 illustrates how to allocate phase manipulation functions in the frequency domain. The sum of manipulation yields the positive side element $\varphi_+(f)$ of the phase of an FVN.

The coefficients $\{c_\varphi[n]\}_1^N$ and the positions $\{f_c[n]\}_1^N$ define the phase $\varphi(f)$ of an FVN, where N represents the number of frequency segments.

$$\exp(j\varphi(f)) = \exp(j\varphi_+(f))\exp(-j\varphi_-(f)), \tag{3}$$

$$\text{where} \quad \varphi_+(f) = \sum_{n=1}^{N} c_\varphi[n] w_p(f - f_c[n]) \tag{4}$$

$$\varphi_-(f) = \sum_{n=1}^{N} c_\varphi[n] w_p(f + f_c[n]) = \varphi_+(-f), \tag{5}$$

where the unit phase manipulation function $w_p(f)$ is an even function. (Note that the length of $w_p(f)$ is generally larger than F_d. Therefore, $w_p(f)$ spans several subdivided segments.) The inverse Fourier transform of the discretized version of the complex-valued odd function $\exp(j\varphi(f))$ yields a real-valued discrete signal $h_{\text{fvn}}[n]$, where $j = \sqrt{1}$ is the imaginary unit. This $h_{\text{fvn}}[n]$ is the impulse response of an all-pass filter because the absolute value of $\exp(j\varphi(f))$ is independent of the frequency.

2. 2 Numerical optimization of parameters

We define the unit phase manipulation function $w_p(f)$ by using the six-term cosine series [7] proposed for anti-aliasing glottal source models made. The following equation defines $w_p(f), (-3c_{\text{mag}}F_d \leqq f \leqq 3c_{\text{mag}}F_d)$.

$$w_p(f) = \sum_{m=0}^{5} a_m \cos\left(\frac{m\pi f}{3c_{\text{mag}}F_d}\right), \tag{6}$$

where c_{mag} is a stretching factor to control the relative spread of the phase manipulation in terms of F_d. We numerically optimized the coefficients $\{a_m\}_{m=0}^5$ based on the Nuttall's procedure [8] by an exhaustive search of 10-digit numbers.

$$\{a_m\}_{m=0}^5 = \{0.2624710164, 0.4265335164, 0.2250165621, 0.0726831633, 0.0125124215, 0.0007833203\}. \tag{7}$$

The Fourier transform of this function is highly localized. The maximum level of the side-lobes is -114 dB from the main lobe level, and their levels decay at the rate of -54 dB/octave [7]. This selection of $w_p(f)$ is the key to make FVN practical and useful because it localizes the resulted impulse response $h_{\text{fvn}}[n]$ without side effects [6]. Truncation of the number of digits up to 6 digits does not change this desirable behavior significantly.

The exhaustive simulation provided the best setting of the stretching factor c_{mag} and the maximum magnitude $\varphi_{\max}$ of phase manipulation. The root mean squared average value (RMS) of the resulted impulse responses made from different sets of random numbers approximates Gauss function well by setting $c_{\text{mag}} >= 2$ and $\varphi_{\max} = \pi/2$ [1]. We use $c_{\text{mag}} = 2$ afterward.

These results make FVN have only one design parameter, F_d, which defines the duration σ_T. The second-order moment of the impulse response $h_{\text{fvn}}[n]$ determines the duration σ_T. Therefore, for given σ_T, a set of simulations provides the following equation to find the relevant F_d [1].

$$F_d = \frac{1}{5\sigma_T}. \tag{8}$$

2. 3 Shaping the time-domain envelope (RMS)

We introduced a non-linear and smooth monotonic function $g(f)$, which warps the frequency axis to design FVNs with frequency-dependent duration $\sigma_T(f)$ [6]. In this article, we make the monotonic function $g(f)$ a function of the frequency derivative of the FVN's phase function $\varphi_+(f)$ to shape the time-domain envelope of the resulted FVN.

The following equation defines the modified phase function $\tilde{\varphi}(f)$.

$$\tilde{\varphi}(f) = \varphi_+(g(f)) - \varphi_+(-g(f)) \tag{9}$$

$$g(f) = \int_a^f \psi\left(\frac{d\varphi_+(\nu)}{d\nu}\right) d\nu + c_0, \tag{10}$$

where $\psi(\nu)$ is positive definite and the function $g(f)$ has to satisfy $(g(0) = 0, \text{and } g(f_s/2) = f_s/2)$. Inverse Fourier transform of $\exp(j\tilde{\varphi}(f))$ provides the temporal envelope-shaped impulse response $\tilde{h}_{\text{fvn}}[n]$. We provide a specific candidate of $\psi(\nu)$ in the appendix of [1].

3. FVN-based test signals

The impulse response of an FVN is a member of TSP and usable for impulse response measurement of acoustic systems. Let $H_{\text{fvn}}(f)$ represent the Fourier transform of the impulse response $h_{\text{fvn}}[n]$ of an FVN. Using the definition of FVN (Eq.(3)), $H_{\text{fvn}}(f)H^*_{\text{fvn}}(f) = 1$ holds, where X^* represents the complex conjugate of X.

High degrees of freedom in designing FVN provide useful test signals introduced below by systematic allocation of individual impulse responses of FVNs. First, we briefly introduce test signals we already proposed in [1]. Then, we introduce to design a test signal which can sound like a piece of music by making use of the envelope shaping and a multi-resolution framework.

3.1 Periodic allocation of impulse responses of FVN [1]

Allocating an impulse response of an FVN repeatedly on the time axis using a constant interval provides a periodic test signal. Convolution of the measured response of an acoustic system driven by the test signal with the time-reversed version of the impulse response (of the FVN) provides the impulse response of the system, which is repeatedly allocated on the time axis. Averaging these repeated responses with time alignment yields the impulse response of the system with reduced measurement noise.

Since this test signal is periodic, the ratios of instantaneous frequencies of harmonic components of the test signal and the recorded response signal provide the time-alignment function of the signals [1]. Using the six-term cosine series $w_p[n]$ for the envelope of the analytic signal provides accurate estimates of the instantaneous frequency of each harmonic component [9].

Allocating different impulse responses from a set of FVNs each time on the time axis using a constant interval provides a random test signal. Interpolation of each phase of this random test signal and corresponding periodic signal made from one impulse response of FVN provides a continuum from random to periodic test signals [6].

3.2 Designing orthogonal set of sequences [1]

Periodic allocation of different impulse responses of FVN using the binary coefficients $b_m^{(k)} \in \{-1, +1\}$ provides a set of test signals. Using a set of mutually orthogonal binary sequences for setting the coefficients $b_m^{(k)}$ enables impulse response measurements of multiple acoustic paths at once [1]. The index k represents the ID of the sequence. The subscript m represents the ID of the location in each sequence.

Assume that the target acoustic paths consist of N loudspeakers and one shared recording microphone. Driving each acoustic path using each test signal yields the signal, which is the sum of all responses. Convolution of the recorded signal and one of the FVNs (for example, k-th) yields the repeated impulse responses of the k-th acoustic path and sum of cross-talks from the different sequences. Averaging of each repeated impulse response using the corresponding coefficient $b_m^{(k)}$ completely cancels the cross-talks from the different sequences [1].

One difficulty of this measurement is the length of the test signal. For measuring N acoustic paths at the same time, the length of the test signal is 2^{N+1} times of the repetition period.

Figure 2 shows the distribution of the maximum cross-correlation between impulse responses of FVN and the selected 16 impulse responses to generate the orthogonal sequences. Figure 2 shows FVNs with a duration of 25 ms.

3.3 Summation of the orthogonal set of sequences [1]

Summation of the N test signals generated using the procedure in 3.2 provides the next type of the test signal. The test signal enables measurement of the effects of non-linear distortion, intermodulation distortion, temporal variation, and other deviations from ideal linear time-invariant (LTI) systems at once without introducing additional devices such as band elimination filters.

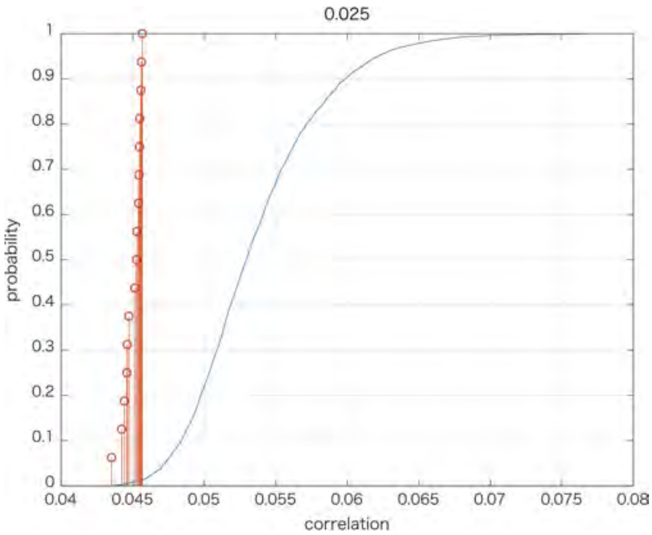

Fig. 2 Distribution of the maximum cross correlation between FVN impulse responses. The orange lines shows the first 16 FVN impulse responses with the least maximum cross correlation.

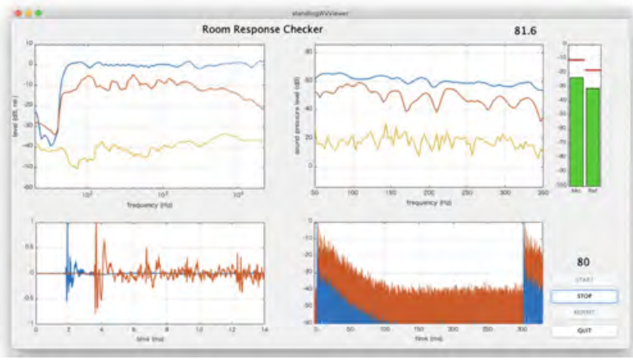

Fig. 3 Snapshot of the GUI of an acoustic measurement tool based on FVN.

If the target acoustic system is an LTI system, impulse responses calculated from using corresponding constituent FVN impulse responses are identical to each other. Therefore, RMS deviations from the average of the estimated individual impulse responses provide an estimate of the combined effects of deviations from LTI systems [1].

3.4 Implementation and tools

We implemented acoustic measurement procedures described in the previous sections using MATLAB and open-sourced [10]. We also prepared an interactive and realtime acoustic measurement tool based on the implemented functions.

Figure 3 shows a snapshot of the tool [1]. This tool simultaneously measures two acoustic impulse responses with error evaluation. The upper panels show the transfer gains of the whole frequency range (in the logarithmic frequency axis) and the lower frequency range (from 50 Hz to 350 Hz in the linear frequency axis). The lower panels show the impulse responses of the initial 45 ms and a whole 300 ms. Currently, these tools are redesigned using Auditory Toolbox and App Designer of MATLAB. The revised tool consists of simultaneous measurement of the effect of nonlinearity as a builtin function.

4. Acoustic measurement using FVN

In this section, we focus on the propagation measurement, in which we found the strange effect discussed in this article. An FVN is a TSP on a broader sense, and itself is applicable to measure acoustic impulse responses. The orthogonalization of periodic sequences using FVN enabled simultaneous measurement of multiple acoustic paths. Moreover, the sum of the sequences enabled the measurement of the nonlinearity effect. For those procedures and details, please refer to [1].

4.1 Propagation delay measurement and the strange effect

The integration of the difference between instantaneous frequencies of the input $f_{i\mathrm{IN}}(t)$ and the output $f_{i\mathrm{OUT}}(t)$ provides the deviation estimate of the propagation delay $\tau_d(t)$. It is reasonable to use a constant frequency test signal and let $f_{i\mathrm{IN}}(t) = f_c$.

$$\tau_d(t) = \int_0^t \left(\frac{f_{i\mathrm{OUT}}(\lambda)}{f_c} - 1 \right) d\lambda. \tag{11}$$

We used a test signal made from a unit FVN repeated at $1/f_c$ and used 10-th to 35-th harmonic components for calculating the instantaneous frequency of each component. We prepared a set of analytic signals centered around harmonic frequencies with the six-term cosine series for the envelope. The length of the envelope is 12-times of the fundamental period. The output frequency $f_{i\mathrm{OUT}}(t)$ is the average of the corresponding fundamental frequency of each harmonic component.

Figure 4 shows examples of propagation delay in various paths. The test signal uses a unit FVN with 25 ms duration (in terms of the second moment) and 221 samples for repeated allocation. The length of the test signal is 45 s, and the sampling frequency is 44100 Hz.

A strange aftereffect happened while doing this test. The tester (the author) was exposed to the test signal during each measurement. When the test signal ends, background noise, including sounds from a distant television program, sounded like altered by the “flanger” effector. This strange altered sound disappeared in a couple of minutes.

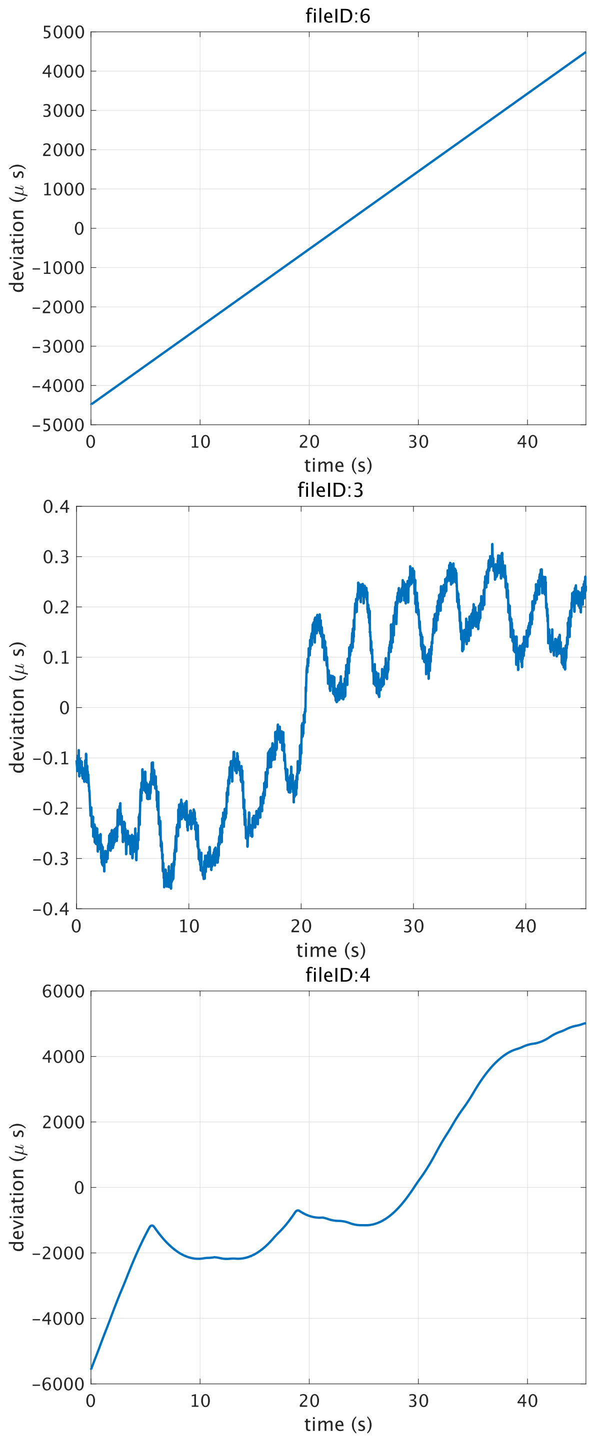

Fig. 4 Propagation time deviations. (Top panel) AD and DA converters use their internal sampling clocks. They are independent each other. (middle panel) The tester was breathing close to the propagation path, and (bottom panel) The loudspeaker uses Bluetooth connection.

5. Discussion

There are several possibilities of the mechanism that causes this strange aftereffect. First, we tried to quantify the strange effect we observed. However, all attempts were not successful. The effects caused by the exposure seem to disappear within several minutes and seem not to leave long-term damage. The test procedure and method for quantification depend on the mechanism, which causes the strange effect. The following is a tentative list.

Fast adaptation to firing misalignment

The finding by Uppenkamp [3] may suggest that the auditory periphery has a function to compensate for the propagation delay on the basilar membrane. The strange aftereffect may suggest that the compensation process keeps updating based on the input sounds with relatively fast adaptation.

Damaging internal firing alignment system

Too strong disturbance of the firing of the auditory nerve may damage the compensation process; in other words, the internal alignment process of firing alignment. The damaged process fails to re-align the firing time caused by everyday sounds, and it results in the strange feeling discussed before.

Strange sensation generator

A “flanger feeling generator” may exist. Exposure to the test signal may activate the sensation generator. While exposing

to the test signal, the generator will be over-activated and in a sensitive state.
We want to discuss these and possible tests.

The procedure of nonlinearity measurement is also applicable. It is to assess the nonlinearity in auditory feedback control of voice fo [11]. We first introduced the TAF (Transformed Auditory Feedback) experiment paradigm and measured impulse response from pitch perception to voice fundamental frequency control [12–14]. In the reference, we used M-sequence for modulating the fundamental frequency of the feedback sound. By replacing this M-sequence with a sum of FVN-generated orthogonal sequences, it enables quantitative evaluation of the effects of nonlinearity in the auditory to voice pitch control process.

The continuum between the random signal to the periodic signal made from FVN sequences [6] also is useful. The continuum is a counterpart of the IRN (Iterative Rippled Noise) used for pitch perception tests [15].

6. Conclusion

We introduced a strange aftereffect caused by the exposure to a periodic signal made from FVN (frequency domain variant of velvet noise). It made everyday sounds sounded like altered using the "flanger" effector. It may provide a new set of tests to investigate the perception of fine temporal structure in sounds. The FVN-based test signals also provide methods to evaluate nonlinearity in the auditory feedback control of vocalization. Tools based on FVNs are open-sourced in our GitHub repository.

Acknowledgement

The author wishes to thank Yutaka Kaneda, professor at Tokyo Denki University, for discussions on Swept-Sine and MLS signals. He also wishes to thank Minoru Tsuzaki, professor at Kyoto City University of Arts, for discussion on the role of aftereffects. KAKENHI (Grant in Aid for Scientific Research by JSPS) 16H01734, 15H03207, 18K00147, and 19K21618 supported this research.

Reference

[1] H. Kawahara, K.-I. Sakakibara, M. Mizumachi, H. Banno, M. Morise, and T. Irino, "Frequency domain variant of velvet noise and its application to acoustic measurements," ASPIPA ASC 2019, pp.1523–1532, Lanzhou, China, 2019.

[2] 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.

[3] 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.

[4] H. Järveläinen and M. Karjalainen, "Reverberation modeling using velvet noise," AES 30th International Conference, Saariselkä, FinlandAudio Engineering Society,, pp.15–17 2007.

[5] 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.

[6] H. Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, T. Tomoki, and T. Irino, "Frequency domain variants of velvet noise and their application to speech processing and synthesis," Proc. Interspeech 2018, pp.2027–2031, Hyderabad India, 2018.

[7] H. Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, T. Toda, and T. Irino, "A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis," Proc. Interspeech 2017, pp.1358–1362, Stocholm, Aug. 2017.

[8] 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.

[9] H. Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, and T. Toda, "Accurate estimation of f0 and aperiodicity based on periodicity detector residuals and deviations of phase derivatives," Proc. APSIPA ASC 2017, pp.1556–1564, 2017.

[10] H. Kawahara, "FVN: Frequency domain variants of Velvet Noise, a flexible building unit for speech synthesis, psychoacoustics and acoustic measurement.," (Last access: 2019-10-22). https://github.com/HidekiKawahara/FVN.

[11] J.A. Jones and K.G. Munhall, "Perceptual calibration of F0 production: Evidence from feedback perturbation," The Journal of the Acoustical Society of America, vol.108, no.3, pp.1246–1251, 2000.

[12] H. Kawahara, "Transformed auditory feedback: Effects of fundamental frequency perturbation," The Journal of the Acoustical Society of America, vol.94, no.3, pp.1883–1884, 1993.

[13] H. Kawahara, "Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies," Journal of the Acoustical Society of Japan (E), vol.15, no.3, pp.201–202, 1994.

[14] H. Kawahara, H. Kato, and J.C. Williams, "Effects of auditory feedback on f0 trajectory generation," Proc. ICSLP 96, pp.287–290, 1996.

[15] 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.