An interference-free representation of group delay for periodic signals

Hideki Kawahara*, Masanori Morise†, Ryuichi Nisimura* and Toshio Irino*
* Wakayama University, Wakayama, Japan
E-mail: {kawahara, nisimura, irino}@sys.wakayama-u.ac.jp
† Ritsumeikan University, Shiga, Japan
E-mail: morise@fc.ritsumei.ac.jp

***Abstract*—This article introduces a new group delay representation for periodic signals. The proposed method yields a group delay representation that is free from interferences due to repetitive excitation. Power spectrum-weighted averaged group delay using shifted copies of the weighted group delay separated by a half fundamental frequency is proven to have the desired property.**

I. INTRODUCTION

Instantaneous frequency [1] and group delay, which are defined as the temporal derivative and the frequency derivative of phase respectively, are better representations than phase itself, because they are physically meaningful and do not require unwrapping, which is a fragile operation [2]. However, abrupt changes and discontinuities, which are caused by interference between constituent components prevented these representations from potential applications. As the final piece of the authors' investigations for providing interference-free representations of power spectrum [3], [4] and instantaneous frequency [5], an interference-free representation of group delay is introduced. It is derived from the group delay representation, analogues to Flanagan's instantaneous frequency representation [6]. The interference-free group delay is defined as the power spectrum weighted average of the shifted pair of group delays 1/2 fundamental frequency apart.

II. CANCELLATION OF PERIODIC INTERFERENCE

This section starts from brief introductions to interference-free representations of the power spectrum and the instantaneous frequency of periodic signals followed by the derivation of the interference-free group delay representation.

A. Interference-free power spectrum

The simplest interference can be represented by taking two neighboring harmonic components of periodic signals. Then, it is general enough to assume one of the components, k-th component this time, is normalized and the $k+1$-th component have a different amplitude α and phase β.

$$x(t) = e^{jk\omega_0 t} + \alpha e^{j(k+1)\omega_0 t + j\beta}, \tag{1}$$

where ω_0 is the fundamental angular frequency and $j = \sqrt{-1}$. Note that $\omega_0 = 2\pi/T_0$, where T_0 is the fundamental period.

Assume that the effective pass-band of the frequency domain representation $W(\omega)$ of the windowing function $w(t)$ to cover up to two harmonic components. It also is general enough to assume $k = 0$. Then, the power spectrum of the windowed signal is

$$\begin{aligned} P(\omega, t) = & |W(\omega)|^2 + \alpha^2 |W(\omega - \omega_0)|^2 \\ & + 2\alpha W(\omega) W(\omega - \omega_0) \cos(\omega_0 t + \beta) \ , \end{aligned} \tag{2}$$

where the third term represents the temporally varying component, the periodic interference in the time domain.

Since it varies sinusoidally, adding a component having the opposite polarity cancels this interference. Equation 2 indicates that two power spectra separated by a half-pitch period have components with phase difference π, opposite polarity. The interference-free power spectrum (TANDEM spectrum) $P_T(\omega, t)$ is defined by the average of these two power spectra

$$P_T(\omega, t) = \frac{1}{2} \left[P\left(\omega, t - \frac{T_0}{4}\right) + P\left(\omega, t + \frac{T_0}{4}\right) \right] \ . \tag{3}$$

B. Flanagan's instantaneous frequency equation

Before investigating interference on the instantaneous frequency, it is worthwhile to revisit Flanagan's instantaneous frequency equation. It starts from the polar form of a complex signal $x(t)$

$$x(t) = a(t) e^{j\theta(t)}, \tag{4}$$

where $a(t)$ is the amplitude and $\theta(t)$ is the phase.

The instantaneous frequency $\omega_i(t)$, defined by the time derivative of phase $\theta(t)$, is also derived as the time derivative of the imaginary part of the log-converted signal $\log(x(t))$.

$$\log(x(t)) = \log(a(t)) + j\theta(t) \tag{5}$$

$$\begin{aligned} \omega_i(t) &= \frac{d\ \Im[\log(x(t))]}{dt} \\ &= \Im\left[\frac{1}{x(t)} \frac{dx(t)}{dt}\right] \\ &= \frac{\Re[x(t)]\Im\left[\frac{dx(t)}{dt}\right] - \Im[x(t)]\Re\left[\frac{dx(t)}{dt}\right]}{|x(t)|^2}, \end{aligned} \tag{6}$$

where $\Re[x]$ and $\Im[x]$ represent the real part and the imaginary part of a complex number x. The last representation is the Flanagan's equation. It does not consist of phase unwrapping, which is necessary when calculating the instantaneous frequency using its original definition.

This work is partly supported by Grant in Aid for Scientific Research of JSPA and advanced research project of Wakayama University Japan.

It is convenient to define two spectral representations $S(\omega,t)$ and $S_d(\omega,t)$ to implement Eq. 6.

$$S(\omega,t) = \int w(\tau - t)x(\tau)e^{-j\omega\tau}d\tau \quad (7)$$

$$S_d(\omega,t) = \frac{dS(\omega,t)}{dt} = \int \frac{dw(\tau - t)}{dt}x(\tau)e^{-j\omega\tau}d\tau, \quad (8)$$

where integration domain $(-\infty,\infty)$ is not explicitly written for visual simplicity (This convention is after Cohen [2].).

Substituting $S(\omega,t)$ and $S_d(\omega,t)$ into Eq. 6 yields the following simpler form:

$$\omega_i(\omega,t) = \frac{\Re[S(\omega,t)]\Im\left[S_d(\omega,t)\right] - \Im[S(\omega,t)]\Re\left[S_d(\omega,t)\right]}{P(\omega,t)}. \quad (9)$$

C. Interference-free instantaneous frequency

Equation 9 simplifies the averaged instantaneous frequency $\omega_{Ai}(\omega,t_1,t_2)$ because the denominator of Eq. 9 is the power spectrum $P(\omega,t)$ and is canceled by the weights used in Eq. 10, where t_1 and t_2 are two window locations.

$$\omega_{Ai}(\omega,t_1,t_2) = \frac{P(\omega,t_1)\omega_i(\omega,t_1) + P(\omega,t_2)\omega_i(\omega,t_2)}{P(\omega,t_1) + P(\omega,t_2)}, \quad (10)$$

Substituting Eq. 9 into Eq. 10 yields the simplified form of the numerator of Eq. 10, $R(\omega,t_1,t_2)$:

$$\begin{aligned} &R(\omega,t_1,t_2) \\ &\quad = \Re[S(\omega,t_1)]\Im\left[S_d(\omega,t_1)\right] - \Im[S(\omega,t_1)]\Re\left[S_d(\omega,t_1)\right] \\ &\quad + \Re[S(\omega,t_2)]\Im\left[S_d(\omega,t_2)\right] - \Im[S(\omega,t_2)]\Re\left[S_d(\omega,t_2)\right], \end{aligned} \quad (11)$$

The first term and the their term share the same form. They are represented by the following equation for the test signal (Eq. 1) input.

$$\begin{aligned} &\Re[S(\omega,t)]\Im[S_d(\omega,t)] \\ &= W(\omega)W_d(\omega) + \alpha\cos(\omega_0 t + \beta) \\ &\quad (W(\omega)W_d(\omega - \omega_0) + W_d(\omega)W(\omega - \omega_0)) \\ &\quad + \alpha^2 W_d(\omega - \omega_0)W(\omega - \omega_0)\cos^2(\omega_0 t + \beta). \end{aligned} \quad (12)$$

Similary, the second term and the fourth term are represented by the following equation for the test signal (Eq. 1) input.

$$\begin{aligned} &-\Im[S(\omega,t)]\Re[S_d(\omega,t)] \\ &= \alpha^2 W(\omega - \omega_0)W_d(\omega - \omega_0)\sin^2(\omega_0 t + \beta). \end{aligned} \quad (13)$$

By using the constant relation ($\sin^2\theta + \cos^2\theta = 1$) the numerator is reduced to the following form.

$$\begin{aligned} &\Re[S(\omega,t)]\Im\left[S_d(\omega,t)\right] - \Im[S(\omega,t)]\Re\left[S_d(\omega,t)\right] \\ &= W(\omega)W_d(\omega) + \alpha^2 W(\omega - \omega_0)W_d(\omega - \omega_0) \\ &\quad + \alpha\cos(\omega_0 t + \beta)\cdot \\ &\quad (W(\omega)W_d(\omega - \omega_0) + W_d(\omega)W(\omega - \omega_0)), \end{aligned} \quad (14)$$

where the third term is the time varying component. This term can be cancelled by averaging terms calculated at one half pitch period apart. Note that the denominator of Eq. 10 is the TANDEM spectrum when this condition holds. Therefore, the averaged instantaneous frequency weighted by the power spectra calculated at two locations one half pitch period apart yields the interference-free instantaneous frequency [4].

D. Group delay without phase unwrapping

In this section, a group delay equation having the similar form to the Flanagan's equation is derived, by starting from the spectral representation $X(\omega,t)$. The group delay is defined by the negative frequency derivative of the phase of $X(\omega,t)$. It is equivalent to calculate the derivative of the imaginary part of the log-converted spectrum $\log(X(\omega,t))$.

$$\begin{aligned} -\tau_g &= \frac{d\ \Im\left[\log(X(\omega,t))\right]}{d\omega} \\ &= \Im\left[\frac{1}{X(\omega,t)}\frac{dX(\omega,t)}{d\omega}\right] \\ &= \frac{\Re[X(\omega,t)]\Im\left[\frac{dX(\omega,t)}{d\omega}\right] - \Im[X(\omega,t)]\Re\left[\frac{dX(\omega,t)}{d\omega}\right]}{|X(\omega,t)|^2}, \end{aligned} \quad (15)$$

To simplify Eq. 15, two spectral representations $X(\omega,t)$ and $X_d(\omega,t)$ are defined.

$$X(\omega,t) = \int w(\tau)x(\tau - t)e^{-j\omega\tau}d\tau \quad (16)$$

$$X_d(\omega,t) = \frac{dX(\omega,t)}{d\omega} = \int w(\tau)x(\tau - t)\frac{de^{-j\omega\tau}}{d\omega}d\tau \quad (17)$$

$$= -j\int \tau w(\tau)x(\tau - t)e^{-j\omega\tau}d\tau, \quad (18)$$

Substituting $X(\omega,t)$ and $X_d(\omega,t)$ into Eq. 15 yields the simpler form.

$$\begin{aligned} &-\tau_g(\omega,t) = \\ &\quad \frac{\Re[X(\omega,t)]\Im[X_d(\omega,t)] - \Im[X(\omega,t)]\Re[X_d(\omega,t)]}{|X(\omega,t)|^2}, \end{aligned} \quad (19)$$

This is the counterpart of the Flanagan's equation for the group delay. Equation 19 also does not require phase unwrapping.

E. Interferences on group delay caused by repetition

When defining a signal $x(t)$ to test interference to the group delay, it is general enough to assume that the one of the excitation is normalized and the amplitude of the following repetition is modified by a constant number α.

$$x(t) = \delta\left(t - \frac{T_0}{2}\right) + \alpha\delta\left(t + \frac{T_0}{2}\right), \quad (20)$$

where T_0 is the interval between the repetitions. The spectral representation $S(\omega,t)$ calculated by the short term Fourier transform is

$$\begin{aligned} S(\omega,t) &= w\left(t - \frac{T_0}{2}\right)e^{-j\omega\left(t - \frac{T_0}{2}\right)} \\ &\quad + w\left(t + \frac{T_0}{2}\right)\alpha e^{-j\omega\left(t + \frac{T_0}{2}\right)}, \end{aligned} \quad (21)$$

where $w(t)$ is the time windowing function.

The power spectrum $P(\omega,t)$ of this test signal is

$$\begin{aligned} P(\omega,t) &= \left|w\left(t - \frac{T_0}{2}\right)\right|^2 + \alpha\left|w\left(t + \frac{T_0}{2}\right)\right|^2 \\ &\quad + 2\alpha w\left(t - \frac{T_0}{2}\right)w\left(t + \frac{T_0}{2}\right)\cos\left(2\pi\frac{\omega}{\omega_0}\right), \end{aligned} \quad (22)$$

where the third term is the spectral interference caused by the repetition. The interference is a sinusoid with the period $\omega_0 = 2\pi/T_0$ on the frequency axis.

This interfering term can be cancelled by averaging two shifted versions of the power spectra $\omega_0/2$ apart on the frequency axis and yields the interference-free power spectrum $P_F(\omega, t)$.

$$P_F(\omega,t) = \frac{P\left(\omega - \frac{\omega_0}{4}, t\right) + P\left(\omega + \frac{\omega_0}{4}, t\right)}{2}, \quad (23)$$

where the suffix F of $P_F(\omega, t)$ indicates that the interference-free power spectrum this time is defined on the frequency axis.

F. Interference-free group delay

Since the group delay equation Eq. 19 has the power spectrum in its demoninator, power spectrum weighted averaging cancels it. Define the weighted average group delay $\tau_{gA}(\omega, t)$ by

$$\tau_{gA}(\omega,t) = \frac{\tau_{g1}(\omega,t)|S_1(\omega,t)|^2 + \tau_{g2}(\omega,t)|S_2(\omega,t)|^2}{|S_1(\omega,t)|^2 + |S_2(\omega,t)|^2}, \quad (24)$$

Simplification starts from the numerator of this equation.

1) Simplification of numerator: Let $S(\omega, t)$ represent the short term Fourier transform of the test signal $x(t)$ using a windowing function $w(t)$

$$\begin{aligned} S(\omega,t) =& w\left(t - \frac{T_0}{2}\right) e^{-j\omega\left(t - \frac{T_0}{2}\right)} \\ &+ w\left(t + \frac{T_0}{2}\right) \alpha e^{-j\omega\left(t + \frac{T_0}{2}\right)} \\ =& w(t_1)\cos(-\omega t_1) + jw(t_1)\sin(-\omega t_1) \\ &+ w(t_2)\alpha\cos(-\omega t_2) + jw(t_2)\alpha\sin(-\omega t_2), \quad (25) \\ &\text{where} \quad t_1 = t - \frac{T_0}{2},\ t_2 = t + \frac{T_0}{2}\ . \end{aligned}$$

The frequency derivative of $S(\omega, t)$ is denoted by $S_d(\omega, t)$

$$\begin{aligned} S_d(\omega,t) =& -j\left[w_d(t_1)e^{-j\omega t_1} + \alpha w_d(t_2)e^{-j\omega t_2}\right] \\ =& -jw_d(t_1)\cos(-\omega t_1) + w_d(t_1)\sin(-\omega t_1) \\ &- jw_d(t_2)\alpha\cos(-\omega t_2) + w_d(t_2)\alpha\sin(-\omega t_2), \quad (26) \end{aligned}$$

Using $S(\omega, t)$ and $S_d(\omega, t)$, the first term of the numerator of the group delay equation yields

$$\begin{aligned} &\Re[S(\omega,t)]\Im[S_d(\omega,t)] = \\ &- w_d(t_1)w(t_1)\cos^2(-\omega t_1) - \alpha^2 w_d(t_2)w(t_2)\cos^2(-\omega t_2) \\ &- \alpha w_d(t_2)w(t_1)\cos(-\omega t_1)\cos(-\omega t_2) \\ &- \alpha w_d(t_1)w(t_2)\cos(-\omega t_1)\cos(-\omega t_2), \quad (27) \end{aligned}$$

and the second term yields

$$\begin{aligned} &\Im[S(\omega,t)]\Re[S_d(\omega,t)] = \\ &w_d(t_1)w(t_1)\sin^2(-\omega t_1) + \alpha^2 w_d(t_2)w(t_2)\sin^2(-\omega t_2) \\ &+ \alpha w_d(t_2)w(t_1)\sin(-\omega t_1)\sin(-\omega t_2) \\ &+ \alpha w_d(t_1)w(t_2)\sin(-\omega t_1)\sin(-\omega t_2). \quad (28) \end{aligned}$$

Using the constant relations $\sin^2\theta + \cos^2\theta = 1$ and $\cos A\cos B + \sin A\sin B = \cos(A - B)$ to simplify the numerator yields

$$\begin{aligned} &\Re[S(\omega,t)]\Im[S_d(\omega,t)] - \Im[S(\omega,t)]\Re[S_d(\omega,t)] = \\ &- w_d(t_1)w(t_1) - \alpha^2 w_d(t_2)w(t_2) \\ &- \alpha\left(w_d(t_1)w(t_2) + w_d(t_2)w(t_1)\right)\cos\left(2\pi\frac{\omega}{\omega_0}\right), \quad (29) \end{aligned}$$

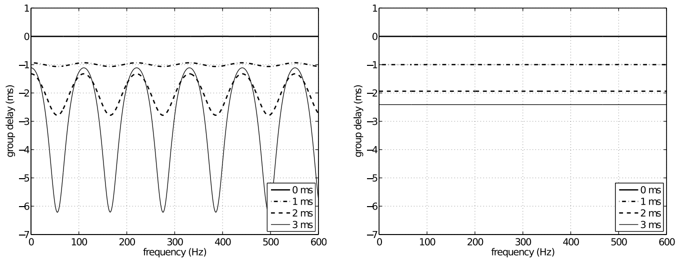

Fig. 1. Group delays calculated by positioning the time window center at 0, 1, 2, 3 ms from one of the pulse position. The test signal is a pulse train with a 110.25 Hz fundamental frequency. (Left plot) Conventional group delay. (Right plot) Interference-free group delay

where the third term is the interfering term on the frequency axis. This sinusoidal variation can be cancelled by calculating the weighted average of the shifted versions $\omega_0/2$ apart on the frequency axis. The denominator in this case is $P_F(\omega, t)$, the interference-free power spectrum on the frequency axis.

2) Interference-free group delay: Based on derivations above, the power spectrum weighted average of two frequency shifted versions of group delay yields the interference-free group delay $\tau_{dF}(\omega, t)$.

$$\begin{aligned} \tau_{dF}(\omega,t) =& -\frac{1}{P_F(\omega,t)} \cdot \\ &[\Re[S(\omega_1,t)]\Im[S_d(\omega_1,t)] - \Im[S(\omega_1,t)]\Re[S_d(\omega_1,t)] \\ &+ \Re[S(\omega_2,t)]\Im[S_d(\omega_2,t)] - \Im[S(\omega_2,t)]\Re[S_d(\omega_2,t)]]\ , \quad (30) \\ &\text{where } \omega_1 = \omega - \frac{\omega_0}{4},\ \omega_2 = \omega + \frac{\omega_0}{4} \end{aligned}$$

3) Implementation: The spectral shift operation $\pm\omega_0/4$ on the frequency axis is implemented by multiplying $e^{\pm j\omega_0 t/4}$ to the windowing function in the time domain.

III. Numerical examples

This section illustrates how the proposed method works for typical signals. Tested signals are a periodic pulse train, a synthetic vowel /a/, and a natural vowel /a/. All test signals are sampled at 44,100 Hz. The Hann windowing function is used in the following tests.

A. Periodic pulse train

A 110.25 Hz pulse train was analyzed using a time window of length 17 ms. Figure 1 shows the results. The center of the windowing function was positioned at 0, 1, 2, and 3 ms behind from one of the pulse. The left plot shows the group delay calculated using Eq. 19, the conventional method. The right plot shows the group delay calculated using Eq. 30, the interference-free group delay. Note that the periodic variation found in the left plot is completely cancelled in the right plot.

B. Synthetic vowel /a/

A synthetic vowel /a/ was generated. The first four formant frequencies (Fn, where n represents the order) and band widths (Bn) used were F1:800 Hz, B1:90 Hz, F2:1,200 Hz, B2:45 Hz, F3:2,700 Hz, B3:120 Hz, F4:3,500 Hz, B4:240 Hz. Figure 2 shows the waveform and the power spectra. The left plot also shows an example of windowing function positioning. The right plot shows the conventional power spectrum (thin line)

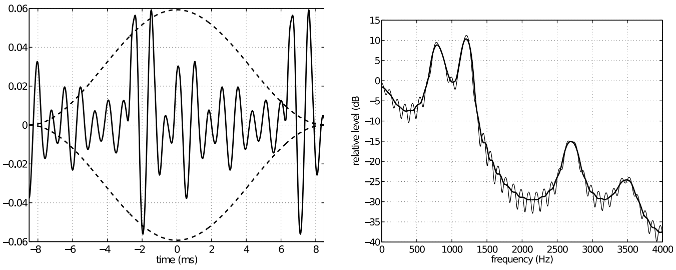

Fig. 2. (Left plot) Waveform of a synthetic speech /a/ and an example of window positioning. (Right plot) Stabilized power spectrum (thick line) using window positioning with 3 ms displacement from the excitation location and usual power spectrum (thin line)

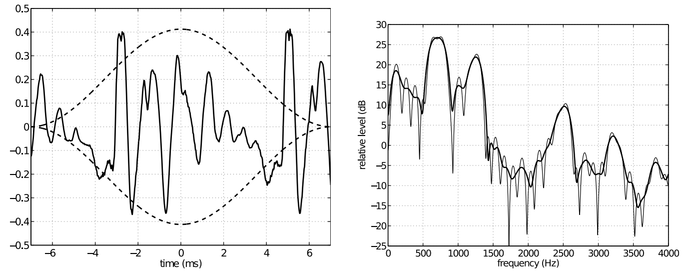

Fig. 4. (Left plot) Waveform of a natural vowel /a/ and an example of window positioning. (Right plot) Stabilized power spectrum (thick line) using window positioning with a 3 ms displacement from the one of the GCIs and usual power spectrum (thin line)

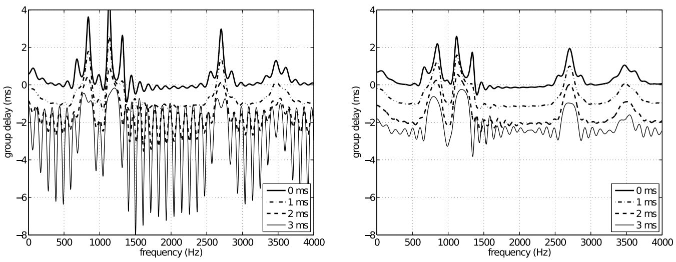

Fig. 3. Group delays calculated by positioning the time window center at 0, 1, 2, 3 ms from one of the excitation pulse position. The test signal is a synthetic vowel /a/ with a 110.25 Hz fundamental frequency. (Left plot) Conventional group delay. (Right plot) Interference-free group delay

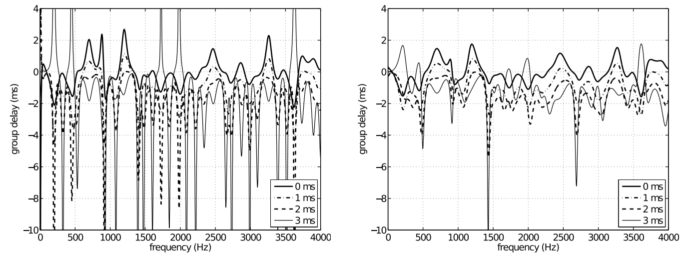

Fig. 5. Group delays calculated by positioning the time window center at 0, 1, 2, 3 ms from one of the GCIs. The test signal is a natural vowel /a/. The fundamental frequency was 127.5 Hz. (Left plot) Conventional group delay. (Right plot) Interference-free group delay

and the (frequency domain) interference-free power spectrum $P_F(\omega, t)$ (thick line). Note that periodic variations found in the conventional power spectrum are virtually completely removed in the interference-free representation.

Figure 3 shows the group delays. The left plot shows the conventional group delay and the right plot shows the interference-free group delay. Note that strong periodic variations in the conventional group delay are effectively suppressed in the interference-free group delay.

C. Natural vowel /a/

A natural vowel /a/ spoken by a Japanese male was analyzed. Figure 4 shows the waveform and the power spectra. The fundamental frequency of the signal was 127.5 Hz. The left plot also shows an example of windowing function positioning. The right plot shows the conventional power spectrum (thin line) and the (frequency domain) interference-free power spectrum $P_F(\omega, t)$ (thick line). The window location in the left plot was set with a 3 ms displacement from the one of the glottal closure instances (GCIs) [7]. Note that periodic variations found in the conventional power spectrum are effectively removed in the interference-free representation.

Figure 5 shows the group delays. The left plot shows the conventional group delay and the right plot shows the interference-free group delay. Note that strong periodic variations in the conventional group delay are effectively suppressed in the interference-free group delay. Note that the left plot suggests that another repetition about 1 ms interval exists.

IV. Conclusions

A new interference-free representation of the group delay is proposed based on a similar form to the Flanagan's instantaneous frequency equation. The proposed method is useful for investigating detailed temporal structure of voice excitation signals and will lead to improvement of the voice quality of synthetic speech signals. The proposed method is also generally applicable to any periodic signals.

Acknowledgment

The authors thank B. Yegnanarayana for valuable discussions on group delay and GCI detections.

References

[1] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. I. fundamentals," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 520 – 538, 1992.

[2] L. Cohen, *Time-frequency analysis*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[3] M. Morise, T. Takahashi, H. Kawahara, and T. Irino, "Power spectrum estimation method for periodic signals virtually irrespective to time window position," *Trans. IEICE*, vol. J90-D, no. 12, pp. 3265–3267, 2007, [in Japanese].

[4] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno. A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation. In *Proc. ICASSP 2008*, pp. 3933–3936. IEEE, 2008.

[5] H. Kawahara, M. Morise, and T. Irino. An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction. In *Proc. ICASSP 2011*, pp. 5420–5423. IEEE, 2011.

[6] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, pp. 1493–1509, November 1966.

[7] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio Speech and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.