Hilbert-Huang Transform of geomagnetic pulsations at auroral expansion onset

Ryuho Kataoka,1 Yoshizumi Miyoshi,2 and Akira Morioka3

Received 4 March 2009; revised 12 June 2009; accepted 24 June 2009; published 4 September 2009.

[1] The waveform of geomagnetic pulsations at auroral expansion onset looks irregular and is hardly resolved by Fourier transform. Here we perform a novel analysis of the Hilbert-Huang Transform (HHT) to address this problem, focusing on the event investigated in detail by Morioka et al. (2008), in which the auroral kilometric radiation (AKR) breakup was clearly identified. From the HHT analysis of high-latitude search coil ground magnetometer data, Pi1, Pc3, and Pi2 pulsations are extracted as the first, second, and third intrinsic mode functions, respectively. Amplification of the Pi1 and Pc3 pulsations is first detected as a clear precursor to the AKR breakup. The Pi1 and Pc3 pulsations show sudden enhancement at the AKR breakup. We suggest that the HHT is capable of automatically extracting the Pi1, Pi2, and PC3 from the irregular high-latitude geomagnetic pulsations, providing a new type of diagnostic tools for understanding the onset mechanism of auroral substorms.


1. Introduction

[2] Auroral substorm is the greatest explosive plasma phenomenon in the terrestrial ionosphere and magnetosphere [Akasofu, 1964]. Two different scenarios are traditionally proposed as the onset mechanism: Magnetic reconnection in the midtail triggers the onset in the Near-Earth Neutral Line model [Baker et al., 1996; Shiokawa et al., 1998], while the disruption of near-Earth tail current triggers the onset in the Current Disruption model [Lui et al., 1992; Ohtani et al., 1999]. The THEMIS project [Angelopoulos et al., 2008] is now going to differentiate between the two onset mechanisms, providing the coordinated observations by multiple spacecraft and ground-based observatory network.

[3] Recently, Morioka et al. [2008] reported a dynamic behavior of the auroral particle acceleration at substorm onsets from the auroral kilometric radiation (AKR) observations. The AKR is emitted as a result of field-aligned electron acceleration and is a good proxy of the evolution of the field-aligned acceleration region [Kaiser and Alexander, 1977; Morioka et al., 1981]. The AKR breakup is defined as the explosive development of a new high-altitude AKR source at the time of an auroral breakup, which suggests the abrupt formation of new field-aligned acceleration region above a preexisting low-altitude acceleration region [Morioka et al., 2007, 2008]. In this study, we focus on the geomagnetic pulsations detected from the search coil magnetometer at auroral expansion onset associated with the AKR breakup to diagnose the onset process in more detail.

[4] The auroral substorm onset has been traditionally identified in ground magnetometer data as Pi2 pulsations (40–150 s) [Jacobs et al., 1964; Saito, 1969], while high-latitude Pi1 (1–40 s) and Pc3 (10–45 s) pulsations possibly contain important information about the underlying physics of the onset mechanism as suggested by Morioka et al. [2008], on the basis of the Fourier band-pass filtering analysis. It is, however, essentially difficult to decompose the mixed waveform using Fourier transform alone because of the overlapping frequency ranges. The rapid change in the frequency and phase poses another challenge. The purpose of this report is to address this problem using a nonlinear analysis of Hilbert-Huang transform (HHT) [Huang et al., 1998]. We demonstrate an example how the HHT works to decompose the nonlinear waveform of high-latitude geomagnetic pulsations into several wave modes such as Pi1, Pi2, and Pc3 pulsations.

2. Geomagnetic Pulssations

[5] Figure 1 shows the time series of the high-latitude geomagnetic pulsations at substorm onset associated with the AKR breakup [Morioka et al., 2008], obtained from the search coil (Figures 1a–1c) and fluxgate (Figures 1d–1f) magnetometers at Tjornes in Iceland (66.5 deg magnetic latitude, L = 6.4, and 0000 UT corresponds to 23.72 magnetic local time). A total of 256 data points are shown starting from 2310 UT on 16 April 1996, with the time cadence of 2.0 s. We would like to note here that the time cadence of 2.0 s cannot resolve the full Pi1 range of 1–40 s. The magnetometer data were recorded in H-D-Z coordi-
nates, with H corresponding to magnetic north, D magnetic east, and Z downward.

The AKR breakup was associated with a sudden reinforcement of the upward field-aligned current (FAC) immediately after the initial low-latitude Pi2 activation [Morioka et al., 2008], and the geomagnetic pulsations associated with the AKR breakup at substorm onset provide a diagnostic to understand the onset mechanism associated with the rapid development of the ionosphere and magnetosphere coupling via FAC. The AKR breakup at $t \sim 170$ s, as identified by Morioka et al. [2008], is indicated by the vertical solid line. Two-step evolution of westward electrojet current is identified in the H component across the AKR breakup. The D and Z components are relatively weaker than H component, and the $-100$ nT negative excursions of the D component at $t \sim 170$ s and $t \sim 250$ s may indicate the enhancement of the FAC twice after the AKR breakup. The high-latitude magnetic pulsations look very irregular as shown in Figures 1a–1c, seem to contain shorter-period wave activities of possibly Pi1 pulsations excited from $t \sim 100$ s, when the electrojet gradually enhanced. Consequently, it is not easy to extract much more information from a quick visual inspection of the geomagnetic pulsations shown in Figure 1. Usually, a filter is applied to the time series which resolves some of these issues to a degree.

3. Hilbert-Huang Transform

The HHT is a nonlinear time-frequency analysis, designed to combine the well-known Hilbert spectral analysis and recently developed empirical mode decomposition (EMD) [Huang et al., 1998]. A key part of the HHT is the EMD with which any complicated data set can be decomposed into a finite and often small number of intrinsic mode functions (IMF), whose Hilbert transform supports the physically meaningful instantaneous frequencies and instantaneous amplitude. The definition of the IMF are as follows: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one, and (2) at any point, the mean value of the envelope

Figure 1. Geomagnetic pulsations at auroral substorm onset associated with the AKR breakup, obtained from (a–c) search coil and (d–f) fluxgate magnetometers at Tjornes in Iceland. The AKR breakup at $t \sim 170$ s, as identified by Morioka et al. [2008], is indicated by the vertical solid line.
defined by the local maxima and the envelope defined by the local minima is zero \cite{Huang et al., 1998}. We applied a noise-assisted data analysis of the ensemble empirical mode decomposition (EEMD) method \cite{Wu and Huang, 2009} to obtain the robust IMF. Appendix A shows the detailed description of the method of analysis, including how we determined the envelope. The instantaneous frequency of each IMF is then calculated, following the method of Duffy \cite{2004} and Barnes \cite{1992}

\begin{equation}
\omega(t) = \frac{1}{2 \Delta t} \tan^{-1} \left[ \frac{h(t - \Delta t)H(t + \Delta t) - h(t + \Delta t)H(t - \Delta t)}{h(t - \Delta t)h(t + \Delta t) + H(t + \Delta t)H(t - \Delta t)} \right],
\end{equation}

where $\Delta t$ denotes the time cadence, $h(t)$ is the IMF, and $H(t)$ is the Hilbert transform of the IMF. The Hilbert spectral analysis itself, or so-called analytic signal analysis, to obtain the instantaneous frequency is not a new method proposed in this study, and essentially rooted in the work of Carson and Fry \cite{1937} and Gabor \cite{1946}, and the Hilbert spectral analysis has been already shown to be useful in space physics \cite{Glassmeier, 1980; Walker et al., 1992}.

Applying the HHT analysis, the IMF and instantaneous frequency of the geomagnetic pulsations are shown in Figures 2 and 3, respectively. The top panel of Figure 2 shows the two horizontal components of search coil magnetometer data, which is the same as shown in Figure 1. The HHT is independently applied to $dH$ and $dD$ components. Hereafter, the $dH$ and $dD$ components are shown by black and red lines, respectively. The original waveform is decomposed into 5–6 IMFs by the EEMD, and the first five IMFs are shown in Figure 2. Corresponding instantaneous frequency of each IMF is shown in Figure 3. From the visual inspections, approximately equivalent to the

\begin{figure}
\centering
\includegraphics[width=\textwidth]{ensemble_empirical_mode_decomposition.png}
\caption{Ensemble empirical mode decomposition of the two horizontal components of the geomagnetic pulsations. The first five intrinsic mode functions (IMFs) are shown. Black and red lines show $dH$ and $dD$ components, respectively. Pi1, Pc3, and Pi2 pulsations are extracted as the first, second, and third IMFs, respectively (see also Figure 3 for the instantaneous frequency). Vertical solid line shows the timing of AKR breakup.}
\end{figure}
Pi1 band (1–40 s) according to Jacobs et al. [1964] is extracted in the IMF-1 as identified in the unstable fluctuations of the periodicity, while approximately equivalent to the Pi2 band (40–150 s) is extracted in the IMF-3 as also identified in the unstable fluctuations with a phase shift at $t = 250$ s as identified by the rapid change in the periodicity. Approximately equivalent to the Pc3 band (10–45 s) is extracted in the IMF-2 as identified by the relatively stable periodicity. The waveform of longer periodicities, possibly reflecting quasi-DC variations of the large-scale current system, also appeared in IMF-4 and IMF-5 in the Pc4 (45–150 s) and Pc5 (150–600 s) ranges, respectively, as identified by the stable periodicity. It is worthwhile to note here that the IMF-3 and IMF-4 have the same period range for the definition of Jacobs et al. [1964], and the fluctuation level of the periodicity is larger in IMF-3 than in IMF-4. Consequently, the HHT successfully extract the wave modes of Pi1, Pi2, and Pc3 pulsations where the stability of the instantaneous frequency provides a new objective criterion to identify the type of geomagnetic pulsations such as Pi and Pc [Jacobs et al., 1964; Saito, 1969].

4. S Transform

The S transform [Stockwell et al., 1996] is introduced as an excellent counterpart method of decomposition to compare with the HHT. The S transform is a Fourier-type algorithm and a natural extension from Fourier, Gabor, and Wavelet transforms:

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t) g(t - \tau, f) e^{-i\omega t} dt,$$

where $\sigma(f) = |f|^{-1}$ is selected to realize the direct frequency-time representation with the best time-frequency resolution. The S transform has been used in space physics as a useful diagnostic tool [e.g., Pulkkinen and Kataoka, 2006; Kataoka and Pulkkinen, 2008]. As with the Fourier transform, the S transform can be applied to a two-component vector time series by constructing a one-dimensional complex valued time series. By looking at the phase difference between the positive and negative frequencies, it is possible to infer complete information of the generalized elliptical motion, including the sense of rotation, the major and minor axes and ellipticity [Stockwell et al., 2004]. More practically, when the polarization is counterclockwise (clockwise), the relative amplitude in positive (negative) frequency is more enhanced.

![Figure 3. Instantaneous frequencies for the first five IMFs of Figure 2. Black and red lines show dH and dD components, respectively. Vertical solid line shows the timing of AKR breakup.](image)
The $S$ transform is applied to the complex time series of $h(t) = dD(t) + idH(t)$, which is shown in Figure 4a. Figures 4b and 4c show the $S$ transform spectra in the positive and negative frequency ranges, respectively. It is not easy to identify the wave modes from the visual inspection of the turbulent spectra themselves. Each spectral peak, however, actually corresponds to the instantaneous frequency of HHT, and the combined display is more useful. White and black diamonds in Figure 4 show the “skeleton spectrum” of instantaneous frequency from IMF-2 of the dH and dD components, respectively; that is, the frequency modulation of the Pc3 pulsation are shown in Figure 3. Tracking the skeleton spectrum from $t/C24 \approx 200$ s to $t/C24 \approx 300$ s, and comparing the relative amplitude of positive/negative $S$ transform spectra along the skeleton spectrum, the polarization of the Pc3 pulsation is identified as switching from L mode to R mode across $t \approx 250$ s. The monochromatic Pc3 waveform with such a clear polarization is difficult to obtain from a random noise signal, indicating a real existence of the Pc3 pulsation hidden in irregular Pi1 and Pi2 pulsations. Consequently, the HHT and $S$ transform complement each other to obtain a comprehensive view to track the fast temporal variation of the frequency and polarization.

5. Discussions

The HHT has been known to be useful in many fields of research (see Huang and Wu [2008] for review). This report adds one more example how the HHT successfully works in space physics to identify particular wave modes in the ground magnetic field variations associated with the explosive plasma phenomena of auroral substorms. Without predetermining the frequency range or any particular modes as usually done in the Fourier transform, the waveform of geomagnetic pulsations is automatically decomposed into small number of IMFs by the HHT, and the wave modes corresponding to Pi1, Pi2, and Pc3 are automatically extracted as shown in Figure 2. From the HHT analysis of the other events listed by Morioka et al. [2008], the high-latitude geomagnetic pulsations associated with the substorm onset are similarly decomposed into 5 to 6 IMFs, indicating a robustness of the results shown above (see also Appendix B).
[12] From the HHT analysis of the irregular high-latitude geomagnetic pulsation (Figure 1), Pi1, Pi3, and Pi2 pulsations are extracted as the first, second, and third IMFs, respectively (Figures 2 and 3). The amplification of the Pi1 and Pi3 pulsations is first detected as a precursor of the onset. The Pi1 and Pi3 pulsations show sudden enhancement at the AKR breakup. After the AKR breakup, the Pc3 pulsation shows a polarization change from L mode to R mode (Figure 4), and the Pi2 pulsation shows a phase shift (Figure 3), both potentially corresponding to a subsequent breakup of AKR. The one-to-one correspondence is, however, not conclusive between each smaller AKR breakup and the Pi2 phase shift or Pc3 polarization change (see Figure 6 of Morioka et al. [2008] for details). The phase shift and polarization change also correspond to the recovery phase of negative bay (Figure 1) and also correspond to the negative excursion of D component which implies the transient enhancement of the FAC.

[13] The HHT analysis identified similar Pc3 pulsations in the other events listed by Morioka et al. [2008] (see also Appendix B). The existence of Pc3-type pulse itself is not a new finding of this paper, and may correspond to PiC pulsations [Arnoldy et al., 1998] and/or long-period Pi1 ULF waves [Milling et al., 2008; Murphy et al., 2009]. Arnoldy et al. [1998] reported that PiC pulsations occurred simultaneously with the Pi1B in about a third of the events, and had its onset at the same time across most of the stations of the Antarctic magnetometer array, while Miling et al. [2008] reported localized onset of the long-period Pi1 and expansion from the “epicenter.”

[14] It is important to note here that the Pi2 signal may be a result of a switch on process, and in that case the deconvolution of the switch on function from the system response would be better treated as done by Behrens and Glassmeier [1986] before applying the HHT or S transform. Although the search coil magnetometer data, i.e., the time derivative of the magnetic field, works to decrease the contamination problem of such a switch on response, a few wave trains spanning the switch on timing can be sometimes affected and therefore be carefully interpreted. Less affected from the contamination is the long-lasting wave trains such as the Pc3 pulsation as identified in this paper.

[15] The Pi1 pulsation detected in this study is bursty as shown in Figure 4, and can be regarded as the Pi1B pulsation which is also used as the substorm onset signatures. Lessard et al. [2006] reported the nature of Pi1B pulsations, suggesting a mode conversion mechanism for the production of the transverse Pi1B pulsations at ground from compressional Pi1B pulsations in the magnetosphere. Traditionally, Pi1B pulsations can be naturally interpreted with ionospheric Alfvén resonator (IAR) [Lysak, 1988] where the strong current pulse coming into the ionosphere excite the Pi1B due to the gradient of Alfvén speed to create the resonant cavity. According to the IAR theory, the Pi1B pulsation detected in this study is interpreted as the front signal of the oncoming strong FAC pulse at the substorm onset. The ionospheric feedback instability may also be involved in the IAR theoretically [Lysak, 1991], which would be potentially important during the simultaneous amplification of Pi1 and electroject around the auroral breakup.

[16] The Pi2s have been shown to have clear polarization and phase shift [Lester et al., 1983, 1984] and without more stations during the same event it is hard to distinguish what types of wave this is. Although multistation magnetometer network data were not available for the present study, it is anticipated that the network monitoring of the spatial extent and the time evolution of Pi1, Pi2, and Pc3 activities, using the HHT as a filter, would provide a new type of diagnostic tools for the onset mechanism and to distinguish the possible role of the ionosphere in the substorm onset process. The analysis method of Glassmeier [1980] to show the contour map of instantaneous amplitude using magnetometer network would be useful for that purpose. It may also be beneficial to maximize the accuracy and science achievement using the coordinated observations of THEMIS. In fact, using the CARISMA and THEMIS/GBO magnetometer data, it has been reported that there is a localization of the long-period Pi1 pulsation or short-period Pi2 pulsations [Milling et al., 2008; Murphy et al., 2009; Rae et al., 2009a], which may be closely related to the Pc3 pulsation detected in this study. These authors also found that the ~10–40 s ULF wave band occurred first in the ionosphere subsequent to substorm expansion phase onset, and in the region defined by auroral onset [Rae et al., 2009b]. Also, as shown in Figure 3, the IMF-3 sometimes actually spans both the Pi1 and Pi2 period bands, and may have similarity to the ULF waves rather than traditional Pi2 associated with the substorm current wedge structure. Although the search coil magnetometer data would be essential to obtain the high signal-to-noise ratio for precisely decomposing the high-frequency geomagnetic pulsations of Pi1 and Pc1–3, the HHT may also work as a denoise filter for the high-frequency pulsations in fluxgate magnetometer data and would be still useful to accurately detect the onset time of the Pi1 and Pc3 waveforms. We are preparing a new analysis of magnetometer network data using HHT technology for diagnosing the onset mechanism in more detail.

6. Conclusions

[17] The HHT is a new type of diagnostic tools, capable of automatically decomposing the high-latitude irregular geomagnetic pulsations obtained from the search coil magnetometer at the AKR breakup onset into Pi1, Pi2, and Pi3 pulsations. The high-latitude Pi1 and Pi3 pulsations as identified from the HHT analysis may be useful as a precursor of auroral breakup. A comprehensive time-frequency spectral view is also obtained from the instantaneous frequency, especially when complemented with the S transform, and the instantaneous frequency provides a new objective criterion to identify the type of geomagnetic pulsations such as Pi and Pc. It would be useful to apply the HHT to all the available both ground- and space-based magnetic data sets across all local times and latitudes for diagnosing the wave activities and underlying physics associated with the substorm onset.

Appendix A: Procedures for HHT Analysis

[18] The procedures of the HHT analysis used in this study are briefly described for code developers. The “sifting” process for EMD can be described as follows: (1) identify all the local maxima and minima, (2) connect the maxima (minima) by cubic spline to form the upper (lower) envelope,
(3) subtract the mean values of the upper and lower envelopes from the data. In this study, to mitigate the "end effect," two more points of extrema are artificially added to each end of the data just before the spline. The time separation and amplitude of the artificial extrema are simply calculated from the mean values of those of the two closest extrema.

[19] To obtain the first IMF, the sifting process is repeated until the number of extrema and the number of zero crossings are (1) equal or at most differ by one and (2) stay the same for three consecutive times, although (3) the number of siftings is limited to 300 at maximum in this study. The first IMF is extracted from the data, and the residual is then treated as new data to obtain an IMF of lower frequency.

[20] Ensemble EMD used in this study contains following steps: (1) add a white noise series to the targeted data, (2) decompose the data with added white noise into IMFs, (3) repeat steps 1 and 2 for enough number of times but with different white noise series each time, and (4) obtain the ensemble means of corresponding IMFs of the decompositions as the final result. In this study, the ensemble number is set to 100, and the amplitude of white noise is taken to be 0.1 of the standard deviation of the targeted data.

Appendix B: Another Example of HHT Analysis

[21] From the auroral breakup events of Morioka et al. [2008], another typical example of the HHT analysis is shown in Figures B1, B2, and B3. The same HHT and S transform analysis are performed for the search coil magnetometer data obtained at Syowa station in Antarctica (~65.6 deg magnetic latitude), with the time cadence of 2.0 s. The magnetic local time and UT are approximately the same at Syowa station.

[22] Comparing with Figures 2, 3, and 4, it is found that similar results are obtained in Figures B1, B2, and B3 as follows: Pi1, Pc3, and Pi2 pulsations are automatically
Figure B2. Instantaneous frequencies for the first five IMFs of Figure B1. Black and red lines show dH and dD components, respectively. Vertical solid line shows the timing of low-latitude Pi2 activation.
Figure B3. The S transform spectra of (a) the combined signal of dD + idH in (b) positive and (c) negative frequency ranges. White and black diamonds show the skeleton spectrums of the instantaneous frequency from the IMF-2 of dH and dD components, respectively. Vertical solid line shows the timing of low-latitude Pi2 activation.
extracted as the first, second, and third IMFs, respectively (Figures B1 and B2). The amplification of the Pc3 pulsations is first detected as a precursor of the Pi2 onset as indicated by vertical solid lines. The polarization of the Pc3 pulsation likely shows a phase shift at $t = 380$ s (Figure B2).

[23] Acknowledgments. The work by R.K. was supported by a research fellowship of Special Postdoctoral Research Program at RIKEN. The geomagnetic data from Tromsø and Søyvata were provided from the Auroral Data Center at the National Institute of Polar Research, Japan. This work was supported by the Grant-in-Aid for Creative Scientific Research fellowship of Special Postdoctoral Research Program at RIKEN.


Acknowledgments. The work by R.K. was supported by a research fellowship of Special Postdoctoral Research Program at RIKEN. The geomagnetic data from Tromsø and Søyvata were provided from the Auroral Data Center at the National Institute of Polar Research, Japan. This work was supported by the Grant-in-Aid for Creative Scientific Research fellowship of Special Postdoctoral Research Program at RIKEN.


