

## **Evaluation of Mental Status by Uttered Voice**

Electronic Navigation Research Institute

Dr. Kakuichi Shiomi

### 1. Introduction

In the early part of 1990s, statements were made in that heart ailment and disorder in the functions of the brain of a subject person to be examined can be found through the analysis of blood flow of the person based on the chaotic theory (Tahara, Tsuda and others). According to their observations, the strange attractor of pulsatory waves measured at a finger tip of the person draws a simple form as the brain disorder such as Alzheimer's disease advances.

In 1998, continuous observation of chaoticity of uttered voice was made and it was reported that the time average values of first Lyapunov exponents were increased before the uttered person recognizes fatigue (Shiomi and Hirose).

As of this writing, we think that the variations in the first Lyapunov exponents of uttered voice are strongly related to the mental status of the uttered person and the first Lyapunov exponents of uttered voice vary in accordance with the status of the brain functions of the person who utters. If the person has brain disorder or he or she is tired, the exponents vary at a low level. However, for the case of a normal person who was appropriately rested, the first Lyapunov exponents of his uttered voice vary corresponding to the "degree of utilization of his brain".

As a matter of fact, we can safely say the following based on the results of experiments made, i.e., brain work can increase the level of the first Lyapunov exponents of uttered voice.

For example, if we were to analyze the speech of a person who is carefully selecting his words, we expect to obtain relatively high first Lyapunov exponent values.

In the case of casual conversation, the first Lyapunov exponents may be varying at a low level on the average. However, locally high level first Lyapunov exponents can be obtained, if the person encounters personal or undesirable subjects in the conversation.

Thus, it is possible to observe the changes in the mental status of a speaker from the changes in the chaoticity of the speaker's voice. Moreover, if we were to more precisely analyze the chaoticity of the speaker's voice, individual keywords for the person can be found.

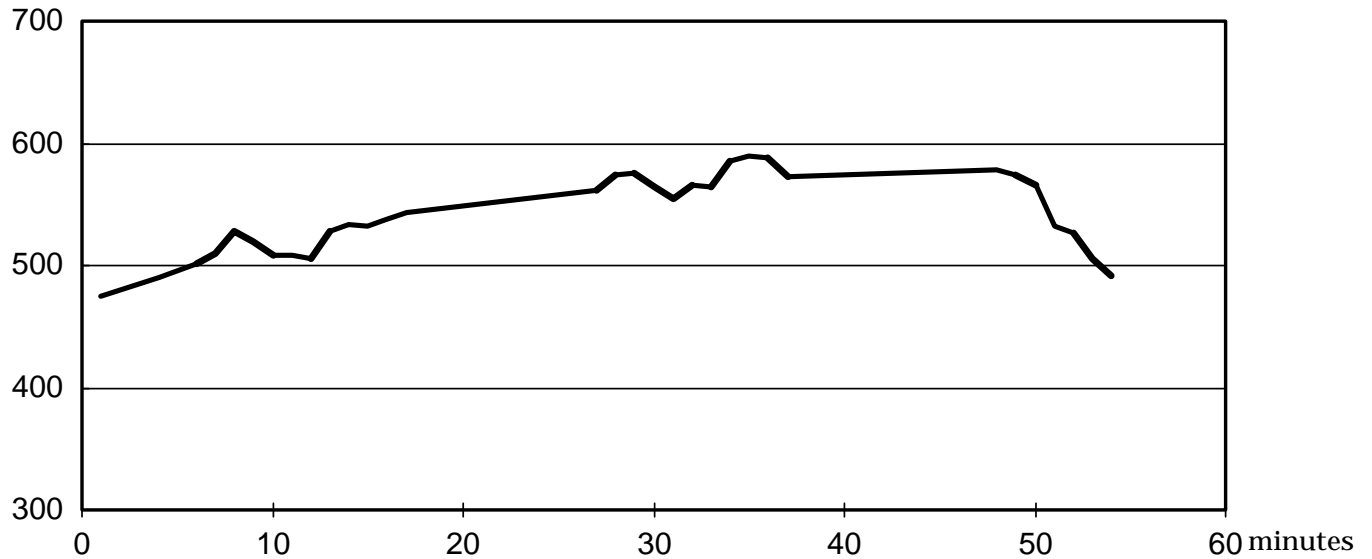
### 2. Evaluation of fatigue

Two persons were asked to read a newspaper aloud and the changes in the first Lyapunov exponents of their voice were computed and the results are shown in Figures 1 and 2. The above results were made in the following manner, i.e., (1) continuously uttered voice was divided into processing units having a time width of one second, (2) first Lyapunov exponents were computed for every second, (3) an averaging time width was set to five minutes to compute the time averages of the first Lyapunov exponents and (4) the changes in these values are plotted.

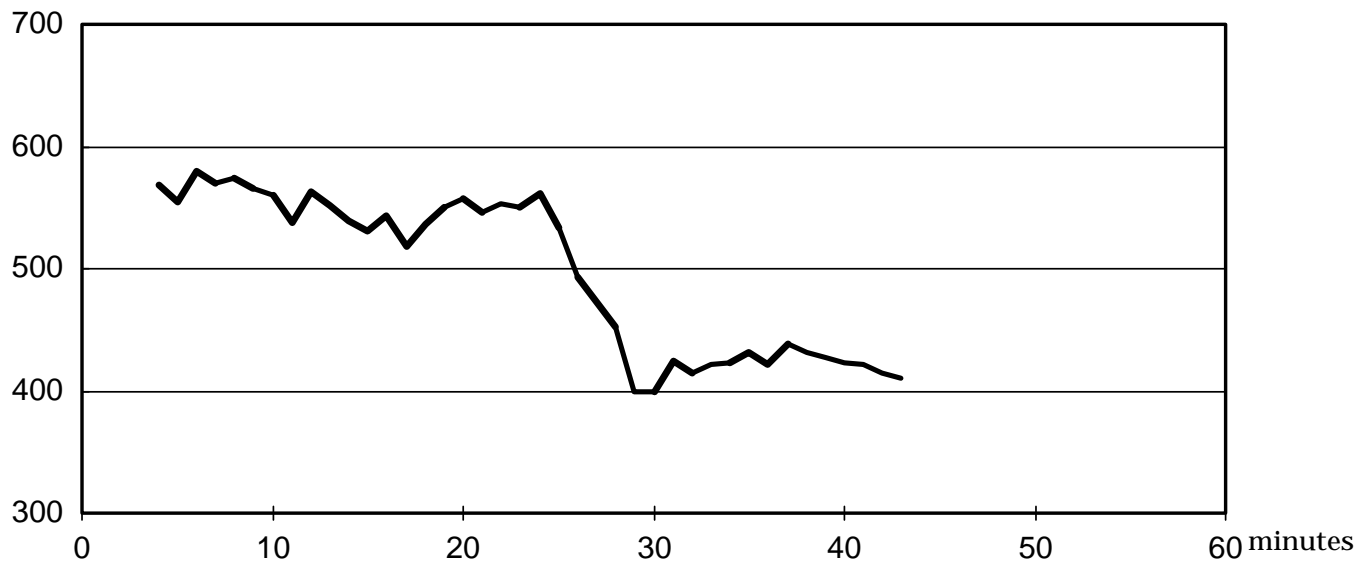
The person A was asked to read political and business sections of the newspaper aloud over

approximately one hour and the computational results are plotted in Figure 1.

The person B was asked to read the political and the business sections of the newspaper for approximately twenty minutes and then was asked to read aloud a sports section and the computational results are plotted in Figure 2.



**Figure 1** Variations in the first Lyapunov exponents in a reading aloud



**Figure 2** Variations in the first Lyapunov exponents in a reading aloud

The person A was very tired at the end of the fifty minutes of the reading aloud to the extent that he was unable to continue the task. However, the person B was able to continue the task and completed his assignment at approximately forty-five minutes.

The first Lyapunov exponents of the uttered voice of the person A were increased from 480 to

580 or approximately 20% over the fifty minute period whereas the first Lyapunov exponents of the uttered voice of the person B was 580 at the beginning of the task. However, due to the change made in the contents of the reading aloud, the exponents were decreased to the 400 level.

In either case, the uttered voice was sampled at the rate of 11.025kHz, the embedding dimension was 4, the embedding delay was approximately 0.9ms (ten times the sample clock time) and the evolution delay time was set to approximately 0.9ms.

The experimental results shown in Figure 1 indicate that the task for the person A i.e., reading of the newspaper aloud resulted in fatigue. On the other hand, the experimental results of Figure 2 show that the reading of sports events task was lot easier than reading the subjects on political and business news.

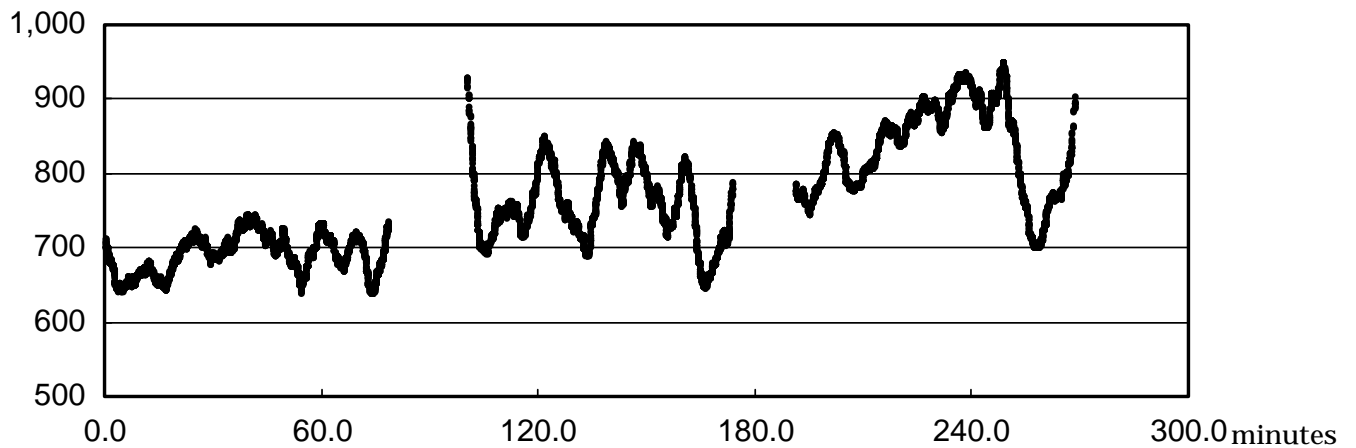
If the first Lyapunov exponents of uttered voice indicates the level of the load being generated in the brain of the speaker, the person A experienced fatigue that was so severe that he was unable to continue the reading because the high work load continued. For the case of the person B, the reading contents shifted from a hard to understand subject to a more simple one resulted in no fatigue generation and he experienced no fatigue after the approximate forty-five minute reading.

As of this writing, several tens of experiments were conducted and no contradictory experimental result was obtained to oppose the hypothesis in which positive correlation exists between the time average values of the first Lyapunov exponents of the uttered voice of a speaker and the work load level generated in the brain of the speaker.

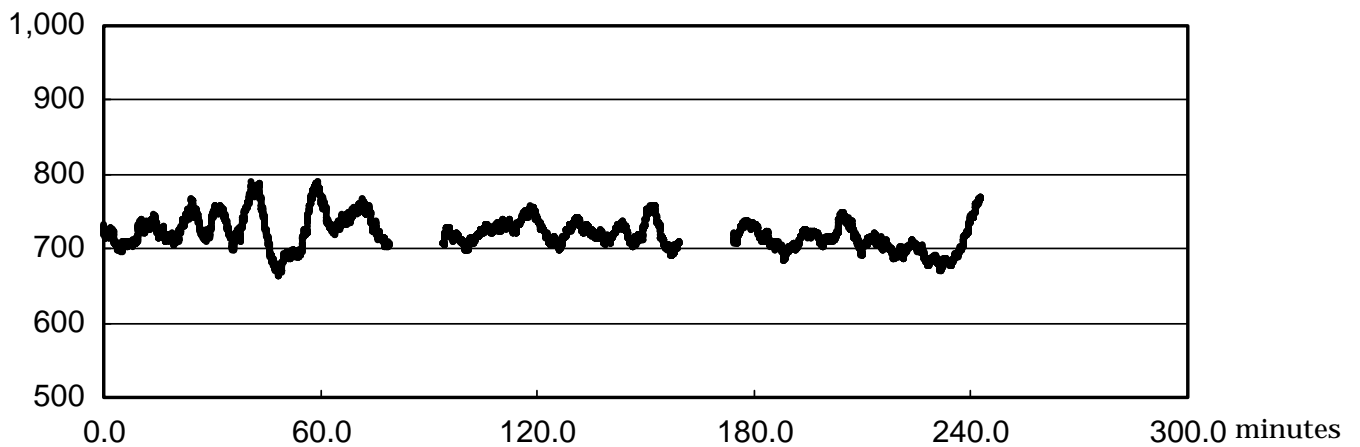
When high first Lyapunov exponents continue, a person being examined will experience fatigue sooner or later. However, when the person was tasked to read a book aloud for few hours and he does not complain about the task, the transition in the first Lyapunov exponents of the utter voice of the person becomes flat.

Figures 3 and 4 show the results of the experiments in which two persons to be examined were asked to read books aloud seventy to eighty minutes with approximately fifteen minutes of resting time.

The examinee C who gave the results shown in Figure 3 felt strong fatigue after the completion of the experiments whereas the examinee D who showed the results of Figure 4 felt no strong fatigue after the completion of the experiments.



**Figure3** Variations in the first Lyapunov exponents in a reading aloud



**Figure4** Variations in the first Lyapunov exponents in a reading aloud

The variations in the first Lyapunov exponents shown in Figures 3 and 4 are thought to indicate the variations in the degree of concentrations to the operations of reading aloud.

The examinee C lost the concentration to the reading aloud in the second reading aloud after the first rest, continued the task, accumulated fatigue and felt heavy fatigue in the third reading aloud.

The examinee D also lost the concentration to the reading aloud in the first reading aloud similar to the case of the examinee C. However, after the rest, the examinee D somehow accomplished a change in his feeling and completed the second and the third reading aloud tasks without being too tired.

If we were to set the average time width to be five minutes, the time average values of first Lyapunov exponents of uttered voice are indicative of the average of the degree of activity of the brain of the speaker.

It is also possible to consider the time average values of the first Lyapunov exponents of

uttered voice to be the stress being exerted onto the brain of the speaker and if relatively high level stress continues for a long time, the speaker realizes fatigue as a result.

### 3. Evaluation of mental status

As mentioned above, it can be said that the time average values of the first Lyapunov exponents of uttered voice indicate the averages of the loading status of the brain of the speaker.

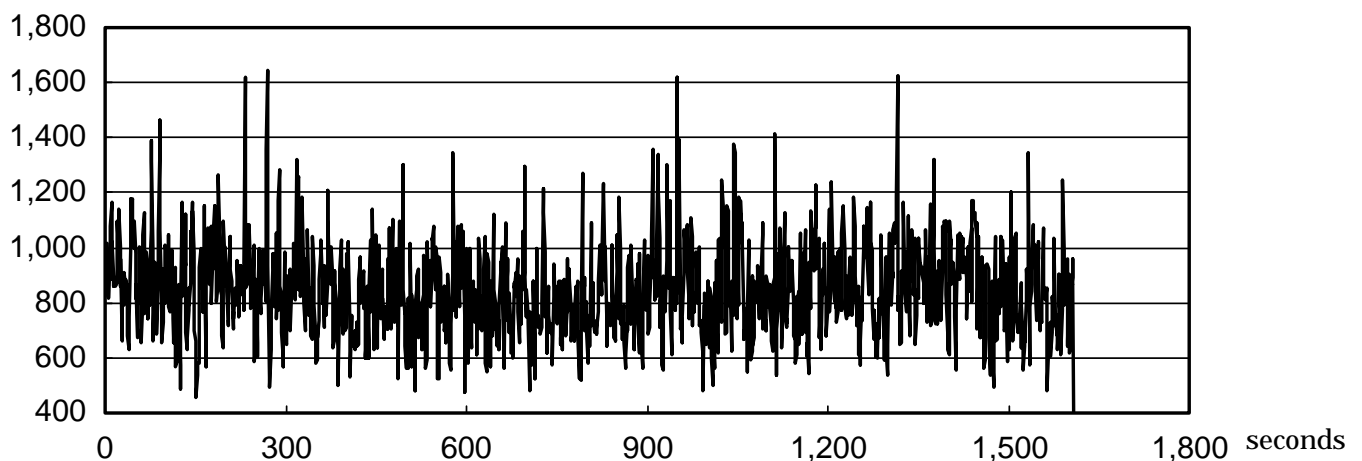
In Section 2 “Evaluation of fatigue”, the averaging time width was set to five minutes and it was said that the evaluation of the fatigue state of a speaker was possible. If the width of the averaging time is sufficiently made small, i.e., even though the averaging time is set to few seconds or little over ten seconds, for example, it is possible to make a sufficiently reproducible evaluation and therefore, stress corresponding to the contents of uttering may be evaluated at that time.

The followings are the experimental results of a software development aiming at to reduce the width of averaging time without adversely affecting the reliability of the value of the evaluation.

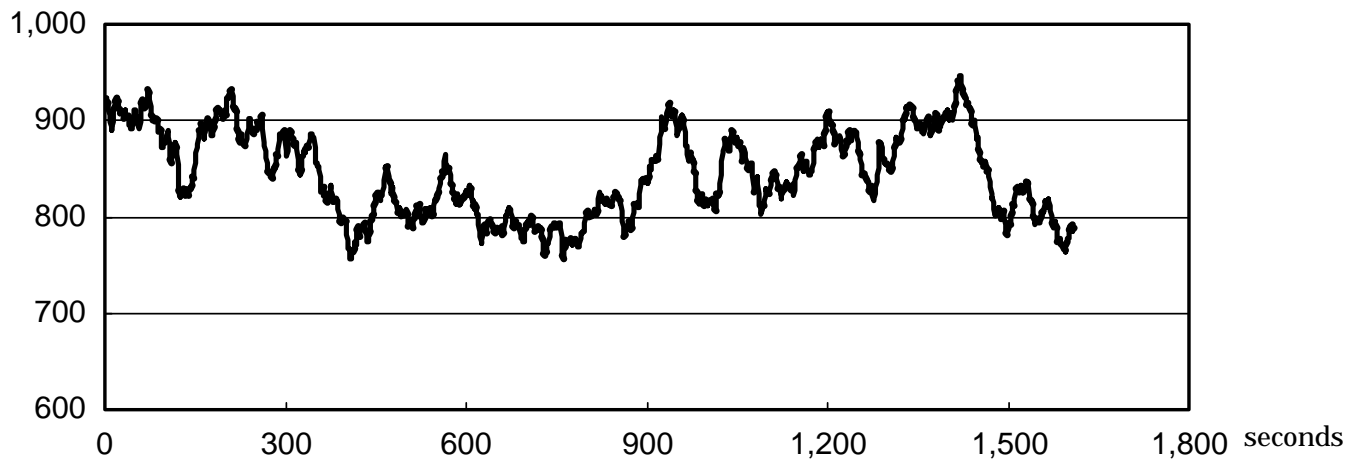
#### 3.1 Problems associated with the conventional methods

Figure 5 and 6 indicate the results the variations in the first Lyapunov exponents of the general-policy speech of the Prime Minister Koizumi delivered on September 27, 2001 processed by the similar method used in the evaluation of fatigue described in the previous sections.

In Figure 5, changes of the first Lyapunov exponents in every second computed from each processing unit are plotted. Figure 6 shows the results of the same process employing a time averaging width that was set to one minute. Note that the uttered voice of the Prime Minister was obtained from the NHK’s broadcast which was video recorded and PCM audio processed.



**Figure 5** Variations in the first Lyapunov exponents in a general-policy speech (processed for every second)

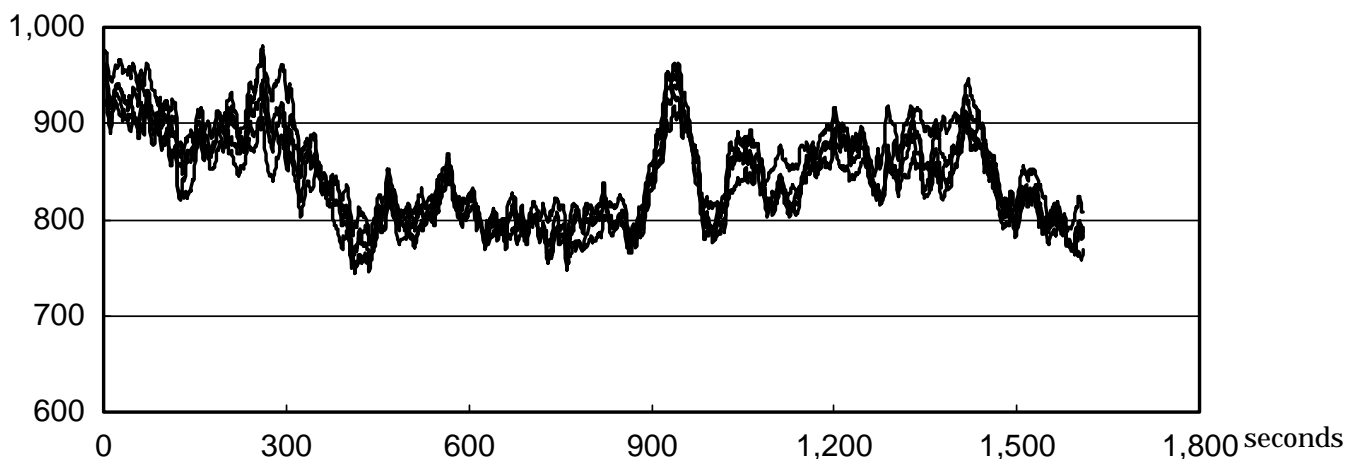


**Figure 6** Variation in the first Lyapunov exponents in a general-policy speech (time averaging width was set to one minute)

In Figures 5 and 6, fine structures are observed and let us find out the meanings of these structures first.

Let us make a hypothesis saying “The structures appeared in these graphs are assumed to be corresponding to the contents of the speech and the structures are preserved even through processing parameters are changed slightly”. Employing the above hypothesis, the cutting timing from recorded data to the processing unit was varied and the changes in the structures of these graphs are observed.

Figure 7 shows the results of the process in which the cutting timing to the process unit of the recorded data of Figure 6 was shifted for 0.1 second.



**Figure 7** Variation in the first Lyapunov exponents in a general-policy speech (time averaging width was set to one minute)

Please note that a portion of the structures observed in Figure 6 is indicated in Figure 7.

However, many structures appear to have no meaning corresponding to the contents of the speech.

One second of processing unit may be sufficient to evaluate the degree of fatigue of a speaker by the five minute average values of the first Lyapunov exponents. However, it appears to be insufficient to extract more detailed information relative to the mental status of the speaker.

### 3.2 Problems associated with the conventional algorithm

In the conventional chaos signal processing, algorithms of Kantz and Rosenstein are used to compute first Lyapunov exponents from time series data in accordance with the processing objectives and the nature of the data to be processed. Moreover, Sano-Sawada algorithms are appropriately used for the computations of Lyapunov spectra.

Each of these algorithms has respective features. However, it is assumed that the dynamics of the system which generates time series data are stable and the first Lyapunov exponents or Lyapunov spectra are given by converging computations.

In a general conversational uttering, the associated dynamics do not have the stability being assumed by the conventional algorithms.

In an everyday Japanese conversation, vocal sounds corresponding to 400 to 600 characters are continuously pronounced every minute, the continuation interval of individual vocal sound is equal to or less than few hundred milliseconds and the dynamics associated with the conversation in the continuation interval are not necessarily stable.

In the results of experiments shown in Figures 1 through 7, the processing unit of uttered voice was set to one second. The length of one second is relatively long compared with the continuation interval of a phoneme in an everyday conversation and therefore, it can be said that the computed first Lyapunov exponents are thought to be the averaged values of the first Lyapunov exponents of plural phonemes (i.e., plural uttering dynamics).

In the vowel sounds of /a/, /i/, /u/, /e/ and /o/ of the Japanese language, each vowel is uttered in different dynamics and it has been confirmed that different first Lyapunov exponents or Lyapunov spectra are obtained by processing the data using the algorithms mentioned above.

Thus, in order to compute the first Lyapunov exponents at the time of the uttering, it is necessary to reduce the processing unit to the approximate continuation interval of individual phoneme at least.

However, if the processing unit of uttered voice is set to 100ms, for example, the waveforms of a vowel are only repeated approximately 20 times, transient conditions related to the connection with other phonemes are generated prior to or after the processing unit and no sufficient time and no stability that have been assumed in the conventional algorithms are expected in that time interval.

If the sampling rate of uttered voice is increased, the apparent size of the data is increased but,

the number of similar waveforms repeated in one phoneme does not change.

When first Lyapunov exponents or Lyapunov spectra are to be computed for time series signals having a definite periodicity, the increase in the sampling rate will result in the increase in the neighborhood point in the vicinity time wise and no chaotic evaluation precision within a processing unit time is improved.

While analyzing ordinary uttered voice, the continuous time of a consonant is short compared with the continuous time of a vowel and the time toward a rise varies. Even though, we limit the cases to be vowels only, the continuation interval and the stability relative to the connection with other phonemes vary due to the differences among phonemes.

In order to obtain chaotic evaluation values having sufficient reliability and reproducibility, it can be said that new process algorithms should be found.

In order to solve the problems mentioned above, a new method will be described in the following sections to compute chaotic evaluation values of uttered voice in which dynamics are continuously changing.

### 3.3 Local time Lyapunov spectrum

In this section, the objects to be processed are limited to uttered voice and the periodicity of the uttered voice is utilized to compute for the local time Lyapunov spectra.

Please note that the algorithms to be described herein are based on the Sano-Sawada algorithms.

When Lyapunov spectra are to be computed by the Sano-Sawada algorithms, we have to (1) set an embedding delay dimension, an embedding delay time and a evolution delay time, (2) determine a neighborhood distance and find a set of the neighborhood point from embedding points generated from time series data, (3) compute evolution point of each point included in the neighborhood point set and (4) Lyapunov spectra are computed from the neighborhood point set and the evolution point set.

In order to properly compute Lyapunov spectra employing the Sano-Sawada algorithms, it is important to properly set a neighborhood distance. If the neighborhood distance is set to be too small, the neighborhood point set does not have a sufficient number of elements and the Lyapunov spectra can not be computed.

If the neighborhood distance is set to be too large, the neighborhood point set will include points having completely different phases on the orbits within an embedding space and the first Lyapunov exponents will have smaller values.

Note that in the Sano-Sawada algorithms, no condition is imposed except “a mutual distance or a distance from a reference point is equal to or less than a neighborhood distance”. Therefore, a neighborhood point must be found from the entirety of processing units and if the processing units become large, the processing time is increased.

Obviously, if the processing units are made small, the processing time of individual processing

unit becomes short. However, if the processing unit is made short, the adverse effect on the processing results caused by the deviation of the dividing timing to each processing unit becomes relatively large. If we are to stabilize the evaluation values by the processes such as an averaging or the like, processes such as cutting out each processing unit from the entire data while overlapping them are required and as a result, the total processing time will be increased.

In the proposed algorithms herein, it is assumed that uttered voice signals have definite periodicity and the neighborhood point exist in synchronism with the periods of voice signals, i.e., the range of the neighborhood point retrieval is limited, the signal processing time is reduced and localized Lyapunov spectra are computed in the following procedures.

In the proposed algorithms, localized Lyapunov spectra similar to Lyapunov spectra are computed employing the algorithms which are obtained by slightly varying the Sano-Sawada algorithms.

In order to distinguish the exponents computed by the proposed algorithms from the conventional Lyapunov exponents or Lyapunov spectra, the exponents computed by the proposed algorithms are called cerebral exponents or cerebral spectra.

In the proposal algorithms, (1) an embedding dimension, an embedding delay time and a evolution delay time are set, (2) the number of elements of a neighborhood point set, a processing frequency band and the bandwidth of the sway of the uttered voice frequencies are set, (3) a set of the neighborhood point is found from embedding points generated from time series data, (4) an evolution point set corresponding to the neighborhood point set is computed and (5) cerebral spectra are computed from the neighborhood point set and the evolution point set.

In the proposed algorithms, (6) a first neighborhood point is searched for as a closest point within the time range computed from a processing frequency band with respect to an embedding reference point, (7) the range of the dynamics of the neighborhood point set is determined from the reference point, the first neighborhood point and the bandwidth of the sway that is beforehand set, (8) neighborhood point are searched and (9) the number of neighborhood point equivalent to the number of the elements of the neighborhood point set beforehand set is found.

The elements of the neighborhood point set determined by the method have beforehand set time gap among them. Therefore, when time series data have clean orbits as strange attractors, it is expected that the elements exist in mutually close locations on the orbits of an embedding space.

Moreover, it is possible to compute the size of a hypersphere, which is cerebral of including a neighborhood point set, with respect to the size of the strange attractor generated by the data to the neighborhood point located at a farthest point from a reference point and hereafter this is called as a cerebral epsilon.

Note that the relationship between the cerebral exponents computed employing the neighborhood point set as a reference and the cerebral epsilon corresponds to the relationship

between the conventional Lyapunov exponents and its neighborhood distance.

In the Sano-Sawada algorithms a neighborhood distance or a neighborhood condition as the ratio with respect to the size of a strange attractor is given and a neighborhood point set is made first and then, cerebral epsilon is computed from the set.

Since a neighborhood point condition is given at first in the Sano-Sawada algorithms the same condition is applied even though a evolution point set exist with respect to a certain neighborhood point set for example, or a next neighborhood point set is built using the center of hypersphere that includes the evolution point set as a reference point and convergence computations having a beforehand set number of operations are conducted.

In the proposed algorithms, the convergence computations are continued only for the case in which a next neighborhood point set generated from a evolution point set has a neighborhood distance that is equal to or smaller than the cerebral epsilon of the neighborhood point set giving the evolution point set (or equal to or smaller than the distance that is made by operating a beforehand set factor to the cerebral epsilon).

Thus, when the time series data to be processed contain a large amount of noise, convergence computations being expected in the Sano-Sawada algorithms are not realized in the proposed algorithms.

When continuously uttered voice is to be analyzed, dynamics of uttered voice are continuously changed by the variation of phonemes.

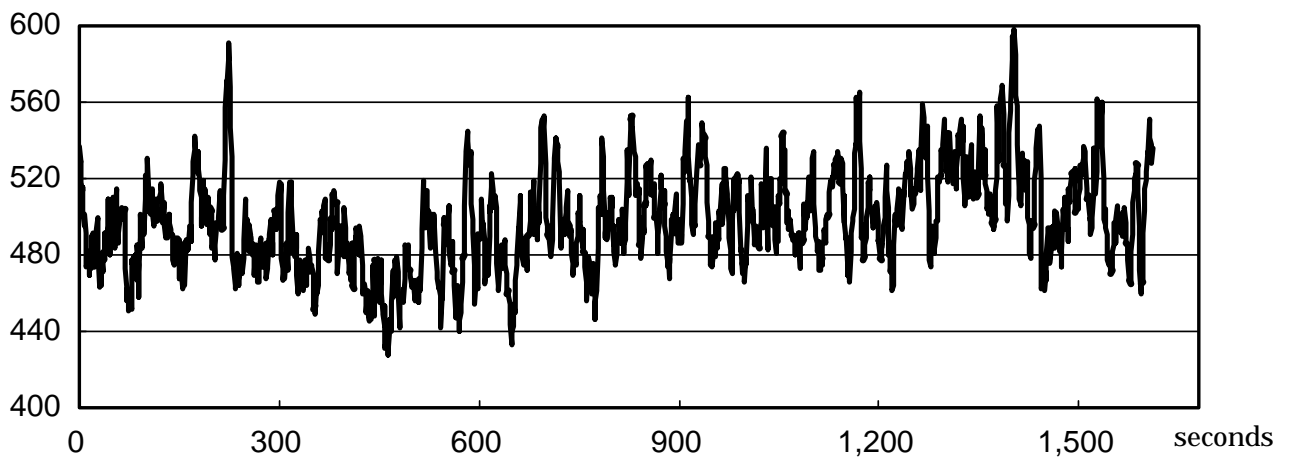
If the dynamics that provide time series data are continuously varied, the convergence computations are terminated at the point where the dynamics are changed, if the proposed algorithms are employed and cerebral exponents corresponding to the dynamics are computed every time.

In the processes of the proposed algorithms, the followings are conducted from the large sized time series signals generated by the dynamics which are changing in a complex manner, i.e., (10) cerebral exponents at each time and the list of cerebral epsilon that provided the exponents are obtained, (11) the range of the cerebral exponent values and their neighborhood distance conditions are used as a filter, (12) the cerebral exponents that meet the condition are taken out and (13) the processes similar to the statistical processes conducted to the conventional Lyapunov exponents including time average value computations are conducted. Then, the time varying of cerebral exponents are visualized similar to the experimental results shown in Figures 1 through 7.

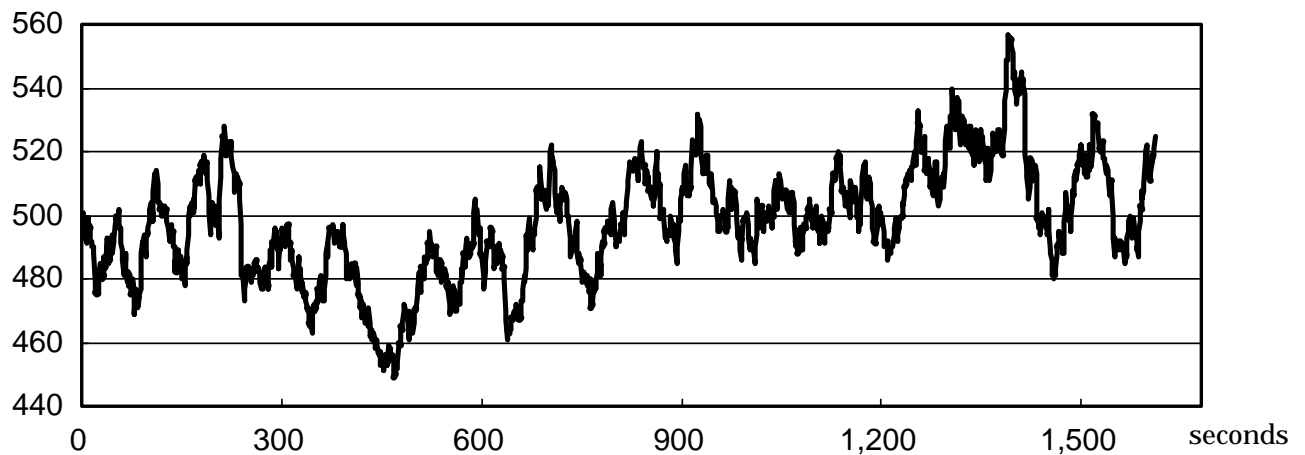
### 3.4 Cerebral exponents

The data which yield the results shown in Figures 5 to 7 are processed by the algorithms described above and the results are shown in Figures 8 to 10 indicating the changes in the cerebral exponents that are the indexes being proposed.

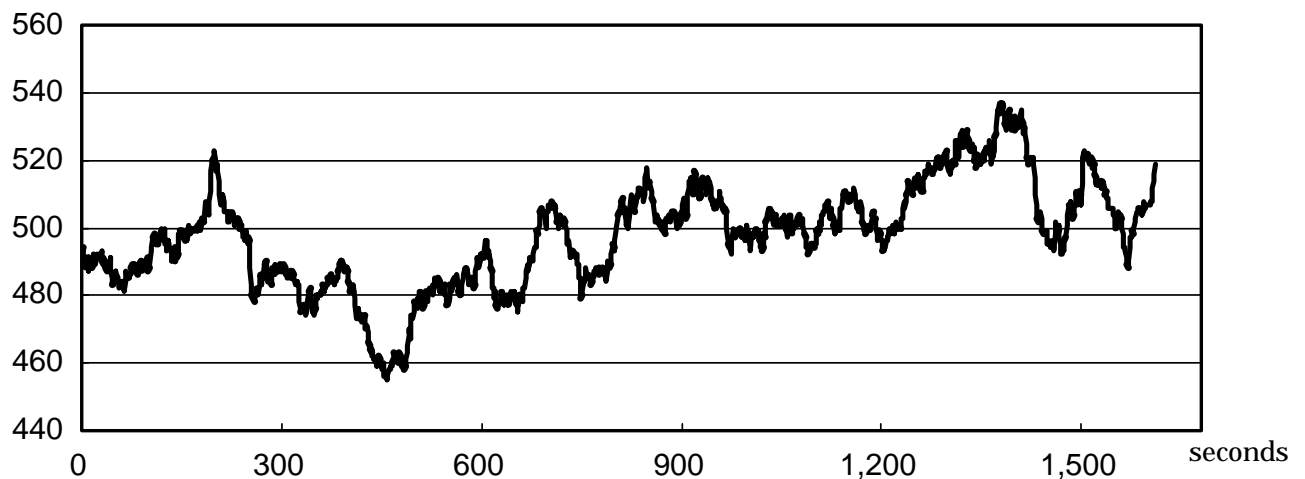
In Figures 8 to 10, the time average widths are 10 seconds, 30 seconds and 1 minute, respectively. In these figures, finer structures are lost as the time average width increases. However, the structure corresponding to the time interval which is longer than a time average width is sufficiently preserved.



**Figure 8** Variations in cerebral exponents in the speech (time averaging width : 10 seconds)



**Figure 9** Variations in cerebral exponents in the speech (time averaging width : 30 seconds)



**Figure 10** Variations in cerebral exponents in the speech (time averaging width : one minute)

We restrain ourselves to make comments on the meanings corresponding to the contents of the uttering for all of the fine structures when the averaging time width is set to 10 seconds.

However, we can say that even though the input timing is shifted for one tenth of a second for the voice data to be processed, no variation is observed for the structures of the graphs.

The structures observed in Figures 8 and 9 are more complex than the structures in Figure 6. However, the structures of the Figures 8 and 9 are much more stable against the changes in the processing parameters.

Thus, we can say that the signal processing algorithms proposed in Section 3.3 “Local time Lyapunov spectrum” are superior than the conventional methods in the process of uttered voice.

Please note that in the above computation of the cerebral exponents, the embedding dimension was set to 4, the embedding delay time was 1ms, the evolution delay time was 1ms, the number of elements of a neighborhood point set including a reference point was 7, the process frequency band was 83 to 250 Hz ( as time interval of a neighborhood point 12 to 4 ms ), the sway band

width was  $\pm 10\%$  with respect to the time interval between the reference point and a first neighborhood point and the continuation condition of the convergence computations is as follows, i.e., “the cerebral epsilon of a next neighborhood point set should be equal to or less than 110% of the cerebral epsilon of the previous neighborhood point set”.

In the discussion given in the previous sections, time averaged values of the first Lyapunov exponents of uttered voice indicate the average degree of activities of the brain of a speaker. In the cerebral exponents, a graph is easily obtained to indicate the tendency of fatigue being accumulated in the person to be examined similar to the case of the first Lyapunov exponents, if a time average width is set to five minutes.

The cerebral exponents have a better sensitivity with respect to a shorter time average width than the case of the first Lyapunov exponents and the cerebral exponents superbly indicate the degree of activities of a brain according to the results of our experiments.

The followings can be observed from the hills and valleys of the graphs which indicate the variations in the cerebral exponents:

- (a) When the speaker was carefully selecting the words for the speech or he was going to mention topics which may cause strong reactions from the general public, the levels of the cerebral exponents are increased and
- (b) When he repeats same subjects frequently or he was talking about subjects for which he could easily predict the reactions from the listener, the levels of the cerebral exponents are reduced or remain at lower levels.

In reality, we will never know the exact thinking of the speaker when he or she makes a speech unless we conduct an experiment such as a performance test of a lie detector.

When a statesman delivers a speech by reading a manuscript which has been examined carefully prior to the speech and he reads the manuscript aloud with his or her special feeling, the judgment made above may have some reliability.

### 3.5 Variations in cerebral exponents during a questions and answers session

Figure 11 is a graph indicating the changes in the cerebral exponents of the uttered voice of a questioner and an answerer in a questions and answers session.

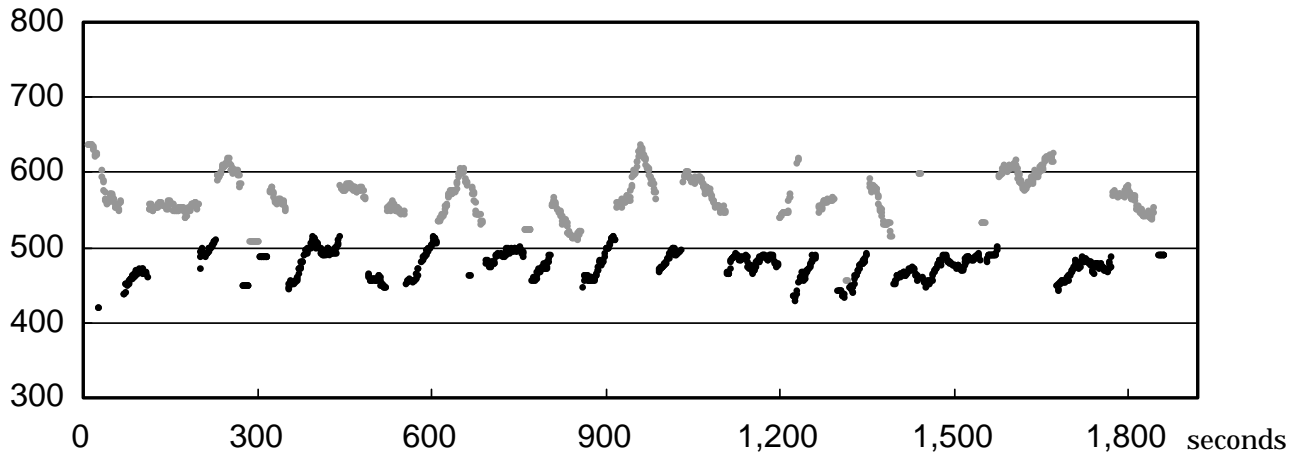
The processed data were taken in the Diet while a member of the Diet of a party out of power was cross-examining a member of the Diet of the party in power. The voices of these members were separated, processed by the proposed algorithms, timely corresponding relationships were given and plotted in the graph.

Please note that the member, who was cross-examining the member of the party in power, was trying to obtain a critical answer from the member of the party in power and therefore, the questioner became tense compared with the answerer. The above situation was the properly reflected in Figure 11.

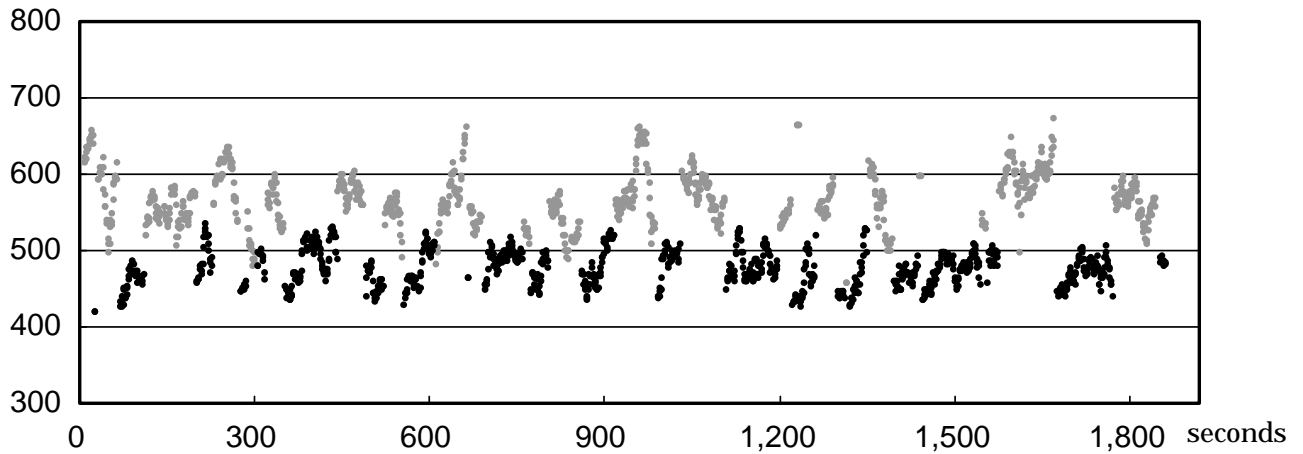
Please also note that in approximately 200 seconds from the beginning of the session, the

answerer was increasing his tension due to the pursuing of the questioner. Moreover, the questioner was increasing his degree of tension in the series of uttering and after changing the topics, his tension was relieved to a certain extent.

As to the answerer, the cerebral exponents were increased from the start of the uttering and he was encountering difficulties because the contents of the questions were different than what he was anticipated prior to the session.



**Figure 11** Changes in cerebral exponents during a questions and answers session (time averaging width : 30 seconds)



**Figure 12** Changes in cerebral exponents during a questions and answers session (time averaging width : 10 seconds)

#### 4. Epilogue

If we can successfully evaluate the mental state of a speaker by analyzing his or her uttered voice and this technology is spread throughout the world, we will be relieved from the necessity of keeping up our appearances even though we told a lie.

A war is the greatest tragedy of humankind. However, the occurrence of a war as a result of mutual distrust can be prevented by a mutual conversation that is free from a lie.

A chaotic uttering voice analysis technology can be considered as a technology to install a tachometer on the brain of a speaker. The proposed signal processing algorithms will greatly enhance the sensitivity of the tachometer compared with the conventional methods.

However, please note that the information obtained by the proposed technology is only related to the level of the work load generated in the brain of a person who is to be examined and uttering and no useful information is obtained unless proper interpretation is made for the values and the changing patterns of the cerebral exponents.

In the example of the questions and answers session above, the interpretation made are reasonable one because the conversations were interpreted based the recorded video.

However, our interpretation is only a plausible one and we know that it is necessary to establish an ethical guideline for the introduction of the proposed technology.

As of this writing, it took approximately two minutes to process voice, which was sampled at 44.1kHz, using a 2GHz clock speed Pentium4.

Therefore, we need to improve the speed of the processing algorithms of the proposed technology and to realize a GUI and a post-processor for a specific application design.