Research Thesis: Thesis

CHAOTIC VOICE ANALYSIS AND MODEL OF CEREBRAL DYNAMICS

Kakuichi SHIOMI

Electronic Navigation Research Institute, 7-42-23 Jindaijihigashi-machi Chofu, Tokyo 182-0012 Japan

Abstract: 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”. 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. 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.

Keywords: *Psychosomatic diagnosis, Voice analysis, Chaos theory*

1. Introduction

In the latter half of the 20th century, a scientific breakthrough has been made in understanding of the basic problems related to the brain. In the 21st century, by understanding each phenomenon related to brain separately, it is expected to clearly show how the brain is functioning as a whole. [Note 1]. In this article, from Chaotic Voice Analysis, which is the result of cerebral functions, the model, which is effective for examining the functional structure of brain and Means of Chaotic Voice Analysis based on chaos analysis study are discussed.

MEG, put in practical use by Coen of MIT in the year 1969, served as a substitute for EEG used since year 1920. MEG shows the degree of utilization of a person’s brain in time resolution of milliseconds and position resolution of less than a few millimeters, while the fMRI and PET used since the year 1980 showed the degree of utilization of the deeper parts of the brain. [Note2].

According to this research using such devices, understanding of brain has rapidly progressed like hardware, but the mechanism, which creates the feelings and consciousness is not at all clear as yet. [Note1].

To develop understanding of the functions of entire brain, to understand the growth of sensitivity in individuals due to the cerebral activities, implementation of software oriented research is required in addition to the above research.

It is considered that existence of multiple languages indicates the software data in the brain and meaningful speech is the outcome of the software functions in the brain. The object of thinking is linguistic, even if not all the information within a brain is linguistic but according to the analysis of speech voice, if it is possible to take out any hardware information of brain, in the research where the brain is taken as whole, hopefully it will be very useful.

By starting the evaluation of speech voice chaoticity, we have started the actual research in the year 1997, to define the possibility of the changes in the psychosomatic state of a speaker. As a result of this, experimental result showing the temporal average value of maximum Lyapunov exponents of speech voice as the exertion status of speaker, was obtained in the year 1999 [Note3].

The author here wants to show that the degree of activity of brain in the speech status can be evaluated from the analysis of speech voice.

In this article, if defined exponent of brain activity is calculated after analysis of speech voice by the techniques using chaos theory, exponent of brain activity is changed in a fixed range, irrespective of the speaker or language.

Compare the speech voice of a relaxed person with the speech voice at the low load environment. The brain activity exponent can be calculated at the higher percent than the speech voice existing in the high load environment.

According to the present author, chaoticity of speechvoice is described, chaotic method used in the analysis of speech voice is explained, after that brain function model and evaluation principles of brain activity are described and the type of brain activity exponent are described on the basis of experiment result and result of reading experiment carried out in the experiment room.

2. Chaotic property of the speech voice

Chaos analysis of the speech voice was very popular in the 90s. Then practical use of the synthesis of emotional voice or assumption of psychosomatic state was expected [Note 4]. However, this was very difficult for the analysis of common speech voice. Therefore no remarkable result was reported by the analysis of simple vowels.

Figure 1 and 2 are the Takensu plot where the “Aa” and “O” sounds are embedded between two phases in same conditions (Embedded dimension is 4 and embedded delay time is about 1 ms). Takensu plot is provided with strange attractor as “aa” and “o” sounds have chaoticity and show different forms clearly.

This is a system with different physical quantity of 2 sounds. In other words, chaoticity is genereated by different systems. The human speech organs have many modes that differ chaotically and it is considered that different sounds are generated by changing the modes.

Even though the Chaos theory so far indicates the complex chaoticity of the generated signal, the prerequisite is that the system generating sound should be chaotically stable and cannot be applicable while changing the chaoticity of the system generating the time series signals that is the target of the process. In other words, simple Chaos theory is not applicable when the signals of many modes differing chaotically are mixed or for the system, which is in transitional state changing from one mode to another.

While applying the chaoticity means to the analysis of common speech voice, some supplementary processes such as processing without the stable portion etc. become necessary and the means for normalizing the chaoticity index value to be calculated from many modes also becomes necessary at the same time. Moreover, it is well known that the human speech voice has the chaotic properties and hence chaoticity cannot be simply considered as “the same item having ideally chaotic time series signal generated from ideal mathmatical model”.

According to the author, the speech voice is the signal of real world, which has the randomness representing the white noise and in addition to this “Chaotic voice signal does not ideally exist where there is no randomness.”

3. Means of chaoticity of time series signals analysis

While analysing time series signals, guidelines such as maximum Lyapunav exponents / Lyapunav spectrum, KS entropy and auto correlation coefficient are genreally used.

Any index value is the value given by convergent calculation, while calulating the numerical value for the sample data. Regarding the algorithm to get this index value, the length of the data should be sufficient for timely convergent calculation in any chaotically stable system [Note 5].

In above mentioned index, maximum Lyapunov exponents are often used since they are comparitively easy to calculate and it is the value, which has been used comparitively frequently as it is considered valid by chaoticity confirmation. Many algorithms are proposed for calculation. Especially, the Sano-Sawada algorithm [Note 6] is considered as “The parameter that must depend on trial and error & can be comparitively set easily” and according to the author it is applicable to the voice signal process. Further, Sano-Sawada algorithm is used for calculating Lyapunov spectrum and the maximum Lyapunov exponents are same as first Lyapunov exponent, which has the maximum elements of Lyapunov spectrum.

Sequential outline to calculate Lyapunov spectrum by using Sano-Sawada algorithm for identifying the difference with SiCECA algorithm is as follows.

(1) Set the embedding dimensions, embedding delay time and expansion delay time

(2) Decide the adjacent condition (adjacent super spehere diameter) and clustered adjacent points are found out from the embedding points created by time series data

(3) Calculate each expansion point included in clustered adjacent points

(4) Calculate the initial value of Lyapunov spectrum from the clustered adjacent points and clustered expansion points.

(5) Create clustered adjacent points for next covergence calculation from clustered expansion points and calculate the convergence value of Lyapunov spectrum carrying out a number of convergence calculations set in advance.

It is necessary to set the adjacent super sphere diameter properly

Figure 1: Takensu plot of sound “a”
(Strange attractor of sound “a”)

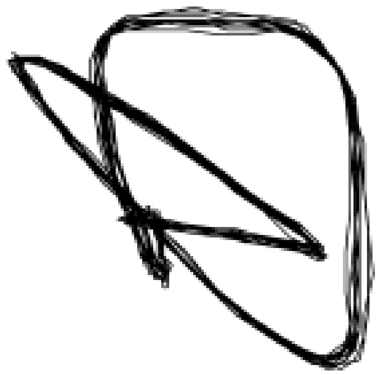


Figure 2: Takensu plot of sound “o”
(Strange attractor of sound “o”)

for calculating the exact Lyapunov spectrum by Sano-Sawada algorithm. If the set diameter is too small, Lyapunov spectrum cannot be calculated due to non-existence of sufficient elements for adjacent points. If it is too large, in the clustered adjacent points, totally different phases are included in the orbit of embedding space and the value of maximum Lyapunov exponents will reduce. Though the algorithm is properly set, the correct maximum Lyapunov exponents may not be calculated.

When it is possible to divide the time series with the sufficient time for the application of Sano-Sawada algorithm from common speech voice, the timely change of that chaoticity can be observed by maximum Lyapunov exponents regarding general speech. However, the sufficient time for evaluating chaoticity of time series depends on the signal ranking by signal quality, intensity of the overlapped randomness etc and it should be tested.

While calculating the maximum Lyapunov exponents from the speech vowel, if the signal with length more than 100 msec can be divided, the value denying the existence of chaoticity in the speech voice is not calculated. In other words, maximum Lyapunov exponent is a finite positive value. The above mentioned length of the signal is not always sufficient, but at present it is considered as sufficient length for the assessment of chaoticity existence.

However, even if the stable signal of 100 msec obtained from continuously uttering “aa” or “i” can be divided, it is not easy to cut the signal of few msec, which seems to be stable, obtained from general speech voice. With respect to the speech signal obtained from the subject reading a Japanese text, eight to ten vowels can be read in 1 sec, on an average the duration to read each vowel is 100msec, however, waveform of the portion with lot of consonants occurring continuously or the portion different from the previous vowels, may be changed in a complicated manner. The time at which stable signals can be obtained is only few seconds as shown in previous Fig.1 and Fig. 2 indicating Takensu plot.

Though it is a usual “aa” sound, Takensu plot shows considerable difference in the beginning, middle and at the end of the signal. The calculated maximum Lyapunov exponent value is changed by more than 1 digit even in case of first 100msecs signals.

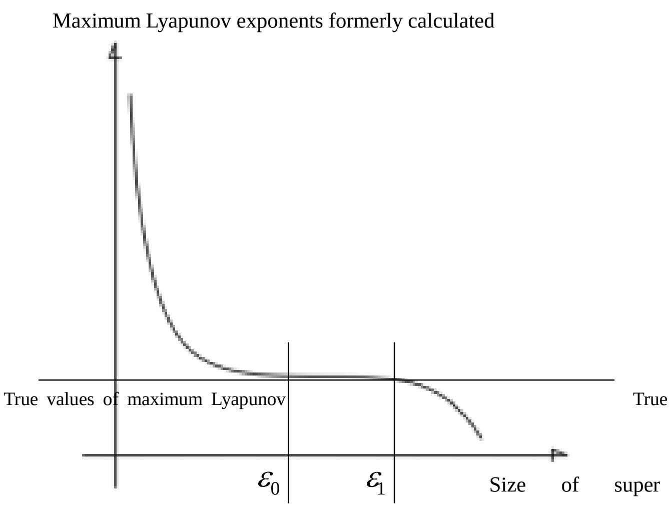
Though, conventionally Sano-Sawada algorithm is appropriate in the processing of general speech voice, it is difficult to calculate meaningful maximum Lyapunov exponents, because the prerequisites are not adequately fulfilled.

We have not considered the “Maximum Lyapunov exponents defined in the present chaos theory” as an evaluation index of speech voice chaoticity and have used the following features of algorithm, which calculates the maximum Lyapunov exponents. We have fixed the relation between chaoticity of time series signal derived from the previously calculated maximum Lyapunov exponents and randomness disturbing this chaoticity and we have considered the value obtained from it as index value, and observed the timely changes in the chaoticity of voice speech.

Regardless of the Sano-Sawada algorithm used for the calculation of maximum Lyapunov exponents, the points embedded in the space, which configure the strange attractor displayed as “Takensu Plot”, are considered as ‘reference points’ and super sphere diameter set in advance is applied as ‘adjacency condition’ and a cluster of adjacent points including reference points is created.

When time series contains chaoticity, the element points configuring reference points are scattered in a specific direction as the time progresses. When it is included in the super sphere with cluster of adjacent points, then the cluster of expansion points after the unit time of expansion from the element points, gives spheroid extended on one side, which is not included in the super sphere.

Moreover, when above maximum Lyapunov exponents are



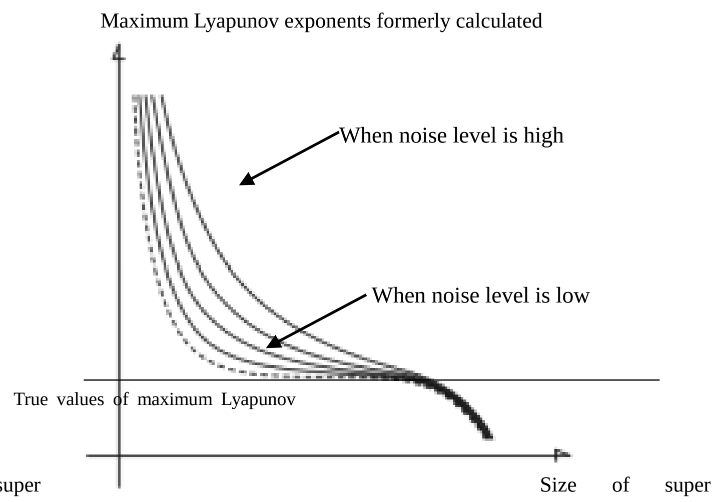


Fig.3: Maximum Lyapunov exponents as a function of the size of super sphere Fig.4: Maximum Lyapunov exponents as a

calculated from the time series signal by using Sano-Sawada algorithm, it is observed that the calculated maximum Lyapunov exponents change as shown in figure 3 depending on the size of super sphere diameter set as adjacency condition.

Moreover, here, the super sphere diameter is the ratio for super sphere diameter excluding the attractor.

When maximum Lyapunov exponents are to be found in the analysis of time series signal obtained experimentally, the relation between the above mentioned super sphere diameter and the calculated maximum Lyapunov exponents is based on the conclusion that "It is very important to decide the size of super sphere diameter properly".

In present situation,the maximum Lyapunov exponent value is considered as True value when there is a minimum change in maximum Lyapunov exponent value, corresponding to change in super sphere diameter. However, as the absolute value of maximum Lyapunov exponents is not very useful in case of experimental data, comparatively small super sphere diameter is set in most of the cases/examples.

When white noise is superposed in time series having chaoticity, the following points can be made clear by numerical simulation related to the severity of white noise (i.e. randomness) disturbing the chaoticity of time series.

According to Sano-Sawada algorithm, when maximum Lyapunov exponents are calculated same as earlier, from time series in which white noise is superposed, the change in the maximum value of Lyapunov exponents calculated for the size of super sphere diameter as per the change in severity of white noise is indicated in Figure 4 [Note 7].

When the randomness of time series is comparatively strong and when super sphere diameter is small, calculated maximum Lyapunov exponent value is greater as compared to the value in case of weak randomness. This property is comparatively strong in time series with chaoticity and it can also be observed by comparing the data size with the required data size and processing it for less than few minutes for acquring fixed point convergence value, while calculating maximum Lyapunov exponent value.

Therefore, even though it is not sufficient for caculating constant maximum Lyapunov exponent value, the speech voice series is divided into partial processing unit and if it is possible to measure the intensity of randomness in these processing units, the change with time in speech voice chaoticity can be observed in common speech series.

And, if it is considered that this index value is showing changes to cope with some psychosomatic state of the uttered person, it is possible to evaluate the psychosomatic state of the uttered person and the changes in it can be measured quantitatively by this index value.

Figure 5 and Figure 6 show the result of processed reading voice and result of analyzed reading voice of the same subject (Man in his 50’s) for ten seconds. Figure 5 shows the result of the voice of the uttered person who read the general printed document and Figure 6 shows the result of the voice of the same uttered person who read the same document by closing his right ear with his right hand.

Reading voice with 16 bit/22.05 kHz is taken as sample, dimension of embedding phase space is taken as 4 and both embedded delay time as well as expansion time is taken as approx.1.0 ms (22 samples/clock).

It is necessary that the super sphere diameter for evaluating the randomness corresponding to chaoticity is small. However, it is taken as 10% of the outward diameter of Strange Attractor formed during the experiment because if the super sphere diameter is too small, it becomes impossible to create the cluster of the adjacent points.

All respective sample data are considered as starting point (reference point), cluster of adjacent points is created from the data for the time in which there will be 6 rounds of attractor and then Sano-Sawada algorithm is applied conventionally and conventional maximum Lyapunov exponents are acquired. As for the data area

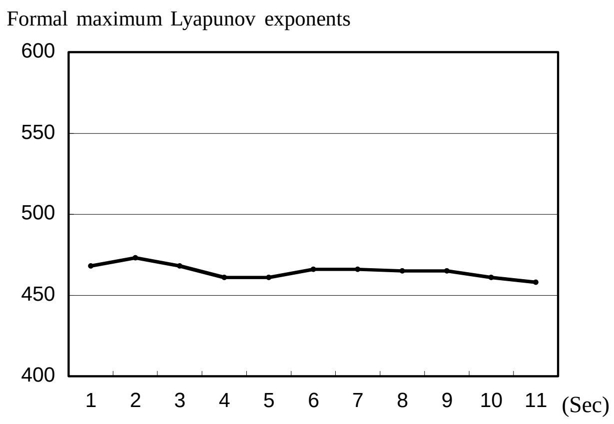


Figure5: Change in average value of formal maximum Lyapunov exponents (In case of usual reading)

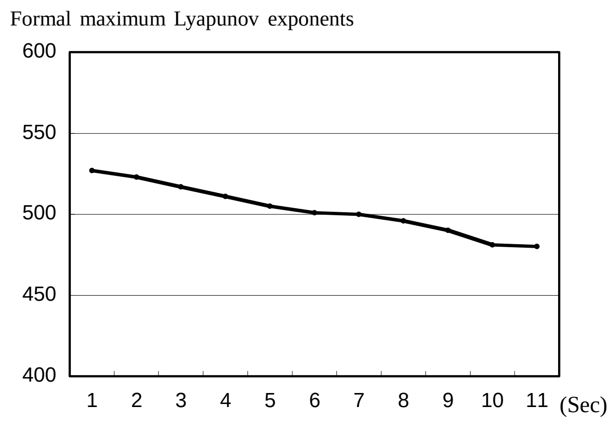


Figure 6: Change in average value of formal maximum Lyapunov Exponent (In case of reading by closing one ear)

with approximately fixed pitch frequency of chaotic voice, point convergence calculation is performed similar to the process as per Sano-Sawada algorithm and conventional maximum Lyapunov exponents are calculated for entire sampling period.

According to the above process, in the experiment by this subject, it is observed that the noise level that disturbs the chaoticity of the voice read by closing one ear is high.

4. Principle underlying the evaluation of model of cerebral dynamics and cerebral activity

Wernicke’s and Broca’s areas in brain, play an important role in human speech and it is considered that some information is exchanged between both the fields during the speech. Moreover, it is also considered that brain processes the input by five senses, thinks by using the memory, also processes various information and there is exchange of information between various areas of brain and other areas.

Morevoer, it is known that each brain part is acquiring the functional difference, however, even though it is possible to observe these differences dynamically by fMRI, the anatomical static differences could not be observed and hence it is considered that brain is highly homogeneous extensive system.

As for the information exchange, it is considered that structure in which each part is shielded or structure with shielded information transmission circuit connecting each part, does not exist and it is assumed that “Various signals are exchanged but there is possibility of some interference by other signals”. Especially regarding speech, it is considered that the signals that are exchanged between other parts interfere with the signals that are exchanged between Wernicke’s and Broca’s areas and signals that are exchanged in the internal part of respective area.

It is possible to investigate the function that is realised by acquired self-organization in the cerebrum by the model given in Figure 7. These functions are considered to be realised physically by the combination of synapse and the type can be observed depending on the PGA Software for the function realized in programmable gate array (PGA) of physical cerebrum.

The authors have considered “leakage of information” has an essential meaning in the model shown in figure 7.

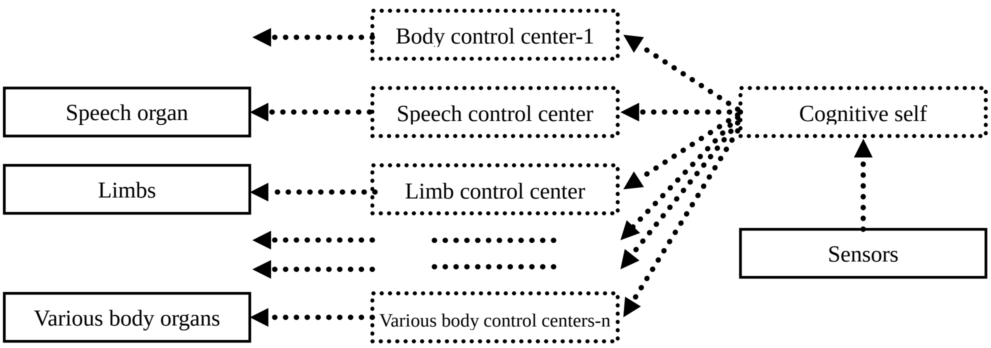
In the design of PGA, gate isolation is a major problem and although present PGA can be used only under the conditions, in which interference does not occur between gates, it is considered that the problem is solved for the cerebrum by PGA in everyday life with the advanced redundancy.

The comparison between usual reading voice and reading voice by closing one ear is that by closing one ear gives an unusual environment to the speaker and by doing this, load may occur in the cerebrum.

For the previous experiment result when the above-mentioned plan is executed for all subjects, it is not observed that the noise level goes up by comparing the usual reading voice with reading voice by closing one ear and by closing both the ears. However, by the above-mentioned method and when load is added by making one's voice feed back with a head set again etc., it is observed that there is a rise of the exponent value to some dergree in 80% of the subjects. When there is a limit on the rise of noise level of reading voice by “closing one ear or both the ears”, it is observed that in 43 subjects out of 47 subjects (91.5%), there is rise of a few percent up to 10% percent in the “average value during 10 seconds of the exponent value.”

In addition, the reading contents used in the above-mentioned experiment were chosen, varying from the official message to the interest of the subject. Moreover, when the subject is man / woman in 20’s to 60's, who came to the Electronic Navigation Research Institute with some purpose, such as business, etc., read by relaxing as much as possible although there was a possibility of biased hearing by them.

In the brain there are different active centers engaged in different activities. Those active centers generate their characteristic signals and exchange the signals with other centers. In such a situation, a signal transmitted from one center to another may be affected by a third signal passing nearby, or conversely the signal may affect the third signal passing nearby. If the occurrence of such crosstalks can be accepted, it will be possible to “monitor the general background



The dotted lines of function boxes and arrows indicate there are leaks of information through those boundaries.

Fig.7: Cerebral dynamics model obtained from the chaos analysis of speech voice

activity of the brain by analyzing a speech voice signal." It was believed that there was enough possibility and the author considers that he can show this possibility by the experiment mentioned above.

When a person makes a speech, although it may not necessarily be Wernicke and Broca's field, his/her "congitive self"governing the brain will send signals to "speech control centers," etc.

In the brain, the "speech control center" is not shielded from active centers such as "limb control centers", "body control centers," etc. Therefore, one part of signals delivered by the "congitive self" to the "speech control center" will affect, more or less like crosstalks in a cable network, the activities of "limb control centers" and "body control centers". Conversely, if signals are delivered by the "Cognitive self" to the "limb control centers" and "body control centers", one part of the signals will affect, more or less like crosstalks in a cable network, the activity of the "speech control center".

The "cognitive self" receives signals from "sensors" to process them, and generqtes signals while it is "thinking something". Those signals and internal signals also have crosstalks with the "speech control center".

Thus, it will be impossible for the "speech control signal" to receive only the signals directed solely to the "speech control signal", as long as "the person keeps on living as a normal human beign. Because the "speech control center" is exposed to such crosstalks at all times, the "speech organ" will produce sounds permanently accompanied with a noise component.

The stronger such cross talks imposed on "speech control center" become, the higher the level of noises accompanying vocal sounds becomes. The increase in this noise level can be determined by the earlier mentioned measurement technique of noise level that disturbs the chaoticity in a chaotic signal.

If the degree of activation of the brain is high when the amount of information exchanged in the brain is abundant, the degree of cerebral activity can be fixed by fixing the chaos included in the chaotic voice.

5. Result of reading experiment

The result of the reading experiment in case of normal reading

Figure 8: Overview of reading experiment

for about 10 to 30 seconds and when one ear or both the ears are closed is as mentioned earlier. A head set is used, and the result of an experiment is shown below when chaotic voice is given as feedback.

As seen in the photo, in Figure 8, the subject wore a head-set, and his/her reading voice was recorded with a digital audio tape-recorder (16bit/44.1kHz or 48.0kHz). Later, the data recorder in DAT was fed as digital signals to a personal computer for storage and analysis.

In the beginning no voice from head set was taken as feedback. Later on, the chaotic voice of the subject with or without special effects like echo or delay was taken as feedback.

Figure 9 is the experimental result of males in their 50's. After reading for about an hour without having voice feedback in between, feedback was taken finally after continous reading of 1 hour. No special sound effects were added to the voice feedback, and only air

condtioning sound was eliminated by digital filter. Further, consonant of strong sound of "sa, shi, su se, so" were made "gentle to the ears" by multi band compressor thereby causing a delay of about 1ms in the actual measurements.

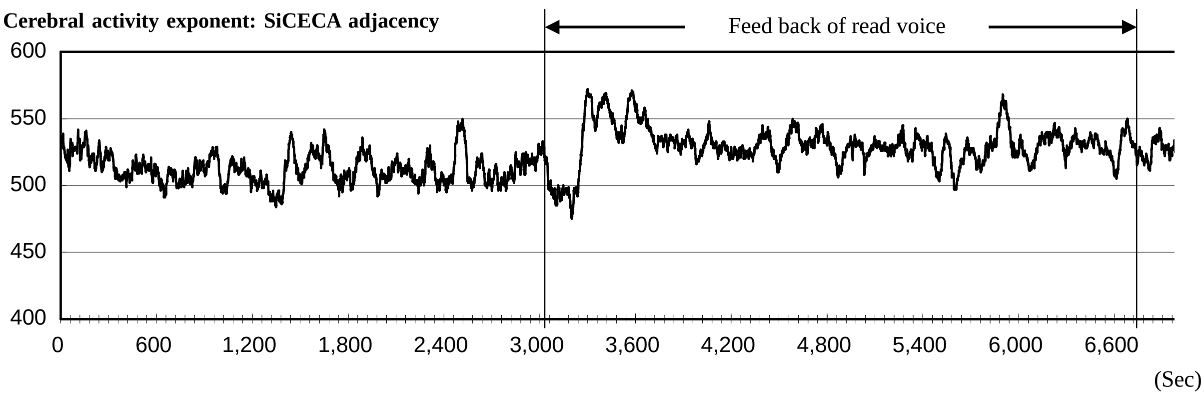


Figure 9: Change in cerebral exponent of read voice during load experiment (Average time span is 30 seconds)

It is understood from the experimental results that the noise level affecting chaoticity including speech voice increased on an average as per the feedback of speech voice. The average value of exponent indicating noise level increased from about 510 to 530.

After completing reading, the subject commented that he did not feel sleepy even though he was tired. He considerably felt sleepy before the sound feedback and awoke with the starting of the feedback.

The experimental result shown in Figure 9 shows the processed items by algorithm “SiCECA: Shiomi’s Cerebral Exponent Calculation Algorithm” that corrected Sano-sawada algorithm (Details are mentioned later. Cerebral exponent, that is the exponent on Y axis, is also mentioned later.).

The feedback of speech voice is taken with the aim of increasing the load on the brain by inculcating the information forcibly through ears. The experimental result whose hidden motive is considered successful is in conformity with the model of human cerebral dynamics and principle underlying the evaluation of brain activity.

In order to impose more load on the brain, reading voice of the subject is fed back with a delay of about 200ms in other experiment, and it is considered effective when further delayed voice is taken as feedback. However, this technique cannot be used commonly by anyone. For many people, continuous speech is difficult. It is reported that the subject who read and listened to delayed voice for about 1 hour could not sleep on the night of the experiment. If the system that can sufficiently respond to the viewpoint of medical field is not introduced, it is clear that the experiment cannot be carried out where more load is imposed on brain as mentioned above. From now on, a system is necessary so that the experiment could be carried out easily.

The experiment result shown in Figure 9 is based on reading experiment of whole 2 hours and the result same as the experiment for several minutes is confirmed

Now on, experiment protocol is decided and is to be implemented continuously and systematically.

6. SiCECA

The Brain activity exponents are defined below and the algorithm SiCECA (Shiomi's Cerebral Exponents Calculation Algorithm) for computing them are developed on the principle of periodicity of chaotic voice. The problems in case of conventional algorithm are eliminated as far as possible and increase in the processing speed is aimed at.

6.1. Problems observed in case of conventional algorithm

Conventionally, in chaos analysis signal processing, when attempting to compute maximum Lyapunov exponents from time series data, they are matched with the processing target and standards of the data to be processed and the algorithms by Kantz and Rosenstein, and Sano sawada algorithm in case of Lyapunov spectrum computation, are used with temporal selection. Although these operation algorithms have their respective unique characteristics, they share following common traits: In order that they work properly, they presuppose that the dynamics of the responsible sytem are stable during the production of time series, which is to be studied, and they determine maximum Lyapunov exponents, or Lyapunov spectrum via convergence calculation [Note 5].

As for the speech voice data, in order to fully use this convergence calculation by the conventional algorithm as presumed, at least several times more data as compared to the continuation time of vowels in common speech, is required. Further, in case of the cluster of adjacent points, the computation of maximum Lyapunov exponents is done using more than several dozens of adjacent points, as in the published application example of conventional algorithm. However, from the data for continuation time of vowels that can be cut from general speech voice, considering that each vowel has sufficient time interval in the orbit in embedded space, it is impossible to extract adjacent points covering several dozen points.

For the time being, if we assume that the vowels require continuation time of 100ms, during that time, the waves of these vowels repeat for a very little number of times. Besides, there is also occurrence of transitional waves when connecting with other phonemes and the adjacent points with the size, as of the conventional algorithm example, cannot be created. It is of course possible to increase the apparent size of data by increasing the rate at which data is sampled from the speech voice series. However, the number of "similar” waves recurrently appearing during the emission of a phoneme is invariable regardless of the sampling rate. If it is required to obtain the maximum Lyapunov exponents from a time series with a clear periodicity, increase of the sampling rate will lead to the increase of adjacent points at a temporal point close to a start point, that is, to the increase of adjacent points disposed along the orbits around an attractor, and thus simple increase of the sampling rate will not result in the improved precision of the chaos analysis of the time series. The problem will not be resolved even if autocorrelation is employed instead of the Lyapunov exponent as a means of evaluating the chaoticity of a time series, because autocorrelation does not include any temporal definition enabling one to resolve unstable dynamics.

As per the conventional techniques, it is impossible to evaluate unstable dynamics or transitional dynamics.

As in case of the actual speech voice, in case of time series signal, whose ideal waves are not known and can only be derived experimentally, in order to evaluate the chaotic stability and instability of each signal from the processing result by numerical calculation, it is necessary to find an index that can be calculated stably for a data size shorter than that assumed in conventional algorithm. If it is possible to define and find such index, local periodical chaotic evaluation becomes possible in shorter time than in the past as a yardstick for the data to which this definition can be applied.

6.2. Calculation of Cerebral activity exponent as per SiCECA

"Cerebral activity exponent" is the measure of size of "Randomness disturbing that chaoticity" included in the signal with chaoticity. Anybody can suggest the appropriate name but we have used this name.

As mentioned earlier in "4. Principle underlying the evaluation of model of crebral dynamics and cerebral activity", cerebral activity exponent can be used in each evaluation of speech voice, as the evaluation of "Randomness disturbing that chaoticity"in the signal with chaoticity, is possible with small amount of data than using stable calculation of maximum Lyapunov exponents.

The cerebral activity exponent is the value calculated from the correlation between the noise level in the time series to be processed and the adjacent super sphere diameter, which has to be set in the calculation of maximum Lyapunov exponents, in the calculation result by the formal application of algorithm for calculating the maximum Lyapunov exponents.

For example, adjacent super sphere diameter is calculated by considering 1% for strange attractor excluding super sphere diameter and then it is calculated in the same way considering it as 2 %. If you want to calculate till 100% and if sufficient chaoticity is saved in the time series to be processed and when the super sphere diameter is considered to be above n %, you can find n indicating features same as in case of the temporary change in maximum Lyapunov exponents (cerebral activity exponent) to be calculated conventionally.

In case of a generated time series of a system having a stable dynamics same as in the conventional chaos theory considered as target or in case of a signal overlapped by some noise in the same time series signal, even though the super sphere diameter is increased to any size, it may be possible to generate clustred adjacent points so that changes over a period of time in the cerebral activity exponent value to be calculated may be equal. But when only small processing unit can be started from common speech voice or when the characteristics of overlapped noise are not very clear, it is necessary to change the super sphere diameter and to confirm that the evaluation of noise level is carried out stably.

Accordingly, computational complexity increases drastically and in the analysis of the common utterence voice, it is very necessary to find the super sphere diameter n for which a stable calculation is performed by changing the super sphere diameter to various sizes.

While calculating as mentioned above by the algorithm of Sano-Sawada, only for the changed super sphere diameters, if you increase the resolution of super sphere diameter n for which the same time series signal is to be processed, the calculation complexity to be processed increases suddenly.

In analyzing a speech voice signal, SiCECA presupposes that the adjacent points move in synchrony with the periodic waves of the speech signal, based on the fact that the speech voice signal consists of distinct periodc waves. By doing so, SiCECA narrows its search for appropriate adjacent points and determines Lyapunov spectra at a finely resolved time scale (to be referred to as cerebral spectrum hereinafter). The microscopic cerebral exponent is the largest element of the Lyapunov spectrum and will be abbreviated as the micro- cerebral exponent hereafter.

While using the algorithm of Sano-Sawada, when the adjacent super sphere diameter is changed, the same calculation should be repeated every time for the identical time series signals, in SiCECA, (c_m, ε_S) can be obtained by one time process, only for all the clustered adjacent points generated from the time series signals as they are generated as per the previous process. In other words, super sphere diameter (ε_S: SiCECA adjacency distance) excluding this is calculated for clustered adjacent points. If the process result is filtered as per the ε_S value, the data necessary for evaluation of noise level at the time when clustered adjacent points are generated in time series signal, can be obtained. In SiCECA, as per above mentioned computational complexity technique, while processing the same time series repetitively by the algorithm of Sano-Sawada, the signal processing speed of 2~3 digits is comparitively improved.

Then in SiCECA, macroscopic cerebral activity exponent is calculated by performing statistical compution process in the micro cerebral activity (c_M : mentioned below as Macro cerebral activity exponent.). Sensitivity evaluation corresponding to the ranking of voice signal to be processed can be realised by adjusting the parameters connecting the processes for calculating micro cerebral activity exponent and macro cerebral activity exponent.

Temporary change in cerebral activity can be observed by transition time average transaction, corresponding to the necessary time resolution.

When maximum Lyapunov exponents are considered as index as per the conventional techniques, time resolution cannot be realized even in 5 minutes but it can be done within 10 seconds in SiCECA. Specifically as per the following procedure, time series is processed in the SiCECA.

[Microscopic cerebral activity exponent (c_m)]

(1) Setting the embedding dimension D, delay time of embedding τ_d, and delay time of expansion τ_e ;

(2) Determining the number of adjacent points N, and implementing the shortest period T_m and longest period T_M as the periodically condition necessary for processing the signal ;

(3) Finding a cluster of adjacent points P from the embedding points obtained from the time series data $P = \{P_1, P_2, ..., P_N\}$;

(4) Determining a cluster of expansion points corresponding to the cluster of adjacent points

(5) Calculating a cerebral spectrum $c_m(t)$ from the cluster of adjacent points and expansion points; and

(6) Executing a convergence computation as is practiced in said determination of the Lyapunov spectrum, if it is possible to generate a cluster of adjacent points that can satisfy the periodicity and proper connection with the next step of the computation.

Incidently, the embedding points $P_1, P_2, ..., P_N$ constituting the cluster of adjacent points in paragraph (3) are temporally spaced from each otherat a period T which is ($T_m \le T \le T_M$).

SiCECA adjacency distance ε_S as the super sphere diameter exluding the cluster of adjacent points, is calculated as per the super sphere diameter excluding strange attractor being determined by time series starting from time t_1 when the embeding point P_1 comes into being, till the time when P_N comes into being, that is, $t_1 + (N-1) \times T + (D-1) \times \tau_d$, or till the time when the expansion point of P_N comes into being, that is $t_1 + (N-1) \times T + (D-1) \times \tau_d + \tau_e$. The proper connection with the next step of the computation mentioned in paragraph (6) may include a condition that "the SiCECA adjacency distance defined with respect to clustered adjacent points obtained from the foregoing cluster of expansion point is equal to or smaller than the SiCECA distance of the clustered adjacent points responsible for the generation of said cluster of expansion points."

For a given speech time series, SiCECA can "generate, for any given sampling time t of the series, a cluster of adjacent points moving from an embedding point that takes data present at time t as the first component, and, if there is a cluster of expansion points corresponding to the cluster of adjacent points, "calculate the following formula (1) for time t.

$$\left(c_m(t), \varepsilon_S(t), T(t)\right) \tag{1}$$

In treating a speech time series $s(t)$, SiCECA produces the following formula (2) via micro cerebral activity exponent calculation process.

$$\underline{CEm}(t) = \left\{\left(c_m(t), \varepsilon_S(t), T(t)\right) \middle| t = 1, 2, ...\right\} \tag{2}$$

[Macroscopic cerebral activity exponent (c_M)]

SiCECA calculates macro-cerebral exponent c_M by performing the following process for micro-cerebral activity exponent c_m obtained as a result of the micro-cerebral activity exponent calculation process.

In a continuous series of speech voice, each of the phonemes constituting the speech, consist of waves having practically constant period T over its duration. In view of this, the following operation is introduced: for $\underline{CEm}(t)$, find a period $[t_0, t_1]$ where $T(t)$ is stable to produce $\underline{CEm}(t | t_0 \le t \le t_1)$, arrange the obtained elements in ascending order of size of $\varepsilon_S(t)$ to produce the following formula (3).

$$\underline{CEm}(i | 1 \le i \le n) = \left\{\left(c_m(i), \varepsilon_S(i), T(i)\right) \middle| 1 \le i \le n, \right.$$
$$\left. \varepsilon_S(i) \le \varepsilon_S(i+1)\right\} \tag{3}$$

When SiCECA adjacency p ($0 < p \le 1$) is introduced as a parameter related to the SiCECA adjacent distance, c_M can be provided by the following equation (4).

$$c_M^p(t | t_0 \le t \le t_1) = \frac{1}{i_{\varepsilon(p)}} \sum_{i=1}^{i_{\varepsilon(p)}} c_m(i) \tag{4}$$

where $i_{\varepsilon(p)}$ is an index to signify $n \times p$'th micro-cerebral activity exponent or $c_m(i_{\varepsilon(p)})$ out of $\underline{CEm}(i | 1 \le i \le n)$ which are arranged in ascending order of $\varepsilon_S(t)$.

The author introduced the process that obtains $\underline{CEm}(i)$ from $\underline{CEm}(t)$ as mentioned above since the intensity of fluctuations of signal is large, thereby complete signal and noise ratio cannot always be secured by speech voice processing and the following formula (5)

$$c_M^p(t | t_0 \le t \le t_1) = \frac{1}{N\left(c_m(t | t_0 \le t \le t_1)\right)} \sum_{t=t_0}^{t_1} c_m(t) \tag{5}$$

in which macro cerebral activity exponent c_M is purely considered as average value c_m cannot be calculated by number of c_m in time $t_0 \le t \le t_1$ for $N\left(c_m(t | t_0 \le t \le t_1)\right)$.

As compared to the change in psychosomatic state of the speaker, changes (Changes in types of phoneme) in speech voice dynamics occur frequently in a short time. This model is introduced assuming that variation in chaoticity of speech voice during the time span from 1 second to less than 10 seconds, is known immediately even though the time stability in minutes is somewhat lost by this process.

Only positive or negative exponents are not allowed in the calculation of maximum Lyapunov exponents by the conventional algorithm aiming at chaotic evaluation of the time series signals to be processed. When comparison of fixed chaoticity is aimed at, various experimental methods are introduced for obtaining stable calculation results since chaotic evaluation of stable system during comparatively longer time than speed voice is aimed at as in case of evaluation of tip pulse waves in biological signal.

For example, in the Sano-Sawada algorithm, in place of "Adjacency clustering is done by taking out only points decided in advance in the order close to the time for the basic point when clustering fulfills the adjacency condition", "Adjacency clustering is done by taking out only points decided in advance in the order from the points closer to the basic points when clustering fulfills the adjacency condition" is performed but such adjacency points clustering technique is not more than an experience for calculating stable maximum Lyapunov exponents comparatively easily.

In the evaluation of chaoticity by the conventional technique, only the process result by similar method can be compared and the translation of process result between different techniques is left as future subject so far.

Therefore, when it is necessary to compare ‘SiCECA process result’ with the result in which algorithm of Sano-Sawada was applied; the author is unable to understand the meaning of comparison and it is possible to obtain macro cerebral activity exponent c_M by formula (5), by simply taking c_m as average value instead of obtaining $\underline{CEm}(i)$ from $\underline{CEm}(t)$.

At present in Chaos signal process, when time series signal is extracted from real world, even if same algorithm is to be applied there are many parts that should be decided, like method of creating cluster of adjacent points, setting of embedding parameter etc. experimentally or practically and at present the author cannot declare the necessity of process of obtaining $\underline{CEm}(i)$ from $\underline{CEm}(t)$. The detail argument regarding simulation result of this point will be conceded in the separate article.

At least in chaotic voice, if macro cerebral activity exponent c_M calculated for the period where periodicity is stable (in above case $t_0 \le t \le t_1$), is to be displayed in relation with SiCECA adjacent ratio p, and it can be displayed as shown in graph in Figure 4 “The relation between size of adjacent diameter and maximum Lyapunov exponents calculated conventionally when noise is superposed in chaotic time series signal”.

If speech signal of higher rank is taken for getting better experimental results, the process results are same as can be obtained by using formula (5).

Moreover, when temporal change in continuous speech voice is observed as change in hourly moving average value of macro cerebral activity exponent mentioned above for more than one minute, remarkable difference in graph trend calculated by Formula (4) or Formula (5) is not found.

7. General properties of analysis result of speech voice and cerebral activity exponent

Analysis of speech voice with SiCECA was carried out for more than 100 sounds. Speech voice includes various things, and it may not always be possible to arrive at some conclustion from analysis of 100 sounds, however, the author obtained the following results from the experiment.

Further, speech voice is processed by SiCECA, after setting the following parameters as base.

(1) Dimension of embedding, 4 ;
(2) Embedding delay time, 1ms ;
(3) Expansion delay time 1ms ;
(4) Number of elements contained in a cluster of adjacent points including a reference point, 7 ;
(5) Frequency band studied, 83~250Hz ;
(6) Frequency band of fluctuations is determined to be ±10% of the temporal interval between the reference point and the first adjacent point ; and
(7) Connection condition for convergence computation includes that “The SiCECA adjacent distance of a cluster of adjacent point is equal to or smaller than the 110% of SiCECA adjacent distance of the previous cluster of adjacent points.”

Further, macro cerebral activity exponents are calculated by Formula (4). As a result of this, for the processing result of speech voice series for more than ten senconds,

(8) When embedding dimension is increased to 5, 6, calculated cerebral activity exponent value also increases. However, remarkable change was not seen in the graph obtained.
(9) Even if embedding delay time and expansion delay time was changed within 0.8~1.2 ms, remarkable change was not found in the calculated value or the graph obtained.
(10) Even though size of cluster of adjacent points was changed to 6 or 8, no remarkable change was found in the calculated value or the graph obtained. However, when the size was changed above 20 or 30 of common settings as per the application of conventional algorithm, the graph became flat.
(11) When other parameters were changed by ±10%, remarkable change was not seen in the calculated value or graph obtained.

When calculating micro-cerebral activity exponents of a continuous series of speech voice, SiCECA uses every sampling time of the speech series signal as a ‘starting point’ of the calculation. Therefore, as far as SiCECA is concerned, a micro-cerebral activity exponent is uniquely determined for each sample fraction of the signal, and there will be no alternative micro-cerebral activity exponents that may appear as a result of sampling implemented with different timings. This also applies to the determination of macro-cerebral activity exponents which are computed based on micro-cerebral activity exponents. True, when properly calculating macro-cerebral activity exponents, it is necessary to choose processing units having an appropriate temporal length. However, it was demonstrated experimentally for a speech voice signal that, as long as macroscopic exponents are averaged over the time width of over 10 seconds, the difference that otherwise might result depending on sampling procedures can be safely canceled out, particularly when the processing unit is sufficiently short as companred with the duration of each phenome. Therefore, the graph showing the temporal course of moving averages of macro-cerebral exponenets remains practically invariable as long as the same time width and same stepwise shift are used for averaging, regardless of the sampling procedures actually practiced.

Thus, in evaluating the psychosomatic state including fatigue of a speaker, based on his/her speech voice signal, SiCECA will be able to provide a far more stable and sensitive result than does the previous method principally aimed at "setting the processing unit time width and calcutating the average psychosomatic state of a speaker based on the first Lyapunov exponent value for each processing unit".

For the analysis of speech voice recorded in general office environment, when SiCECA adjacency ratio is less than few percentage points, the shape of the graph for showing output greatly changes and so it is considered that stable calculation will not be carried out. When the adjacency ratio is more than 10%, and since the cerebral activity exponent value decreases with the increase in adjacency ratio, the shape of the graph that shows changes will be almost constant.

Absolute value depends on the applicable hourly standards, same as the maximum Lyapunov exponent value. In the standards set by the author, when recorded frequency band of speech voice is between 20Hz~20kHz, the cerebral activity exponent value of speech voice flctuates between 350~700. If the minimum value of recorded frequency band of speech voice increases, the cerebral activity exponent value to be calculated reduces comparitively.

From all the process results, following points were made clear for the cerebral activity exponent.

(12) For all the process results, the fluctuations in the value are within the restriced range but the difference between minimum and maximum value cannot be obtained.

(13) The range of cerebral activity exponent fluctuations do not change much for a Japnese or for a foreigner and the fluctuations are within the same range for one's mother tongue.

(14) For the range of cerebral activity exponent fluctuations, the difference between age and gender is not seen.

The details of the above mentioned experiment results will be presented in a separate article.

8. Conclusion

At the beginning of the 1990s, Tawara, Tauda and others reported that it is possible to discover the cardiovascular or cerebral disorders of patients by applying chaos analysis to the bloodflow of those patients. Sence then, various biological signals have been analyzed by chaos analysis techniques (see note 8).

Some authors insisted that the dynamics of the systemresponsible for speech are so unstable that applying chaos analysis to a speech voice signal would be useless. However, in 1998, the present authors continuously traced the chaos measures of a speech voice signal, and found that the moving averages of first Lyapunov exponents started to increase before the speaker became conscious of his/her exertion (Note 3). Moreover, from the results of the experiments performed by author, today it can be said that brain work raises the maximum Lyapunov exponent of a speech voice, and the "Cerebral activity exponent" shown in this paper, [notes 3 and 8].

Presently, we cannot say anything about the cerebral activity exponent, whether it really indicates the cerebral activity or not and whether this exponent can determine the fatigue and stress.

However, it is an obvious fact that, in many people, the cerebral activity exponent rises if spoken by closing the ears with finger.

As a next step of this research, the author wants to obtain the observation data from MEG and fMRI when spoken by closing ears with finger.

If it is possible to check the correspondence between the fluctuations in the exponent value of the speech voice and fluctuations in the cerebral blood flow and even if the correlation between both the fluctuations is not checked, a new cerebral activity model can be built and this research can be continued.

According to the Old Testament, Cain, or a son born between Adam and Eve, killed his brother Abel, and became the first killer in human history. It is said, the Lord knew Cain's crime from the voice of the soil soaked with the blood of Abel. In addition to the voice from the soil, the Lord would have also noticed noises in the answering voice of Cain which would disturb the chaosity of the voice.

We can not hear a voice from the soil that tells us a truth in a convincing manner, nor can we distinguish noises from the chaosity in a speech voice signal solely dependent upon our auditory system. However, because high performance microphones and computers are now available to us, it is possible to evaluate a noise component contained in a speech voice signal that interferes with the chaosity of the signal.

At present, though there are still some insufficiencies in all the aspects, such as restrictions or temporary resolution in usage environment, the author thinks that clues for realizing "a system that can evaluate psychosomatic state of the speaker as well as the stresses imposed on a speaker, by analyzing a speech voice signal from the speaker." are obtained at last.

Notes

[1] S. Greenfield, "BRAIN STORY", BBC, 2000.

[2] http://www.ctf.com/home_j.html

[3] Shiomi and Hirose: "Trial for detecting sleepiness and fatigue of a speaker from his/her voice." presented at 37th Airplane Symposium, 1999.

[4] Aihara: Guide of chaotic theory, Teaching material of University of the air, 2001.

[5] http://www.mpipks-dresden.mpg.de/~tisean/

[6] Sano, M. and Sawada, Y.: Measurement of the Lyapunov Spectrum from a Chaotic Series, Physical Review Letters, 55, 10, 1082-1085, 1985.

[7] T. Ikeguchi and K. Aihara: Lyapunov Spectral Analysis on Random Data, International Journal of Bifurcation and Chaos, 7, 6, 1267-1282, 1997.

[8] http://www.toshiba.co.jp/magazine/sci_talk/cont98/caos/index_j.htm
[9] Takaoka, Shiomi, et al., "On the possibility of diagnosing the psychosomatic state of a speaker from his/her spoken voice," General Health Check, Vol.29, No. 3, 2002.