A Study of the Nature of Sentence-Based Listening

KAWASHIMA Hirokatsu

要約

文レベルのリスニングの特徴を理解するための調査を日本人英語学習者を対象に行った。文レベルのリスニングとして、パーシャルディクテーションと文に含まれる音声 / 文法 / 意味上の間違いを認識するリスニングタスクを用い、それぞれのパフォーマンス及び総合的リスニング能力との関係を調べた。調査の結果、１）意味上の間違いを認識するリスニングタスクのパフォーマンスが相対的に高い、２）パーシャルディクテーションは総合的リスニング能力に最も深く関与している可能性が高い、３）音声 / 意味上の間違いを認識するリスニングタスクと総合的リスニング能力の間には非線形的な関係が認められる、４）総合的リスニング能力の 43％は、パーシャルディクテーションの能力と音声 / 意味上の間違いを認識する能力で説明することができる、こと等が明らかにされている。

Background

Rost (2011: p. 204) claims that "in the area of listening, assessment is particularly important because receiving adequate feedback is essential for increasing the learner's confidence and for designing instruction that addresses learners' apparent weaknesses – or the weaknesses in the curriculum." It must be noted, however, that it is not yet clear exactly what listening sub-skills, which are indispensable for feedback-based assessment, are composed of and what their interrelationships should be like, and consequently that the nature of listening assessment has not been fully understood, especially at the classroom diagnostic level.

In order to improve this research situation, the author launched a research project in 2007 and has made attempts to elucidate the nature of listening sub-skills, in which a series of empirical studies have been conducted. Kawashima (2009), who deals with sentence-based partial dictation as a listening sub-skill, reports, for example, that partial dictation in the middle constituent of the sentence is most significantly related to general listening proficiency ($r = .55, p < .01, n = 79$). As far as Kawashima (2010), in which sentence-based error recognition is defined as a listening sub-skill, is concerned, it is reported, for example, 1) that phonetic and semantic error recognition competence are related with statistic significance to general listening proficiency, 2) that the strength of the relationship of semantic error recognition competence with general listening proficiency ($r = .53, p < .01, n = 45$) is greater than that of phonetic error recognition with general listening proficiency ($r = .41, p < .01, n = 45$), and 3) that the combination of sentence-based phonetic and semantic error recognition competence generates the highest predictive power with general listening proficiency ($R2' = .34, p = .00, n = 45$).

The results of these two studies focusing upon sentence-based listening can be considered to have made some contribution to the elucidation of the nature of listening sub-skills. It must be pointed out, however, that much remains unclear and unexplored regarding the nature of sentence-based listening itself, although it can be explicitly stated that automatic sentence-based listening plays an important role in successful listening. Unfortunately, the predictive power of the combination of the competences of sentence-based partial dictation and sentence-based error recognition with general listening proficiency is unknown, for example. In order to explore such a predictive power, investigations must be carried out paying more comprehensive attention to sentence-based partial dictation and error recognition in the same research context.

Current Study

A review of the literature of studies of sentence-based listening shows that few investigations have been carried out at least with Japanese learners of English from the perspective of sentence-based partial dictation and error recognition in the same research context. The current study, which is also a follow-up to one of the two investigations reviewed above, aims at offering a more comprehensive map of the nature of sentence-based listening from such a perspective.

Its research designs, following those of the two previous studies, are summarized below, and then some of the major findings are reported and discussed.

1. Research Designs

1.1 Research Questions

The current study involves four main research questions:

1) What type of sentence-based listening (cf., 1.2 below) is the most/least difficult in performance?

2) What type of sentence-based listening is the most/least related in performance to comprehension-based general listening proficiency?

3) Are there any non-linear relationships between sentence-based listening and comprehension-based general listening proficiency? If so, how large are they?

4) What is the best combination of types of sentence-based listening that may account for the variance of comprehension-based general listening proficiency?

1.2 Types of Sentence-Based Listening

In order to measure the performance of sentence-based listening, four types of sentence-based listening were targeted:

1) partial dictation, in which uttered sentences are processed and understood, and the designated blanks in them are filled in

2) phonetic error recognition, in which uttered sentences are processed and understood, and it is reported whether they include phonetic errors or not

3) grammatical error recognition, in which uttered sentences are processed and understood, and it is reported whether they include grammatical errors or not

4) semantic error recognition, in which uttered sentences are processed and understood, and it is reported whether they include semantic errors or not

1.3 General Listening Proficiency

In order to measure the performance of general listening proficiency, the listening sections of standardized objective tests (*STEP Grade 2*), widely and frequently administered in Japan, were used.

1.4 Subjects

42 first-year students of the general education course at a university in Japan participated in the current investigation.

1.5 Materials

1.5.1 Sentence-Based Partial Dictation

As the materials for measuring sentence-based partial dictation (referred to as SPD hereinafter) performance, 12 fifteen-word sentences were prepared, which involved three dictation blanks in each of the right/middle/left parts of the sentence. The following SPD is an example:

SPD) (Misaki) (is) (planning) to study the (history) (of) (England) which is full (of) (interesting) (stories).

All the sentences were recorded onto CD at a self-selected normal speaking rate by a male native speaker of English.

1.5.2 Sentence-Based Error Recognition

As the materials for measuring sentence-based phonetic/grammatical/semantic error recognition (referred to as PER, GER and SER hereinafter) performance, 20 ten-word and 20 fifteen-word sentences were prepared, respectively.

The following PER1 and PER2 are examples of sentence-based phonetic error recognition tasks, in which the subjects listen to ten-word short sentences and judge if they are phonetically correct:

PER1) The center of the city is far away from here. (phonetically correct)
PER2) I don’t sink this Saturday is good to go out. (phonetically incorrect)

The following GER1 and GER2 are examples of sentence-based grammatical error recognition tasks, in which the subjects listen to ten-word short sentences and judge if they are grammatically correct:

GER1) The student said she was interested in international volunteer work. (grammatically correct)
GER2) This song is very popular between girls in Asian countries. (grammatically incorrect)

The following SER1 and SER2 are examples of sentence-based semantic error recognition tasks, in which the subjects listen to ten-word short sentences and judge if they are semantically correct:

SER1) I started going to school when my parents became rich. (semantically correct)
SER2) Kaori enjoyed swimming yesterday, because she had a bad cold. (semantically incorrect)

All the sentences, 120 in total, were recorded onto CD at a self-selected normal speaking rate by the male native speaker of English.

1.5.3 General Listening Proficiency

As the materials for measuring general listening proficiency (referred to as GLP hereinafter), two sets of listening sections of *STEP Grade 2* tests were used, which had been originally designed to match the level of high school graduates in general. The tests were administered in June 21 and October 8, 1998 and June 18, 2000. Each set had 20 four-option multiple-choice test items, and 60 test items were used in total.

1.6 Data Collection and Procedure

The investigation was carried out during regular English classes, which lasted from the beginning of April to the end of July in 2008. The main goal of these classes was to improve the subjects’ overall listening proficiency.

1.6.1 Measuring SPD

The subjects took four SPD tests as part of their regular classroom listening activities, in which they carried out sentence-based partial dictation tasks and checked their weaknesses in listening to English. About 10 minutes were allocated for each class, after which the subjects immediately checked their answers on the distributed investigation sheets and checked their SPD performance.

1.6.2 Measuring PER/GER/SER

The subjects took eight error recognition performance tests as part of their regular classroom listening activities, in which they carried out sentence-based phonetic/grammatical/semantic error recognition tasks and checked their weaknesses in listening to English. Prior to measuring their error recognition performance, which started in the middle of May in 2008, the subjects had received general instructions about error recognition performance tests as classroom listening activities and taken three similar practice tests so that they could get accustomed to their forms and contents. These three practice tests were considered to serve to increase the subjects' awareness of the significance of grasping their error recognition performances and to make them fully prepared for the investigation and its procedures.

During each of the first four classes, the subjects listened to 5 ten-word and 3 ten-word sentences for each of the sentence-based phonetic/grammatical/semantic error recognition tasks. Likewise, during each of the next four classes, the subjects listened to 5 fifteen-word and 3 fifteen-word sentences for each of the sentence-based phonetic/grammatical/semantic error recognition tasks. About 20 minutes were allocated for each class, after which the distributed investigation sheets were collected, and then the subjects immediately checked with their sub-investigation sheet if their answers were correct and checked their PER/GER/SER performance.

1.6.3 Measuring GLP

The subjects took three GLP tests at certain intervals in order to examine their general listening proficiency and monitor their progress periodically: in the middle of April, at the beginning of June, and at the end of July. About 20 minutes were allocated for each test, after which the distributed computer-scored investigation sheets were collected, and then the subjects immediately checked with their sub-investigation sheet if their answers were correct and grasped their general listening proficiency by the totaled score.

2. Scoring and Processing of the Data

All the investigation sheets were collected after each of the classes was over, and then the raw data were scored, examined, and processed for analysis.

2.1 Scoring

2.1.1 Scoring SPD

First, the correctness of each dictation item on the investigation sheets was carefully checked with the same item scores (0, 1), and then the total score of each of the three sentence constituent parts of the fifteen-word sentences (right/middle/left) was calculated with the item scores (0, 1, 2, 3). If the blanks are filled by a subject as in the following, for example, he/she gets six points in total [2 (right) + 3 (middle) + 1 (left) = 6]:

Example Subject's Answers:
(Misaki) (is) (?) to study the (history) (of) (England) which is full (?) (interesting) (?).

Correct Answers:
(Misaki) (is) (planning) to study the (history) (of) (England) which is full (of) (interesting) (stories).

Slight spelling mistakes were expected to be made (e.g., planing), but the current study did not regard them as incorrect answers unless they would cause serious semantic confusion.

2.1.2 Scoring PER/GER/SER

As far as the sentence-based phonetic/grammatical/semantic error recognition tasks were concerned, the correctness of each test item on the investigation sheets was carefully checked by the author, in which the correctness of each test item was provided with the item scores (0, 1) representing correct and incorrect answers, respectively.

2.1.3 Scoring GLP

With regard to the GLP tests, the computer-scored investigation sheets were read and processed by an optical mark reader (*SR-3500*, Sekonic) and a mark reader computer software (*SS kun II*, Software for Education), in which the correctness of each test item was provided with the item scores (0, 1) representing correct and incorrect answers, respectively.

2.2 Examining Internal Consistency Reliability

The scored data were then examined in terms of internal consistency reliability using the *Cronbach Alpha* coefficient[1)]. First, the internal consistency reliability coefficient of the three GLP tests (the total number of test items is 60) and those of SPD/PER/GER/SER tests (the total number of test items is 36, 40, 40, and 40, respectively) were measured. Table 1 presents their results:

Table 1: Internal Consistency Reliability by *Cronbach Alpha* Coefficient (Original Test Items)

	GLP	SPD	PER	GER	SER
Number of Test Items	60	36	40	40	40
Cronbach Alpha Coefficient	.76	.84	.29	.02	.60

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition

As is quite obvious from this table, the internal consistency reliability of the performance of sentence-based phonetic/grammatical/semantic error recognition is not high. A number of reasons are deemed to lie behind this, but it must be noted that the small number of subjects and the inappropriateness of some of the sentence-based phonetic/grammatical/semantic error recognition tasks seem most likely to have caused this kind of poor internal consistency reliability. Higher internal consistency reliability may be obtained with a greater number of subjects, but since the number of subjects is uncontrollable after the investigation, the current study made some attempts to raise the internal consistency reliability of sentence-based phonetic/grammatical/semantic error recognition tasks by paying careful attention to each of the test items used and reconsidering what should constitute those tasks.

It is generally assumed that *Cronbach Alpha* coefficient should exceed at least .7 for reliable analysis, so the current study has expunged a number of “unsuitable” test items from each test item list of sentence-based phonetic/grammatical/semantic error recognition tasks so that *Cronbach Alpha* coefficients might get as closer to .7 as possible. Table 2 presents the results of measuring the internal consistency reliability coefficients of sentence-based phonetic/grammatical/semantic error recognition tasks whose original test items have been restructured:

Table 2: Internal Consistency Reliability by *Cronbach Alpha* Coefficient (Restructured Test Items)

	GLP	SPD	PER	GER	SER
Number of Test Items	60	36	20	20	20
Cronbach Alpha Coefficient	.76	.84	.72	.62	.75

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition

It is found from this table 1) that a great number of test items have been removed from the sentence-based phonetic/grammatical/semantic error recognition tasks, and consequently 2) that their internal consistency reliability has been greatly increased as a whole, although the *Cronbach Alpha* coefficient of the semantic error recognition performance (.62) did not exceed .70. With regard to the first point, the current study assumes that although the degree at which the original test items were reduced may be viewed as quite large, downgrading the construct validity, the number of the reduced test items, 20 for each of the sentence-based phonetic/grammatical/semantic error recognition tasks, should stay within an acceptable range of construct validity while admitting that it may not be the best. As far as the second point is concerned, the current study more or less takes Dörnyei (2007: p.207)'s view that "somewhat lower *Cronbach Alpha* coefficients are to be expected" owing to "the complexity of the second language acquisition process", and assumes that although the *Cronbach Alpha* coefficient of the sentence-based grammatical error recognition test (.62) may fall below .70, it can be used to a certain degree while paying attention to its limits in terms of internal consistency reliability.

2.3 Examining Normal Distribution

Lastly, the restructured data of the SPD, PER/GER/SER and GLP tests was examined in terms of normal distribution, upon which the statistical analyses of the current study are based. *Shapiro-Wiki* tests, whose α value had been set at .01, were conducted for this examination[2]. Table 3 presents the results:

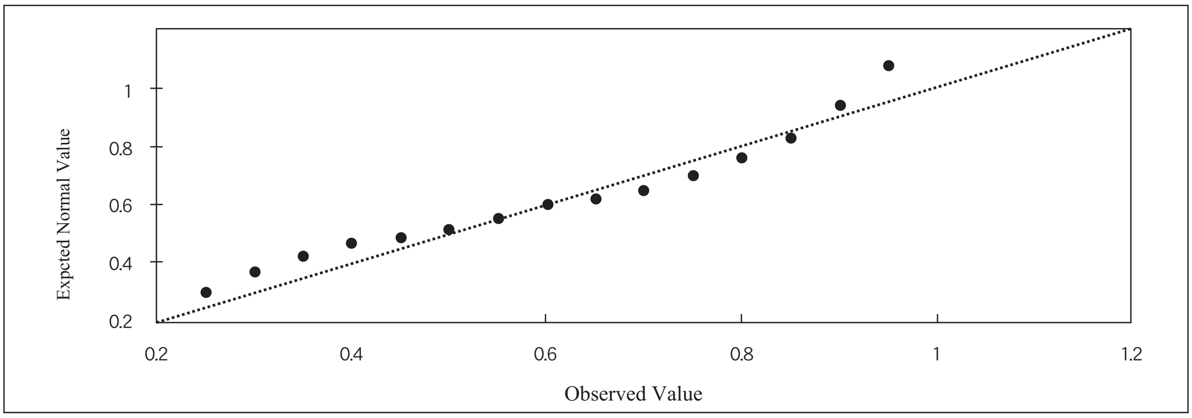
Table 3: Normal Distribution of the Restructured Data of the SPD/PER/GER/SER/GLP Tests

	GLP	SPD	PER	GER	SER
W	.98	.95	.98	.97	.92
p-value	.59	.09	.52	.34	.01

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition α= .01

It is statistically found from this table that much of the restructured data is normally distributed, in which each p-value is greater than .01, except for that of the SER test. A purely statistical stance may claim automatically that since p-value is equal to .01, the restructured data of the SER test should not be normally distributed, but it must be noted that the difference between both α-value and p-value was quite small and that careful attention is needed before becoming conclusive. The current study examined normal distribution of this restructured data by looking at the following Normal Q-Q Plot:

Figure 1: Normal Q-Q PLOT for the Data of the SER Test



The straight line presents what the data would look like if it were perfectly normally distributed, and the actual data is represented by the small squares plotted along this line. The closer the small squares are located to the line, the more normally distributed the data is plotted. The current study assumes that most of the small squares fall along the line though not perfectly, and that the restructured data of the SER test as a whole is not greatly deviant from normal distribution.

3 Data Analysis

The pre-examined data above were then processed for analysis[3)].

3.1 Descriptive Statistics

First, the minimum, maximum, mean score and standard deviation of the processed data of each of the GLP/SPD/PER/GER/SER Tests were calculated by the percentage. The results are presented in Table 3:

Table 3: Descriptive Statistics for the GLP/SPD/PER/GER/SER Tests

	GLP	SPD	PER	GER	SER
Min.	.24	.34	.25	.25	.25
Max.	.76	.80	1.00	.90	.95
Mean	.49	.55	.64	.54	.69
S.D.	.12	.12	.18	.17	.18

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition

This table shows that the subjects' performance may vary across the four types of sentence-based listening, and that it may be the lowest and highest in sentence-based grammatical and semantic error recognition (mean: .54 and .69, respectively).

3.2 Performance of Sentence-Based Listening

In order to verify the above tentative results with statistic significance, the current study first examined the data employing GLM Repeated Measures. The results are presented in Table 4 and Table 5:

Table 4: *Mauchly Test of Sphericity*

					Epsilon		
Within-Subject Effect	*Mauchly*'W	Approximateative Chi-Square	df	P -Value	Greenhouse -Geisser	Huynh -Feldt	Infimum
Types of SBL	.76	10.95	5	.05	.88	.94	.33

SBL: sentence-based listening

Table 5: Test of Within-Subject Effect

		Sums of Squares: Type III	df	Mean Square	F-Value	P-Value
Type	Assumption of Sphericity	.64	3	.21	9.24	.00
	Greenhouse-Geisser	.64	2.63	.24	9.24	.00
	Huynh-Feldt	.64	2.82	.23	9.24	.00
	Infimum	.64	1.00	.64	9.24	.00
Error	Assumption of Sphericity	2.84	123	.02		
	Greenhouse-Geisser	2.84	107.63	.03		
	Huynh-Feldt	2.84	115.66	.03		
	Infimum	2.84	41.00	.07		

As Table 4 shows, the assumption of sphericity is valid for data analysis ($p = .05$). Table 5, based upon this statistical validity, claims that the subjects' performance of the SPD/PER/GER/SER tests varies significantly with the four types of sentence-based listening [($F(3, 123) = 9.24$, $p = .00$)].

Multiple comparisons by *Bonferroni*, which could be used for repeated measured data, were then conducted in order to examine whether the subjects' performance of the semantic error recognition test was the highest among the four types of sentence-based listening. The results are presented in Table 6:

Table 6: Multiple Comparisons by *Bonferroni* for the Data of the SPD/PER/GER/SER Tests

Combination	Difference	MSe	P-Value
SPD-PER	-.09	.03	.04
-GER	.01	.03	1.00
-SER	-.14	.03	.00
PER-SPD	.09	.03	.04
-GER	.10	.04	.09
-SER	-.05	.04	.88
GER-SPD	-.01	.03	1.00
-PER	-.10	.04	.09
-SER	-.15	.04	.00
SER-SPD	.14	.03	.00
-PER	.05	.04	.88
-GER	.15	.04	.00

SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition

This table shows 1) that although sentence-based semantic error recognition is equal in performance to sentence-based phonetic error recognition (MSe = .04, P = .88), there is a statistically significant difference in performance between sentence-based semantic error recognition and either sentence-based grammatical error recognition or partial dictation (MSe = .04, P = .00 and MSe = .03, P = .00, respectively), and consequently 2) that the subjects' performance of sentence-based semantic error recognition is relatively quite high.

There are other differences and similarities in performance among the four types of sentence-based listening like this finding, such as the closeness in performance between sentence-based partial dictation and grammatical error recognition (MSe = .03, P = 1.00). Figure 1, based upon the above multiple comparisons, illustrates the results on a horizontal line:

Figure 1: Performance of the Four Types of Sentence-Based Listening

(GER = PER)

Low GER = SPD < PER = SER High

(SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition)

3.3 Linear Relationships between General Listening Proficiency and Sentence-Based Listening Competence

In order to explore linear relationships between general listening proficiency and sentence-based listening competence, their simple linear correlation coefficients were first computed. Table 7 presents the results:

Table 7: Simple Correlation Matrix (Relationships among Performances of the GLP/SPD/PER/GER/SER Tests)

	GLP	SPD	PER	GER	SER
GLP	1	.58**	.34*	.05	.23
SPD		1	.24	.38*	.27
PER			1	.07	.24
GER				1	.07
GER					1

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition * < .05 ** < .01

This table shows 1) that some types of sentence-based listening competence are related with statistic significance to general listening proficiency: namely, sentence-based partial dictation and phonetic error recognition competence, and 2) that the strength of the relationship of sentence-based partial dictation with general listening proficiency (r = .58 , < .01) is greater than that of phonetic error recognition competence (r =.34 , < .05).

It must be noted, however, that these relationships may be superficial and unstable, because it is possible that they are mere reflections of some unrevealed third relationships (known as pseudo correlations), subsistent in the nature of sentence-based listening competence. In order to closely examine these relationships between general listening proficiency and sentence-based listening competence, partial regression analysis was conducted, in which partial regression coefficients of direct relationships between two variables were computed, thereby eliminating the influences of the rest. Table 8 presents the results:

Table 8: Partial Correlation Matrix (Relationships among Performances of the GLP/SPD/PER/GER/SER Tests)

	GLP	SPD	PER	GER	SER
GLP	1	.57**	.24	-.23	.03
SPD		1	.01	.43**	.17
PER			1	.04	.17
GER				1	-.03
GER					1

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition * < .05 ** < .01

This table shows that while sentence-based error recognition competence is not related, sentence-based partial dictation competence is related with statistic significance to general listening proficiency (r = .57 , < .01). This result may need careful consideration, but considering the figures in Table 8 and the nature of statistical significance in a broad sense, the relative strength of relationships in performance between general listening proficiency and sentence-based listening competence might be delineated on a horizontal line as in Figure 2:

Figure 2: Relative Strength of Relationships between GLP and SPD/PER/GER/SER competence

Weak GLP-GER = GLP-SER = GLP-PER < GLP-SPD Strong

(GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition)

3.4 Non-Linear Relationships between General Listening Proficiency and Sentence-Based Listening Competence

The simple/partial linear regression analysis in the above 3.3 has been conducted to determine if there is a statistically significant relationship between general listening proficiency and sentence-based listening competence, and has shown its strength. It must be pointed out, however, that even if such a relationship is not recognizable, it does not directly mean that there is no relationship between the two variables, because the analysis presupposes that relationships between general listening proficiency and sentence-based listening competence are linear, and therefore because it may fail to recognize non-linear relationships, such as quadratic ones. In order to better understand relationships between general listening proficiency and sentence-based listening competence, they must be viewed also from the perspective of non-linearity. There are a number of ways to explore the non-linearity of relationships between two variables. The current study, following Kawashima (2010), has focused upon exploring quadratic and cubic relationships between general listening proficiency and sentence-based listening competence.

3.4.1 Quadratic and Cubic Relationships

First, quadratic relationships were examined between general listening proficiency and sentence-based listening competence, and their multiple coefficients and coefficients of determination calculated. The results are shown in Table 9:

Table 9: Multiple Coefficients and Coefficients of Determination in Quadratic Relationships between General Listening Proficiency and Sentence-Based Listening Competence

		SPD	PER	GER	SER
GLP	R'	.57	.31	.08	.13
	$R^{2'}$	.33	.09	.01	.02
	p-value	.00	.06	.33	.28

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition R': multiple coefficient adjusted for the degree of freedom R2': coefficient of determination adjusted for the degree of freedom

This table shows that a quadratic relationship is found only between general listening proficiency and sentence-based partial dictation (R'= .57, p = .00).

Next, examinations were likewise conducted in order to know if there were cubic relationships between general listening proficiency and sentence-based listening competence, in which their multiple coefficients and coefficients of determination were calculated. The results are shown in Table 10:

Table 10: Multiple Coefficients and Coefficients of Determination in Cubic Relationships between General Listening Proficiency and Sentence-Based Listening Competence

		SPD	PER	GER	SER
GLP	R'	.56	.37	-	.43
	$R^{2'}$	.32	.13	-	.18
	p-value	.00	.04	.46	.01

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition GER: grammatical error recognition SER: semantic error recognition R': multiple coefficient adjusted for the degree of freedom R2': coefficient of determination adjusted for the degree of freedom

This table shows that although they are not very strong, cubic relationships are also recognizable between general listening proficiency and sentence-based partial dictation and phonetic/semantic error recognition competence (R’= .56, p = .00 , R’= .37, p = .04, and R’= .43, p = .01, respectively).

Careful attention must be paid in order to determine which statistically significant non-linearity may represent in a more accurate and comprehensive manner the relationships between general listening proficiency and sentence-based listening competence. The current study, paying attention to the distribution of the actual data, assumes as a first step 1) that the strength of coefficient of determination adjusted for the degree of freedom may be employed for such a purpose, and 2) that a higher coefficient of determination adjusted for the degree of freedom indicates that the relationships between general listening proficiency and sentence-based listening competence can be represented better by its non-linearity.

It follows from this assumption 1) that no quadratic or cubic relations are found between general listening proficiency and sentence-based grammatical error recognition competence, 2) that relationships between general listening proficiency and sentence-based partial dictation are more quadratic (33% of the variance of GLP can be explained) than cubic (31% of the variance of GLP can be explained), and 3) that relationships between general listening proficiency and sentence-based error recognition competence are more cubic (13% and 18% of the variance of GLP can be accounted for by PER and SER, respectively) than quadratic (no variance of GLP can be accounted for by either PER or SER).

3.4.2 Directness of Non-Linear Relationships

As has been stated above in 3.3, these non-linear relationships may be superficial, unstable and mere reflections of some unrevealed third relationships. In order to confirm that they are intrinsic and stable so that the nature of relationships between general listening proficiency and sentence-based listening may be understood more comprehensively, partial regression analysis was conducted, for which the linear data for sentence-based partial dictation was used because it was able to explain the variance of general listening proficiency (34%) better than the above quadratic data (33%), and for which the above non-linear cubic data was transferred into linear data. Table 11 presents the results:

Table 11: Partial Correlation Matrix (Relationships among Performances of the GLP/SPD/PER/GER/ SER Tests with Linear Transformation

	GLP	SPD	PER(C)	GER	SER(C)
GLP	1	.43**	.32*	-.16	.29
SPD		1	.11	.43**	.26
PER(C)			1	-.01	-.05
GER				1	-.04
SER(C)					1

GLP: general listening proficiency SPD: sentence-based partial dictation PER(C): phonetic error recognition (the cubic data was transformed into a linear one) GER: grammatical error recognition SER(C): semantic error recognition (the cubic data was transformed into a linear one) * < .05 ** < .01

This table shows 1) that sentence-based partial dictation and phonetic error recognition are related with statistic significance to general listening proficiency, and 2) that the strength of the relationship of sentence-based partial dictation competence with general listening proficiency (r = .43, < .01) is greater than that of phonetic error recognition competence (r = .32, < .05). Based upon these partial regression results and the nature of statistical significance in a broad sense, the relative strength of relationships in performance between general listening proficiency and sentence-based listening competence may be delineated on a horizontal line as in Figure 3:

Figure 3: Relative Strength of Relationships between GLP and SPD/PER/GER/SER Competence with the Transformation of Non-Linear Data

Weak　GLP-GER,　GLP-SER　<　GLP-PER　<　GLP-SPD　Strong

(GLP: general listening proficiency　SPD: sentence-based partial dictation　PER: phonetic error recognition　GER: grammatical error recognition　SER: semantic error recognition)

3.5 Relationships between General Listening Proficiency and Sentence-Based Listening Competence Based upon Multi-Regression Analysis

The Figure 3 above shows one aspect of relationships between general listening proficiency and sentence-based listening competence. The current study has also investigated them in a comprehensive manner by conducting multi-regression analysis. First, the predictive power of each combination of two of the three types of sentence-based listening (SPD, PER and SER) and that of all the three types with general listening proficiency were examined employing their linear data used above in 3.3, in which the unpromising data for the GER shown in Table 8 was excluded. The results are shown in Table 12:

Table 12: Predictive Power of Sentence-Based Listening with General Listening Proficiency Based upon the Linear Data

		SPD& PER	SPD & SER	PER & SER	SPD, PER& SER
GLP	R’	.59	.56	.30	.58
	R2’	.35	.31	.09	.33
	p-value	.00	.00	.06	.00

GLP: general listening proficiency　SPD: sentence-based partial dictation　PER: phonetic error recognition　SER: semantic error recognition　R’: multiple coefficient adjusted for the degree of freedom　R2’: coefficient of determination adjusted for the degree of freedom

This table shows 1) that the best combination of two of the three types of sentence-based listening competence, which generates the highest predictive power with general listening proficiency, is SPD & PER, 2) that this combination can account for 35% of the variance of general listening proficiency (p = .00), and 3) that this accountability is slightly better than that created by all the three types of sentence-based listening competence: 33%.

Next, the predictive power of each combination of two of the three types of sentence-based listening competence (SPD, PER and SER) and that of all the three types with general listening proficiency were likewise examined employing their non-linear cubic data transformed for linear multi regression analysis used above in 3.4. The results are shown in Table 13:

Table 13: Predictive Power of Sentence-Based Error Listening with General Listening Proficiency Based upon Non-Linear Cubic Data

		SPD& PER	SPD & SER	PER & SER	SPD, PER& SER
	R'	.62	.61	.58	.65
GLP	$R^{2'}$	.39	.37	.33	.43
	p-value	.00	.00	.00	.00

GLP: general listening proficiency SPD: sentence-based partial dictation PER: phonetic error recognition SER: semantic error recognition R': multiple coefficient adjusted for the degree of freedom $R^{2'}$: coefficient of determination adjusted for the degree of freedom

This table shows 1) that the best combination of two of the three types of sentence-based listening competence, which generates the highest predictive power with general listening proficiency, is also SPD & PER, 2) that this combination can account for 39% of the variance of general listening proficiency (p = .00), which is not much greater than the linear result above: 35%, and 3) that the highest accountability is created by all the three types of sentence-based listening competence (43%), which is 10% greater than the linear result above (33%) shown in Table 12.

4. Summary & Discussion

The above analyses have made clear several important points with regard to the nature of sentence-based listening, some of which confirm and reinforce the results of the author's two previous studies.

4.1 Performance of Sentence-Based Listening

As Table 6 above shows, there are a number of differences and similarities in performance among the four types of sentence-based listening (partial dictation, phonetic, grammatical, and semantic error recognition), in which it is extremely difficult to illustrate and understand the order of their performance difficulty in a simple unified manner, as is shown in Figure 1. There are, however, some obvious findings from two perspectives: relationships among the three types of sentence-based error recognition and those between sentence-based partial dictation and each type of sentence-based error recognition.

With regard to the first perspective, it has been found that the performance of sentence-based semantic error recognition is relatively high, though not the highest, and likewise that that of sentence-based grammatical error recognition is relatively low, though not the lowest. This result is consistent partially with that of Kawashima (2010: p.228) which claims:

1) The processing of spoken sentences may be more semantic-focused than phonetically/grammatically-focused considering its time-restricted nature (uttered sentences have to be processed and comprehended along the time axis) and their elements of general daily-like conversations (in which usually relatively less attention is directed at their phonetic and grammatical errors than semantic ones).

2) The subjects thus may have found it more difficult to recognize phonetic and grammatical errors than semantic ones, and their performance of sentence-based semantic error recognition was therefore better than that of sentence-based phonetic and grammatical error recognition.

If performance difference had been found between sentence-based semantic and phonetic error recognition, these claims would have been directly applied to the result of the current study, but it may at least be possible to modify them to interpret the result as follows:

3) The processing of spoken sentences may be more semantically/phonetically-focused than grammatically-focused considering its time-restricted nature (uttered sentences have to be processed and comprehended along the time axis) and their elements of general daily-like conversations (in which usually relatively less attention is directed at their grammatical errors than semantic/phonetic ones).

4) The subjects thus may have found it more difficult to recognize grammatical errors than semantic/phonetic ones, and their performance of sentence-based semantic error recognition was therefore better than that of sentence-based grammatical error recognition.

With the limited data, it is hard to correctly and comprehensively interpret the equality in performance between sentence-based semantic and phonetic error recognition, which was not found in Kawashima (2010), but one possible interpretation for this result discrepancy is that sentence-based phonetic error recognition may be more influenced by some kind of interactive semantic/phonetic processing, and consequently that its performance may tend to fluctuate to a certain degree. Further studies must be conducted in order to determine which result may represent the nature of the performance relationship better.

As far as the second relational perspective is concerned, it has been found that although sentence-based partial dictation is equal in performance to sentence-based grammatical error recognition, its performance is lower than that of sentence-based phonetic/semantic error recognition. It also is hard to correctly and comprehensively interpret this research finding with the limited data, but a possible partial interpretation for it can be attempted taking into account the general nature of sentence-based partial dictation. Its essence may be summarized as follows:

5) The success of sentence-based partial dictation depends upon a well-balanced skill of listening and writing, which is considered to have increased the subjects' processing burden. This may explain the reason why its performance is lower than that of sentence-based phonetic/semantic error recognition, which does not impose such an "extra" burden on the subjects.

As far as the case in which the performance of sentence-based partial dictation does not differ from that of grammatical error recognition is concerned, it may be possible to claim as follows:

6) Sentence-based partial dictation involves elements of the-- same kind of grammatical processing as sentence-based grammatical error recognition, as the partial regression coefficient of the direct relationship between sentence-based partial dictation and grammatical error recognition ($r = .43$, $p < .01$) may indicate in Table 8. This "overlap" of processing elements may "surface" and show the same performance.

Future empirical studies would examine the validity of the above interpretation, in which more of the nature of the performance of sentence-based partial dictation and grammatical error recognition would be elucidated.

4.2 Sentence-Based Listening Competence and General Listening Proficiency

With regard to relationships between sentence-based listening competence and general listening proficiency, a number of findings have been made by the above analyses, some of which are consistent with the results of Kawashima (2010: p.229). Their essence would be summarized as follows:

1) Relationships between sentence-based listening competence and general listening proficiency are multi-faceted and complicated, mostly embracing non-linear cubic elements.

2) Overall, sentence-based partial dictation is most related with statistic significance to general listening proficiency as a single variable, whereas sentence-based grammatical error recognition is hardly related.

3) The combination of sentence-based partial dictation and sentence-based phonetic and semantic error recognition competence generates the highest predictive power with general listening proficiency (explaining 43% of the variance) when the non-linear cubic elements of sentence- based phonetic and semantic error recognition competence are combined.

It is likewise hard to correctly and comprehensively interpret these research findings with the limited data, but some possible interpretations can be made taking the general nature of sentence-based listening into account as in 4.1 above. Usually relationships between two variables, such as semantic error recognition competence and general listening proficiency, are assumed to be linear, in which correlation coefficients, the strength of the linear association between two variables, are measured. The current study, however, challenged this assumption again, and a finding like 1) above has also been obtained, which seems to deserve attention. In general, non-linear relationships are not easy to interpret, but considering a property pertaining to the cubic curve, the non-linearity of 1) above may be overall understood as follows:

4) Sentence-based phonetic/semantic error recognition competence increases sharply as general listening proficiency does.

5) Sentence-based phonetic/semantic error recognition competence then increases less sharply or stays at a certain level for a certain period even though general listening proficiency goes up.

6) Sentence-based phonetic/semantic error recognition competence starts to increase sharply again as general listening proficiency does.

These "stages" or "changes" in the relationships between sentence-based phonetic/semantic error recognition competence and general listening proficiency are understandable considering non-linear aspects or developments of language-related competence and proficiency that average learners often experience. Such "changes" may represent better the nature of relationships between sentence-based phonetic/semantic error recognition competence and general listening proficiency. It may even be possible to claim that many of the relationships between language-related variables are more non-linear than linear.

With regard to the research finding 2) above, it must be noted that its first part (sentence-based partial dictation is most related with statistic significance to general listening proficiency as a single variable) may provide a greater insight into the nature of sentence-based listening than the result of Kawashima (2010) which claims that the processing of spoken sentences may be more semantically-focused and consequently that the relationship between semantic error recognition competence and general listening proficiency should be the strongest. It is likewise hard to correctly and comprehensively interpret this new research finding with the limited data, but one simple but obvious interpretation for this perspective is that sentence-based partial dictation may embrace more elements of processing spoken sentences than sentence-based semantic error recognition, which are also parts of general listening proficiency, and consequently that sentence-based partial dictation should be more closely related to general listening proficiency than sentence-based semantic error recognition.

It has been found that the latter part of the research finding 2) (sentence-based grammatical error recognition competence is not related to general listening proficiency) is the same as that of the previous study. It is difficult to interpret the same result, but it cannot be denied as the previous study claims that a number of the "unsuitable" materials to measure sentence-based grammatical error recognition were reduced in order to raise the internal consistency reliability of the target data, and therefore that its competence was not measured as properly as the other two. The number of the research materials to measure sentence-based phonetic/semantic error recognition was reduced as is shown in Table 2, so it may not be possible to explain the research finding 2) by such a claim alone. Further studies must be conducted to find a systematic explanation for this finding.

According to the research finding 3) above, sentence-based partial dictation and phonetic/semantic error recognition competence may be independent important elements of general listening proficiency. The highest predicative power (explaining 43% of the variance of general listening proficiency), approximately 10% higher than that of the previous study (explaining 34% of the variance of general listening proficiency), may be worthy of attention considering that the single predicative powers of sentence-based partial dictation and phonetic/semantic error recognition competence with general listening proficiency are 32% and 33%, respectively, as shown in Table 10 and Table 13. It must be noted, however, that the core of this predicative power may lie in sentence-based partial dictation competence as is understood from Table 13, in which the two-type combinations of sentence-based listening including sentence-based partial dictation generate higher predicative power of general listening proficiency. It may be possible to contend that sentence-based partial dictation may involve more common elements of general listening proficiency than sentence-based error recognition competence. It is extremely hard to correctly and comprehensively interpret this research finding with the limited data, but all that can be maintained at this moment is that certain combinations of different types of sentence-based listening, which are considered to interact with one another in complicated manners, may contribute to a higher predictive power of general listening proficiency, and that the use of some different types of sentence-based listening, which embrace more of the essence of general listening proficiency, may greatly increase its predicative power.

Future empirical studies would examine the validity of the above interpretation, in which more of the nature of the relationships between sentence-based listening competence and general listening proficiency would be elucidated.

Concluding Remarks

The current study, which followed the same research framework of the two studies that the author conducted in 2009 and 2010, is a small-scale classroom activity-based investigation attempting to elucidate the nature of sentence-based listening. Some of the results above have turned out to be consistent with those of the previous studies, such as in non-linear cubic elements of sentence-based error recognition, and others have not, such as in the predicative power of general listening proficiency. It must be noted, however, that the results reported and discussed in the current study are still tentative and inconclusive in a number of respects, such as the control of such experimental factors as internal consistency reliability and construct validity of research materials and of various types of “noise” in collecting data. Future studies, taking these points into account, will take us closer to a complete map of the nature of sentence-based listening.

Notes

1) *SPSS* (Version 16.0: SPSS Inc.) was used for this examination.
2) *XLSTAT-PRO* (Version 2009: Addinsoft Inc.) was used for this examination.
3) *EXCEL STATISTICS* (Version 5.0: Esumi Inc.), *TAHENRYOU-KAISEKI* (Version 5.0: Esumi Inc.), and *SPSS* (Version 16.0: SPSS Inc.) were used for the analyses.

Acknowledgements

This study is supported by Grant-in-Aid for Scientific Research (1952539) from the Japan Society for the Promotion of Science. The author is grateful to Dr. Richard Hodson, an English lecturer at University of Nagasaki, Siebold for his full support for making the SPD/PER/GER/SER tests and proofreading the first draft.

References

Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge University Press.

Buck, G. (2001). *Assessing Listening*. Cambridge University Press.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.

Field, J. (2008). *Listening in the Language Classroom*. Cambridge University Press.

Flowerdew, J. & Miller, L. (2005). *Second Language Listening*: *Theory and Practice*. Cambridge University Press.

Hio, Y. (1983). Dictation as a Testing Device –Based on Error Analysis–. *Language Laboratory*, 20, 2-16.

Icho, H. (2008). Understanding the Differences between Successful and Unsuccessful Japanese High School EFL Listeners: Focusing on Explanatory Factors. *ARELE*, 19, 1-10.

Kawashima, H. (2009). Understanding Sentenced-Based Partial Dictation. *The Journal of Nagasaki University of Foreign Studies*, 13,185-196.

Kawashima, H. (2010). Understanding the Nature of Sentence-Based Error Recognition. *The Journal of Nagasaki University of Foreign Studies*, 14, 213-232.

Mackey, A. & Gass, S. M. (2005). *Second Language Research: Methodology and Design*. Mahwah, N. J.: Lawrence Erlbaum.

Nation, I.S.P. & Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking*. Routledge, Taylor and Francis.

Rost, M. (2002). *Teaching and Researching Listening*. Longman.

Rost, M. (2011). *Teaching and Researching Listening* (Second Edition). Longman.

Seliger, H. W. & Shohamy, E. (1989). Second Language Research Method. Oxford: Oxford University Press.

Sugawara, Y. (1999). Dictation and Listening Comprehension – Does Dictation Promote Listening Comprehension? –. *Language Laboratory*, 36, 33-50.

Takeno, J. & Takatsuka, S. (2007). Factors Affecting Listening Comprehension Ability of Japanese Learners of English. *ARELE*, 18, 1-10.

---

kawashima@tc.nagasaki-gaigo.ac.jp