論　文

Interaction of the Factors Affecting the Accuracy of L2 Pronunciation — based on the acoustic analysis of English voiced plosives in word-final position

Toru Isono

要　旨

第二言語学習者の発音は，学習者の母国語音がそのまま使用されているような母国語なまりの発音から，その第二言語の母語話者に近い発音まで，実に多種多様である。また，同じ学習者の発音だけをみた場合でも，状況によってその正確さに大きな差が観察され，この点においても学習者言語は多様であるといえる。中間言語研究において，この「多様性」の研究は主要課題の1つとされ，主に，「学習者が自分の第二言語発音に注ぐ注意の度合い」と「発音の正確さ」の関係から議論されてきた。本論文では，「注意の度合い」に，「第二言語音自体の難しさ」，「学習者の学習レベル」という2つの要因を加え，それらがどのように関係しあいながら第二言語学習者の発音の正確さに影響を与えているのかを，日本人英語学習者から得た英語破裂音の音響分析データを踏まえながら，実証的に考察していく。

キーワード：L2 pronunciation（第二言語発音），plosives（破裂音），interlanguage（中間言語），the amount of attention（発音への注意の度合い），intrinsic difficulty（言語音自体に備わる自然言語的観点からみた発音の難しさ）

1. Introduction

In the acquisition process of Second Language (L2) pronunciation, as Nemer (1971) reported, L2 learners tend to have the stages to produce many examples of elements which do not belong to either First Language (L1) or L2 sound systems before they master the accurate L2 pronunciation. We also often hear the examples of L2 learners who can produce native-like L2 sounds in some cases but cannot in other cases. These facts support the idea that "learners do not progress from a state of non-acquisition to a state of acquisition, but rather pass through a series of stages" (Ellis, 1994, 96). In the developmental process, L2 learners are supposed to construct and develop their own regular and systematic linguistic systems, and it causes the production of the above pronunciation particular to L2 learners (See details in Isono, 2005). The linguistic systems which lie midway between L1 and L2 are named interlanguage, and the aim of interlanguage study is to clarify the characteristics of the developmental process of the linguistic systems.

The concept of interlanguage has been revised and developed as the result of being affected by various trends of Second Language Acquisition (SLA) research. Yamaoka (1997) summarises the characteristics of the current accounts of interlanguage into the six features presented, as referred to in the studies of Hatch (1983) and Richards (1985).

1. Systematicity	2. Permeability	3. Transitionality
4. Universality	5. Variability	6. Fossilisation

(Hatch, 1983; Richards, 1985, cited in Yamaoka, 1997: 75-76)[1]

Among the above six characteristics, this paper focuses on variability in interlanguage, and examines what factors affect the accuracy of L2 pronunciation, based on the acoustic analysis of English plosives produced by Japanese learners.

2. Variability

When SLA researchers face the issue as to how they explain the variability of interlanguage theoretically, their views differentiate depending on whether they

approach it from the standpoint of the *homogeneous competence model* or the *heterogeneous competence model*. The difference between the two approaches is whether the source of the variability of interlanguage refers to non-linguistic factors affecting a learner's linguistic performance or a learner's linguistic competence itself.

In the homogenous competence model, linguistic competence is separated from linguistic performance. To sum up, all speakers' efforts at performance, and the fact that they do not always manifest the ideal system in the way they speak are dismissed as irrelevant to an understanding of the ideal system (Chomsky, 1965, cited in Ellis, 1985). Linguistic competence as the ideal system is regarded as a homogeneous one, so variability of interlanguage is considered as the phenomenon in the process of performance, and all variability is classified as non-systematic variability. Since the homogeneous model distinguishes linguistic competence from linguistic performance and mainly focuses on the former only, the variability of interlanguage has seldom become an object of study for SLA researchers supporting this model. This notion is condensed in Gregg's (1990: 379) comment, "since variability is a part of performance, it is perfectly possible that one could formulate an interesting theory of IL variation as part of a theory of IL performance. But to try to force variability into a theory of the acquisition of competence by claiming that competence is itself variable is self-defeating". In this sense, as Ellis (1985) noted, the homogeneous competence model discounts stylistic variability or rather treats it only as one of the aspects of performance.

In contrast to the homogenous model, in the heterogeneous model, the variability is explained at the level of linguistic competence. In short, in the model, a learner's linguistic competence is regarded as being heterogeneously formed by various kinds of linguistic knowledge, and this fact is considered as a reason of the occurrence of the variability. Some scholars (e.g., Bialystok, 1982; Bialystok & Sharwood Smith, 1985; Hulstijin, 1990) present various theoretical frameworks within this model to explain the variability in interlanguage.

As far as the experimental research in L2 pronunciation which was conducted within the heterogeneous model is concerned, it has become common to explain the variability in interlanguage phonology in relation to the amount of attention that L2 learners are thought to have given to language production. In short, it is predicted that the more attention is given, the greater the accuracy observed.

A study which became the basis of SLA research was Labov's (1970) experiment which investigated the variability in a natural language. He investigated the speech patterns of New Yorkers in the following five tasks: (1) casual speech; (2) careful speech; (3) reading aloud; (4) the reading of word lists aloud; and (5) minimal pairs. There was an assumption that little attention was paid in task 1, but greater attention was given in task 5. He found that the more attention was given, the more instances of [θ], which is a prestige sound in New York English, were observed, but the less attention the subjects paid, the more instances of non-prestige variants, such as [t], were found.

For SLA research, Dickerson (1975) investigated the production of the English [z] in the speech of 10 Japanese learners three times in a nine month period using the following three tasks: (1) free speech; (2) reading dialogues aloud; and (3) reading word lists aloud. She found that the correct [z] variant was observed most frequently in task 3, followed by task 2, and least often in task 1. Similarly, other researchers also reported on the finding of higher frequencies of native-like pronunciation in tasks involving reading aloud or imitating a model sound, and of less native-like pronunciation in spontaneous speech (e.g., Tarone, 1982, 1983; Beebe & Zuengler, 1983).

The main aim of the current research which will be outlined in later sections is to examine the relationship between the amount of attention and the accuracy (= variability) of L2 pronunciation, and to propose some other factors which should be taken into account, in order to clarify the interaction of the factors affecting the accuracy of L2 pronunciation.

3. The Outline of the Research

3-1. Subjects

Isono (2004) investigated Japanese learners' preferred strategies (whether it was 'epenthesis' or 'devoicing' or 'deletion') for English plosives in word-final position, and clarified that the preference might vary according to learners' learning proficiency. The analysis was based on the data which were produced by 30 Japanese subjects consisting of 3 groups (the US, the PS, and the AS groups) and 8 native speakers of

English who had been teaching English in Essex, East Anglia, England (the NS group). Both Isono (2004) and the current research were conducted in a series of the study on English plosives in word-final position, so the same subjects who participated in Isono (2004) were also enrolled in the current research.

The US groups (Male = 4, Female = 6) consisted of 3rd and 4th year university students in Japan, whose major was English and American literature. In this group, no one had studied abroad, except 1 subject who stayed in England for a month. The PS group (Male = 4, Female = 6) was made up of postgraduate students in a Japanese university, who majored in linguistics and literature[2]. 2 of the subjects had experience studying abroad. One studied in America for a year and the other in Australia for a year. The AS group (Male = 3, Female = 7) was comprised of postgraduate students who had been studying in University of Essex, England. All of them were in the Language and Linguistics Department, and had stayed in England for more than 3 years[3]. According to the result of the reading-aloud test, it was clarified that the Japanese groups were at different levels in their pronunciation proficiency, which means, the most advanced Japanese subjects in order were the AS group, then the PS group, and the US group (See details in Isono, 2004).

3-2. Method and Procedure

The subjects were enrolled in the following two tasks. In the first task, the subjects were asked to produce sentences including the following six target words: *cap* – *cab*; *pat* – *pad*; and *back* – *bag*, which manifested the [CæC] syllable, and contained the target English plosives which had voiceless—voiced contrasts in word-final positions. According to Yavas (1993, cited in Yavas, 1994), final voiced plosives are more accurately produced when the preceding vowel is a low vowel rather than a high vowel, so these target words were relatively suitable for producing accurate voicing of a voiced plosive in word-final position. In this task, when the test sentences were constructed, the following three environments were conditioned to elicit the speech data in equal conditions as done in the second task, described below, as far as possible. Firstly, the final phoneme of a word which preceded the target word was always /ɪ/ (including diphthongs /aɪ/ and /eɪ/). Secondly, the first phoneme of a word which followed the target word was always /ə/. Finally, the test sentences were constructed by using words which were normally learnt before the 2nd-year in high school in Japan. In addition,

the test sentences were constructed so that they made sense whichever one of the minimal pairs was used, in order to prevent the subjects from sensing the target words and to minimise the effect of intonation on the target words. Two different sentences were made for each pair, so we had four possible sentences for each pair (See Appendix). In each test sentence, the upper sentence was included in Test A, and the lower sentence in Test B. Half of the subjects in each group were asked to read Test A, and the other half were instructed to read Test B. In both tests, the sentences were mixed with some other sentences for another experiment. The subjects were asked to repeat the test sentences three times in a row.

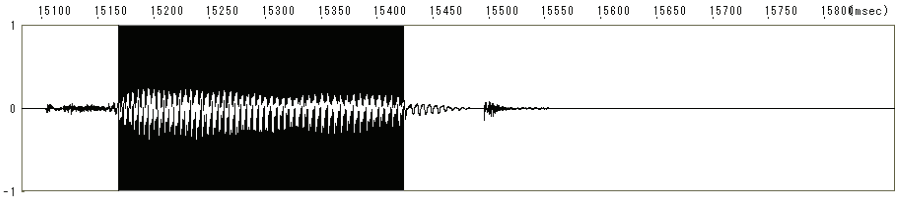
In the second task, the subjects were instructed to produce the target words in the carrier sentence 'I say _____ again.'. The target words were shown to the subjects before starting the task. Each target word was randomly presented at regular intervals to each of the subjects three times.

3-3. Acoustic Measurements

The speech production data were digitally recorded, and 1368 different productions (38 subjects × 2 tasks × 6 words × 3 repetitions) were elicited. According to Isono (2004), most of the Japanese subjects had a great difficulty with voicing English voiced plosives in word-final position. On the other hand, they succeeded in devoicing English voiceless plosives in word-final position. Therefore, for the aim of the current research, we focus on the 684 productions of voiced plosives in word-final position (38 subjects × 2 tasks × 3 words × 3 repetitions).

By using a digital waveform editor and a digital sound spectrogram, each of the productions was characterised according to the following two acoustic features: 1. Stop closure duration time (SD); 2. Voiced closure duration time in the stop closure duration time (VD). 'SD' was regarded as the interval from the end of a vowel to the release of a final plosive. 'VD' was defined as the interval from the end of a vowel to the point where voicing energy was no longer detected. Based on the two features, 'degree of voicing' ('VD' divided by 'SD') was calculated.

Figure 1. An example of 'pad' [pæd] in English



|VD |
|SD |

The mean score of the three repetitions for each of the tasks was regarded as each subject's representation for a target word in each task.

4. Research Questions and Hypothesis

Past studies (Ohala & Riordan, 1979; Yavas, 1993, cited in Yavas, 1994) show that a velar plosive [g] is the most intrinsically difficult plosive to voice in word-final position. In addition, the alveolar plosive [d] is assumed to be more difficult to voice than the bilabial plosive [b] in word-final position. The reason of this hierarchy in the intrinsic difficulty of voicing is attributed to the different effects of aerodynamics, which are caused by the different places of articulation, as Yavas (1994: 273) explains as follows: "the larger the supraglottal area is, the better it can accommodate glottal flow for some time before oral pressure exceeds subglottal pressure and stops the vocal cord vibration".

Based on the above discussion, as far as the degree of voicing is concerned, it is expected that the bilabial plosive [b] would be produced more accurately than the alveolar plosive [d], and, in turn, the alveolar plosive would be more accurately produced than the velar plosive [g]. However, a more interesting question is how these differences in the intrinsic difficulty affect the accuracy of the productions exhibited by L2 learners who are at various levels of learning. In other words, the research question is that what different kinds of patterns of improvement emerge according to the differences in the intrinsic difficulty, when comparing advanced learners' productions with less advanced learners' productions.

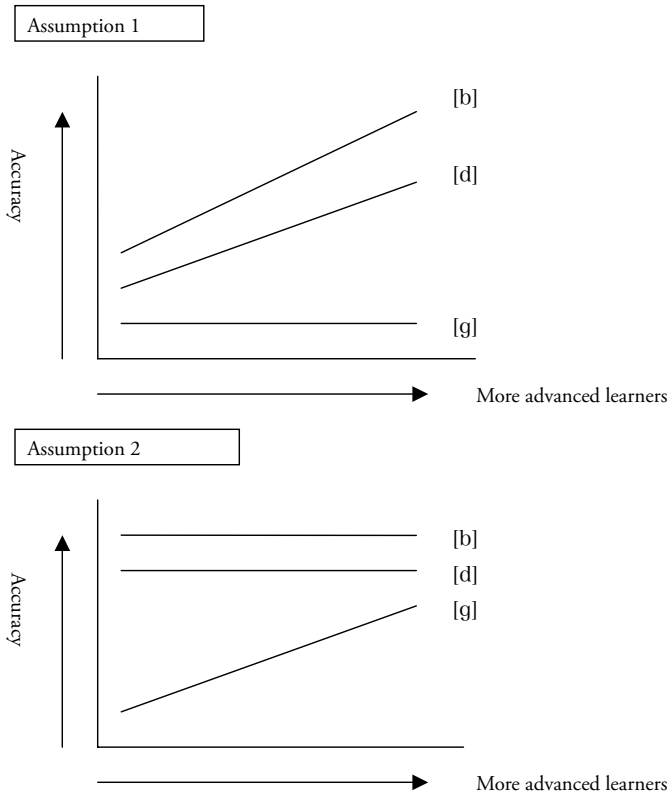
The most likely assumption is that a significant improvement would be observed

as the intrinsic difficulty eases. To put it in an extreme way, great and noticeable improvements would be obtained for [b] and [d], but no significant improvement would be found for [g] because of the intrinsic difficulty.

However, taking the large similarity in voiced plosives between English and Japanese into consideration, we can propose another assumption which predicts that a learner's learning level would be a significant factor only for the accuracy of [g], but not for [b] and [d], since the latter voiced plosives would be accurately produced irrespective of a learner's learning level. In this case, a significant improvement would be found only for [g], but not for [b] and [d], although the latter plosives would constantly have high degrees of accuracy.

As far as the data gathered on Japanese learners are concerned, these two assumptions are illustrated as Figure 2.

Figure 2. Descriptions of two hypotheses for the accuracies of voiced plosives



This paper hypothesizes that whether we find the first assumption or the second assumption depends on how much a learner's attention is drawn to the voicing contrasts. In casual speech style, we would find the first assumption since it is the style in which little attention is paid to form, and the essence of the linguistic rule would be directly reflected in the productions. On the other hand, in formal speech style, in which a large amount of attention is given to form, we would expect the second assumption. This is due to the fact that revision, which is caused by attention to form, in the accuracy of [b] and [d] is activated regardless of a learner's learning level, while revision for that of [g] is activated only by advanced learners, because of the intrinsic difficulty of the motor control.

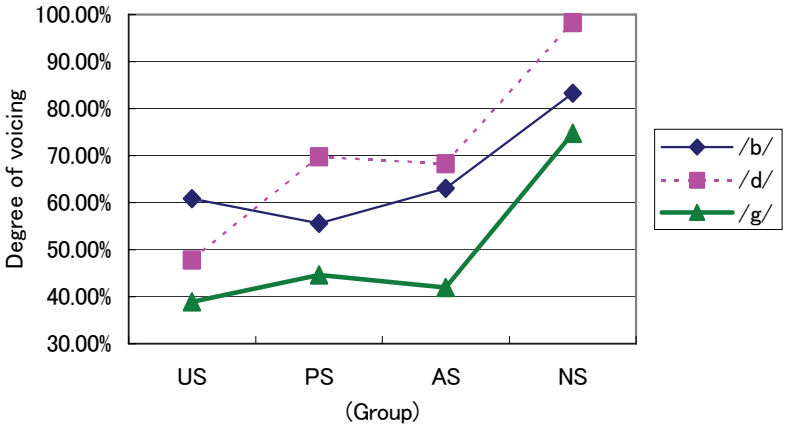
To put it more concretely, the hypothesis predicts that the results in the first task in which the target words were hidden will look like the graphs in the assumption 1 in Figure 2, meaning that a significant improvement in the degree of voicing is observed in [b] and [d], but not in [g], in the Japanese subjects' data. On the other hand, the results in the second task in which the target words were not hidden are assumed to be similar to the graphs in the assumption 2, which suggest that a significant improvement is found only in [g], but not in [b] and [d].

5. Results

In order to investigate the effect of the interaction among the amount of attention, the intrinsic difficulties of English plosives in word-final position, and the subjects' learning levels on the differences in the accuracy, the productions in the first and the second tasks were analysed. As explained earlier, the target words in both tasks were placed according to the same condition in which the last phoneme of a word preceding a target word was /ɪ/ and the first phoneme of a word following a target word was /ə/.

Figure 3 shows the results of the first task. The degree of voicing of [b] in word-final position is represented by a thin solid line, that of [d] by a thin dotted line, and that of [g] by a thick solid line. This also applies to Figure 4.

Figure 3. Results for Task 1



Firstly, it should be noted that the degree of voicing by the NS group was significantly higher, compared with the results of past studies (e.g., Yavas, 1993, cited in Yavas, 1994). These results might suggest that this experiment sufficiently attained the aim that desirable conditions were provided for voicing in voiced plosives in word-final position.

Looking at the differences in the degree of voicing among the voiced plosives, we find that, as predicted, the degree of voicing for the velar plosive [g], which is said to be the most difficult plosive to voice, was constantly lower than for the other two voiced plosives in all groups.

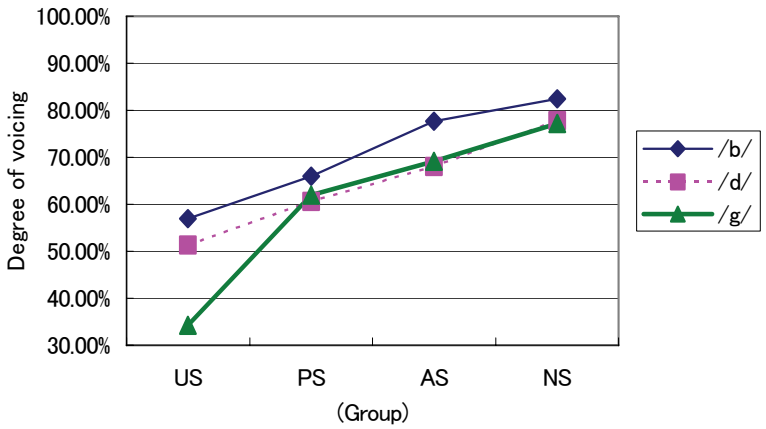
We direct our attention to the issue of how the degrees of voicing of these voiced plosives were improved according to the learning levels of the subjects. As Figure 3 indicates, for the Japanese groups, a much sharper fluctuation is found in [d] than in [b] and [g]. A series of one-way ANOVAs revealed that a significant group difference was found for [d] at the .01 level, and for [g] at the .05 level, but it was not obtained for [b] at the .05 level. According to Bonferroni's Post Hoc Tests, the [d] values of the PS and AS group were significantly greater than the value of the US group, although they were also significantly smaller than the value of the NS group. On the other hand, no significant group difference for both the [b] and [g] values was found among the Japanese groups at the .05 level. The difference in the results between [b] and [g] values was that there were no significant differences between the Japanese groups' [b]

values and the NS group's [b] value; however, all the Japanese groups' [g] values were significantly smaller than the value of the NS group, at the .05 level.

The above statistical results confirm the hypotheses for the acquisition processes of the degree of voicing of [d] and [g], but not for the hypothesis regarding [b], in the case of the first task. The developmental process of the voicing of [b] was assumed to be a process such that the sharpest fluctuation would be found from the US group to the AS group, because it was intrinsically the easiest phoneme to voice among the voiced plosives. However, as it was revealed by the above statistical results, there was no significant difference in the degree of voicing of [b] between the Japanese groups and the NS group. This result clarifies that, in the degree of voicing, even the least experienced Japanese group could attain a performance which was not distinguished from the native speakers, and this could be interpreted to mean that the voicing of [b] is comparatively easier than that of the other two voiced plosives. The reason why the degree of voicing in [b] was lower than that in [d] in most of the groups will be discussed by referring to the results in the second task.

Figure 4 shows the results in the second task in which the target words were not hidden.

Figure 4. Results for the second task



Concerning the intrinsic difficulties of voicing [b] and [d], we find a relatively clearer picture in the second task than in the first task that the voicing of [b] might be

easier than that of [d]. This result also suggests that the reason why low degrees of voicing in [b] were obtained in the first task can be partly attributed to the possibility that the target words containing [b] in word-final position were placed in the environments where it was difficult to voice them, despite of the effort of fixing the last phoneme of words preceding the target words and the first phoneme of words following them. For example, in the test sentence 'I saw a grey *cab* at the right-hand corner.', some subjects might have made a pause between 'cab' and 'at', and generally, the possibility of devoicing a word-final voiced plosive becomes high when it is followed by a pause compared with the case when it is followed by a low vowel (Yavas, 1994). However, even though we acknowledge this possibility, it is still not clear why these conditions affected only the productions of the PS, AS and NS groups, but not those of the US group.

Regarding the developmental processes in the voicing from a less advanced to a more advanced group, we find that all the graphs showed a rise toward the right-hand side, which means that the voiced plosives were more accurately voiced as the subjects became more advanced learners. However, going from the US group through to the PS group, a much sharper fluctuation is found in [g] compared with [b] and [d]. One-way ANOVA revealed that a significant group difference was not found for [b] and [d]: $F(3, 34) = 2.53$, $p > .05$; $F(3, 34) = 1.83$, $p > .05$, respectively. On the other hand, a significant difference was obtained for [g]: $F(3, 34) = 5.86$, $p < .01$, and according to Bonferroni's Post Hoc Tests, the value of the US group was significantly smaller than the other groups at the .05 level. These results for the second task confirm the hypothesis that the subjects' learning levels are a significant factor in the improvement of [g], but not for [b] and [d], in the formal condition, in which the amount of attention given to the items is relatively large.

6. Conclusion

To sum up the above results, firstly, the degrees of intrinsic difficulty in voicing English voiced plosives in word-final position were confirmed, which is that the velar plosive [g] is more likely to be devoiced than the alveolar plosive [d], and it is also more likely to be devoiced than the bilabial plosive [b]. Regarding the question of how these differences in intrinsic difficulties contribute towards characterising the

productions of Japanese learners, the current experiment suggests that the effect varies according to the amount of attention being paid to the final voiced plosives, as well as learners' learning levels. Concretely, in the results, when little attention was given, the accuracy in the voicing of the alveolar plosive [d] was the only feature distinguishing advanced learners from less advanced learners. This was due to the fact that, for Japanese learners, the bilabial plosive [b] was easy, and the velar plosive [g] was difficult to voice, regardless of their learning levels. On the other hand, when a large amount of attention was given, the accuracy of the velar plosive [g] became the distinguishing feature. This was due to the fact that less advanced learners could not accurately produce it regardless of how much attention was given to it, but more advanced learners could according to the amount of attention. On the other hand, concerning [b] and [d] in word-final position, which are easier plosives to voice than [g], they were accurately voiced regardless of learners' learning levels provided sufficient attention was given to the features.

As previously discussed, it is said that the variability of interlanguage is conditioned by the amount of attention paid to an item. In short, the amount of attention given to an item had been assumed to be directly reflected in the accuracy of production. However, the current study implies that this is not always true, and that the effect of the amount of attention devoted to the improvement in accuracy is conditioned by something resembling a 'filter', which is the intrinsic difficulty of an L2 item. When the filter is thin, which means the target item is intrinsically easy, the amount of attention tends to effectively affect the accuracy of production. On the other hand, when the filter is thick, which means the target item is intrinsically difficult, the effect of the amount of attention is minimised since the effect tends to be obstructed by the thick filter. In both cases, the filter is supposed to be made thinner according to a learner's proficiency level, to some degree. This paper suggests that, in order to clarify the mechanism of the variability of interlanguage, not only the amount of attention but also at least a learner's proficiency level and the intrinsic difficulty of the L2 item should be taken into account, since the latter are the features controlling the effect of learners' attention on the accuracy of their L2 productions.

Notes

1. The translation from Japanese to English was made by the author of this paper.
2. No one had studied a subject connected with any kind of pronunciation as their majors.
3. In the same way as the PS group, only the subjects whose majors were not relevant to any kind of pronunciation study were selected.

References

Beebe, L. and Zuengler, J. 1983. Accommodation Theory: an explanation for style shifting in second language dialects. In Wolfson, N. and Judd, E. (eds.), *Sociolinguistics and Second Language Acquisition*. Rowley, Mass.: Newbury House.

Bialystok, E. 1982. On the Relationship between Knowing and Using Linguistic Forms. *Applied Linguistics*, 3 (3): 181–206.

Bialystok, E. and Sharwood Smith, M. 1985. Interlanguage Is not a State of Mind: an evaluation of the construct for second language acquisition. *Applied Linguistics*, 6: 101–117.

Dickerson, L. 1975. The Learner's Interlanguage as a System of Variable Rules. *TESOL Quarterly*, 9: 401–407.

Ellis, R. 1985. *Understanding Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Gregg, K. 1990. The Variable Competence Model of Second Language Acquisition, and Why It Isn't. *Applied Linguistics*, 11: 364–383.

Hatch, E. 1983. *Psycholinguistics: a second language perspective*. Rowley, Mass.: Newbury House.

Hulstijin, J. 1990. A Comparison between the Information-processing and the Analysis/Control Approaches to Language Learning. *Applied Linguistics*, 11: 30–45.

Isono, T. 2004. A Study of Japanese Learners' Preferred Strategies for English Plosives in Word-final Position.『文明21』, 13: 127–139.

Isono, T. 2005. The Acquisition Process of L2 Pronunciation – based on the acoustic analysis of English vowels produced by Japanese learners.『言語と文化』, 12: 1–16.

Labov, W. 1970. The Study of Language in Its Social Context. *Studium Generale*, 23: 30–87.

Nemser, W. 1971. Approximative Systems of Foreign Language Learners. *International Review of Applied Linguistics*, 9: 115–123.

Ohala, J. and Riordan, C. 1979. Passive Vocal Tract Enlargement during Voiced Stops. In Wolf, J. and Klatt, D. (eds.), 1979, *Speech Communication Papers*. New York: Acoustical Society of America.

Richards, J. 1985. Error Analysis, Interlanguage, and Second Language Acquisition: a review. In Richards, J. (eds.), *The Context of Language Teaching*. Cambridge: Cambridge University Press.

Tarone, E. 1982. Systematicity and Attention in Interlanguage. *Language Learning*, 32: 69–82.
Tarone, E. 1983. On the Variability of Interlanguage Systems. *Applied Linguistics*, 4: 142–164.
Yamaoka, T. 1997.『第二言語習得研究』, Tokyo: Kirihara Shoten.
Yavas, M. 1994. Final Stop Devoicing in Interlanguage. In Yavas, M. (eds.),1994, *First and Second Language Phonology*. San Diego, CA: Singular.

Appendix

Test sentences for [p] – [b] in word-final position

★ Target words – cap & cab

Test A ● I saw a grey *cap* at the right-hand corner.
Test B ○ I saw a grey *cab* at the right-hand corner.

Test A ● He moved his dirty *cab* aside.
Test B ○ He moved his dirty *cap* aside.

Test sentences for [t] – [d] in word-final position

★ Target words – pat & pad

Test A ● I *pat* a chair.
Test B ○ I *pad* a chair.

Test A ● I found my *pad* at the door of my house.
Test B ○ I found my *Pat* at the door of my house.

Test sentences for [k] – [g] in word-final position

★ Target words – back & bag

Test A ● Her pretty *back* appealed to everyone in the class.
Test B ○ Her pretty *bag* appealed to everyone in the class.

Test A ● He said he saw my *bag* at the station.
Test B ○ He said he saw my *back* at the station.