CHAPTER 5

TOKEN DURATION

5.1 INTRODUCTION TO CHAPTER 5

As noted in Chapter 1, this thesis is investigating two main points—the effects of the variables in the study on whether or not a given vowel was devoiced, and, for the voiced vowels only, the effects of the variables on the duration of voicing activity. These two points will be dealt with in this and the next chapter.

First it is necessary to ask whether the use of the three computer-controlled display times elicited three adequately distinct sets of data to represent three distinct SRs. This is necessary because of the lack of known previous studies using computer-controlled display times to elicit SR differences. The question of whether or not the program adequately produced 3 distinct data sets will be addressed in the first section of this chapter through rigorous examination of the data generated by the experimental setup. It will be shown that the answer to this question is yes, the program was quite successful.

Further, it is important to determine any effects that the setup of the study had on the data collected. As an example, the participants of the study included both females and males. Previous research (Yuen & Hubbard 1997) has shown that in Japanese, like other languages, males tend to have faster SRs than females. It is therefore necessary to show that the use of both genders in the study did not compromise the elicitation of the 3 data sets; i.e. that the males did not produce significantly shorter duration tokens than the females even though they were performing the same elicitation task. In addition to checking for effects of the variable GENDER on mean token durations (MTDs), checks for effects of the variables PARTICIPANT, REPETITION, BLOCK, TOKEN, and #DEV were also made.

The outline of this chapter is as follows. §5.2 provides detailed descriptive statistics of the token duration data sets for each SR, including discussion of the usability of the data for further statistical checks for effects of variables. §5.3 discusses the setup of the ANCOVA models used to check for effects of the variables controlled for in the study, while §5.4 presents the results of those checks and their interpretation. Finally, §5.5 summarizes the chapter.

As noted in Ch. 1, this chapter and the next two contain background statistical information in addition to the results of the statistical checks. This is for the benefit of those who, like the author, have phonological backgrounds that include little or no exposure to statistics.

5.2 DESCRIPTIVE STATISTICS OF THE DATA

This section presents descriptive statistics and distribution characteristics for the data collected. In particular, checks for *normalcy* (§5.2.1 below) and *homogeneity of variance* (§5.2.2 below) of the data are important, since the statistical procedures being utilized are based on the assumptions that the normalcy and homogeneity of variance of the data are within acceptable limits.

Statistical checks are often made on the main body of a data set, ignoring data points that are far from the mean value of the data set, the so-called *outliers*. (A typical criterion for inclusion of a data point is that it fall within 95% of the data points closest to the mean value of the data set.) However, on the advice of Milton (1992: 31-32), outliers were not removed from the data set before measures of normalcy and homogeneity of variance were made; instead statistical values are given both with and without the outliers. The extent of the outliers can be seen in the distribution plot of the entire data set below, where the most extreme outlier, the long production at TOKEN DURATION = 867 ms in the slow data set, lies well outside the range of the rest of the data.

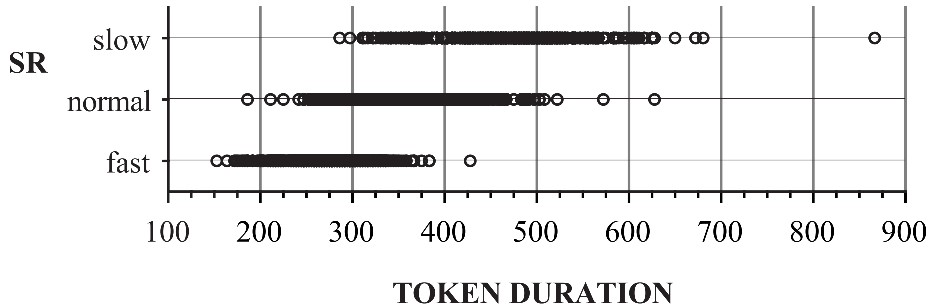

Figure 5.1 Distribution plot of TOKEN DURATION vs. SR; all data points (n = 1800).

The base assumptions underlying the statistical checks will now be presented, focusing on the normalcy and homogeneity of variance of the three data sets.

5.2.1 NORMALCY OF THE DATA

The first of the base assumptions of most statistical calculations is that the data sets being compared be normally distributed, that it follow a *normal curve,* also known as a *Bell curve*. To check normalcy, plots known as *histograms* are used to show the frequency distribution of the data. For the current data set, the histograms show the distribution of the values for the token durations (TOKEN DURATION). These histograms therefore show how many tokens occur in each percentage of the total range of token durations.[1]

The data are separated into three sets corresponding to the three SRs since the carrier sentence display program artificially imposed three separate SRs on the participants' productions. Figures 5.2 to 5.7 present a separate histogram for each SR, Figures 5.2 to 5.4 for all data and 5.5 to 5.7 with the outliers removed. The axes have been kept constant to show the relative distribution of the data for each SR, and a normal curve has been fitted to each histogram to aid visual inspection.

[1]While statistical measures such as the *skewness* and *kurtosis* of the data curve and the modified Kolmogorov-Smirnov (K-S) test are available for quantitatively checking normalcy, the statisticians consulted for this project were of the opinion that visual checks for approximation to the normal curve are sufficient due to the robust nature of the ANOVA and ANCOVA procedures. Statistical measures of normalcy are therefore not presented.

The size of the groups of data represented by the bars in the graph was determined by a statistical procedure known as *Sturges' rule*;[2] the entire range of the data is divided into a number of even duration intervals for comparison with the normal curve. For the data sets presented in this chapter, 10 groups were used. (Not all 10 divisions are visible in the first 3 histograms due to the small number of tokens in the data ranges for the very short and very long durations; the number of data points in these lowest and highest groups are so small as to be indistinguishable from the x-axis.)

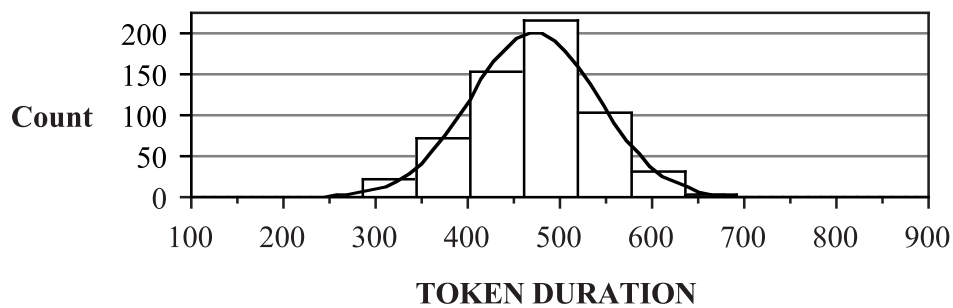

Figure 5.2 Histogram for TOKEN DURATION, *slow* SR; n = 600 (all data points).

---

[2]Sturges' rule (Sturges 1926: 65-66; as quoted by Milton 1992: 18-19) is based on the base 2 logarithm of the data count for data counts greater than 16:

# of categories = truncate ($\log_2$ (data count)) + 1

i.e. the next highest integer above the whole number component of the base 2 exponent for the number of data points. For the current data sets the data count = 600, so the number of categories is given by truncate ($\log_2$ (600)) + 1 = truncate (9.2) + 1 = 10. The range of data for each of the three data sets was therefore divided into 10 equal duration ranges, and counts of data points were made within each duration range.

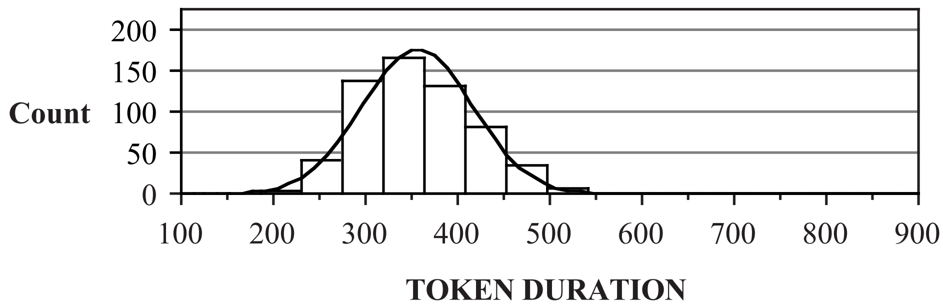

Figure 5.3 Histogram for TOKEN DURATION, *normal* SR; n = 600 (all data points).

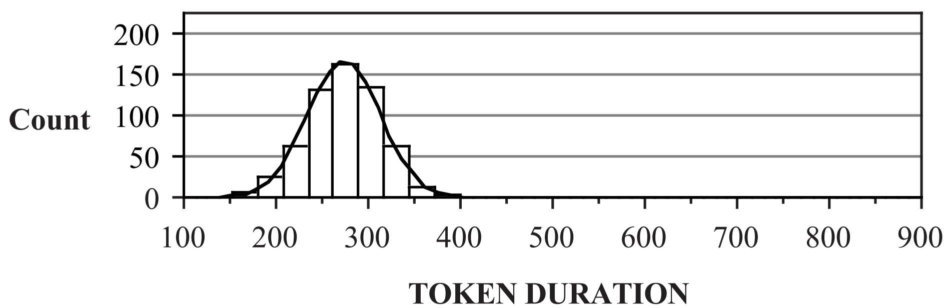

Figure 5.4 Histogram for TOKEN DURATION, *fast* SR; n = 600 (all data points).

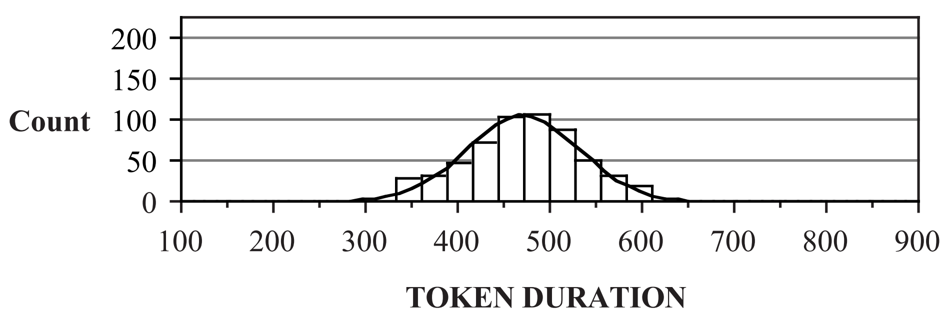

Figure 5.5 Histogram for TOKEN DURATION, *slow* SR; n = 576 (24 outliers removed).

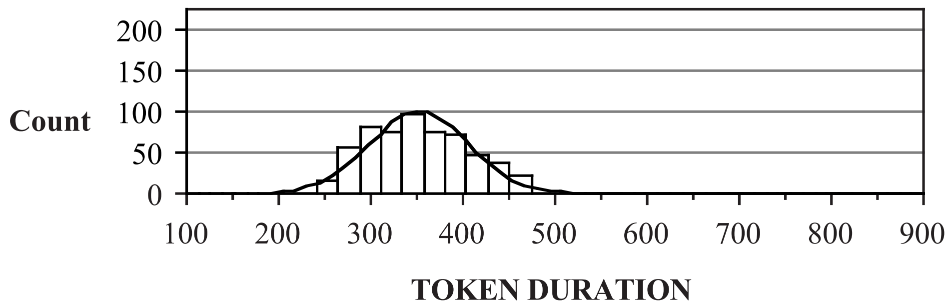

Figure 5.6 TOKEN DURATION ,*normal* SR; n = 579 (21 outliers removed).

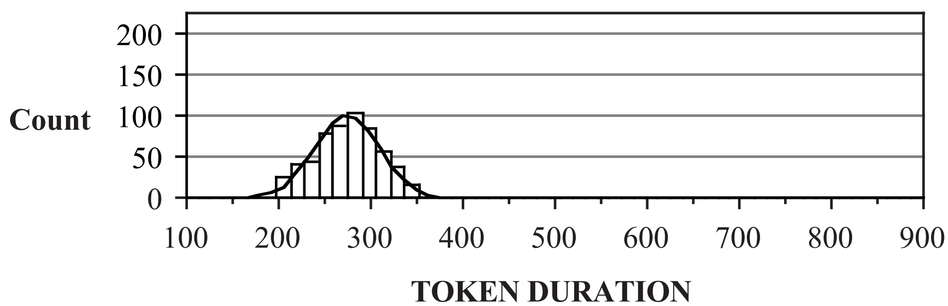

Figure 5.7 TOKEN DURATION, *fast* SR; n = 573 (27 outliers removed).

As can be seen from the figures, the data for the normal and fast SRs approximate the fitted normal curves fairly well both with outliers included and outliers removed. The data sets for TOKEN DURATION are therefore considered to fulfill the requirement of normalcy for statistical tests.

As a side note, it can also be seen that the elicitation program was successful in targeting a different mean token production at each SR—the center of the fitted normal curve shifts toward a shorter duration value from the slow to the normal to the fast SR. The values associated with the centers of the fitted normal curves, averaged across all data points for all 10 tokens used in the study (Figures 5.2 to 5.4), are 472 ms for the slow SR, 356 ms for the normal SR, and 273 ms for the fast SR. These values represent 12%, 13% and 16% of the total sentence display times, respectively—approximately an equal percentage of the total sentence display time, indicating that the change in SR resulted in approximately the same amount of time being spent on producing the tokens at the beginning of the stimuli sentences.

5.2.2 Homogeneity of variance of the data groups

The more important base assumption of most statistical tests is that the data in different groups being compared display *homogeneity of variance*—the data of each group being compared should be distributed about the mean of that group in approximately the same way. That is to say, in order to make observations about the differences in the mean values for groups of data, the *distribution* of values about those means of the data should be about the same. The *variance* of a data set is a measure of this distribution of the values about the means; i.e. a measure of how the data varies about the mean value for the data set. The variance of the three data sets corresponding to the three SRs are given below, along with other statistical measures.

Table 5.1 Summary statistics the TOKEN DURATION data, split by SR; all data points.

	TOKEN DURATION fast	***TOKEN DURATION normal***	***TOKEN DURATION slow***
Mean	273	356	472
SD	39.678	60.076	69.054
SE	1.620	2.453	2.819
⇒ ***Variance***	***1574***	***3609***	***4768***
Count	600	600	600
Minimum	152	186	287
Maximum	428	629	867
Range	276	443	580

Table 5.2 Summary statistics for the TOKEN DURATION data, split by SR; outliers removed.

	TOKEN DURATION FAST	***TOKEN DURATION NORMAL***	***TOKEN DURATION SLOW***
Mean	274	352	470
SD	34.568	53.613	60.330
SE	1.444	2.228	2.516
⇒ ***Variance***	***1195***	***2874***	***3640***
Count	573	579	575
Minimum	198	241	333
Maximum	352	474	608
Range	154	233	275

It is important to note that both the range of values of each data set and the distribution of data throughout each range (as indicated by the variance) increase from the fast to the normal to the slow data set.

The relationship between the mean and variance of the three subsets of data can be seen more clearly in what is known as a box plot. Each box indicates the main body of data points for each data set; the line through the center of each box is the mean value of that data set. The outer bars of each box give the standard deviation, while the extreme values, the outliers, lie above and below each box. The height of each box and range of the outliers therefore indicate the distribution of the data in each set.

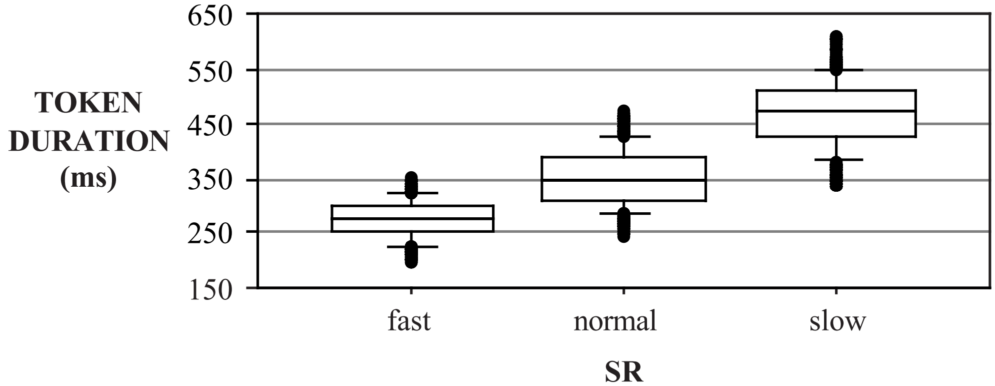

Figure 5.8 Box plot for TOKEN DURATION, separated by SR.

It can be seen that the mean of each data set increases as SR decreases; i.e. the MTD of each data set increased as the presentation time of the stimuli sentences increased. The mean of the *fast* data set, corresponding to a sentence presentation time of 1700 ms, had a MTD of 273 ms; the *normal* data set, corresponding to a presentation time of 2700 ms, had a MTD of 356 ms; and the *slow* data set (presentation time of 4000 ms) had a MTD of 472 ms. The increase in MTD roughly parallels the increase in sentential display duration.

More importantly, the range of most of the data in the three subsets, indicated by the height of the boxes and the spread of the points outside the boxes, again shows an increase in variance as the SR decreases (i.e. the range of the data is greatest for the slow SR). Gravetter & Wallnau (1996) state that for large samples ($n \geq 10$), there is reason for concern if the variance of one set of data is more than twice the variance of another. Following this rule of thumb, non-homogeneity is a problem for the 3 SR data sets for TOKEN DURATION. As can be seen from the tables above, the variance of the slow group of data is about three times the variance of the fast data, and the variance of the normal group of data is over twice the variance of the fast data for both the full set of data and the data with the outliers removed.

These differences can also be seen in the different shapes of the fitted normal curves in Figures 5.2 to 5.7 above. The shape of the fitted normal curve for the slow data is much broader than the shape of the fitted normal curve for the fast data, indicating that the slow data is distributed throughout its range farther away from the data set mean. In addition, the width of the bars in the graph also indicates the overall range, since they are equal to 1/10th of the total range. The bars are much wider at the slow SR than at the fast, again indicating a wider distribution of data at the slow SR. The differences can be seen just as well in the box plot in Figure 5.8 in the increasing height of the boxes, which also shows the greater range of the data at the slow SR.

The statisticians consulted for this project pointed out that this increase in variance indicates a relationship between mean token duration and variance—variance increases along with mean token duration. This relationship is thought to be due the slower SR allowing the participants more freedom in adjusting their sentence production to the display duration of the target sentence. Since the tokens were the first words in all stimuli sentences, they may have been more subject to variation than other parts of the sentence, with 'fine-tuning' adjustments coming later in the production after the participant had time to see whether their initially chosen SR for a given production would actually match the display duration or not.

A relation like this where the variance changes along with the dependent variable often indicates a logarithmic relationship between the two variables (Abacus Concepts 1994: 305). This type of relationship is quite commonly found in the natural sciences, where a period of slow increase is followed by rapid growth or vice versa. The following figure represents this type of relationship.

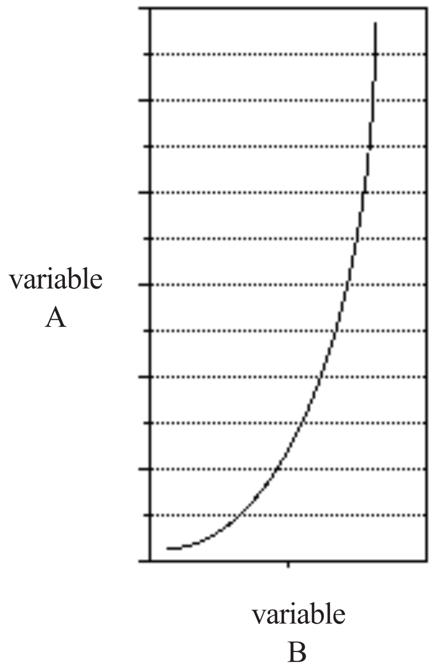

Figure 5.9 Representation of a logarithmic relationship between two variables; linear y-axis.

In order to show the relationship of the two variables represented on the x- and y-axis (in the example figures, variables A and B; in the current data set, variables SR and TOKEN DURATION), the data can be presented on a plot that has a logarithmic axis. In this type of plot, the logarithmic increases are accounted for by adjusting the scaling of the axis. This is equivalent to plotting the logarithmically adjusted values on linear y-axis, but it preserves the units (i.e. the units for the current data set will still be in ms, rather than in logarithm values). The example plot given above would look something like the following when presented on a plot with a logarithmic y-axis.

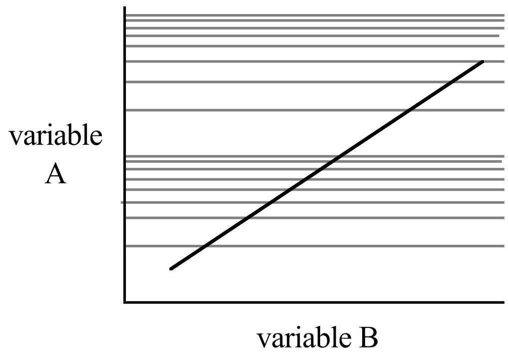

Figure 5.10 Representation of a logarithmic relationship between two variables; logarithmic y-axis.

On the consultants' advice, and as recommended in Abacus Concepts (1994: 305), the data was adjusted with the formula ADJUSTED VALUE = log (TOKEN DURATION). Descriptive statistics were then generated to check the variance of the adjusted values. The variance of the three SRs is given again in the table below, along with the variance of the log adjusted values for comparison.

Table 5.3 Descriptive statistics of the data, before and after logarithmic transformation.

	TOKEN DURATION **fast**	TOKEN DURATION **normal**	TOKEN DURATION **slow**
Mean	273	356	472
SD	39.7	60.1	69.1
⇒ ***Variance***	1574	3609	4768
Mean (log trans.)	2.432	2.546	2.669
SD (log trans.)	.065	.073	.065
⇒ ***Variance (log trans.)***	.0043	.0053	.0042

The adjustment of the variance by the log transformation can be seen in the last row of the table. Whereas the linear data showed a continually increasing mean and variance as the SR decreased, the log transformed data showed a more linear increase in mean and an increase and then decrease in variance as the SR decreased. The largest variance, that of the normal SR, is no longer twice as large as the variance of the other two data subsets.

Again, the information is perhaps easiest to see in the form of a box plot. The following box plot shows the same data given in Figure 5.8 plotted on a logarithmic y-axis.

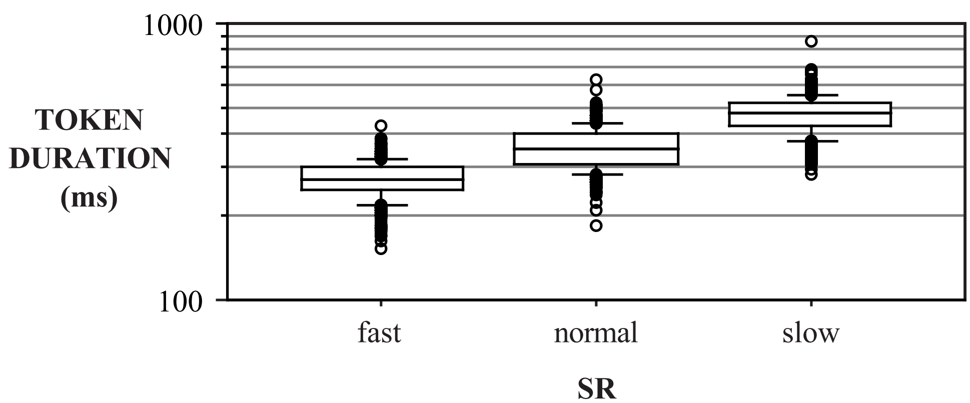

Figure 5.11 Box plot for TOKEN DURATION at each SR; linear axis.

The height of the boxes and the spread of the data points outside the boxes are now fairly uniform, indicating the variance of the three log-transformed data subsets is much more uniform (i.e. the distribution of the data in the three subsets is much more homogeneous).

However, the transformation is by no means completely satisfactory, either. If there were a true logarithmic relationship between the 3 data subsets, the variance of all 3 sets should be fairly equal. The larger variance of the normal data subset suggests that some other more complex relationship is at work here.

Also, it is suspected that this logarithmic increase in variance will not hold as SR slows even further. The most likely scenario is that once the stimuli sentences are displayed for a long enough duration, the differing variation in the distribution of the token duration data will level out for data collected at even longer stimuli sentence display durations. That is, it is expected that while the token durations will continue to fluctuate around the mean value within a given range; after a certain slow SR is reached the range over which the data fluctuates is expected to remain fairly constant as display duration is further increased. The data is therefore suspected to be best represented by some complex relationship (i.e. logarithmic up to a certain slow SR, then plateauing out to some fairly constant level of variance) that will need to be investigated in further research.

For the present study, there are two courses that further analysis can take. First, further statistical checks can be made on the logarithmically adjusted data, either with or without the inclusion of the outliers. This would ensure that the statistics generated by those tests would not have a hidden error due to differing variances of the data subsets. A second option, however, is to simply analyze all of the raw data and consider it unlikely that the differences in variance will cause an erroneous statistic to be generated.

Rosenthal & Rosnow (1991) state that while homogeneity of variance is an important assumption of ANOVA tests, the likelihood of erroneous results stemming from non-homogeneous variances is minimized by using samples of the same size. In their words,

> "Only if the two population variances are very different *and* if the two sample sizes are very different is the violation of this assumption [homogeneity of variance] likely to lead to serious consequences."
>
> Rosenthal & Rosnow (1991: 315); emphasis in the original

In addition, with the possible exception of the most extreme outlier in the slow data set, none of the productions are by any means performance errors; all productions were very normal sounding utterances. Their varying duration is simply a function of the inherent variability of the speech act. In addition, the differences in variance are caused by only one factor in the experimental setup, the factor of SR; all other factors are the same for all 3 data sets. Notably, the effects of this factor are incredibly large for both the raw data and the transformed data; the only difference in preliminary investigations using both raw and transformed data were in the effects of other, less significant factors.

The second option is therefore adopted in this thesis. The untransformed data will be used for all statistical checks and the plots showing the relationships being discussed. In the interests of completeness, however, the statistical checks were run a second time on the data with outliers removed, outliers again being defined as those points falling within the 95% of the data closest to the mean of that data set. Both results will be included in the discussion of the results.

5.3. DESIGN OF THE ANCOVA MODELS FOR TOKEN DURATIONS

This section details the effects that the variables to be discussed had on the mean token durations (**MTDs**), and the interaction among those variables.[3] Significant interactions will be judged at the 95% level of confidence.

Due to the exploratory nature of this investigation, it was decided to include multiple checks for various combinations of variables. Because of the relationship between PARTICIPANT and GENDER, checks could not be made on these two variables at the same time. The gender of the participants did not change during the study (and to the best of our knowledge still has not) so no comparisons involving both variables could be made.

Effects of the inherent segmental material in the tokens and the following syntactic clitic are not directly included in the following model. While there may well be interaction between the segmental material of the tokens themselves, the location of that material in either the first or second mora, and the syntactic clitic following a token (either *ka* or *to),* the interaction among all of these factors is quite difficult to interpret in the current study because the experimental setup did not adequately control these factors—the tokens used in the study did not cover all possible combinations of CV material in each mora and with each clitic. Therefore the only check that will be made involving segmental material will be for the factor TOKEN. All

[3]The effects of any two or more independent variables must begin with their interaction. For example, if one studies the effects of air temperature on human discomfort without taking into account humidity, it might seem as if people living in Tokyo are much more sensitive to heat than those living in Phoenix. Air temperature and humidity interact in their effect on discomfort; i.e. how much one affects human discomfort depends on the value of the other. The effects of heat must therefore be studied within the contexts of humidity and vice versa.

of the segmental effects mentioned above are therefore combined in this one variable. The interaction among the component effects will be left for further, more controlled research.

Since the number of devoiced vowels is suspected to have a significant effect on mean token duration, the variable #DEV was included in the models as well. As mentioned in §4.4, #DEV is actually a regressor in these models, and the models themselves therefore become *analysis of covariance* (ANCOVA) to reflect the fact that the effect of two dependent variables are being checked for effects on each other.

Finally, as was discussed in Chapter 3, because accented vowels are regularly being devoiced in the Tokyo dialect, and because the decision to devoice an accented vowel or to manipulate the accent location to avoid devoicing an accented vowel appears to be in free variation, the variable PITCH was not included in the models.

The ANCOVA model constructed by the statistical consultants for this project was therefore expanded to include both GENDER and PARTICIPANT. The two models run are as follows:

TOKEN DURATION = SR * TOKEN * #DEV + PARTICIPANT * BLOCK +
PARTICIPANT * REPETITION (BLOCK)

TOKEN DURATION = SR * TOKEN * #DEV + GENDER * BLOCK +
GENDER * REPETITION (BLOCK)

where SR * TOKEN * #DEV indicates all 3 variables by themselves and all possible combinations, PARTICIPANT * BLOCK indicates those two variables by themselves and the interaction between them, and PARTICIPANT * REPETITION (BLOCK) indicates those two variables by themselves and the interaction between them where REPETITION is nested within BLOCK (i.e. 3 repetitions were produced in each of the two repetition blocks).

SR, TOKEN and #DEV were grouped together since these are what is known as *fixed effects,* effects that are controlled in or defined by the experiment—an investigation is being made into the effects of these particular 3 SRs, these particular 10 tokens, and these particular 3 numbers of devoiced vowels (0, 1 or 2). On the other hand, PARTICIPANT and GENDER are combined with REPETITION and BLOCK because these are what is known as *random effects,* effects that are not specifically controlled by the experiment—an investigation is not being made into the effects of these particular 10 participants, but we wish to extend the results to the larger population of young Japanese speakers of the Tokyo dialect. Similarly, we are not interested in the effects of the 3 repetitions in each of these particular 2 repetition blocks, but we wish to extend the results to any number of such repetition blocks.

All models were run as multivariate repeated measures models, since the measured MTDs were measured 6 times for each participant in each experimental condition.

The following sections will discuss the significant effects found, their interpretation, and their implications for further research.

5.4 RESULTS OF THE ANCOVA MODELS FOR TOKEN DURATIONS

The results of both ANCOVA are given in the two tables below after non-significant variables and interactions were removed one by one.[4]

Table 5.4 Results of ANCOVA including the factor GENDER.

Variable(s)	***df***	***F*-ratio**	***p*-value**
SR	2	1700	.0001
TOKEN * #DEV	9	4.5	.0001
#DEV	1	210	.0001
TOKEN	9	7.6	.0001
GENDER	1	37	.0001
REPETITION (BLOCK)	5	5.6	.0001
residual[5]	1772		

[4]*F*-ratios are rounded to 2 significant digits; *p*-values are given as generated by the statistical program used for analysis, SuperAnova™.

[5]The residual of an ANOVA model is the source of variation that is not due to the variables themselves; i.e. the variation that is due to the individual participants and the experimental error in measuring those differences.

Table 5.5 Results of ANCOVA including the factor PARTICIPANT.

Variable(s)	*df*	*F*-ratio	*p*-value
SR * TOKEN * #DEV	20	2.2	.0019
TOKEN * #DEV	9	3.3	.0006
SR * TOKEN	18	2.1	.0059
SR	2	660	.0001
#DEV	1	85	.0001
TOKEN	9	5.6	.0001
PARTICIPANT * REPETITION (BLOCK)	45	2.5	.0001
PARTICIPANT	9	98	.0001
REPETITION (BLOCK)	5	8.7	.0001
residual	1681		

As can be seen in Table 5.4, when GENDER was included as a variable only the interaction between TOKEN and #DEV was significant (i.e. the *F*-ratio is accompanied by a *p*-value of less than .01). In other words, the effect that the number of devoiced vowels in the repetitions of a given token had on MTDs differed for each token. However, when PARTICIPANT was included as a variable in the model rather than GENDER, the 3-way interaction among SR, TOKEN and #DEV was significant, as well as the 2-way interactions between TOKEN and SR, TOKEN and #DEV, and PARTICIPANT and REPETITION (BLOCK). This means that the value of MTDs differed not only depending on the number of devoiced vowels in repetitions of a given token, but also differed at each SR for each TOKEN/#DEV combination. MTD value also differed not only for each participant, but also for at least some repetition blocks for at least some participants.

The difference in the significant interactions in the two models is due to the way the data is divided up in the two models when the statistical calculations are performed. Checks for effects of the variable GENDER cause the data to be split up into two groups, the data generated by the six males in the study and the data generated by the four females. Those two sub-groups are then further subdivided into groups corresponding to each of the other possible values of each other variable in the model. In contrast, checks for effects of the variable PARTICIPANT causes the data to be split up into 10 groups, one for each participant. Those 10 sub-groups of data are then further subdivided. The differences in how the data was grouped when the ANOVA calculations were performed cause the different statistical results.

It will also be noted here that removing the outliers from the data sets before running the models (see §5.2 for discussion) resulted in the *F*-ratio for SR increasing slightly, while the *F*-ratios for the other variables decreased slightly. The only significant change in all models is that in all models, REPETITION (BLOCK) was no longer significant; instead, the variable BLOCK was significant as a main effect ($F_{1,\ 1699} = 26$, $p = .007$ for the model including GENDER and $F_{1,\ 1608} = 20$, $p = .011$ for the model including PARTICIPANT). The significance of this will be discussed in §5.4.2.2 in conjunction with the effects of the interaction PARTICIPANT * REPETITION (BLOCK).

The results of the two ANOVA will now be discussed in conjunction with each other, beginning with the effect of the variable SR.

5.4.1 Effects on MTDs involving SR

As noted earlier, discussion of the effects of variables necessarily begins with the effects of any interactions. However, since SR is the variable being investigated in this chapter, was controlled externally by the program via the use of 3 stimuli sentence display times, and was significant by itself in the model including the variable GENDER, it is worth taking a look at the effects of this variable on its own.

5.4.1.1 Effects of SR on MTDs

As just mentioned, the effect of SR on MTDs is being used to determine the success of the program in generating 3 distinct data sets. In other words, it is of interest to determine whether or not the 3 display times used resulted in productions that had significantly different MTDs. The success of the program can be judged from the size of the effect of the variable SR, which reflects the differences between the participants' SRs as they attempted to match their productions of the carrier sentences to the display time of the stimuli sentences. Again, a longer production time corresponds to a slower SR, resulting in a longer MTD.

As can be seen from Tables 5.4 and 5.5 above, the effects of the variable SR on the MTDs in the model containing GENDER was, in a word, enormous. An *F*-ratio of 1700 seen in Table 5.4 is by itself quite large for any ANOVA model; the *F*-ratio of 660 seen in Table 5.5 is similarly quite large. In addition, the magnitude of these effects is much greater than that of any other variable or interaction in either model. This indicates that the computer program used to elicit tokens of varying duration was quite successful. That is, the display times of 1700, 2400 and 4000 ms corresponding to the fast, normal and slow SRs, respectively, elicited 3 different groups of tokens with highly significant different mean token durations.

The large effect of SR can be seen in the following *interaction line plot,* which shows the interaction between variables via joined line segments.

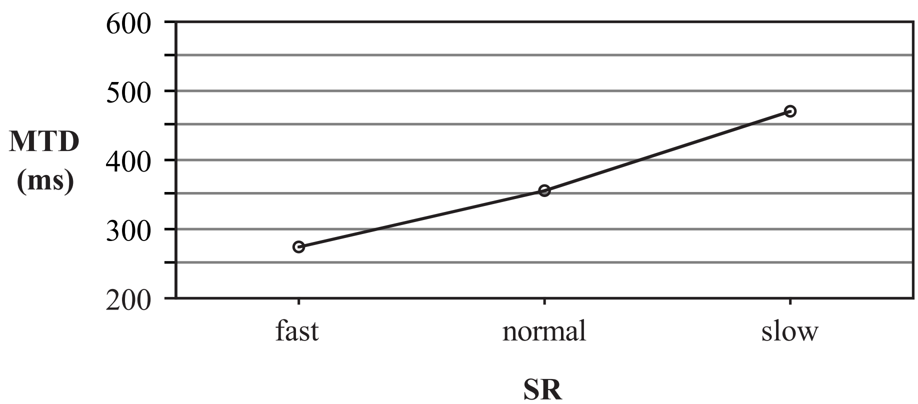

Figure 5.12 Interaction line plot for the effects of SR on MTD.

Each of the 3 data points is the mean token duration for all repetitions of all tokens at that SR; i.e. 600 repetitions at each SR. The large effect of SR, highly significant in all statistical checks, can be seen in the vertical separation of these 3 data points. This large effect of SR indicates that the computer program used to elicit the sentence repetitions, and hence the token repetitions, was quite successful at eliciting 3 distinct sets of data.

However, as previously discussed, it is also necessary to ensure that the other variables in the study did not compromise the effect of SR. The fact that the values of the 3 data points in Figure 5.12 above are the means of 600 data points obscures the difference in MTDs caused by these other variables. The combined effects of SR and these other variables can be observed in the significant interactions in Tables 5.4 and 5.5 above, and can be seen in interaction line plots like the one in Figure 5.12 above.

We will now turn to these interactions involving SR in the PARTICIPANT model. As will be seen, the effect of SR seen in Figure 5.12 above did vary depending on the other variables that were checked at the same time.

*5.4.1.2 Effects of the 3-way interaction SR * TOKEN * #DEV*

In the ANOVA model including the variable PARTICIPANT, SR interacted significantly with TOKEN and #DEV ($F_{20, 1681} = 2.2$, $p = .0019$). That is to say, the effect of SR on MTD was greater for some combinations of TOKEN and #DEV than for others. The effect of this 3-way interaction was not large, as indicated by the small *F*-ratio, but the chance of a similar size effect being found in similar data is quite high, as indicated by the small *p*-value.

Of more concern than the small size of the *F*-ratio of the effect of the interaction is the much larger size of the effect of SR itself; the *F*-ratio of the effect of SR ($F_{2, 1681} = 660$, $p \leq .0001$) is larger by an order greater than two magnitudes than the size of the effect of the interaction. Therefore the concern that the large effect of SR has ‘bootstrapped’ the effect of the interaction—that the enormous effect of SR has caused the effect of all other variables and interactions in the model to be artificially inflated due to the statistical calculations used to determine *F*. The interaction line plot showing the 3-way interaction therefore needs to be examined carefully when judging the significance of the interaction.

The differences in the effects of combinations of these 3 variables can be seen in the following interaction line plot.

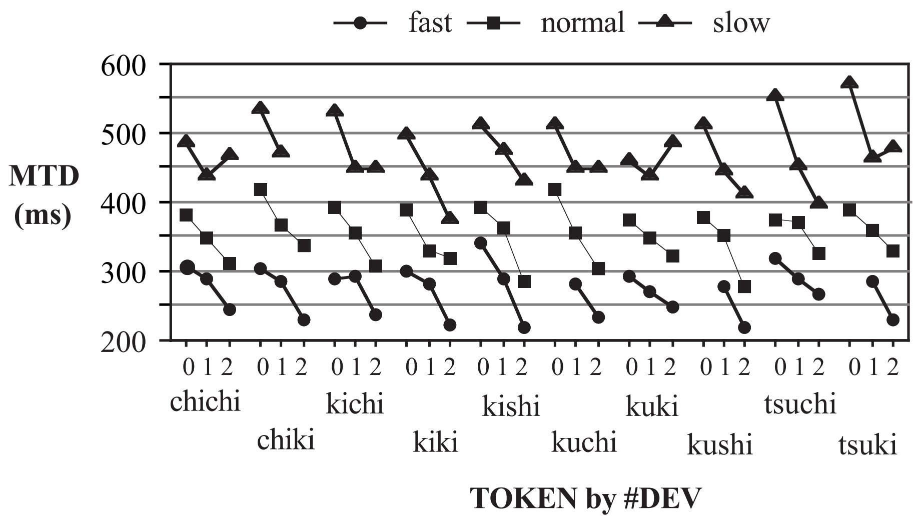

Figure 5.13 Interaction line plot for effects of SR * TOKEN * #DEV on MTD.

The effect of SR can be seen in the vertical separation of the 3 bands of segmented lines marked with the triangles (slow SR), squares (normal SR), and circles (fast SR). The effect of the different tokens can be seen in each column of 3 segmented lines, with tokens being marked on the x-axis. The effect of the differing number of devoiced vowels in the repetitions of a given token at a given SR can be seen in each segmented line: the left endpoint of each segmented line is the MTD of the repetitions containing 0 devoiced vowels, the middle data point of each segmented line is the MTD of the repetitions containing 1 devoiced vowel, and the right endpoint of each segmented line is the MTD of the repetitions containing 2 devoiced vowels. The number of devoiced vowels for each token is also marked on the x-axis.

It can be seen that no repetitions of *chiki* at the slow SR contained 2 devoiced vowels, and that no repetitions of *kuchi, kushi* or *tsuki* contained 0 devoiced vowels at the fast SR. Also, although the effects of shorter durations as the number of devoiced vowels increased held constant for most combinations of SR, TOKEN and #DEV, there were exceptions. For the repetitions of *chichi* at the slow SR, circled on the plot above, the MTD of the repetitions containing 2 devoiced vowels was actually longer than the MTD of repetition containing either 0 or 1 devoiced vowel.

However, it must also be noted that data points on the graph do not each correspond to the mean of the same number of repetitions. This is because not all participants devoiced the same number of vowels in a given token at a given SR. For the *chichi* values in the plot above (the 3 connected data points in the upper-most left corner), there were only 2 repetitions of *chichi* at the slow SR that contained 2 devoiced vowels. In contrast, there were 33 repetitions that contained 1 devoiced vowel and 25 repetitions that contained 0 devoiced vowels. Therefore the data point for repetitions of *chichi* at the slow SR containing 2 devoiced vowels is highly suspect—2 repetitions of anything are not enough to base any judgments on. Since 6 data points needed in each group when comparing group means; that number will be applied here.

This same is true for other combinations of SRs, tokens and number of devoiced vowels; there are 13 data points that represent fewer than 6 repetitions. The interaction plot for the 3-way interaction is presented again with these suspect data points removed.

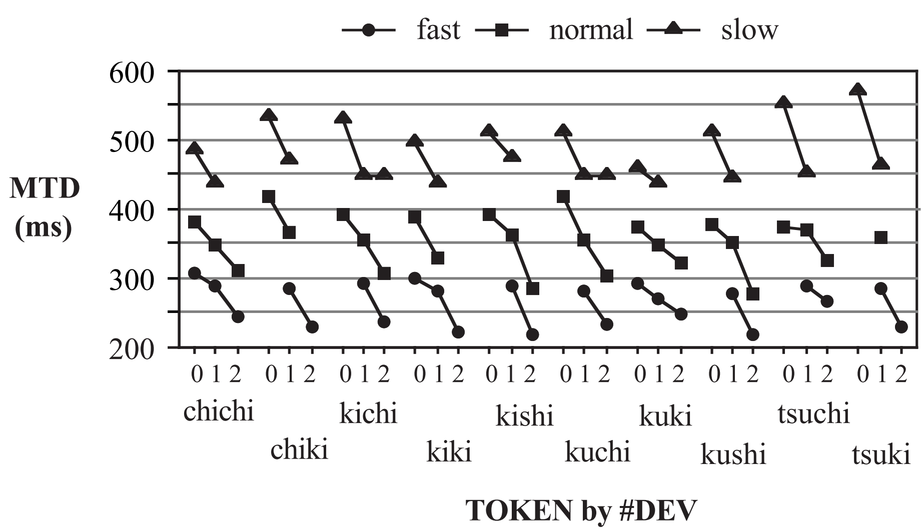

Figure 5.14 Interaction line plot for effects of SR * TOKEN * #DEV on MTD; means of < 6 data points removed.

Removing the data points representing fewer than 6 repetitions of a token at a given SR and number of devoiced vowels shows that in general the inverse relationship between number of devoiced vowels and MTD holds for the data set for all combinations of TOKEN and #DEV present in the data set.

The 3-way interaction between SR, TOKEN and #DEV in the ANOVA model containing the variable PARTICIPANT can be observed in the inconsistencies in the distribution of the data in Figure 5.9. However, taking into consideration the number of repetitions represented by each data point, combined with the small *F*-ratio in the model containing the variable PARTICIPANT, the lack of significant effect in the model containing the variable GENDER, and the much larger effect of SR in both models, it is judged that the significant 2-way interactions in the model containing the variable PARTICIPANT can be analyzed on their own. This is as opposed to analyzing the effects of each 2-way interaction within the contexts of each possible value of the other variable (i.e. the effects of SR * TOKEN for 0 devoiced vowels, 1 devoiced vowel, and 2 devoiced vowels).[6]

There are 3 cases where the stated relationship is not as strong, indicated by line segments with nearly 0 slope: *kichi,* slow SR, 1 and 2 devoiced vowels; *kuchi,* slow SR, 1 and 2 devoiced vowels; and *tsuchi,* normal SR, 0 and 1 devoiced vowels. Even though these variant data points do represent at least 6 repetitions (9, 13 and 11, respectively), the differences seen in those cases are not consistent for all SRs for the four tokens. It can be noted, however, that in all 3 of the cases the segmental material of the 2nd mora is the same; i.e. *-chi*. It remains for further research to see if this is the reason for the similarity between the repetitions of these tokens or not.

Because the 3-way interaction between SR * TOKEN * #DEV is no longer being considered, the ANOVA using the model including PARTICIPANT was continued by removing the 3-way interaction and any resulting non-significant interactions and variables. The new ANOVA tables for the models are given below.

---

[6]The usual procedure for analyzing 2-variable ANOVA results is to first check the interaction between the two variables to see if its effect was significant (e.g. whether the interaction between humidity and temperature is significant). If the interaction is significant, than each variable of the interaction is known as a *simple main effect,* and is checked for effects within the context of the other variable (e.g. the effect of humidity is checked at each value of temperature). If the interaction is not significant, than each variable is known as a *main effect,* and is checked for effects across all values of the other variable. In the current analysis this procedure has been extended to the 2-way interactions comprising the significant 3-way interaction.

Table 5.6 Results of ANCOVA including the factor GENDER; 3-way interaction not considered.[7]

Variable(s)	*df*	*F*-ratio	*p*-value
SR	2	2500	.0001
TOKEN * #DEV	9	5.0	.0001
#DEV	1	100	.0001
TOKEN	9	8.8	.0001
PARTICIPANT * REPETITION (BLOCK)	45	2.4	.0001
PARTICIPANT	9	97	.0001
REPETITION (BLOCK)	5	8.9	.0001
residual	1719		

The ANOVA table for the model including GENDER is repeated for the reader's convenience.

Table 5.7 Results of ANCOVA including the factor GENDER (repeated).

Variable(s)	*df*	*F*-ratio	*p*-value
SR	2	1700	.0001
TOKEN * #DEV	9	4.5	.0001
#DEV	1	210	.0001
TOKEN	9	7.6	.0001
GENDER	1	37	.0001
REPETITION (BLOCK)	5	5.6	.0001
residual	1772		

[7]Again, when the outliers are removed from the models, REPETITION (BLOCK) was no longer significant but instead BLOCK was ($F_{1,\ 1646} = 23$, $p = .0085$ for the model including GENDER and $F_{1,\ 1699} = 26$, p = .007 for the model including PARTICIPANT).

As can be seen from the tables, the results for SR and those involving TOKEN and #DEV are somewhat different in the 3 models; again, this is due to the different data grouping caused by the inclusion of the variables PARTICIPANT and GENDER. That is to say, there were larger differences in MTDs at each SR due to individual variation than due to gender differences; in the same fashion there were larger differences in MTD for each number of devoiced vowels in a token due to gender differences than due to individual differences.

The interaction line plots of the 2-way interactions in the above tables will now be examined, beginning with the interaction between TOKEN and #DEV.

5.4.1.3 Effects of the 2-way interaction between TOKEN and #DEV

The effects of the significant interaction common to both models, TOKEN * #DEV, can be seen in the interaction line plot given below.

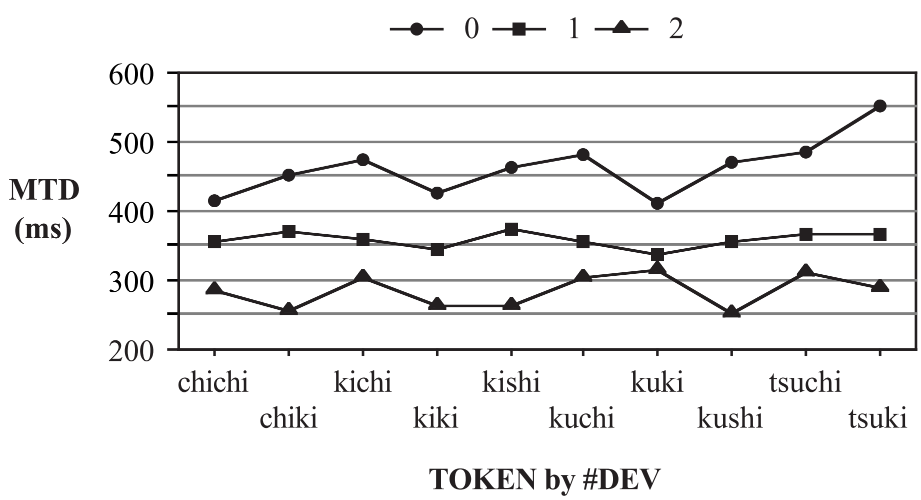

Figure 5.15 Interaction line plot for TOKEN * #DEV.

The effects of the number of devoiced vowels in a set of repetitions of a given token can be seen in the vertical separation of each set of 3 data points for each token. In this case, the data points marked with circles are those repetitions that contained no devoiced vowels, the data points marked with squares those repetitions that contained 1 devoiced vowel, and those marked with triangles those with 2 devoiced vowels. The effect of different tokens can be seen in the differences in relative location of the data points for each set of 3 points for each token. It is also worth noting that all data points on this plot represent MTDs of 9 or more repetitions of each token containing a given number of devoiced vowels; the previous concern about the data points representing the mean of too few repetitions is not applicable here.

It can be seen from the plot above that there were large differences in the MTDs of token repetitions with differing number of devoiced vowels. However, the data associated with each SR is still distinct in that no mean values for the 3 SRs at any given token overlap.

The differences also do not seem to be due to the segmental material of the tokens or the following clitics. This can be seen when the plot is presented again with the tokens segregated by 1st mora content, 2nd mora content, and content of the following syntactic particle.

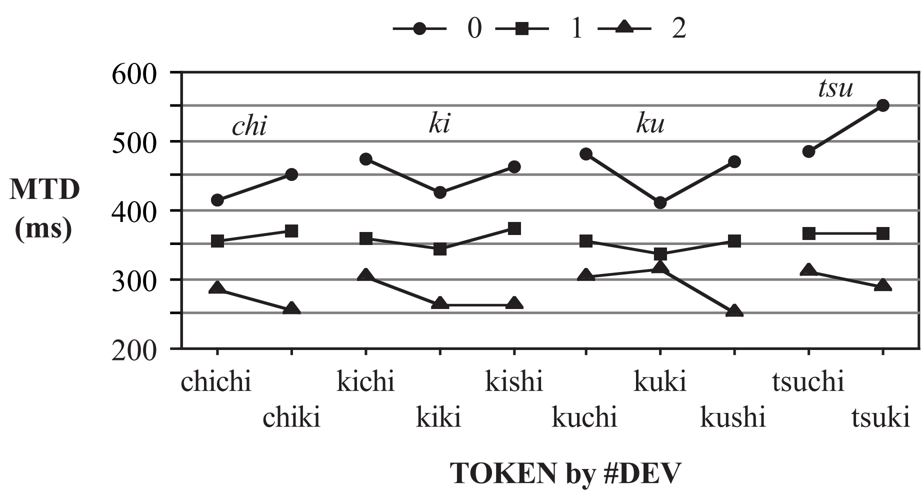

Figure 5.16 Interaction line plot for effect of TOKEN * #DEV on MTD, segregated by segmental content of the 1st mora.

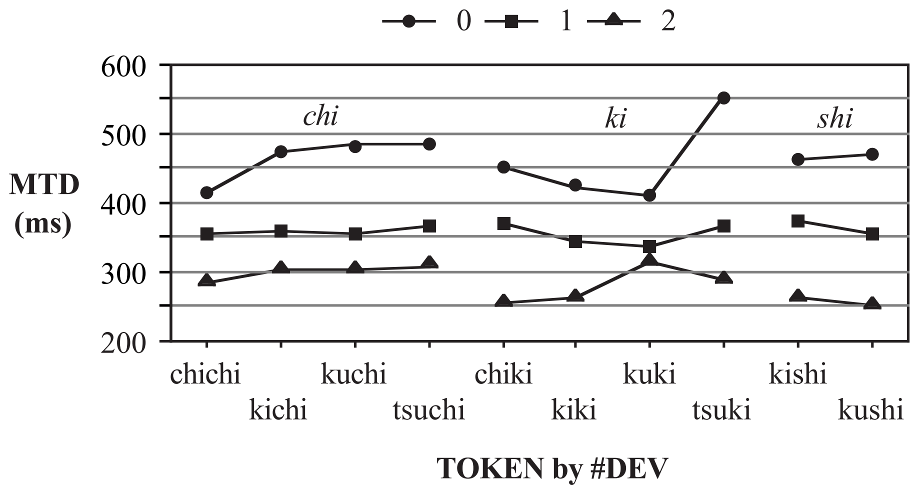

Figure 5.17 Interaction line plot for effect of TOKEN * #DEV on MTD, segregated by segmental content of the 2nd mora.

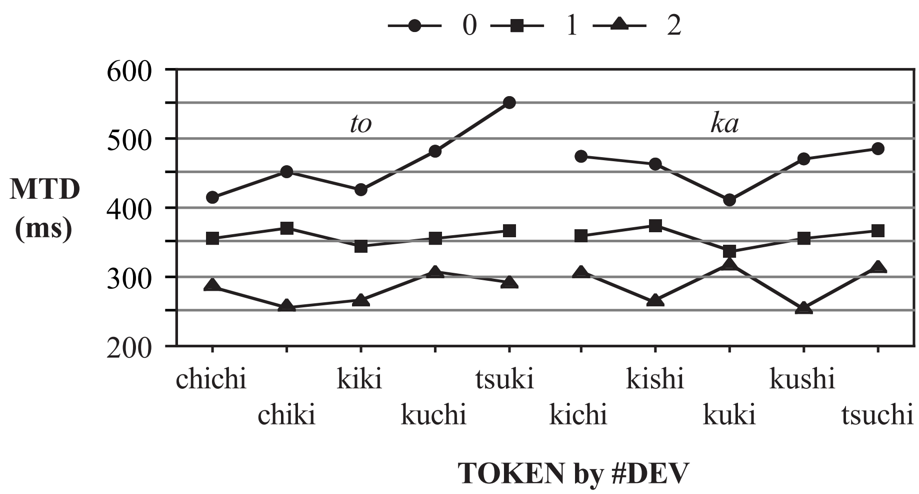

Figure 5.18 Interaction line plot for effect of TOKEN * #DEV on MTD, segregated by segmental content of the following syntactic particle.

There are similarities among the two groups (i.e. the fairly regular effect of the number of devoiced vowels in the tokens *kishi* and *kushi* which share the 2nd mora *shi)*, but the data do not appear to warrant any definite statements about the source of the variation due to the segmental material of the tokens and following clitics. Phonetically balanced stimuli are needed to pinpoint the source of the effects and so this is left for future work.

5.4.2 Effects on MTDs involving gender and participant

As noted earlier, the two models run produced somewhat different results depending on whether the included variable for the participants was GENDER or PARTICIPANT. The effects of these two variables will now be discussed.

5.4.2.1 Effects on MTD involving GENDER

As can be seen in Table 5.4, in the model containing the variable GENDER, both GENDER and REPETITION (BLOCK) (repetitions nested within blocks; see §4.4 for discussion) were significant as main effects; i.e. they both had a significant effect on MTD but they did not interact. The effect of GENDER ($F_{1, 1772} = 37$, $p \leq .0001$) was larger than the effect of REPETITION (BLOCK) ($F_{5, 1772} = 5.6$, $p \leq .0001$).

The effect of the variable GENDER on MTDs can be seen in the following interaction line plot.

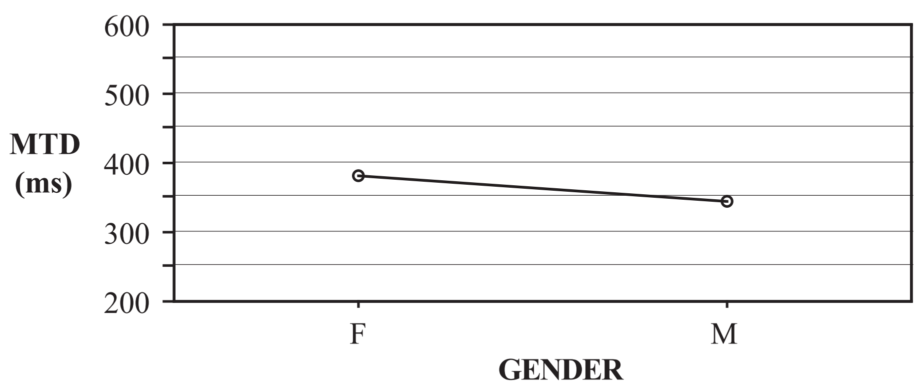

Figure 5.19 Interaction line plot for GENDER.

Note that the effect of GENDER shows a faster SR for the males than for the females. This is consistent with the findings of Yuen & Hubbard (1997), although the effect in the current data set does not seem to be as pronounced as in that work. This is thought to be due to the control on SR imposed by the program used for data elicitation. However, the effect is still small enough in comparison to the effect of SR so that the inclusion of both genders in the study did not in any way compromise the elicitation of 3 distinct data sets with 3 distinct MTDs.

5.4.2.2 Effects on MTD involving PARTICIPANT

The model containing the variable PARTICIPANT was somewhat more complicated, as can be seen from Table 5.6. In that model PARTICIPANT and REPETITION (BLOCK) interacted with each other; i.e. the effect of the different repetitions on MTD ($F_{45,\ 1719} = 2.4$, $p \leq .0001$) were different depending on which participant produced the repetitions. The interaction line plot for the PARTICIPANT * REPETITION (BLOCK) interaction is given below.

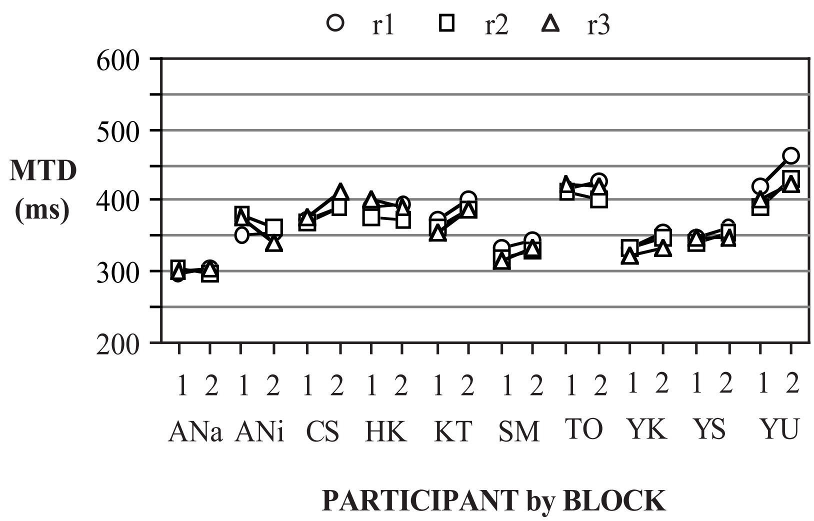

Figure 5.20 Interaction line plot for PARTICIPANT * REPETITION (BLOCK).

The effect of PARTICIPANT can be seen in each set of connected data points marked on the x-axis by the participants' initials. The effect of REPETITION (BLOCK) can be seen in the slope of the line segments in each set of points; the 3 data points on the left of each set are the MTDs of the 3 repetitions in the 1st repetition block, marked on the x-axis by a "1", and the 3 data points on the right of each set are the MTDs of the 3 repetitions in the 2nd repetition block, marked on the x-axis by a "2".

The source of the significant interaction can be seen by comparing each set of data points for each participant. In general, participants produced longer tokens in the 2nd repetition block. This can be seen in the increasing slopes for most of the line segments. However, participants ANa, ANi, HK and TO showed some variation, producing at least one set of repetitions of shorter mean duration in the 2nd block (i.e. the segments had decreasing slopes). As far as a consistent effect of REPETITION (BLOCK) is concerned, participants CS, KT and YS in particular consistently produced longer tokens in the 2nd repetition block.

The effect of differing token durations in each repetition block is thought to be due to the participants anticipating the 2nd repetition block. As the participants began the 2nd repetition block, the fact that it was the 2nd time they had performed this same task may have resulted in their having adopted some strategy for matching their productions to the sentence display time. Some participants seem to have generally slowed down their SR at the outset of their sentence productions, resulting in longer MTDs for the tokens at the beginning of the sentences, while other participants seem to have increased SR at the beginning of the sentence productions, resulting in shorter MTDs for the tokens.

However, the fact that there is no consistent effect from participant to participant indicates that the differences were due to individual variation, not experimental setup. The effect of the interaction is judged not to have compromised the production of 3 distinct data sets.

In the model containing the variable GENDER, REPETITION (BLOCK) did not interaction with any other variable but was instead significant by itself as a main effect. The effect of the variable REPETITION (BLOCK) on MTDs in the model containing the variable GENDER can be seen below.

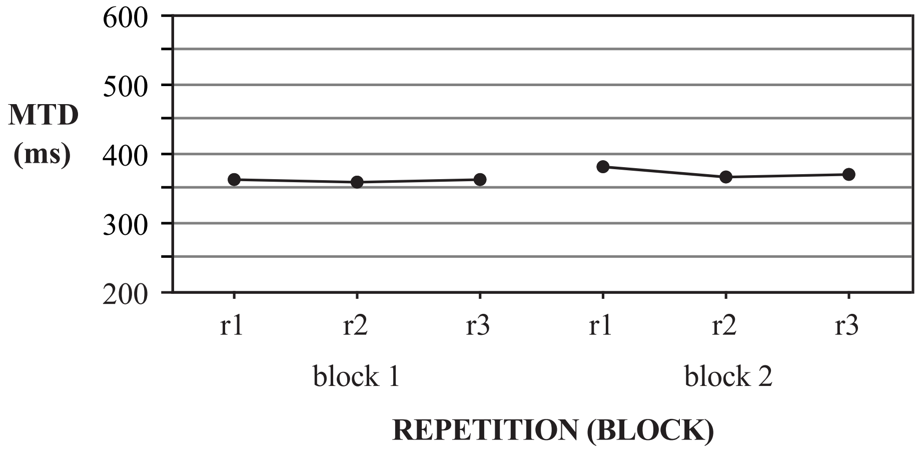

Figure 5.21 Interaction line plot for REPETITION (BLOCK).

The effect of REPETITION (BLOCK) is of some note in that there was a jump in the MTD at the beginning of the 2nd repetition block; i.e. the MTD of the first repetition in the 2nd block. It also appears that the MTDs of the repetitions in the 2nd repetition block were slightly longer than those in the 1st block. Whether or not this difference is statistically significant can be checked with what is known as a *post hoc* test for significance.

In a *post hoc* test for significance, all possible combinations of values of a variable (in this case, reps 1 and 2, reps 1 and 3, etc.) are compared to see how large the differences between the values are. In this way the source of significant differences can be pinpointed. The results of the Scheffé's *post hoc* S test for the variable REPETITION (BLOCK), as generated by the program StatView™, showed that none of the combinations of repetitions were close to the target .05 level of significance. In other words, although the slight increase in MTD for the 1st repetition of the 2nd repetition block can be seen on the plot, the difference between it and the MTD for any other repetition is not statistically significant.

The overall interpretation for the source of the variation caused by this variable is that the use of multiple repetitions segregated into two blocks, while noticeable in the increase in MTD of the 1st repetition of the 2nd block, did not compromise the generation of 3 distinct data sets. There do not appear to have been any significant learning effects as participants became more used to the experimental setup; there was no gradual increase or decrease in MTDs as the repetitions progressed through the 1st repetition block. This will be discussed more fully below along with the model containing the variable PARTICIPANT.

5.5 SUMMARY OF CHAPTER 5

The clear message from the results of the statistical checks for effects of the variables on MTD is the enormous effect of SR. That is to say, the use of three different display times for the stimuli sentences did produce 3 distinct data sets corresponding to the slow, normal, and fast SRs. Further, it was found that none of the other variables checked (TOKEN, #DEV, GENDER, PARTICIPANT, REPETITION (BLOCK)) compromised the production of these 3 data sets. While each of these did have a significant effect on MTDs, examination of the interaction line plots involving these variables showed that the effects were not strong enough to compromise the effect of SR—the effects of using different participants, genders, tokens, and repetitions segregated into 2 repetition blocks did not compromise the elicitation of 3 distinct data sets. In addition, the fact that there were differences in the number of devoiced vowels in the productions of all participants likewise did not compromise the data generation.

The use of different tokens in the study also affected MTDs, but the exact cause of the different effects (i.e. the differing segmental material in the first and second mora and the following clitic) is left for further, more controlled studies. In particular, the effect of the number of devoiced vowels in a given token repetition (#DEV) proved to be fairly consistent. The cause of the effect is thought to be a shortening of the duration of the mora a devoiced vowel is found in, with a resulting shortening of the overall token duration.

Also, the effect of the variable GENDER indicated slightly faster SRs by the male participants in the study. This is consistent with the finding of Yuen & Hubbard (1997). The effect of PARTICIPANT was likewise significant, and is attributed to individual variation.

Finally, the significant effect of REPETITION (BLOCK) seen in the first repetition of the second repetition block is thought to be due to participants' anticipation of producing another repetition block. However, the production strategy resulting from that anticipation appears to have differed from participant to participant—some began their productions at a faster SR, resulting in shorter duration tokens, while others began their productions at a slower SR, resulting in longer duration tokens.

The next chapter will examine the effect that changes in SR had on the frequency of the application of High Vowel Devoicing. It will examine how often vowels were devoiced, and the effect of the variables of the study on whether or not a vowel was devoiced.