CHAPTER IV: RESULTS

4.1 General Overview

Chapter three gave the most relevant information regarding the procedure of data collection. It presented the characteristics of the participants involved in the present study and the way in which recordings and ratings were carried out. One of the objectives of this chapter is to present the results obtained for the data analysis resulting from the scores given to NNESs' speech samples by NESs' raters in terms of intelligibility, comprehensibility and its relation with foreign accent.

For each dependent variable, the following information will be presented:

a) Statement of question /problem
b) Statement of null and alternative hypothesis
c) Output of statistical test
d) Statement of t-score and its significance
e) Interpretation of the Result Stating the rejection or not of the null hypothesis
f) Summary of the Intelligibility Results

Finally, the results concerning the correlation existing (if any) between comprehensibility and foreign accent will be presented.

4.2 Intelligibility Scores

4.2.1 Statement of Question/Problem

The research question regarding the improvement of intelligibility is cited below from Chapter 1:

> *Will students from the experimental group be more intelligible at time 2 than at time 1 compared to students from the control group?*

In order to be able to answer this question paired t-tests were used for the data obtained from the control and experimental group. One of the assumptions underlying this question was that there would be an improvement in terms of intelligibility in the speakers of the experimental group, who are the ones who received the explicit pronunciation instruction over a period of 12 weeks. On the contrary, and since the participants from the control group did not receive any type of pronunciation training, little to no improvement was expected in terms of intelligibility from the pre-test to the post-test for them.

4.2.2 Homogeneity of Both Groups before the Experiment in terms of Intelligibility

In order to assure that the experimental and control groups were comparable at the beginning of the study a t-test was carried out. This t-test compared the scores obtained during the pre-test of both groups. The hypotheses for this test were the following:

Null Hypothesis: The mean intelligibility scores for the pre-test of the control and experimental group are the same. H_o: $\mu_{pre\text{-}control} - \mu_{pre\text{-}experimental} = 0$

Alternative Hypothesis: The mean intelligibility scores for the pre-test of the control and experimental group are different. H_a: : $\mu_{pre\text{-}control} - \mu_{pre\text{-}experimental} \neq 0$

Table 7 shows the samples of the control (18 students) and the experimental (16 students) groups. It also shows the mean intelligibility scores for both groups, where it can be observed that the estimated difference is -0.16, indicating that both groups were homogeneous and that any improvement in terms of intelligibility can be attributed to the presence of pronunciation training.

The third column shows the standard deviation, which shows how spread out the data is from the mean. As observed the high scores indicate that data are spread along the curve.

	N	Mean	StDev
Pretest Control	18	82.7	21.8
Pretest Experimental	16	82.8	16.9
Estimate for difference: -0.159236 95% CI for difference: (-13.906171, 13.587699) Tc = 2.04 (critical value for t) Ts= -0.02 (obtained t-score) DF = 32			

Table 7 – Two-Sample t-test and Confidence Interval for Mean intelligibility scores of the Control and Experimental Group during the pre-test

In the case of a two-tailed decision, if the t-score obtained is higher than the critical value for *t*, the null hypothesis should be rejected. If the absolute

value of the t-score obtained (-0.02) is 0.02 and this one is lower than the critical t-value (2.04), the null hypothesis cannot be rejected. *Since the t-score obtained is smaller than the critical value for t, there is no difference between the mean intelligibility scores of the control and the experimental group during the pre-test*. Therefore, a comparison within groups can be carried out to see if the intelligibility scores remain the same during the pre and post-test (in the case of the control group) or if there was any improvement (in the case of the experimental group) as a result of lack or presence of pronunciation training.

The following section presents the results obtained after comparing the mean intelligibility scores collected during the pre and post-test for the control group.

4.2.3 Intelligibility Scores of Control Group

As a reminder to the reader, the intelligibility task consisted of orthographical transcriptions of each audio stimulus. As expected, the five speech samples produced by the NESs got perfect intelligibility scores. The mean intelligibility scores resulted from the adding of each score divided among the 8 listener-rates. It was used a 100-scale, where 100 equals 100% intelligible and 0 means that the speakers was not intelligible at all.

The intelligibility scores for the participants of the control group during the pre-test ranged from 62.6% to 100%. However, during the post-test, the scores ranged from 5% to 98.8%. In terms of intelligibility, a surprising decrease can be noticed. This affirmation is made under the observation of the mean scores from the pre-test (82.68%) and the post-test (69.48%), with a difference of 13.2%.

Not very surprisingly, the 3 speakers who got the highest scores during the pre-test were the same who got the highest scores during the post-test.

4.2.3.1 Statement of Null and Alternative Hypothesis

The hypotheses for this statistical test can be stated as follows:

Null hypothesis: there is no difference among the mean scores of the pre-test and those of the post-test. H_o: $\mu_{pre} - \mu_{post} = 0$

Alternative hypothesis: There is a difference in the mean scores of the group between the pre-test and the post-test. H_a: $\mu_{pre} - \mu_{post} \neq 0$

Since the students from the control group did not receive explicit pronunciation instruction, it was expected to see the same mean scores during the pre-test and post-test. And because I cannot be certain whether the results from the post-test will be better or worse than those from the pre-test, the procedure for testing the null-hypothesis for the control group (only) requires a two-tailed decision.

The following table shows us the t-test carried out on the mean intelligibility scores of the control group.

	N	Mean	StDev
Pretest	18	86.82	11.47
Posttest	18	69.43	28.82
Difference	18	17.39	29.77
95% CI for mean difference: (2.59, 32.20)			
Tc= 2.11 (critical value for *t)*			
Ts= 2.48 (obtained t-score)			

Table 8 - Statistical Test and Confidence Interval for Mean Intelligibility Scores of the Control Group

Table 8 shows that the control group had a sample of 18 students. The second column presents the mean intelligibility scores during the pretest (86.88) and the posttest (69.43), as well as the difference between the tests' mean scores (17.39). The following column, under the heading of standard deviation, indicates the spread of the data around the mean score. As observed in the data from this column, high scores point out that the data is spread out along the curve, especially for the scores of the post-test.

With 95% confidence, the true mean difference between the two tests falls between 2.59 and 32.20 values. This shows that the mean intelligibility score related to the pretest is higher than the mean intelligibility scores of the posttest. These results, not only indicate that the mean scores for the pre and post-test are not the same, but also that the intelligibility scores during the post test decreased.

The interpretation of the t-score, which will be used in order to reject or accept he null hypothesis, is presented in the following figure.

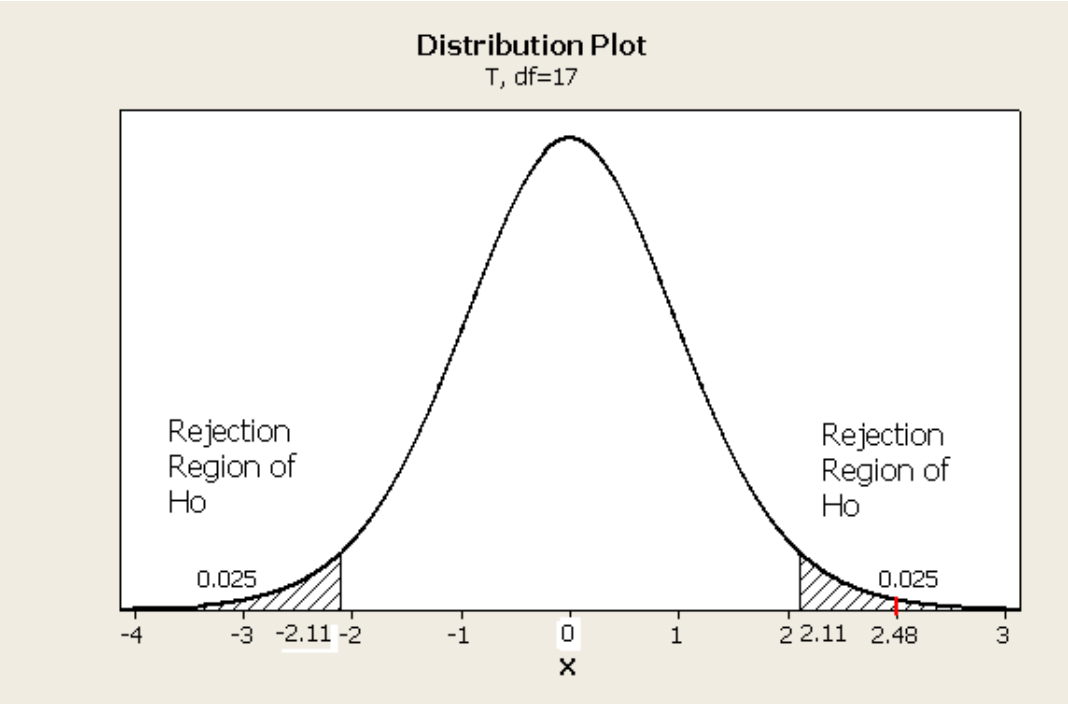

Figure 3 – Distribution Plot of 2-tailed t-test Intelligibility Scores for the Control Group

The figure above shows the results of the statistical test carried out for the control group. With a degree of freedom of 17 we have a critical value for *t* of 2.11 at the 95% confidence level (α= .05). According to the procedure for interpreting the results of the t-test, the null hypothesis should be rejected if t_s is higher than t_c. Since the *t* score yielded is 2.48, which is higher than t_c (2.11), the obtained result is statistically significant and the null hypothesis is rejected, which means that *the mean intelligibility scores for the control group are different between the pre and post-test*. From the t-score obtained I can also observe that the scores during the post-test are lower.

4.2.3.2 Conclusion for Intelligibility and Control Group

As shown in Table 8 and Figure 3, the results are statistically significant and we must *reject the null hypothesis*, which states that the mean intelligibility scores of the control group are the same during the pretest and the posttest. Therefore, it can be said that *the intelligibility scores found during the pretest were not the same as the ones from the posttest*, something which was not expected. However, the results also indicate that there was not only no improvement in terms of intelligibility in the students from the control group but rather a worsening. Although an improvement was not expected, but rather similar scores in both tests, it was not contemplated to observe a worsening in terms of intelligibility.

4.2.4 Intelligibility Scores of Experimental group

The same procedure applied to the data from the control group, was applied to the scores obtained in terms of intelligibility of the speakers from the experimental group. The intelligibility scores for the participants of the experimental group during the pre-test ranged from 50% to 98.8%, which, in comparison with the speakers from the control group, is significantly lower (62.6% and 100% respectively). This shows us, that the participants from the experimental group were less intelligible overall, than the ones from the control group at the beginning of the study. However, statistically speaking and as shown from the simple t-test both groups were still comparable at the beginning of the study.

On the other hand, during the post-test the scores ranged from 46.5% to 98.5%. It is noticeable that the scores obtained from the latter are higher in respect to the scores from the control group.

4.2.4.1 Statement of Null and Alternative Hypothesis

Null hypothesis: there is no difference between the mean scores of the pre-test and those of the post-test. H_o: $\mu_{pre} - \mu_{post} = 0$

Alternative hypothesis: students will score higher on the post-test than on the pre-test. H_a: $\mu_{pre} - \mu_{post} < 0$

A one-tailed decision will be taken into account for this Hypothesis test, since, as described in the alternative hypothesis I am expecting to observe an improvement in terms of intelligibility during the post-test in the students of the experimental group.

The following table shows us the t-test carried out on the mean intelligibility scores of the experimental group.

Table 9 shows that the experimental group had a sample of 16 students. The second column presents the mean intelligibility scores during the pretest (82.23) and the posttest (81.02), as well as the difference in scores from one test to the other. With this, it is observed that the mean intelligibility score from the pretest is slightly higher than the one obtained during the posttest. The column under the heading of standard deviation indicates how far the data is from the mean score. As observed in the data from this column, the high scores point out that the data is spread out along the curve.

	N	Mean	StDev
Pretest	16	82.84	16.89
Posttest	16	81.11	14.78
Difference	16	1.72	17.11
95% upper bound for mean difference: 9.22913 T_c= 1.75 (critical value for *t)* T_s= 0.40 (obtained t-score)			

Table 9– Statistical Test and Confidence Interval for Mean Intelligibility Scores of Experimental Group

With 95% confidence, the true mean difference between the two test results falls below 9.22. This indicates that the there is not enough information to say that the intelligibility scores during the posttest were higher than those from the pretest. The interpretation of the obtained t-scores, which will be used in order to reject or accept the null hypothesis, is presented in Figure 4.

Figure 4 shows the results of the statistical test carried out for the experimental group. With a degree of freedom of 15 we have a critical value for t of 1.75 at the 95% confidence level (α=.05) . Since the statistical test showed a t-score of 0.40, which falls below the critical value of t (0.40 < 1.75), the result is not significant (i.e., it falls outside the rejection region of Ho), the null hypothesis must be accepted, which means that the mean intelligibility scores obtained during the pre and post-test were the same.

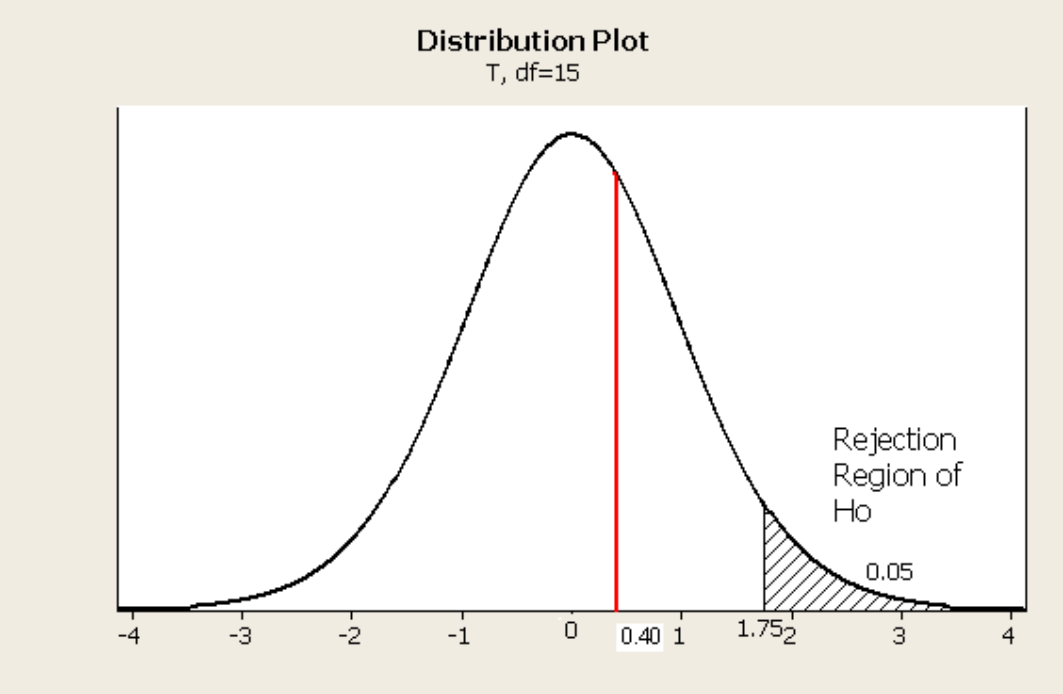

Figure 4 - Distribution Plot of 1-tailed t-test intelligibility Scores for the experimental group

4.2.4.2 Conclusion for Intelligibility and Experimental Group

As shown in table 9 and Figure 4, the results are not statistically significant. For this reason, *the null hypothesis fails to be rejected*, which means that the mean intelligibility scores of the experimental group during the post test are not higher than the scores obtained during the pretest. This indicates that there was not an improvement in terms of intelligibility in the students from the experimental group.

4.2.5 Summary of the Intelligibility Results

After having examined the intelligibility scores obtained by students from the experimental group, it was observed that, contrary to my expectations, the experimental group did not show any improvement of this variable. Therefore, I was not able to perform a 2-sample t-test, as previously planned, in order to show that the students from the experimental group had improved in terms of intelligibility compared to the students from the control group. Only if the experimental group had shown an improvement on this variable for the within-group t-test, it would have been possible to compare it across groups to the control group.

On the contrary, however, it was observed that both the experimental and the control groups received lower scores during the post-test. Even though the scores obtained by the control group during the post-test were much lower than those of the experimental group post-test, it was not my intention to use a t-test to find out which group's decrease in scores was less worse.

4.2.6 Orthographic Transcriptions

The orthographic transcriptions of the NESs were completely free of errors, indicating that the directions were clearly understood by the listener-raters and that the quality of the audio files was clear and good. As a reminder to the reader, orthographic transcriptions of short audio stimulus of NNESs and NESs were made by NESs. The purpose of this task was to see how intelligible NNESs were, the more accurate the transcription was, the more intelligible the speaker.

A frequency of the various types of transcription errors cannot be presented with numbers since the collected data does not lend itself to give it a quantitative treatment; instead, a qualitative data analysis of the types of errors and the possible reasons underlying each one will be explained. Two subsections will be found below used to classify the types errors are: chunk predictions and perception of –ed in regular verbs.

4.2.6.1 Chunk Phrases

This type of error is related to the action of predicting what the speaker is saying just by paying attention to the word in context. More specifically, to the ability of listener-raters as NESs to guess the words uttered after listening to the whole audio file. This was very common with the use of some prepositions and verb tenses.

Regarding the use of prepositions, the following example is presented:

Speaker 3: I am planning to stay <u>in</u> home with my family
Listener 1: **<u>Stay at home</u>** with my family
Listener 2: I plan to stay in home with my family
Listener 3: I am planning to **<u>stay at home</u>** with my family
Listener 4: I wanted to stay in home with my family
Listener 5: I wanted to **<u>stay at home</u>** with my family
Listener 6: I am going to stay in home with my family
Listener 7: I am going to stay in home with my family
Listener 8: I am going to stay in home with my family

Table 10– Orthographic transcription of Audio file no. 3

As shown in Table 10, 3 out of 8 listeners wrote *stay at home*, instead of *stay in home*. The use of the preposition *at* could be triggered by the use of *stay*, which is a chunk phrase.

The following example is also related to the use of prepositions. The use of the preposition *of* by the speaker could be a result of L1 interference.

I don't see anyone of my family much
Listener 1,2,3,6,8: I don't see **anyone in my family** much
Listener 4, 5: I don't see **anymore of my family** much
Listener 7: I don't see **anyone of my family** wort

Table 11 – Orthographic Transcription of Audio File no. 52

As observed, it is grammatically correct to say *anyone in my family* rather than *anyone of my family*.

The following transcriptions are related to the cases where the pronoun predicted the use of certain verbs or auxiliaries. The frequency of this error is 3 out of 8.

Speaker 21: My sister is married she have one son
Listener 1: My sister is my … **she has** one son
Listener 2,3,4: My sister is married, **she has** one son
Listener 5: My sister is myreed **she has** one son
Listener 6, 8:My sister is married **she have** one son
Listener 7: My sister is married **she had** one son

Table 12– Orthographic Transcription of audio file no. 21

As Table 12 shows, it is more likely that the auxiliary verb ‘has’ would be following the third person singular and not ‘have’, which was the word used by the speaker.

The next error transcriptions also present transcription errors as a result of phrase clusters and the use of the present perfect.

All of my brothers <u>are get</u> married
Listener 1, 3,6: All of my brothers **<u>have get</u>** married
Listener 2,4,5,8: All of my brothers are get married
Listener 7: All of my brothers **<u>are got</u>** married

Table 13- Orthographic transcription of audio file no. 01

As the majority of these cases have shown, it is assumed that the raters did not necessarily understand each word uttered by the speakers, which leaves room to question how trustworthy it is to use NESs as raters, especially for an intelligibility task as the one used in the current study. As a reminder to the reader, NESs first listened to the audio file and then proceeded to make the transcription, not forgetting that as described in chapter 3 these audio files were short enough to avoid memory problems. As a result, if the listeners could not understand the function words, such as prepositions and auxiliaries, but if they understood the content words with no problem, they might have guessed the function words used according to the content words of the utterance.

4.2.6.2 Perception of ‘-ed’ in Regular Verbs

During the transcription analysis, it was also noted the differences in production on behalf of speakers and the transcription made by NESs regarding the ‘–ed’

of the past tense of regular verbs. There were 9 utterances that included the use of a regular verb in the past tense.

There were two situations in which the perception and transcription of 'ed' were present. The first one is related to the writing of an utterance with a regular verb in past tense, even though this was not produced by the listener. This case was noticeable in 5 out of 9 of these utterances:

The speaker said:	The listener-rated transcribed (frequency):
I *try* to do exercise	I *tried* to do exercise (1/8)
In my last vacations I'm *visit* to my family in Tlaxcala	In my last vacations I *visited* my family in Tlaxcala (1/8)
I *study* Psychology because I like the human mind	I *studied* Psychology…(2/8)
When I *start* the university I stopped the gym	When I *started* the university… (4/8)
I don't know we *search* for an activity	I don't know we *searched* for an activity (3/8)

Table 14 –Perception of –ed when it was not produced by speakers

By looking at some of these examples the reader might get the impression that these error transcription were driven by the triggering of some content word that indicated the use of the past tense, such as *In my last vacation* or *when*. However, not all of the utterances have such content words in the utterance. In fact, there is not enough information within the same utterance to make such an inference. Also, the listener-raters were not informed about the topics that the speakers had to talk about. In this sense, they did not have any information of the content of the audio files but still, heard an '-ed' where there was not any.

The second situation of the perception of '-ed' is related to the actual production of this segment by the speakers and the lack of perception on the part of the listeners. The transcription of what the speaker said and what the listeners transcribed are presented below.

The speaker said:	The listener-rated transcribed (frequency):
Recently I *stayed* in my house	Recently I *stay* in my house (8/8)
I *used* to spent like three hundred dollars	I *use* to spent like three hundred dollars (1/8)
In my during last vacations I *worked*	In my during last vacations I *work* (3/8)
When I start the university I *stopped* the gym	When I start the university I *stop* the gym (3/8)

Table 15– No perception of –ed when produced by speakers

As can be seen from table 15, the four verbs used by the speakers in past tense contain the use of the phoneme /t/ (stopped and worked) and /d/ (used, stayed) as variations of the '-ed'. None of them include the production of /Id/, which let us to assume that due to the similarity of the orthographic transcription and the production of these verbs (i.e., wanted, visited), there is no transcription error of the spoken form and its transcription. In other words, verbs which are read the same way they are written do not cause any pronunciation problems to Spanish speakers, they are straightforward.

Overall, the error transcriptions demonstrated by the listeners can be labeled into different categories such as: omissions of function words, substitutions of words that did not alter the meaning of the utterance, and addition of function words.

The question underneath the whole issue of intelligibility remains: Can we assume these transcription errors are due to a lack of intelligibility? Could the listener get the main idea of the utterance? This can only be known through the data analysis of comprehensibility and its correlation to intelligibility, which is presented in the following section.

The error transcriptions presented above may give the impression that speakers were not intelligible and this may lead to the conclusion that they were not comprehensible either. In this regard, it may sound logical to think that if listener-raters transcribed the utterances incorrectly (i.e., with a lot of mistakes), as a consequence the speaker would have a low score on intelligibility. However, it is important to understand that intelligibility and comprehensibility are two concepts that although they co-exist, they don't necessarily relate to one another because, as I explained in chapter 2, intelligibility refers to the speaker's ability to identify the words within an utterance while comprehensibility stands for the ability to understand the main idea of the words uttered. Therefore, the listener-rater may not have been able to identify each word as spoken by the speaker, but he/she may have realized that the speaker was talking about (i.e. the number or children the speaker's sister has). Thus, comprehensibility scores rely on the speakers' ability to decode the message. The results concerning the analysis of the comprehensibility scores are presented in the following section.

4.3 Comprehensibility Scores

The present section shows the results obtained by the speakers in terms of comprehensibility. In this sense, comprehensibility is defined as the subjective assessment of ease or difficulty of a message (Derwing, Munro, and Wiebe, 1998). The results within this variable are presented according to the group to which the speakers belonged: the control group or experimental group. In a second part, a t-test will be carried out in order to compare the improvement made by the speakers of the control and experimental group.

As a reminder to the reader, the speakers were rated in terms of comprehensibility with the help of a 4-level Likert scale (1-very easy to understand, 2-a bit difficult to understand, 3-very difficult to understand, 4-impossible to understand). It is important that the reader of this document has this in mind when interpreting the tables below. This translates to the following rule: the lower the score, the better the performance of speakers which means better comprehensibility. For this reason, it was expected to observe 3s and 4s during the pretest and 1s and 2s (which are lower scores) during the posttest. Throughout the following section, I will follow the same organization as above, presenting for each dependent variable, the following information:

a) Statement of question /problem
b) Statement of null and alternative hypothesis
c) Output of statistical test
d) Statement of t-value and its significance
e) Interpretation of the Result Stating the rejection or not of the null hypothesis

f) Summary of the Comprehensibility Results

4.3.1 Statement of the Problem/Question

It is expected that the students from the control group will not attain any improvement in terms of comprehensibility in relation to the pre-test. In other words, equal scores during both the pre and post-test are expected to be found. On the contrary, a significant improvement is expected to be observed in the speakers of the experimental group during the post-test as a result of the explicit pronunciation instruction they received.

The research question that will be answered from this data analysis is cited below:

Will students from the experimental group be more comprehensible at time 2 than at time 1 compared to the students from the control group?

Even though the students from the control group were not exposed to an explicit pronunciation instruction, it is important to know how they scored in terms of comprehensibility. These results will help determine whether or not the students improved, or if they scored the same in the post-test and the pre-test.

4.3.2 Homogeneity of Both Groups in terms of Comprehensibility

In order to see whether the experimental and control group were comparable at the beginning of the study a t-test was carried out. This t-test compared the

scores obtained during the pre-test of both groups. The hypotheses for this test were the following:

Null Hypothesis: The mean comprehensibility scores for the pre-test of the control and experimental group are the same. H_o: $\mu_{pre\text{-}control} - \mu_{pre\text{-}experimental} = 0$

Alternative Hypothesis: The mean comprehensibility scores for the pre-test of the control and experimental group are different. H_a: : $\mu_{pre\text{-}control} - \mu_{pre\text{-}experimental} \neq 0$

Table 16 shows the samples of the control (18 students) and the experimental (16 students) groups. It also shows the mean intelligibility scores for both groups, where it can be observed that the estimate difference is -0.000903, indicating that both groups were homogeneous at the beginning of the study and that any improvement in terms of intelligibility can be attributed to the presence of pronunciation instruction.

	N	Mean	StDev
Pretest Control	18	1.767	0.524
Pretest Experimental	16	1.768	0.632
Estimate for difference: -0.000903 95% CI for difference: (-0.411016, 0.409210) T_c= 2.04 (critical value for *t)* T_s= -0.00 (obtained t-score) DF = 29			

Table 16 – Two-Sample t-test and Confidence Interval for Mean Comprehensibility scores of the Control and Experimental Group during the pre-test

The third column shows the standard deviation, which shows how spread out the data is from the mean. As observed the high scores indicate that data are spread along the curve.

In the case of a two-tailed decision, if the critical value for *t* is higher than the t-score obtained, the null hypothesis should be accepted. If the absolute value of the t-score obtained (-0.00) is 0.0 and this one is lower than the critical t-value (2.04), the null hypothesis cannot be rejected. *Since the t-value obtained is smaller than the critical value for t*, *there is no difference between the mean comprehensibility scores of the control and the experimental group during the pre-test.* Therefore, a comparison within groups can be carried out to see if the comprehensibility scores remain the same during the pre and post-test (in the case of the control group) or if there was any improvement (in the case of the experimental group) as a result of lack or presence of pronunciation training.

The following section presents the results obtained after comparing the mean comprehensibility scores collected during the pre and post-test for the control group, followed by the comprehensibility scores obtained by the participants of the experimental group.

4.3.3 Comprehensibility Scores of Control Group

The range of mean comprehensibility scores per speaker went from 1 to 2.625 during the pre-test. On the other hand, the range of mean comprehensibility scores during the post test ranged from 1.125 to 4. The mean comprehensibility score for this group during the pre-test was 1.763, and during the post-test, 2.229. Overall, and without having carried out any statistical test, it is noticeable

that the mean comprehensibility scores during the post test are higher than those of the pre-test, which means that there was no improvement in terms of comprehensibility for the control group.

4.3.3.1 Statement of Null and Alternative Hypothesis

Null hypothesis: there is no difference among pairs of measurements in the population (i.e., student scores will not differ from the pretest to the posttest). H_o: $\mu_{pre} - \mu_{post = 0}$

Alternative hypothesis: There is a difference in the mean scores of the group between the pre-test and the post-test. H_a: $\mu_{pre} - \mu_{post} \neq 0$

As mentioned earlier (section 4.2.3.1) due to the design of the null and alternative hypothesis for the control group (only) the following Hypothesis testing requires a two-tailed decision.

	N	Mean	StDev
Pretest	18	1.764	0.525
Posttest	18	2.229	0.923
Difference	18	-0.465	1.061
95% CI for mean difference: (-0.993, 0.063) T_c= 2.11 (critical value for *t*) T_s = -1.86 (obtained t-score)			

Table 17- Statistical Test and Confidence Interval for Mean Comprehensibility Scores of Control Group

Table 17 shows that the control group had a sample of 18 students. The second column presents the mean comprehensibility scores during the pretest

(1.764) and the posttest (2.229), as well as the difference in scores from one test to the other (-0.465). The following column under the heading of standard deviation indicates how far the data is from the mean score. As observed in the data from this column, the high scores point out that the data is spread out along the curve. However, compared to the intelligibility scores presented above, these scores are much closer to the mean, hence, there is less spread along the curve.

With 95% confidence, the true mean difference between the two tests falls between -0.993 and 0.063 values. The interpretation of the t-score, which will be used in order to reject or fail to reject the null hypothesis, is presented in the following figure.

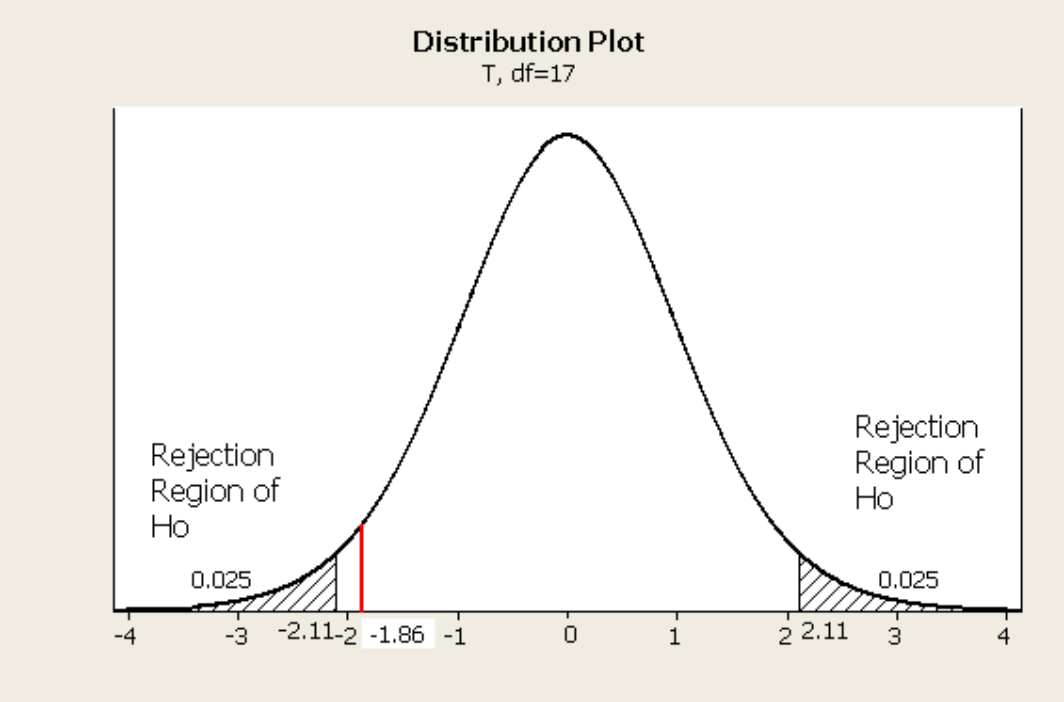

Figure 5 - Distribution Plot of 2-tailed t-test Comprehensibility for the Control Group

Figure 5 shows the results of the statistical test carried out for the control group. With a degree of freedom of 17 we have a critical value for t of 2.11 at the 95% confidence level (α=.05). According to the procedure for interpreting the results of the t-test, the null hypothesis should be rejected if t_s is higher than t_c. Since the *t* score yielded is -1.86, which is lower than t_c (2.11) the null hypothesis is not rejected, which means that can be said that the mean intelligibility scores for the control group are not different during the pre and post-test.

4.3.3.2 Conclusion for Comprehensibility and Control Group

As shown in table 17 and Figure 5, there is not enough confidence to say that the results are statistically significant and we must accept *the null hypothesis*, which means that the mean comprehensibility scores of the control group during the post test are not different to the scores obtained during the pretest. This indicates that the performance of the participants did not vary from the pretest to the posttest. Statistically speaking the students from the control group performed the same during both tests, which is an expected result since this group of participants did not receive any pronunciation training.

4.3.4 Comprehensibility Scores for Experimental Group

The scores obtained during the pre-test per speaker ranged from 1 (easy to understand) to 2.75 (closer to 3 – very difficult to understand). The mean comprehensibility score for the pre-test was 1.76, which tells us that students were not really incomprehensible before they were instructed in pronunciation.

During the post-test, after receiving the explicit pronunciation instruction, the mean scores per speaker ranged from 1(easy to understand) to 3.125 (very difficult to understand). Overall, the mean comprehensibility score during the post-test was 1.97, indicating no improvement in comprehensibility.

4.3.4.1 Statement of Null and Alternative Hypothesis

The hypotheses for the statistical test for the comprehensibility scores of the participants of the experimental group are the following:

Null hypothesis: there is no difference among pairs of measurements in the population (i.e., student scores will not differ from the pretest to the posttest). H_o: $\mu_{pre} - \mu_{post} = 0$

Alternative hypothesis: students will score higher on the pre-test than on the post-test. H_a: $\mu_{pre} - \mu_{post} > 0$

A one-tailed decision will be taken into account for this Hypothesis test, since, as described in the alternative hypothesis I am expecting to observe an improvement in terms of comprehensibility during the post-test in the students of the experimental group. It is worth noting here that an improvement in terms of comprehensibility will be translated in lower scores during the post-test.

	N	Mean	StDev
Pretest	16	1.766	0.632
Posttest	16	1.977	0.567
Difference	16	-0.211	0.653
95% lower bound for mean difference: -0.497 T_c= 1.75 (critical value for *t*) T_s=-1.29 (obtained t-score)			

Table 18 - Statistical Test and Confidence Interval for Mean Comprehensibility Scores of Experimental Group

Table 18 shows that the experimental group had a sample of 16 students. The second column presents the mean comprehensibility scores during the pretest (1.766) and the posttest (1.977), as well as the difference in scores from one test to the other (-0.211). The following column under the heading of standard deviation indicates how far the data is from the mean score. As observed in the data from this column, the high scores point out that the data is spread out along the curve.

With 95% confidence, the true mean difference between the two test results falls above -0.497. This indicates that the there is not enough information to say that the comprehensibility scores during the posttest were lower than those from the pretest, indicating no improvement in terms of comprehensibility. The interpretation of the t-score obtained, which will be used in order to reject or accept the null hypothesis, is presented in Figure 6.

Figure 6 shows the results of the statistical test carried out for the experimental group. With a degree of freedom of 15 we have a critical value for *t* of 1.75. Since the statistical test showed a t-score of -1.29, which falls below the critical value of *t* (-1.29 < 1.75), the result is not significant (i.e., it falls outside the rejection region of Ho), the null hypothesis must be accepted. Therefore it cannot be said that there was an improvement in terms of comprehensibility for the students in the experimental group. Hence, statistically speaking the mean comprehensibility scores during the pre-test and post-test were the same.

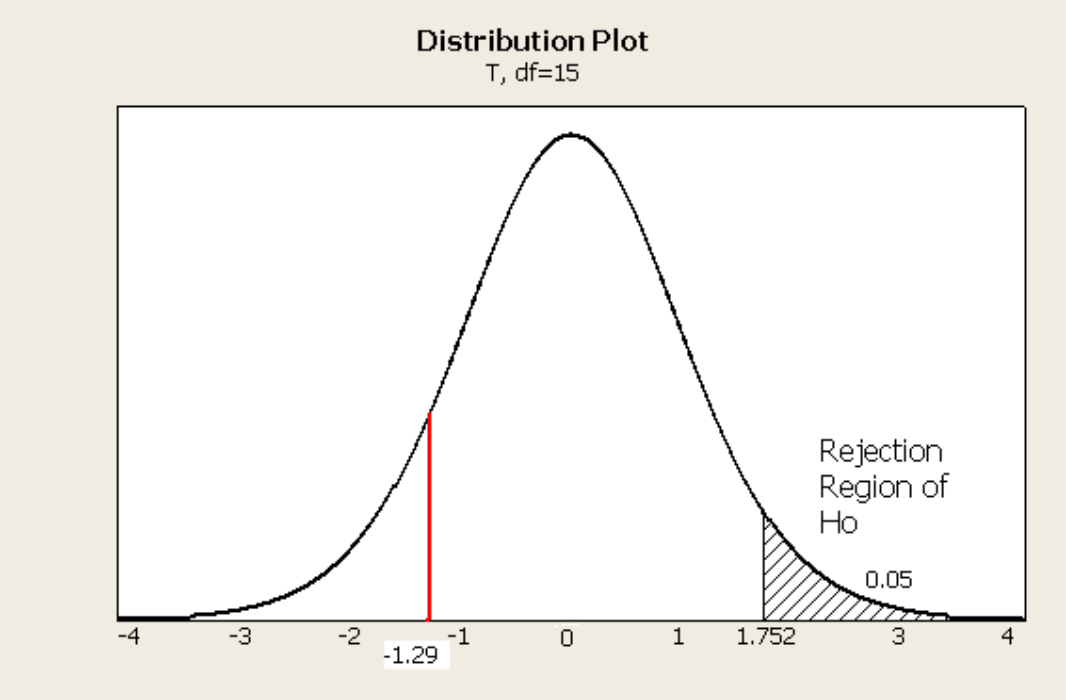

Figure 6 - Distribution Plot of 1-tailed t-test Comprehensibility for the Experimental Group

4.3.4.2 Conclusion for Comprehensibility and Experimental Group

As shown in Table 18 and Figure 6, contrary to my expectations, the results are statistically significant to accept *the null hypothesis*, which means that the mean comprehensibility scores of the experimental group during the post test are not different from the scores obtained during the pretest.

4.3.5 Summary of the Comprehensibility Results

Similarly to the discussion in section 4.2.5, which refers to the lack of improvement observed in terms of intelligibility in the students of the experimental group, contrary to my expectations, I was not able to perform a two sample t-test, previously planned, in order to show that the students from the experimental group had improved in terms of comprehensibility compared to the students from the control group.

So far, each of the dependent variables of this study has been analyzed separately. The main objective of this analysis was to show if there was an improvement in terms of intelligibility and comprehensibility from the pretest to the posttest, with ~~a~~ special attention given to the performance of the experimental group. Although it was not the aim of the current study to see if there was a reduction of the perceived foreign accent after explicit pronunciation training, the following section presents the results obtained in perceived foreign accent. The reason why I have decided to present it is twofold. First, I will be able to contrast and compare my results to those of Derwing et al. (1998), in which an improvement of foreign accent was perceived in the students of the

three groups (global, segmental and no treatment). Second, this data will be useful in order to find a correlation between this variable and comprehensibility.

4.4 Perceived Foreign Accent

Foreign Accent is presented in this section, in order to show if there was any improvement for the students of the experimental group as a consequence of the pronunciation instruction. However, the present section will not answer directly any of my research questions. The reason why I decided not to focus on the reduction of foreign accent was due to my belief that having a foreign accent does not affect comprehensibility. Besides, in my pronunciation instruction I never consider teaching either segmentals or suprasegmentals in order to reduce my students' foreign accent, but to improve intelligibility. However, its results will be useful to determine the correlation existing between the latter and comprehensibility.

4.4.1 Statement of the Problem

Derwing et al. (1998) found that foreign accent decreased in the speeches of their participants, who were enrolled in a full-time ESL program studying in a University in Canada. Among the three groups that participated in their study, it was found that all of them, even the students from the group that had no explicit pronunciation training, had reduced their perceived degree of foreign accent.

Foreign Accent scores were elicited by the same NESs who rated comprehensibility. They used a 4-point scale to perform such task where 1 – no

foreign accent, 2- mild foreign accent, 3- strong foreign accent and 4- very strong foreign accent. As can be seen from the scale, the lower the score the better. Therefore it was expected to see lower scores during the post-test.

The scores obtained during the pre-test per speaker ranged from 1.75 (mild foreign accent) to 3.13 (closer to 3 – strong foreign accent). It was noticeable that none of the speakers during the pre-test got a score of 4 (very strong foreign accent). The mean comprehensibility score for the pre-test was 2.50.

During the post-test, the mean scores per speaker ranged from 1.75 (close to mild foreign accent) to 3.88 (very strong foreign accent). Overall, the mean foreign accent score during the post-test was 2.55, indicating no improvement.

The following section presents the results of the two-sample t-test carried out with the purpose of establishing that the experimental and the control groups were homogeneous.

4.4.2 Homogeneity of Both Groups in terms of Perceived Foreign Accent

The two-sample t-test carried out included the analysis of foreign accent scores obtained during the pre-test from the experimental and the control group. The hypotheses for this test were the following:

Null Hypothesis: The mean accentedness scores for the pre-test of the control and experimental group are the same. H_o: $\mu_{pre\text{-}control} - \mu_{pre\text{-}experimental} = 0$

Alternative Hypothesis: The mean accentedness scores for the pre-test of the control and experimental group are different. H_a: : $\mu_{pre\text{-}control} - \mu_{pre\text{-}experimental} \neq 0$

Because it is expected to see that the groups are homogeneous, the null hypothesis should be accepted. The results of the t-test yield the following results:

	N	Mean	StDev
Pretest Control	18	2.503	0.410
Pretest Experimental	16	2.408	0.591
Estimate for difference: 0.094653 95% CI for difference: (-0.267965, 0.457271) T_c= 2.04 (the critical value for *t*) T_s= 0.54 (obtained t-score) DF = 26			

Table 19 – Two-Sample t-test and Confidence Interval for Mean Foreign Accent scores of the Control and Experimental Group during the pre-test

As can be observed from Table 19, it describes the number of participants of the control group (18) and the experimental group (16). It also presents the mean foreign accent scores obtained by the control group (2.503) and the experimental group (2.408). Likewise, this table shows how the foreign accent scores are spread along the curve through the standard deviation. As observed from table 18, the difference between the mean foreign accent scores is statistically significant which means that both groups were in equal conditions at the beginning of the experiment. Therefore, any improvement in terms of

accentedness will most likely be due to the presence of pronunciation training (in the case of the experimental group)

In the case of a two-tailed decision, if the critical value for *t* is higher than the t=score obtained, the null hypothesis should be rejected. Since the absolute value of the t-score obtained is 0.54 and it is lower than the critical t-value (2.04), the null hypothesis cannot be rejected. *This means that the mean accentedness scores of the control and the experimental group during the pre-test are the same*. Therefore, a comparison within groups can be carried out to see if the foreign accent scores remain the same during the pre and post-test (in the case of the control group) or if there was any improvement (in the case of the experimental group) as a result of lack or presence of pronunciation training.

The following section presents the results obtained after comparing the mean foreign accent scores obtained during the pre and post-test for the control group.

4.4.3 Foreign Accent Scores for Control Group

4.4.3.1 Statement of Null and Alternative Hypothesis

The hypotheses for the statistical test for the accentedness scores of the participants of the control group are the following:

Null hypothesis: there is no difference among pairs of measurements in the population (i.e., student scores will not differ from the pretest to the posttest).

Ho: $\mu_{pre} - \mu_{post} = 0$

Alternative hypothesis: There is a difference in the mean scores of the group between the pre-test and the post-test. H_a: $\mu_{pre} - \mu_{post} \neq 0$

	N	Mean	StDev
Pretest	18	2.503	0.097
Posttest	18	2.550	0.138
Difference	18	-0.047	0.155
95% CI for mean difference: (-0.373, 0.279) T_c= 2.11 (critical value for *t*) T_s= -0.31 (obtained t-score)			

Table 20 - Statistical Test and Confidence Interval for Mean Foreign Accent Scores of Control Group

Table 20 shows that the control group had a sample of 18 students. The second column presents the mean comprehensibility scores during the pretest (2.503) and the posttest (2.550), as well as the difference in scores from one test to the other (-0.047). The following column under the heading of standard deviation indicates how far the data is from the mean score. As observed in the data from this column, the high scores point out that the data is spread out along the curve.

With 95% confidence, the true mean difference between the two groups fall between -0.373 and 0.279. This shows that there is not evidence to suggest that any of the mean scores is higher than the other, which means that they are the same. From this table the result of the t-test can also be observed.

The interpretation of the t-score, which will be used in order to reject or fail to reject the null hypothesis, is presented in the following figure.

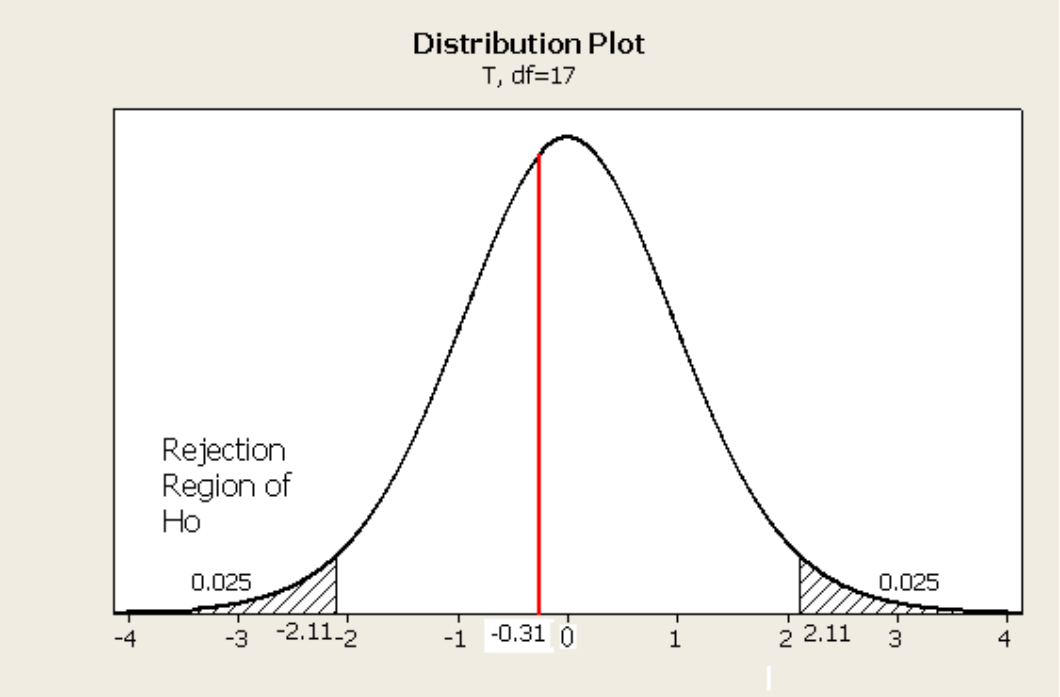

Figure 7- Distribution Plot of 2-tailed t-test Foreign Accent for the Control Group

Figure 7 shows the results of the statistical test carried out for the control group. With a degree of freedom of 17, we have a critical value of t of 2.11. Since the statistical test showed a value of -0.31, which falls outside the rejection region of Ho, the null hypothesis must be accepted. Statistically speaking, the mean accentedness scores obtained during the pre-test and the post-test are the same, which means that there was no improvement for the students of the control group.

4.4.3.2 Conclusion for Foreign Accent and Control Group

As shown in Table 20 and Figure 7, *the null hypothesis fails to be rejected*, which means that the mean foreign accent scores of the control group during the post test are the same as the scores obtained during the pretest. *In the case of the control group,, this is an expected result* because the students of the control group did not receive any type of pronunciation training.

4.4.4 Foreign Accent Scores for the Experimental Group

The scores obtained during the pre-test per speaker ranged from 1.25 (no foreign accent) to 3.38 (closer to 3 – strong foreign accent). The mean comprehensibility score for the pre-test was 2.40.

During the post-test, after receiving the explicit pronunciation instruction, the mean scores per speaker ranged from 1.75(close to 2- mild foreign accent) to 3.25 (strong foreign accent). Overall, the mean foreign accent score during the post-test was 2.54, indicating no improvement in foreign accent. However, the mean scores from the pretest and the posttest are not significantly apart from each other, there is a difference of 0.131.

In order to support the statement posed earlier, a paired t-test was carried out.

4.4.4.1 Statement of Null and Alternative Hypothesis

The hypotheses for the statistical test for the accentedness scores of the participants of the experimental group are the following:

Null *hypothesis*: there is no difference among pairs of measurements in the population (i.e., student scores will not differ from the pretest to the posttest). H_o: $\mu_{pre} - \mu_{post} = 0$

Alternative hypothesis: students will score higher on the pre-test than on the post-test. H_a: $\mu_{pre} - \mu_{post} > 0$

Since the use of the scale indicates that the higher the score the worse, if students from the experimental group obtain lower results during the post test (1- no foreign accent or 2-mild foreign accent), it will mean that they improved in terms of perceived foreign accent.

The results obtained after carrying out the paired t-test are presented in the table below.

	N	Mean	StDev
Pretest	16	2.408	0.591
Posttest	16	2.541	0.099
Difference	16	-0.1331	0.0969
95% lower bound for mean difference: -0.3030 T_c= 2.11 (critical value for *t*) T_s = -1.37 (obtained t-score)			

Table 21 - Statistical Test and Confidence Interval for Mean Foreign Accent Scores of Experimental Group

Table 21 shows that the experimental group had a sample of 16 students. The second column presents the mean comprehensibility scores during the pretest (2.408) and the posttest (2.541), as well as the difference in scores from one test to the other (-0.1331). The following column under the heading of standard deviation indicates how far the data deviates from the mean score. As observed in the data from this column, the low scores point out that the data is close to the mean score, which means that almost all the participants of this group scored the same.

With 95% confidence, the true mean difference between the two groups fall above -0.3030. This shows that there is not evidence to state that any of the mean scores is higher than the other, which shows that there was not any improvement in terms of foreign accent.

The interpretation of the t-value, which will be used in order to reject or fail to reject the null hypothesis, is presented in the figure below.

Figure 8 shows the results of the statistical test carried out for the control group. With a degree of freedom of 15 and a one-tailed decision we have a critical value for *t* of 1.752. Since the statistical test showed a t-score of -1.37, which falls outside the rejection region of Ho, the null hypothesis is accepted. Statistically speaking, the mean accentedness scores during the pre-test and the post-test are the same, which leads us to the conclusion that there was no improvement in terms of accentedness for the speakers of the experimental group despite having received pronunciation training.

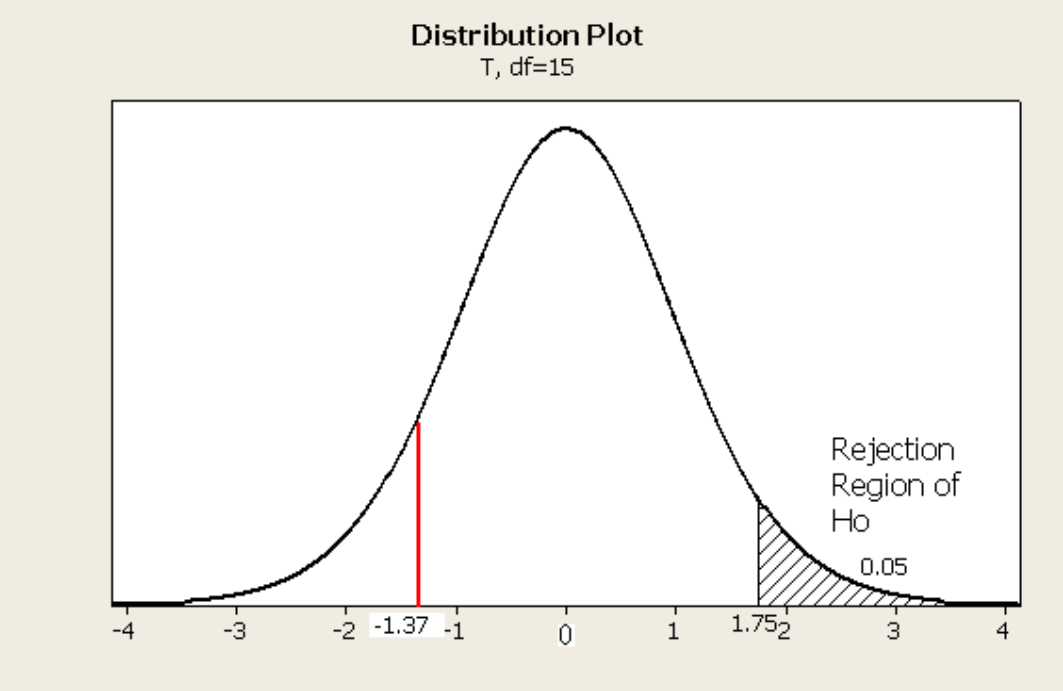

Figure 8- Distribution Plot of 1-tailed t-test foreign accent for the experimental group

4.4.4.2 Conclusion for Foreign Accent and Experimental Group

As shown in Table 21 and Figure 8, *the null hypothesis fails to be rejected*, which means that the mean foreign accent scores of the experimental group during the post test are not different from the scores obtained during the pretest, showing no improvement.

As observed from this analysis, neither the control nor the experimental group showed any improvement in terms of perceived foreign accent during the post-test. Like I mentioned earlier it was not a goal for my study to find such an improvement here. However, these results only support my idea of believing that Derwing et al (1998)'s participants showed an improvement due to the

constant input they received for living in a country where the target language is spoken.

Finally, the last section of this chapter will answer the third question of research. This section follows a similar format to the results presented for each of the dependent variables acknowledged as: intelligibility and comprehensibility.

4.5 Correlation Between the Dependent Variables

4.5.1 Statement of the Question/Problem

This section will show by an analysis of the data collected if there is a correlation between the dependent variables of comprehensibility and foreign accent. This analysis derives from the last research question of this study.

Does the degree of foreign accent affect the experimental and control group's comprehensibility?

For this question, the independent variable of time (pre-test and post-test) was not considered to be an important factor that would determine the relationship between foreign accent and comprehensibility. This question seeks support for the belief that a person with a strong foreign accent can be comprehensible. Therefore, a negative correlation between the degree of perceived comprehensibility and foreign accent is expected to be found. This

would be translated to speakers rated as being very easy to understand and rated as having a very strong foreign accent.

4.5.2 Relationship between Comprehensibility and Foreign Accent

As mentioned earlier, time was not considered an important factor that would help determine an existing correlation. Therefore, the mean scores of the pre-test and the post-test scores of comprehensibility and foreign accent of the control group were used to calculate the Pearson r coefficient yielding an r= 0.878, demonstrating a positive linear correlation between comprehensibility and foreign accent for this group. This means that when the speaker was easy to understand s/he was also rated as having no-foreign accent or mild foreign accent. Likewise, if the speaker was very difficult or impossible to understand (being graded as a 3- Very difficult to understand or 4-impossible to understand), it was likely they had received a similar score in the perception of foreign accent (3- strong and 4-very strong foreign accent).

Likewise, a linear correlation between the independent variables according to the mean comprehensibility and accentedness scores of the experimental group yielding an r= 0.834, was found. Once again, this demonstrates that there was a positive correlation among the variables; if the comprehensibility score improves so does the score related to foreign accent and vice versa.

Therefore, it can be concluded that there is a relationship between these variables; if the comprehensibility scores go up or down, so does the accentedness score. This translates into a high correlation, which does not

support the hypothesis I stated above where I suggested that there was a negative correlation between foreign accent and comprehensibility. As a consequence, my hypothesis is not supported by these results.

In sum, Chapter 4 has presented the most relevant findings yielded after analyzing the data obtained from this experiment. No improvement in terms of intelligibility or comprehensibility for the speakers of the experimental group during the post-test was found. However, regarding the last research question, a strong correlation between comprehensibility and accentedness was observed. In chapter 5, I will discuss the implications of these findings; I will also present a discussion relating these results to the ones obtained by Derwing et. al (1998).