An Acoustical Analysis of a Japanese Speaker's Production of English /r/ and /l/

Susan Kuzniak Kinnaird and Jennifer Zapf
Indiana University

Abstract

In general, native Japanese speakers have difficulty perceiving the English /r/ and /l/ phonemes due to the fact their native language does not have these two sounds as contrasting phonemes (Logan et al., 1991; Lively et al., 1993, 1994). Although much has been written on L1 Japanese with regards to the English /l/ and /r/, little has addressed the acoustical differences between speakers of Japanese and speakers of English as they produce the English liquids. This paper discusses an experiment in which these acoustical differences were described and analyzed. The study investigated the differences between the second and third formants produced by a native speaker of Japanese and a native speaker of English as they pronounced a series of words containing either an /r/, an /l/, or both. The position of the liquid within the word was also taken into consideration. The study found substantial differences between the F3 values for /l/ in every word position, and smaller differences between the F2 values of /l/ and the F3 values for /r/. This evidence provides support for the idea that perception and production may be closely linked and, thus, calls for most acoustical analysis of the productions of native Japanese speakers.

An Acoustical Analysis of a Japanese Speaker's Production of English /r/ and /l/

Susan Kuzniak Kinnaird and Jennifer Zapf
Indiana University

Introduction

The Contrastive Analysis Hypothesis predicts that second language learners will have the most trouble with phonemes that do not exist in their native language(s). This has been borne out in research on L1 Japanese speakers' production and perception of the English /r/ and /l/. It is extremely difficult for most native speakers of Japanese to distinguish between the English liquids /r/ and /l/, both in listening to native speakers of English and in trying to produce the sounds themselves (Logan et al., 1991, Bradlow et al, 1997, Riney et al., 2000). Because of this difficulty, numerous studies have been done on the perception and production of the English /l/ and /r/ by native speakers of Japanese. "In the pedagogical literature that addresses Japanese ESL pronunciation, the English and Japanese liquids have probably received more attention than any other segments" (Riney & Anderson-Hsieh, 1993, as quoted in Riney et al., 2000, p. 717). While native Japanese speakers are able to train their mouths and tongues to produce these two sounds (see, for example, Riney et al, 2000), Japanese speakers are often unable to differentiate between /r/ and /l/ in words spoken by a native speaker of English (Logan et al., 1991). In fact, Japanese speakers who record their own speech and listen to it played back to them often cannot distinguish the /r/s and /l/s, even when a native English speaker can.

Luckily for native Japanese speakers, most speech is conducted in context. Otherwise, a sentence such as "Take a right at the light" could pose quite a problem. In this case, a native speaker of Japanese might hear *right* and *light* as the same word, but the context would help them determine that this sentence is supposed to be "Take a right at the light" and not "Take a light at the right." However, an L1 Japanese instructor who could not differentiate between the words *correct* and *collect* could really confuse his or her students with the statement, "I'm going to _____ your homework now."

The research which has focused on this /r/ and /l/ confusion thus far has mainly concerned native Japanese speakers' perception of the English /r/ and /l/ (Lively, et al., 1993, Logan et al., 1991). In a classic study, Logan et al. (1991) investigated native Japanese speakers' perception of the English /r/ and /l/ phonemes using various phonemic contexts, including /r/ and /l/ occurring word initially, intervocalically, and word-finally. Interestingly, the results suggest that perception performance was most accurate for /r/ and /l/ word finally (both as a singleton and in consonant clusters) and least accurate for /r/ and /l/ word initially. The next logical question, then, would be, "Is this the case for production as well?" Unfortunately, few researchers have focused on the issue of production and most have remained interested in the questions surrounding perception.

In one such study, which did extend the perception findings to production, Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999) trained eleven native Japanese speakers on the perception of English /r/ and /l/ over the course of three months. The results demonstrate that even though the speakers were training merely on perception of /r/ and /l/, their production of the /r/ and /l/ phonemes also improved dramatically (based on judgements by L1 English speakers). The researchers suggest "the underlying

mechanism that facilitates the transfer and retention of learning in the perceptual domain to the production domain is due to training-induced modifications to a common mental representation that underlies both speech perception and speech production" (p. 983). With so much research on native Japanese speakers' productions of the English phonemes /r/ and /l/, this attempt at bridging the gap between perception and production surely seems promising. However, before one can get at the underlying mechanism behind perception and production, it is necessary to have a clear idea of how native Japanese speakers' productions differ from L1 English speakers' productions acoustically.

Interestingly enough, there is little research available on the *acoustics* of the Japanese speakers' production of the English /r/ and /l/. The study outlined in this paper is a preliminary attempt to fill that gap. In this study we compared the formant values for the second and third formants (F2 and F3) of the /r/s and /l/s produced by a native speaker of Japanese with the formant values of a native speaker of English, when the /r/s and /l/s appeared in different contexts. We chose to focus on the values for the second and third formants (F2 and F3) because this is a primary way to distinguish between /r/s and /l/s. /r/s are typically known for particularly low F3 values, while /l/s characteristically display a low F2 coupled with a high F3 (Ladefoged, 2001, Johnson, 2003). We chose to investigate different word positions for the liquids based on the previous studies which noted that native speakers of Japanese have more difficulty perceiving and producing /r/s and /l/s in certain contexts than in others. Our hypotheses were the following:

1. The spectrograms for the two speakers (Japanese and American) will show significant differences in F2 and F3 values. In other words, the differences that native English speakers perceive in Japanese-accented English will become obvious on a spectrogram: the F2 and F3 values for the native speaker of Japanese will be significantly different from the F2 and F3 values for the native English speaker when /r/s and /l/s are pronounced.
2. The most difficult conditions for the Japanese subject will be /r/s and /l/s appearing in consonant clusters, since the Japanese language does not contain consonant clusters (Eckman, 1991).
3. The easiest context for the Japanese speaker would be Word-Initial singleton, because this condition would give the Japanese subject the maximum amount of time to place her articulators in the proper position to produce these two different sounds.

Method

Subjects

Subjects for this study were one native speaker of Japanese (J1) and one native speaker of English (E1). J1 was a thirty-four-year-old female originally from Niigata, Japan. She had studied English for eight years before coming to the United States, and she had been living, working, and studying in the United States for nine years. At the time of this study, she was pursuing a Ph.D. in Psychology at Indiana University and was employed as a research scientist in the psychology department. We chose a speaker of

Japanese who was very fluent in English with the hope that her speech would provide the best clues as to whether and how a Japanese speaker's liquids sound different than those of a native speaker of English, even after much practice and exposure to the target language.

The native English speaker used in this study was a twenty-nine-year-old female pursuing a Ph.D. in the field of TESOL and Applied Linguistics. Although she was born and raised in the Southern United States, she claimed to have a standard American accent developed for her job as a teacher of English to speakers of other languages. We chose a female E1 so that the formant frequencies of the English speaker would be closer to those of our female Japanese subject.

Materials

The target words used in the experiment appear in Table 1. The target phonemes analyzed in this report were taken from several conditions: word-initial, intervocalic, consonant cluster, and word-final. The words focusing on both the /r/ and /l/ phonemes ("One segment distance" and "More than two /l/, /r/ phonemes") were also recorded, but were not analyzed in the present paper.

Table 1. Target Words

Word-initial	**Intervocalic**	**Consonant Clusters**	**Word-final**	**One-segment distance**	**More than two /l/, /r/ phonemes**
Right/light	Arrive/alive	Crime/climb	Fire/file	Liar	Reliable
Red/led	Correct/collect	Bread/bled	Bear/bell	Umbrella	Parallel
Road/load	Aroma/alone	Froze/flows	More/mole	Floor	Roleplay

The target words consist primarily of minimal pairs with an alternating /r/ and /l/. The vowels (/aɪ/, /ɪ/, and /oʊ/) were kept constant as much as possible, although the subjects did not always pronounce the words with the vowel we intended. For example, the word

correct was supposed to be pronounced /kɚrɛkt/, as a minimal pair for *collect* (/kəlɛkt/), but the native speaker of English changed the first vowel to something closer to /ɹ/ in some instances. Also, the Japanese subject sometimes eliminated the first vowel altogether (/krɛkt/). In addition to vowel choice, much care was taken to select words that remained constant over the variables of stress and number of syllables.

Each speaker recorded the entire set of 30 words three separate times, for a total of 90 tokens per speaker. For each repetition (3 total), the order of the words was randomized in order to eliminate any order effects on the participants' productions.

Procedure

Each participant was tested individually and began by being seated in front of a computer in a quiet room. She was asked to wear headphones with an attached microphone. First, she was asked to record her name and the date to ensure the computer was registering her response. Second, she was asked to read a list of words which were presented to her on a sheet of paper. This list consisted of the 30 words, presented in random order, that appear in Table 1. The participant was instructed to read both the number of the word (which remained constant throughout repetitions) and the word itself. Once finished with the first repetition, the participant was given the second list of the same words (in a new random order) to read and subsequently the third and final list (in the final random order).

Coding

Subjects were recorded on a computer using the WaveSurfer program. Once the speakers had been recorded, the researchers listened to each instance of a target word to determine the onset of the /r/ or /l/. Formant values were collected at this point in time

for each repetition of each target word, using the "formant tracker" option in the WaveSurfer program. To use the formant tracker, we changed the "FFT window length" to 1024, the "Cut spectrogram at" value to 6000 Hz, and the LPC order to 10. We chose to focus on the values for the second and third formants (F2 and F3) because this is a primary way to distinguish between /r/s and /l/s. /r/s are typically known for particularly low F3 values, while /l/s characteristically display a low F2 coupled with a high F3 (Ladefoged, 2001, Johnson, 2003). The two researchers worked together to agree on the onset time for each liquid. Once the onset of each /r/ or /l/ was determined, the formant values for that point in time were transferred to a spreadsheet program to be analyzed.

Results

This paper focuses on the results for the target words in the first four conditions (Word-Initial, Intervocalic, Consonant Cluster, and Word Final) of the experiment. The productions of both J1 and E1 over all tokens in these four conditions appear in Figure 1 (see Appendix A). The scatter plot shows that, overall, there was more variation (i.e., inconsistency) in the production of the /r/s and /l/s in the speech of the Japanese subject than in the speech of the American subject. It is also important to note that the /r/ and /l/ plots of F2 and F3 intersections for E1 rarely overlap, whereas for J1, there is much more overlap. In other words, E1 makes much more of a distinction between /r/s and /l/s. Table 2 reveals that the average F2 and F3 values for both speakers are very similar, except in the case of the F3 value for /l/. However, J1 shows a much greater range of values among the formants. It is interesting to note that J1's range of F2 values for /r/ is very similar to the range for E1; however, the range of F3 values for /r/ is much greater

than E1's F3 value range. J1's F2 and F3 values for /l/ show a larger range than E1's values for /l/, but they are still not as great as the difference in the ranges for the F3 value of /r/.

Table 2. Summary of Data- All Conditions

		Japanese		English	
		Average (Hz)	Range	Average (Hz)	Range
/r/	F2	1444	954 – 2312	1465	1003 – 2187
	F3	2363	1655 – 3223	2304	1906 – 2986
/l/	F2	1242	626 – 2051	1210	829 – 1847
	F3	3020	2524 – 4045	3457	2710 – 3835

Figures 2-5 (Appendix A) show scatter plots of each of the four conditions analyzed. Among these target categories, the most variation within an individual speaker (for both J1 and E1) occurred in the Word Final category, especially among the F2 values. This is particularly apparent in the scatter plot in Figure 5 (Word Final Condition). The next greatest variety occurred in the Word-Initial category (Figure 2). Both subjects showed the most consistency and had the most similar results (the Japanese –speaker's formant values were most similar to the formant values of the native English-speaker) in the Consonant Cluster category (Figure 3). The tables below show the range and averages of the second and third formant values for /l/ and /r/ for each speaker in each of the four conditions.

Table 3. Summary of Data- Word Initial Condition

		Japanese		English	
		Average (Hz)	Range	Average (Hz)	Range
/r/	F2	1387	1074 – 1672	1260	1003 – 1599
	F3	2590	2097 – 3223	2112	1945 – 2421
/l/	F2	1367	1103 – 1564	1234	1110 – 1536
	F3	2895	2524 – 3224	3334	2710 – 3580

Table 4. Summary of Data- Consonant Cluster Condition

		Japanese		English	
		Average (Hz)	Range	Average (Hz)	Range
/r/	F2	1532	1319 – 1737	1569	1362 – 1778
	F3	2520	2117 – 2900	2494	2010 – 2986
/l/	F2	1343	1083 – 1573	1320	1115 – 1477
	F3	3124	2592 – 4045	3462	3208 – 3816

Table 5. Summary of Data- Intervocalic Condition

		Japanese		English	
		Average (Hz)	Range	Average (Hz)	Range
/r/	F2	1256	1072 – 1830	1190	1102 – 1368
	F3	2094	1940 – 2283	2104	1906 – 2355
/l/	F2	922	626 – 1087	1065	829 – 1300
	F3	3176	2607 – 3893	3527	3308 – 3759

Table 6. Summary of Data- Word Final Condition

		Japanese		English	
		Average (Hz)	Range	Average (Hz)	Range
/r/	F2	1623	954 – 2312	1841	1388 – 2187
	F3	2232	1655 – 2847	2505	2335 – 2670
/l/	F2	1336	772 – 2051	1223	848 – 1847
	F3	2885	2738 – 3090	3504	3050 – 3835

These charts show that the closest match between the Japanese speaker and the native English speaker were for the Consonant Cluster /r/s (both F2 and F3) and the F2 for /l/ in this condition. In each of these cases, the average for the Japanese speaker was less than 50 Hz above or below the average for the native English speaker. The Intervocalic /r/ was also very close to the native English speaker's (less than 70 Hz difference). These charts also show that the biggest difference between the Japanese subject and the American subject was for F3 values in general. With the exception of the F3 values for the Intervocalic and Consonant Cluster /r/s, all the average F3 values for the Japanese

speaker were at least 270 Hz higher or lower than the values for the native English speaker. Also, it is important to note that all of the Japanese subject's average F3 values for /l/s were lower than the average F3 values for the English-speaking subject. This is important in light of the fact that the typical English /l/ is known by its high F3 value. The information contained in these charts also demonstrates, in general, the Japanese subject was much better at producing English /r/s than English /l/s.

Figures 6- 13 (Appendix A) contain the spectrograms of J1 and E1 producing /r/s and /l/s in various contexts. Figure 6 shows the actual spectrogram for the first repetition of the word "red" as spoken by E1. Figure 7 shows the spectrogram for the same repetition for the word "red" as spoken by J1. On both Figures, the onset of the /r/ phoneme is marked. As can be seen, E1 shows a classic "dip" in the first three formants to produce the low F2 and F3 for an English /r/, and then a rise into the following vowel. J1s formant values remain much more constant from the /r/ into the following vowel. While the F3 values for both speakers are relatively similar in these words, there is a much greater difference in the F2 values (almost 370 Hz).

Figures 8 and 9 compare E1's first repetition of "correct" (Figure 8) with J1's first repetition of the same word (Figure 9). On both Figures, the onset of the /r/ phoneme is marked. These spectrogram images look very different for several reasons. First of all, E1 displays the typical low formants ("dip") for the /r/ and she holds the /r/ long enough for a distinct "hole" (caused by the /r/'s low F2 and F3) to appear in the spectrogram. In Figure 9, no "hole" appears because the F2 and F3 for J1's /r/ remain at approximately the same level from the /k/ through the /r/ and into the following vowel (the word was pronounced /krɑkt/, with a consonant cluster word-initially instead of the CVC).

Turning to /l/s, we expect to find much higher F3 (characteristic of /l/s) than we saw in the target words containing /r/s. Figure 10 shows the spectrogram for the first repetition of the word "bled" as spoken by E1. Figure 11 shows the spectrogram for the same repetition for the word "bled" as spoken by J1. These images show a large distance (just over 2200 Hz) between the F2 and F3 at the onset of the /l/ spoken by E1, while a much smaller difference (less than 1350 Hz) in the word uttered by J1. Figures 12 and 13 demonstrate this difference even more clearly. In Figure 12, we see the word "bell" as spoken by E1, with an obviously rising F3 and falling F2. In Figure 13, we see that the F3 for speaker J1 remains relatively flat, with a more modest falling of the F2. Thus the difference between J1's F2 and F3 is not as great as the distance between E1's F2 and F3 in this particular word.

Finally, the graphs in Figures 14-25 (Appendix A) plot the formant values for the two speakers saying minimal pairs for each of the four conditions in each of the three primary vocalic categories. From these graphs we can see that in most cases, J1 was able to make a distinct difference between her own /r/s and /l/s, even if those /r/s and /l/s did not match the /r/s and /l/s of the native English speaker. J1 was most successful in the Intervocalic condition, where her /r/s and /l/s were very separated. The Word Final category was the next easiest, and the Consonant Cluster category was not much more difficult. She had the most difficulty separating the /r/s and /l/s in the Word-Initial category (one instance of "right" and one of "red" had higher F3s than all of her *light*s and *led*s).

Discussion/Conclusion

The results of this experiment were intriguing. The Japanese subject recorded for this study has very good English pronunciation in general, as judged by the two researchers. Even so, there was often a distinct difference in the formant values of her pronunciation and that of the native English speaker's. The F3 dimension proved especially difficult, particularly when producing /l/s. There was not a clear-cut line between the two phonemes for the native Japanese speaker as there was for the native English speaker. Although this study produced a plethora of interesting results, our hypotheses proved incorrect. We predicted that the Word-Initial context would be the easiest for the Japanese subject but in fact it proved to be one of the most difficult. It seems apparent that the Japanese subject needed a context in order to produce a more native-English-like /r/ or /l/. We also predicted that the Consonant Cluster condition would be the most difficult because Japanese does not have consonant clusters, but this condition allowed the Japanese speaker to get the closest to the native English speaker, presumably because it provided a context for the liquid.

The Word-Initial category and the Word-Final category showed the most variation in the results for both speakers. This could have been due in part to the difficulty in determining the onset of the liquid in these circumstances. In the case of the Word-Initial tokens, on occasion the formants were already rather high before the researchers were able to perceive the onset of voicing for the /r/ or /l/, especially when our Japanese subject spoke very softly. In the case of the Word-Final tokens, it was often difficult to determine where the preceding vowel ended and the liquid began, particularly with words like "fire" and "bear," which are often transcribed with an r-colored vowel.

The vowel and the /l/ in the word “mole” were also particularly difficult for the researchers to separate. Finally, it is important to note that there was a larger variety of vowels preceding the liquids in the Word-Final category, which no doubt had an effect on the onset formants of the liquids. The results for this condition, then, should be read with some caution.

Overall, the evidence provides support for the idea that perception and production may be closely linked, as native Japanese speakers appear to have difficulty with both perception and production of the two phonemes. However, the findings from this study are not entirely generalizable as only one native Japanese speaker was tested. It would be interesting to replicate this study with multiple native speakers of Japanese, in order to make the results more generalizable and to rule out any idiosyncrasies in the pronunciation of one particular speaker. Japanese speakers with various levels of English proficiency could also make the study more useful. In addition, varying the sex of the speakers could produce novel results, as Hagiwara (1994) argues that female speakers show a larger variation of formant values for /r/ than do male speakers. Finally, researchers could investigate the speech of Japanese subjects reading sentences, as opposed to the word list used in this study, or even record unscripted conversation for analysis.

Note: The authors would like to thank Ken de Jong, Kyoko Nagao, Noah Silbert, and Kyoko Okamura for their helpful insights and comments on this project and earlier drafts of this paper. Of course, any errors are the sole responsibility of the authors.

Bibliography

Bradlow, A., Pisoni, D, Akahane-Yamada, R, and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101(4),* 2299-2310.

Bradlow, A., Akahna-Yamada, R, Pisoni, D, and Tokhura, Y. (1999). Training Japanese Listeners to identify English /r/ and /l/: Long term retention of learning perception and production. *Perception and Psychophysics 61(5),* 977-985.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13,* 23-41.

Hagiwara, R. (1994). Sex, syllabic [r], and the American English vowel space. *UCLA Working Papers in Phonetics, 88,* 63-90.

Hagiwara, R. (1995). Acoustic realization of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics, 90,* 2-187.

Johnson, K. (2003). *Acoustic and Auditory Phonetics*. Blackwell Publishers.

Ladefoged, P. (1993). *A course in phonetics, fourth edition.* Harcourt Brace Jovanovich.

Lively, S. E., Logan, J. D., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076-2087.

Logan, J. D., Lively, S. E., and Piston, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874-56.

Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly, 34,* 711-737.

Appendix A

Figure 1.

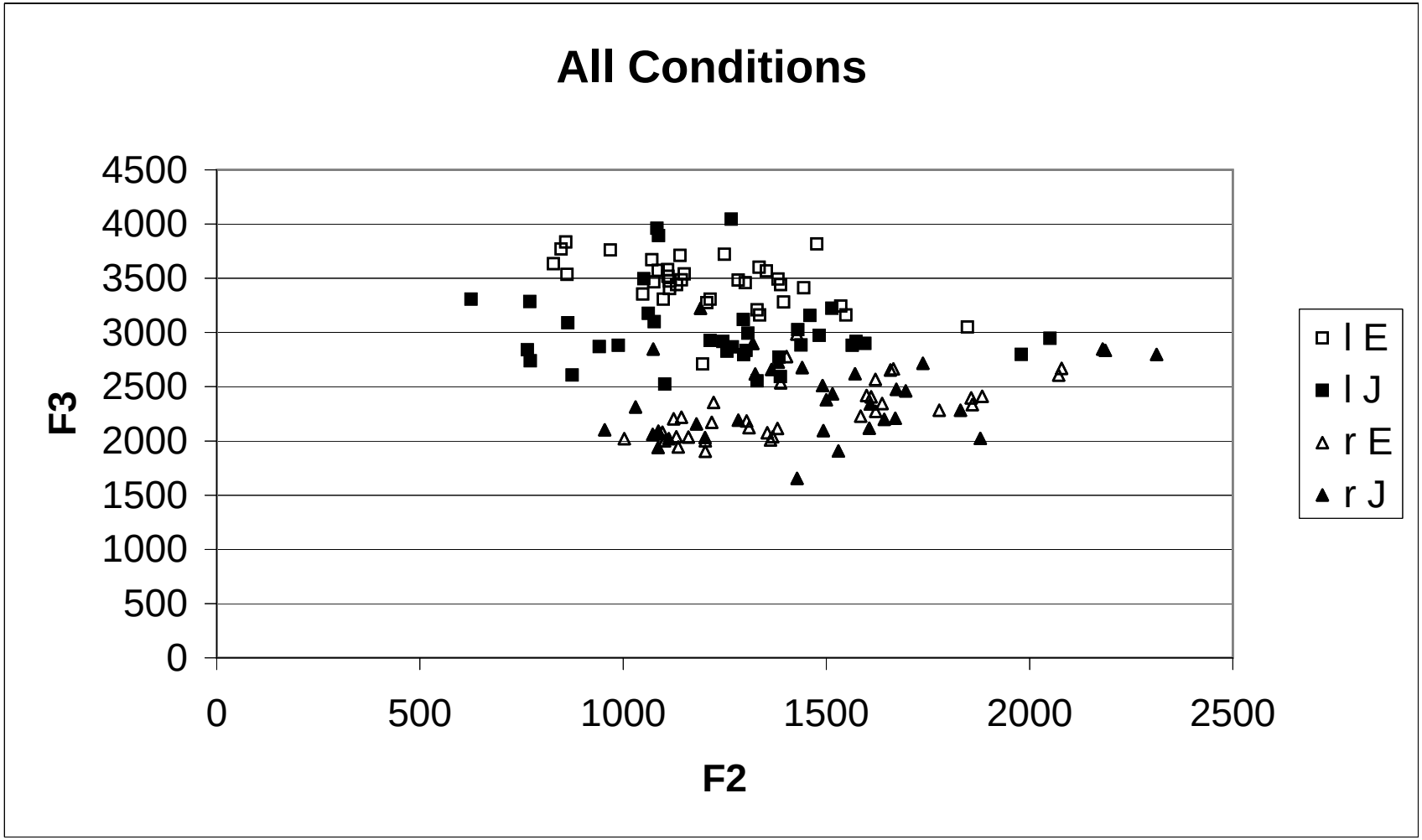
All Conditions
4500
4000
3500
3000
2500
2000
1500
1000
500
0
F3
0
500
1000
1500
2000
2500
F2
l E
l J
r E
r J

Figure 2.

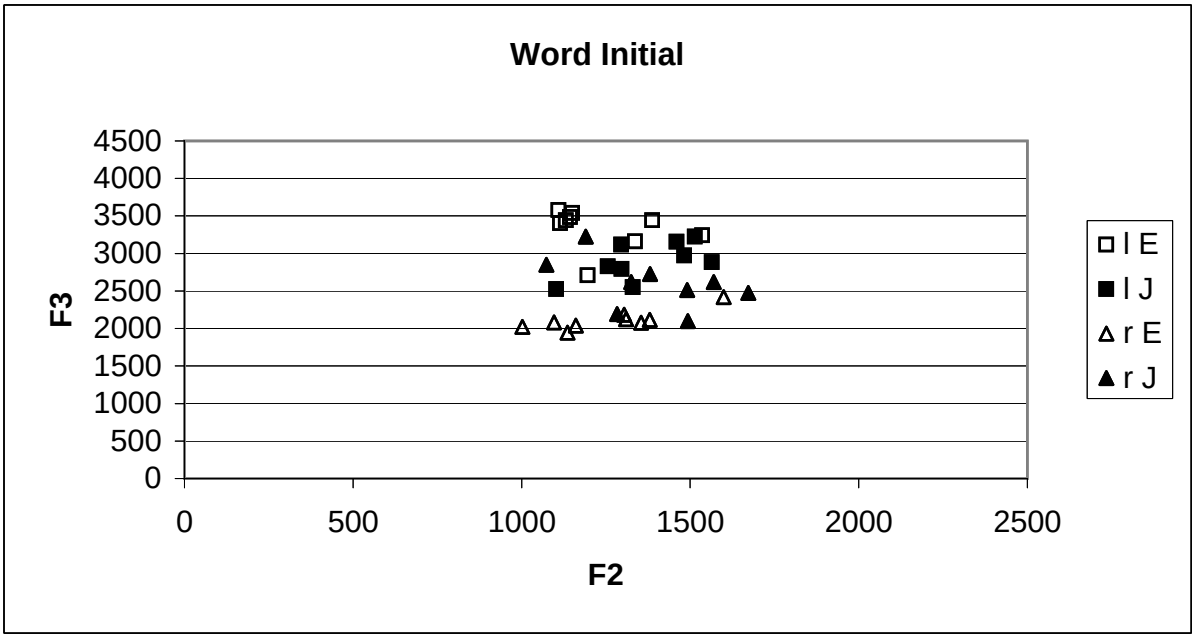

Figure 3.

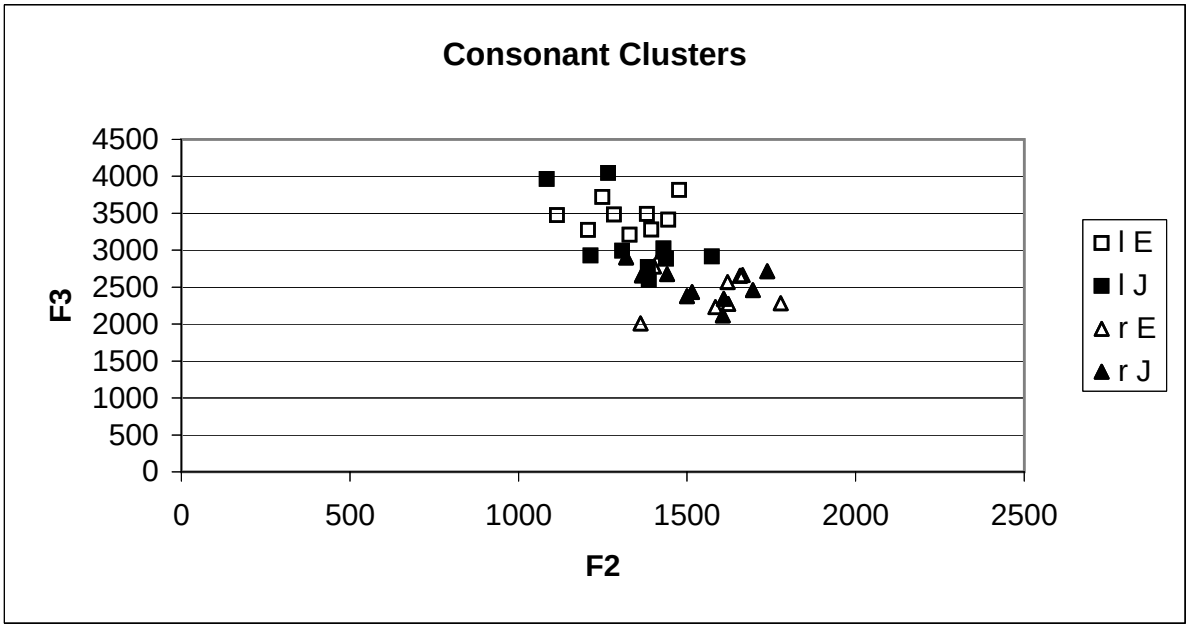

Figure 4.

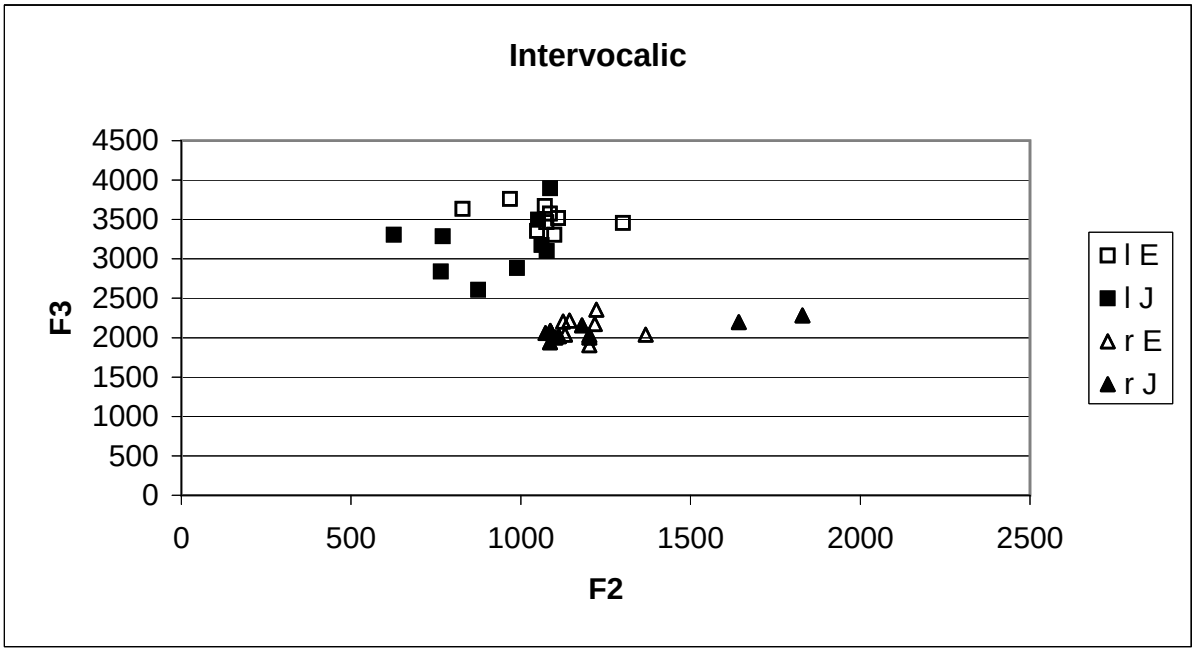

Figure 5.

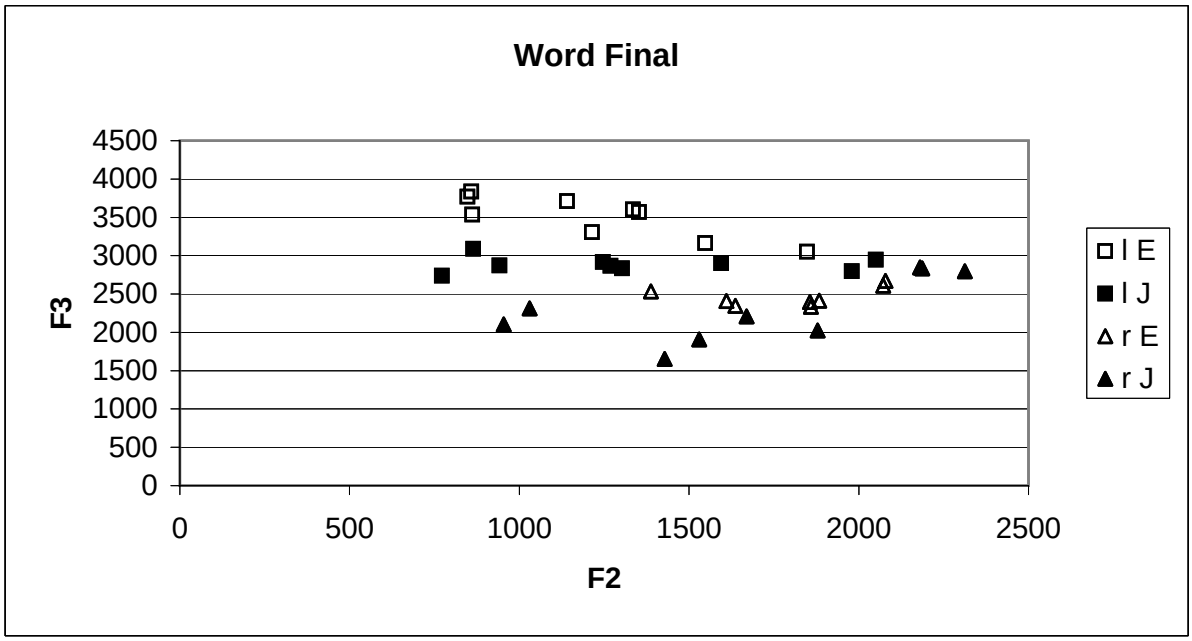

Figure 6. Spectrogram with Formants, English speaker saying “red”

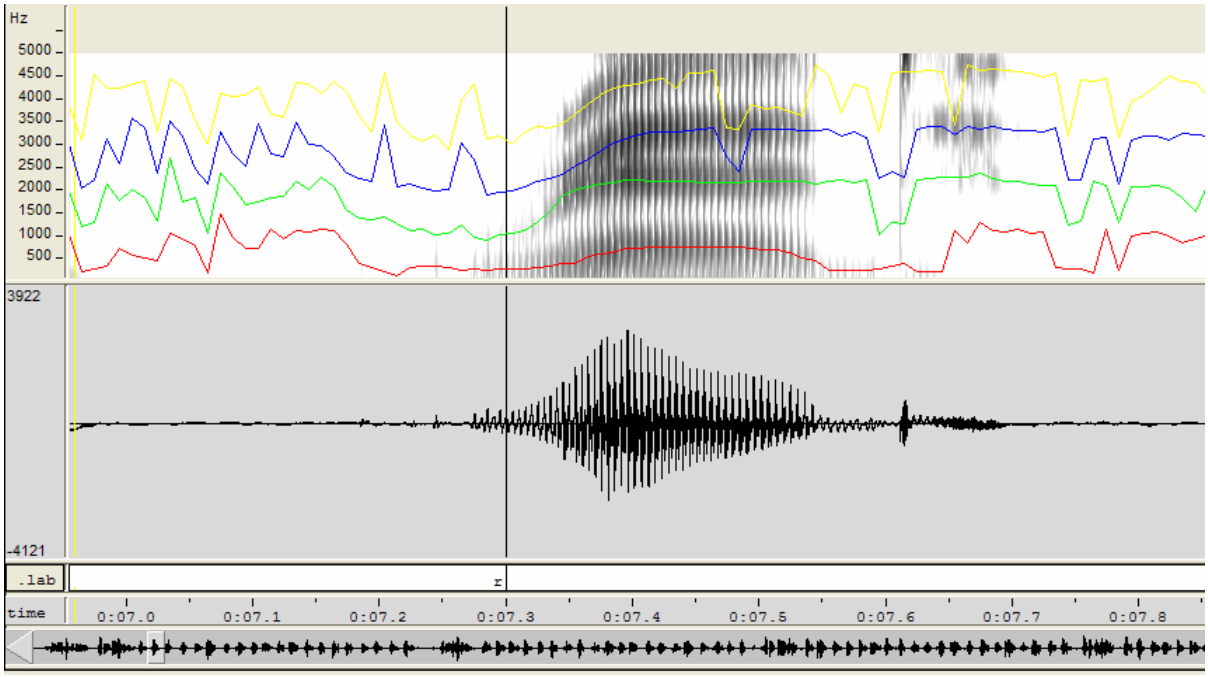

Figure 7. Spectrogram with Formants, Japanese speaker saying “red”

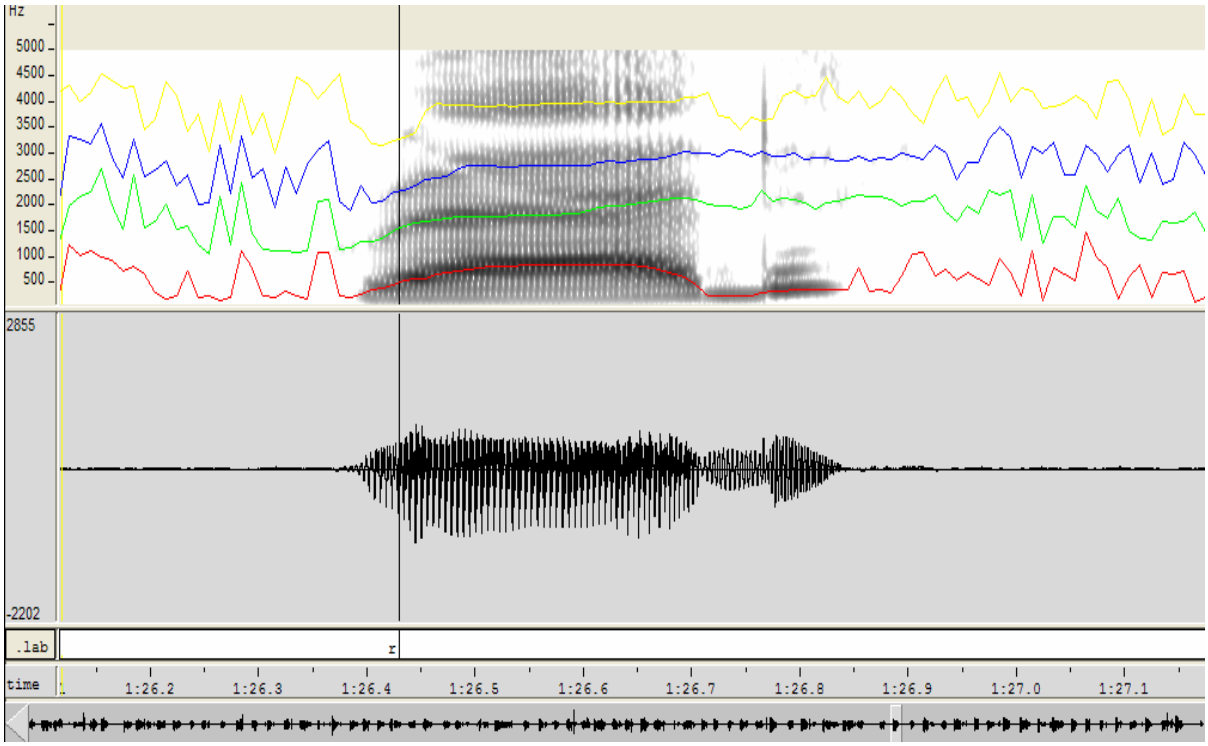

Figure 8. Spectrogram with Formants, English speaker saying “correct”

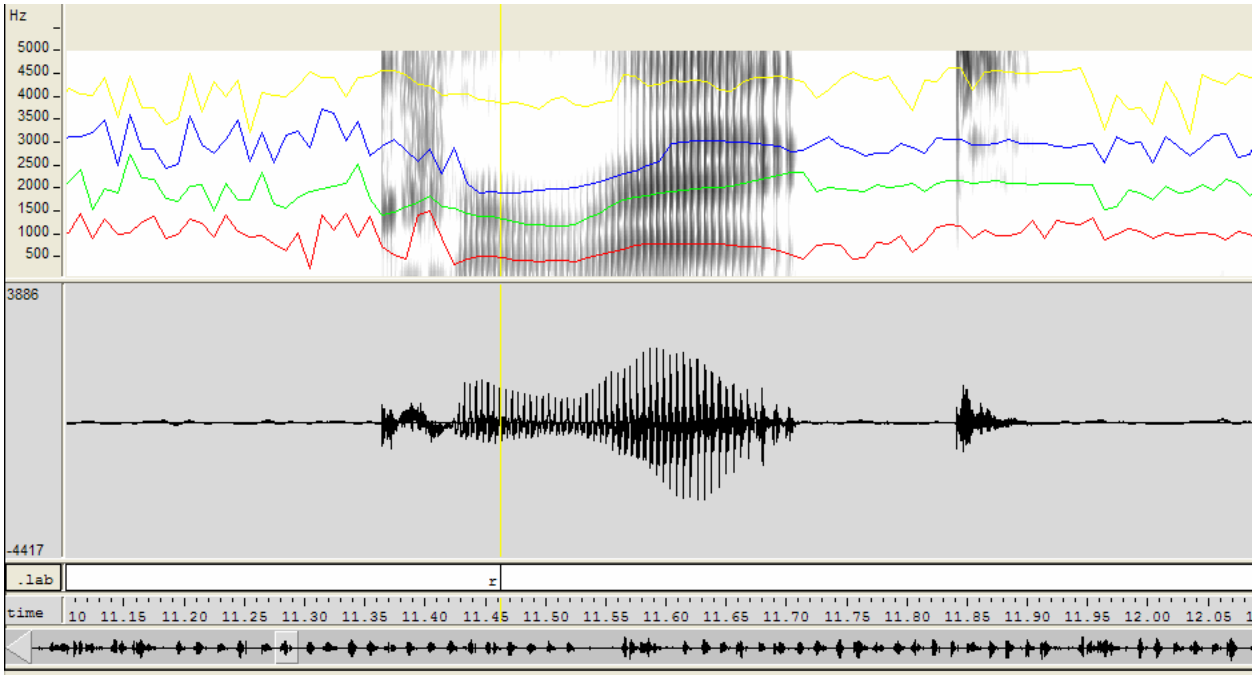

Figure 9. Spectrogram with Formants, Japanese speaker saying “correct”

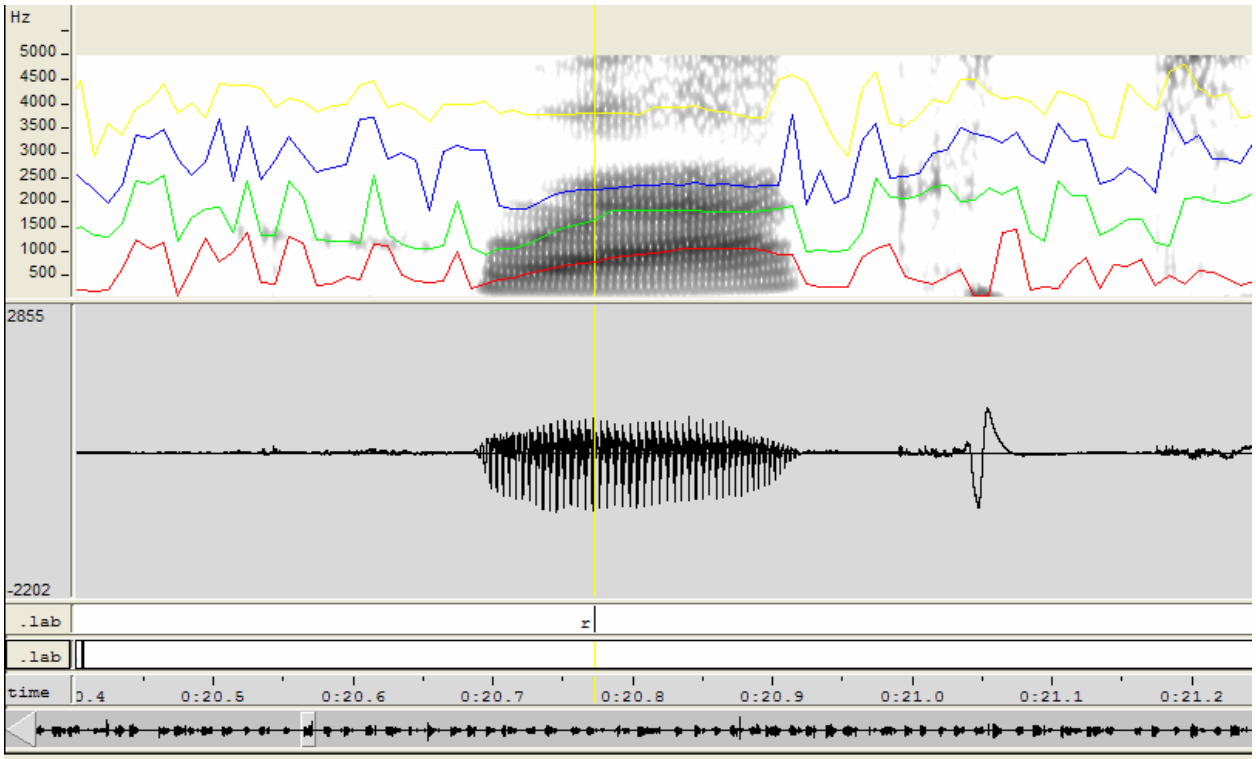

Figure 10. Spectrogram with Formants, English speaker saying “bled”

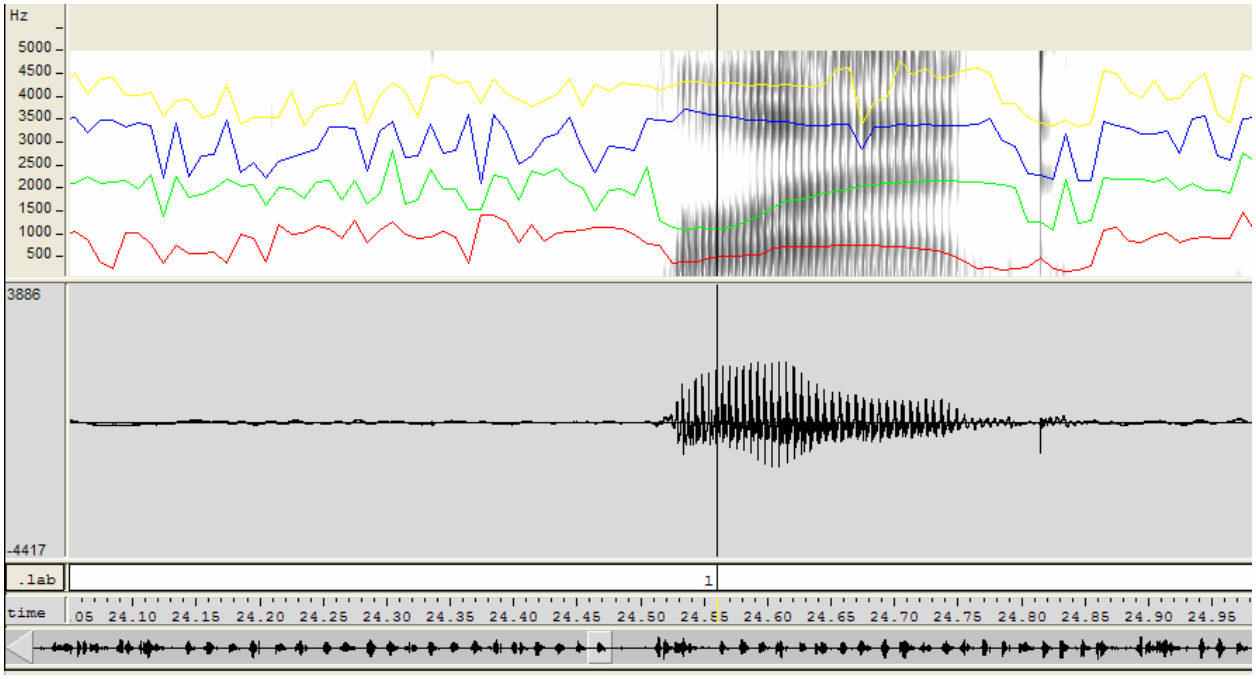

Figure 11. Spectrogram with Formants, Japanese speaker saying “bled”

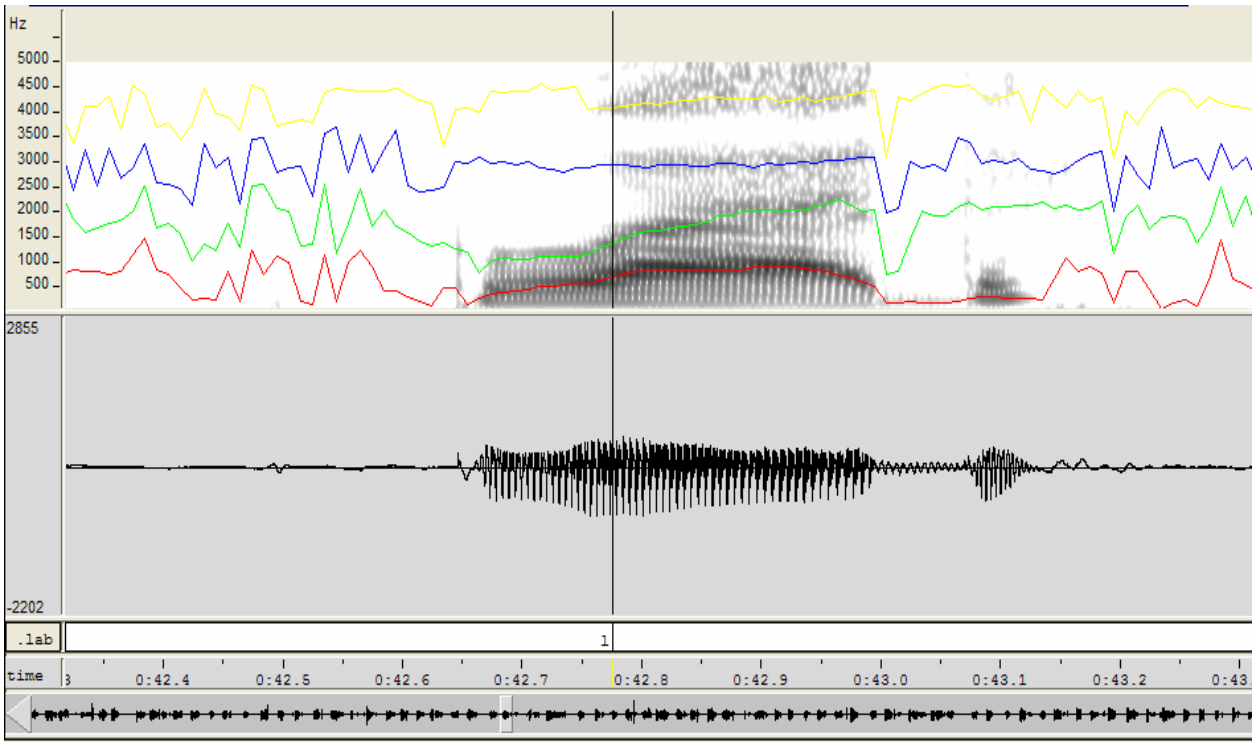

Figure 12. Spectrogram with Formants, English speaker saying “bell”

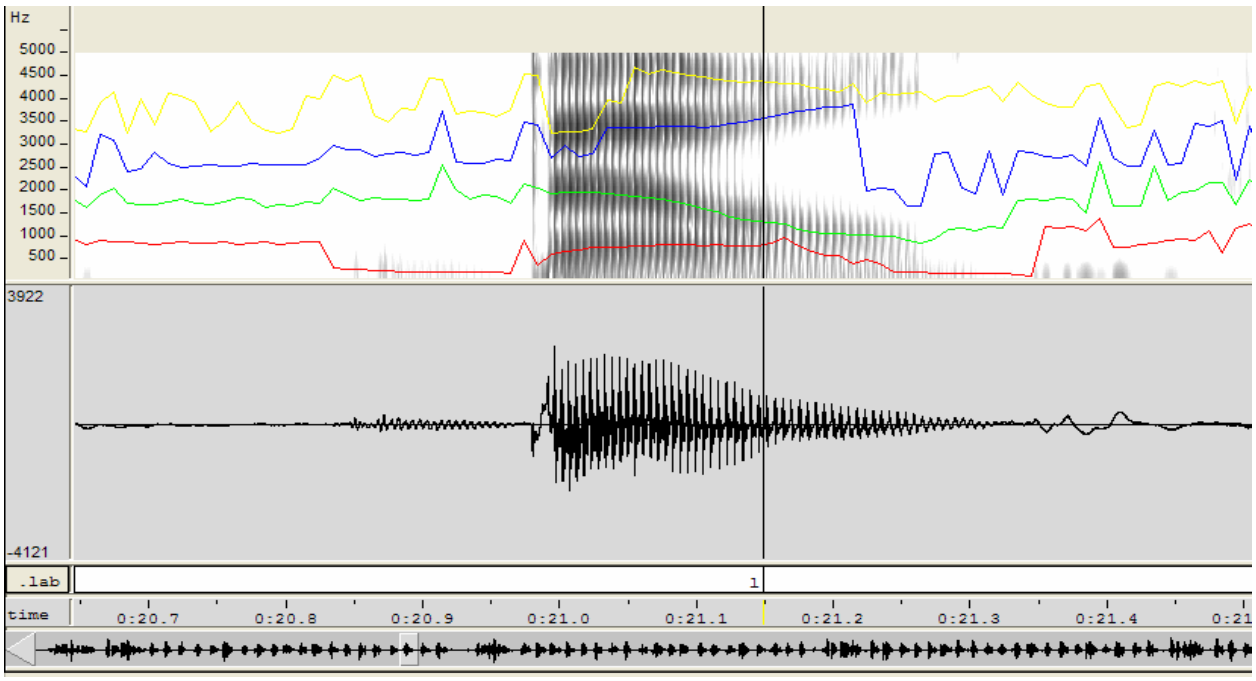

Figure 13. Spectrogram with Formants, Japanese speaker saying “bell”

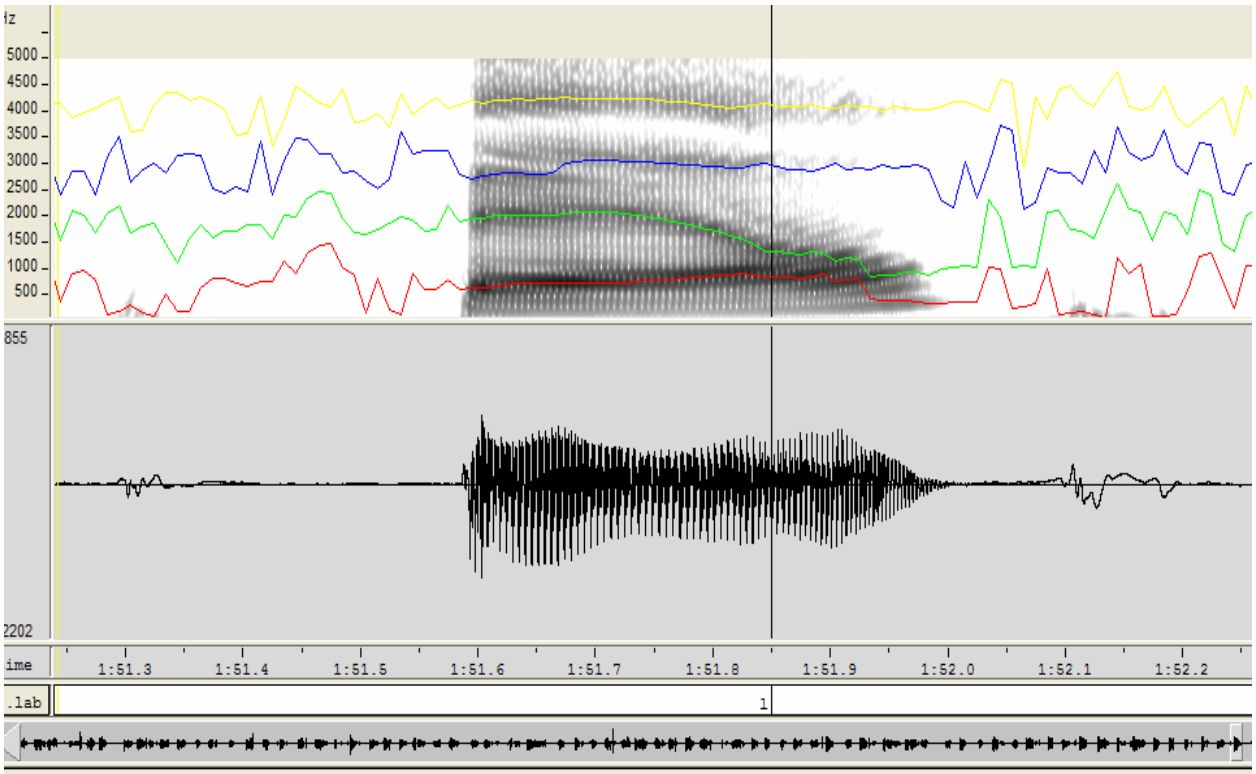

Figure 14.

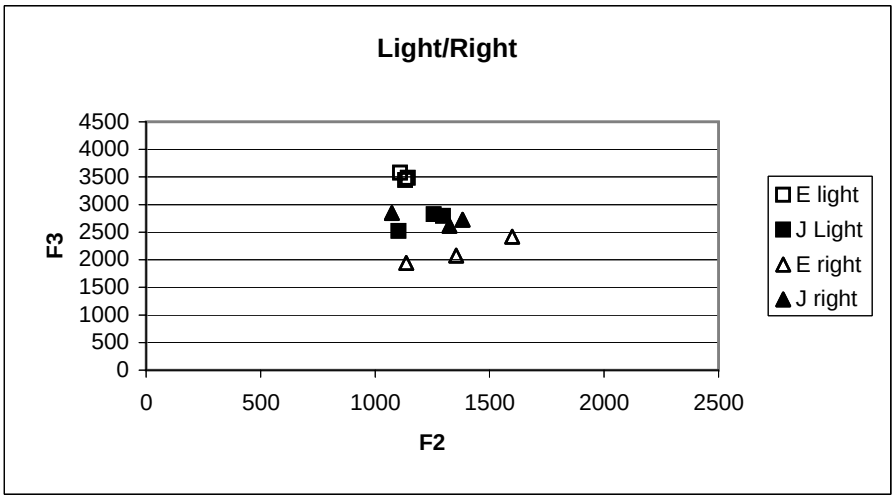

Figure 15.

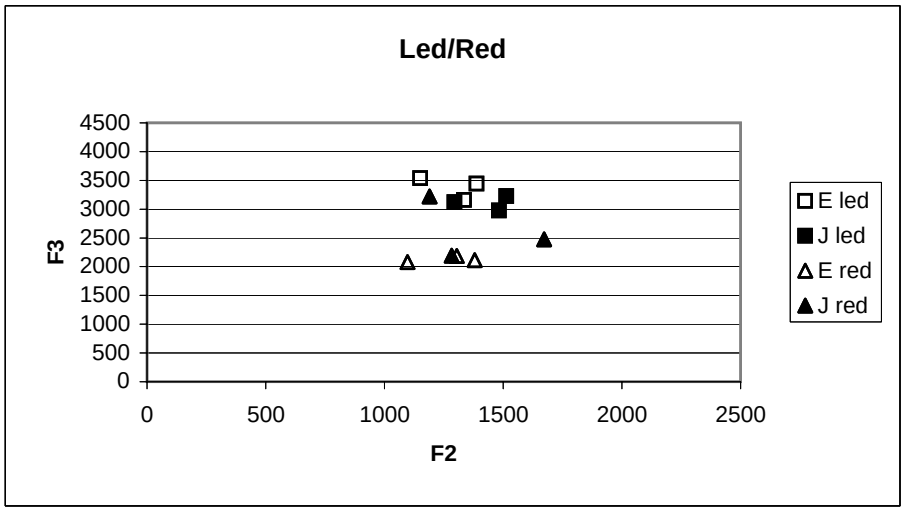

Figure 16.

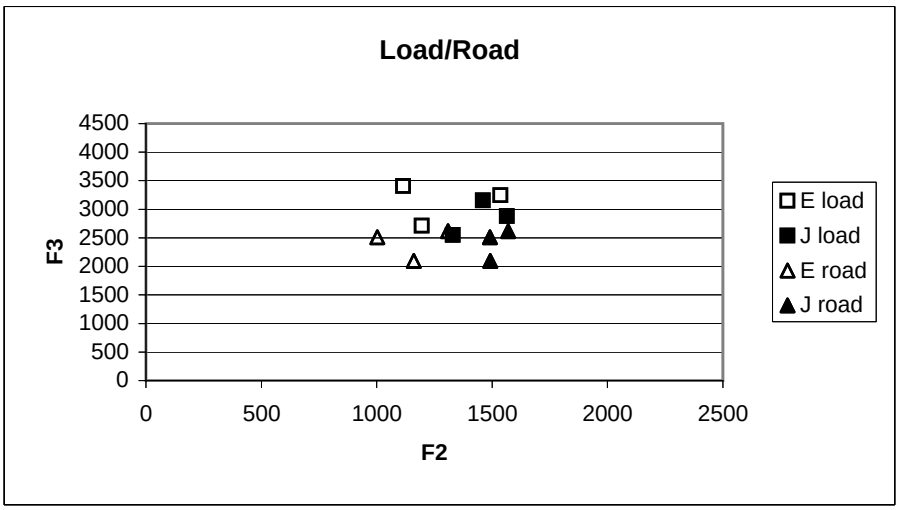

Figure 17.

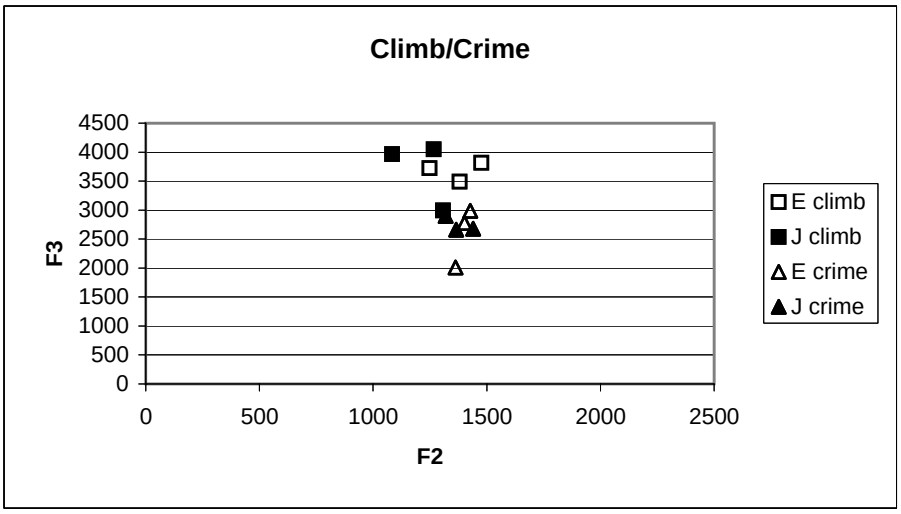

Figure 18.

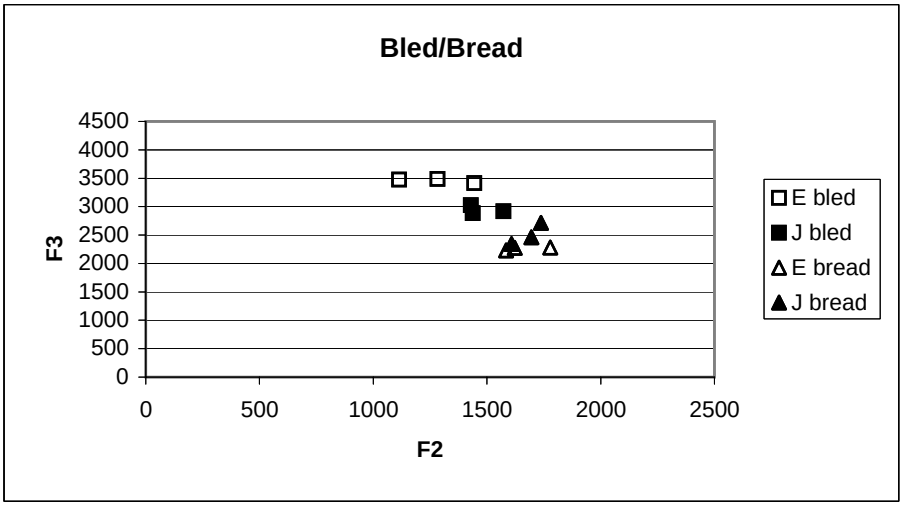

Figure 19.

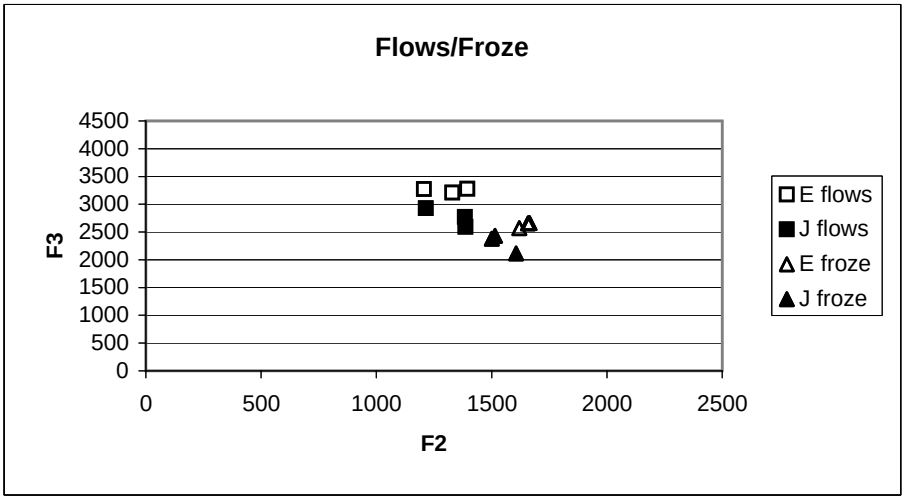

Figure 20.

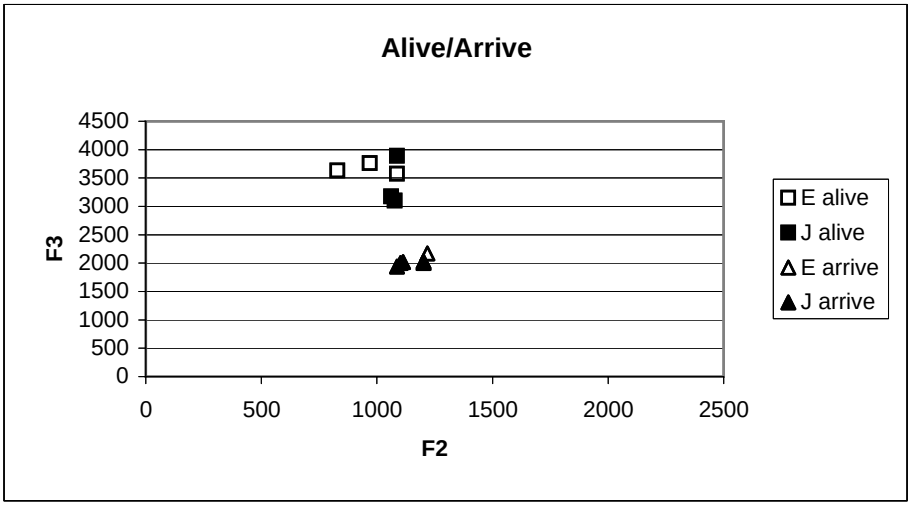

Figure 21.

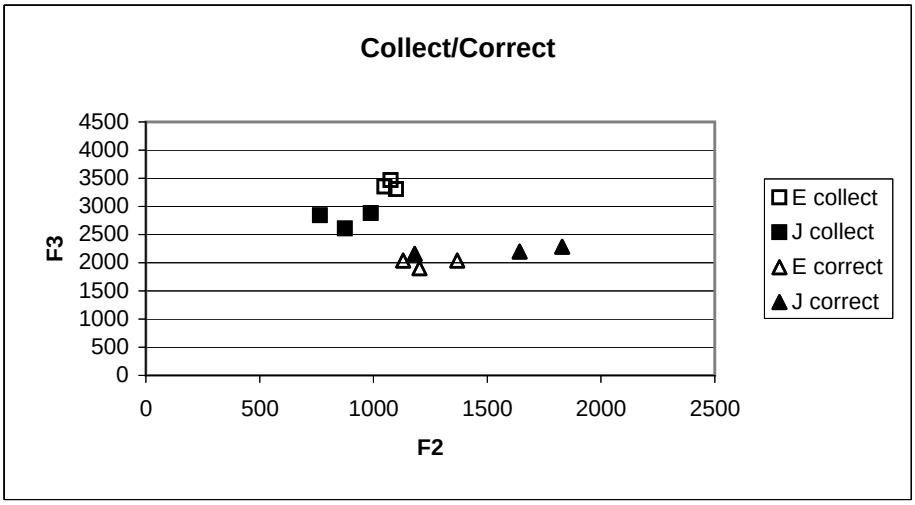

Figure 22.

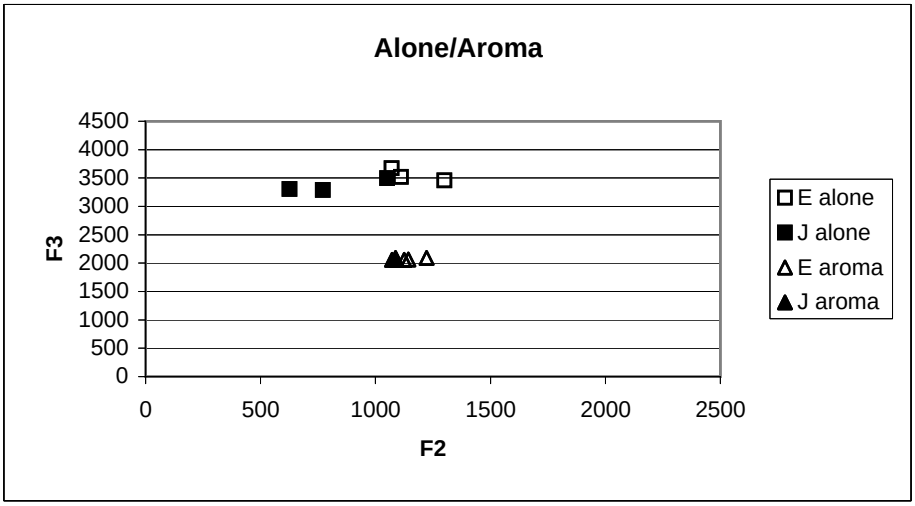

Figure 23.

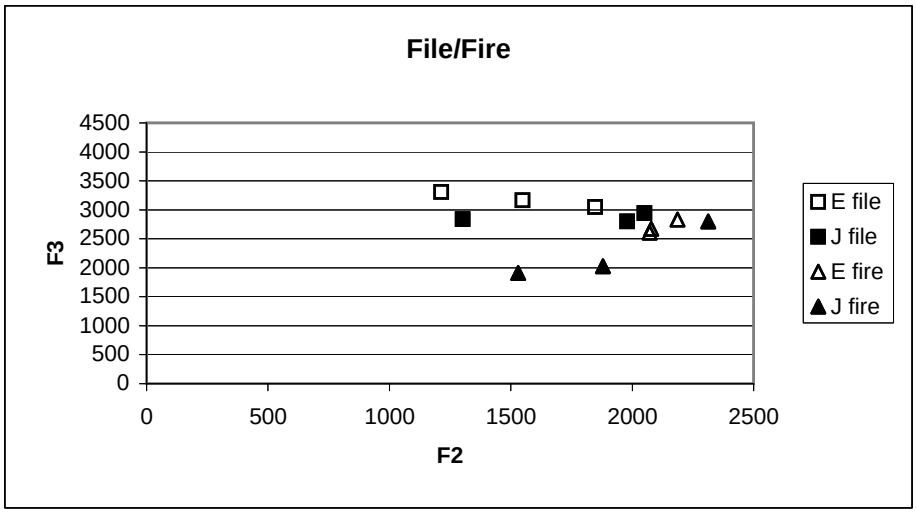

Figure 24.

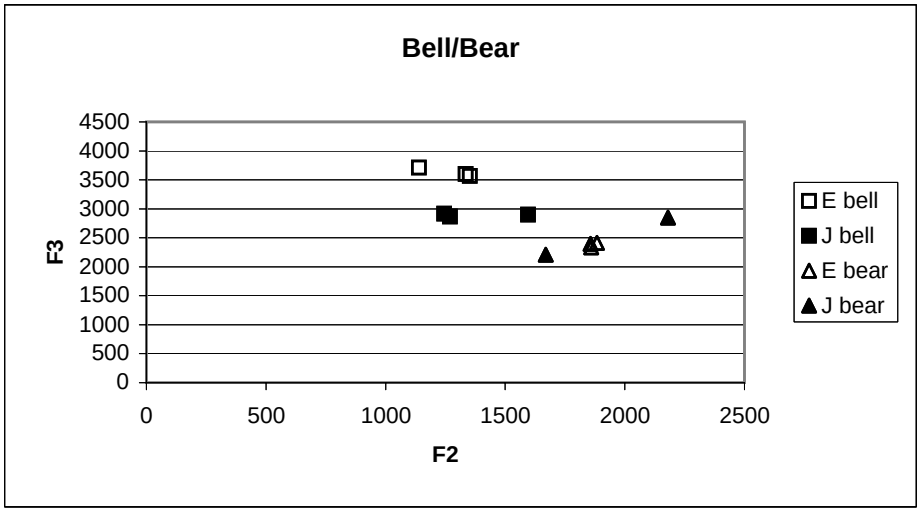

Figure 25.