2000

The relation between intelligibility and the acoustic characteristics of deaf speech: An acoustic analysis of the vowels /i/ and /aw/, and nasals

Jennifer Micheletto

THE RELATION BETWEEN INTELLIGIBILITY AND THE ACOUSTIC CHARACTERISTICS OF DEAF SPEECH: AN ACOUSTIC ANALYSIS OF THE VOWELS /i/ AND /aw/, AND NASALS

by

Jennifer Micheletto

An independent study submitted in partial fulfillment of the requirements for the degree of

Master of Science in Speech and Hearing

Emphasis in Audiology

Washington University
Department of Speech and Hearing

May 26, 2000

Approved by: Rosalie Uchanski, Ph.D., Independent Study Advisor

1. Abstract of Study

Two types of acoustic measures were examined for their possible relation to the overall speech intelligibility of children with cochlear implants. First, each child produced five words containing the vowel /i/ and four words containing the vowel /aw/, and an average F2/F1 and A2/A1 was calculated for each subject from these multiple productions of each vowel. These average scores were then plotted against the speakers' overall intelligibility. Second, a measure of nasal goodness was calculated and, similarly, this result was compared to the overall intelligibility of the speaker. Analysis of the vowel measurements reveals little correlation between the formant ratios and amplitude ratios of the vowels and intelligibility. The nasal goodness measure yields a moderate correlation with overall speaker intelligibility.

2. Introduction

When learning speech and language, deaf children are at a grave disadvantage as they cannot hear the speech of others or themselves. Deaf children rely heavily on visual cues. However, the vowels /i/ and /aw/, and the nasals are characterized by tongue movements and vocal tract postures that are not always visible, thus causing difficulty for the deaf speaker to imitate. Therefore, one research question is, "Do cochlear-implanted children receive spectral information that enables them to produce intelligible speech in the absence of important visual cues?"

The source-filter theory, as it pertains to vowel production, states that "the output energy is a product of the source energy and the resonator (or filter)" (Kent & Read, 1992, p. 18). In the case of speech production, the source is the vocal folds and the filter is the vocal tract configuration. When the vocal folds vibrate at various frequencies, a spectrum is produced with energy at various multiples of the fundamental frequency. The second aspect of the theory, the filter, shapes the formants, or "the natural mode of vibration of the vocal tract" (Kent & Read, 1992, p. 18). Each vocal tract configuration corresponds to a set of formant frequencies; therefore, any deviation from normal vocal tract posture in the production of vowels will produce different formant frequencies and thus, alter the characteristics of the intended vowel.

Vowels differ with respect to their formant structure. As Kent and Read (1992) describe in their book *The Acoustic Analysis of Speech*, "each formant can be described by two characteristics: center frequency and bandwidth" (p. 20). If a vowel is produced with an incorrectly shaped vocal tract, its formant characteristics will deviate from normal and may thus affect that vowel's intelligibility.

Vowels can be classified by place of production. The vowel /i/ is labeled on the vowel quadralateral as a "high, front, unrounded vowel" (Borden, et al., 1994, p. 101). This means that the oral cavity is made smaller by fronting and raising the tongue toward the alveolar ridge. The lips remain spread. These production characteristics result in an average low F1 value of 270 Hz and a high F2 value of 2290 Hz for an adult male /i/

(Borden, et al., 1994, p. 101). By contrast, the vowel /aw/ does not have such a wide frequency separation between F1 and F2. This vowel is produced in the back of the mouth, thus making the oral cavity larger. In addition, the vowel /aw/ is produced with the tongue being lowered. As a result, for an adult male, typical values for the vowel /aw/ are a high F1 (570 Hz) and a relatively low F2 (870 Hz).

Nasals are produced by an open nasal cavity accompanied by a closed oral cavity. In the case of the nasal consonant /m/, the lips seal off the oral tract so sound energy resonates through the nasal passageway. In the nasal consonant /n/, the "blade or tip of the tongue touches the upper alveolar ridge of the hard palate, with the back sides of the tongue touching the upper molars" (Borden, et al., 1994, p. 120). This occlusion of the oral tract at the alveolar ridge still permits sound to escape from the nasal cavities.

Unlike vowels, which have high frequency components, nasals are characterized by low frequency energy. This is due to two characteristics. First, the "addition of the nasal branches to the vocal tract creates a larger, longer resonator," and thus lower frequency components (Borden, et al., p. 120). Second, antiresonances are introduced when the oral tract is occluded, thus contributing to the lower frequency of nasals. Antiresonances can be described as "frequency regions in which the amplitudes of the source components are severely attenuated" (Borden, et al., 1994, p. 120). On spectrograms, antiresonances are seen as the absence of energy in the high frequencies.

The vowels /i/ and /aw/ were chosen in this study because they have the highest and lowest F2 values of all the American English vowels. So, a good F2 separation between these two vowels might be indicative of generally good vocal tract postures during vowel production. Randall Monsen (1978), in his paper entitled "Toward Measuring How Well Hearing-Impaired Children Speak," notes that "when hearing-impaired speakers produce vowels poorly, the phonological space defined by the maximum and minimum frequencies of the first two formants is small" and the vowel ultimately acquires characteristics of the neutral schwa (p. 206). If the formants of the intended vowels /i/ and /aw/ are not near the expected frequency values found for normal-hearing talkers, then the result may be the perception of a completely different vowel, and thus, the intelligibility of a deaf child may be affected. With nasals, the deaf speaker may have difficulty controlling the muscle movements necessary to occlude the vocal tract and to open the velopharyngeal port. This may cause energy to spread to the higher frequencies and thus, the nasal consonant loses its distinct low frequency characteristics.

3. Methods

a). Subjects

The subjects of this study were children who participated in a National Institute of Health research project being conducted through Central Institute for Deaf in the summers of 1997 and 1998. The 1997 camp contained 47 children; the 1998 camp contained 45 children (CID Research Progress Report, 1999, p. 5). In the 1997 camp, the children were "between the ages of 8 years 0 months and 9 years 11 months, and had used a cochlear implant for between 4 years 0 months and 6 years 11 months" with the average duration of use being 5.5 years (CID Research Progress Report, 1999, p. 5). Similarly, for the 1998 camp, the children ranged in age from 8.0 years to 9.94 years, with the mean duration of implant use being 5.5 years. Approximately one-half of the children are in oral education programs, with the remaining half in total communication programs (Research, p. 5). Five normal hearing children of similar age serve as a control group.

b). Acoustic Measures/Materials

Vowels: The children in each cochlear implant summer camp and the children in the control group produced a series of sentences, via imitation (Appendix A). Each sentence was produced twice. The vowel and nasal measurements came from selected words in these sentences (Appendix A). An LPC spectral analysis generated the F1, F2, A1, and A2 values of the vowels /i/ and /aw/ after the vowel midpoint was determined by visual inspection.

Nasals: The nasal boundaries were marked by a researcher, and 50 msec of the succeeding vowel were added automatically. Using a formula devised by Richard Goldhor (1995) in his study entitled, "The Perceptual and Acoustic Assessment of the Speech of Hearing Impaired Talkers," a nasal metric was calculated for each nasal produced. In some instances, a nasal was replaced by a different consonant. All consonants were considered and measured except glides. That is, most consonant substitutions were allowed. However, the production was omitted from analysis if the beginning nasal or consonant was completely omitted and the word, instead, began with a vowel. The factors in this metric include:

Factor D: if the segment duration is greater than 40 msec, then D is assigned the value of 2. If the segment is between 10-40 msec, in duration, D is determined by the equation 1+(segment duration-25 msec)/30 msec. Last, if the segment is shorter than 10 msec in duration, D is assigned the value 0 (Goldhor, 1995, p. 536).

Factor L: the term "log power" is synonymous with "energy per unit of time." The factor L examines the difference between the total energy of the nasal and the energy in the nasal below 750 Hz. If the energy difference is greater than 10 dB, L is assigned the value 0. Otherwise, the value of L is determined by the equation 2.0-0.2*energy difference in dB (Goldhor, 1995, p. 537)

Factor R: this factor compares the average power in the vowel below 750 Hz to the average power in the consonant below 750 Hz. A high value is given to R if there is very little difference between the average power of the consonant and the average power of the vowel. A score of 0 is assigned if the total energy between the two variables is less than -25 dB. This would occur if, for example, a very weak consonant such as an /h/ is produced in lieu of a nasal. If, however, the total energy difference is greater than –5 dB, R is assigned the value of 2. If the energy is between –25 dB and –5 dB, the equation (2.5+.1* energy difference in dB) is used to compute R (Goldhor, 1995, p. 537).

A computer program was written, using ensig software, to compute the necessary durations, filtered speech segments, and power values. Once Factors D, L, and R have been calculated, the variables are plugged into the following nasal goodness equation: Goodness=cuberoot (D*L*R)-1.0 (Goldhor, 1995, p. 537). The highest score possible, indicating a clear nasal production, is 1. The lowest possible score is –1.

c). Intelligibility Data

The intelligibility of each subject was determined by collaborators in Texas. Each subject's intelligibility score is based on inexperienced listeners' ability to understand the words in the McGarr sentences (Appendix B). For each content word correctly perceived, the intelligibility score increases. For the subjects with normal hearing, an intelligibility score of 100% was assumed.

Results:

Vowels: Scatter plots were made to show any correlation that may exist between intelligibility and the ratios of A2/A1 and F2/F1 for both deaf speech and the speech of normal hearing children. In the plot of intelligibility versus the average amplitude ratio of the vowel /aw/ (Figure 1), there does not appear to be any correlation. For the normal hearing speakers, the average amplitude ratio falls in the range of .9-1.1, with intelligibility scores of 100%. By contrast, the speech of deaf children yields varying results. Some deaf speakers have amplitude ratios near those of the normal hearing children and yet may have exceptionally low intelligibility scores. In addition, there are some deaf speakers with similar amplitude ratios and near normal intelligibility. The correlation between the average amplitude ratio and intelligibility for the vowel /aw/ is 0.259. Therefore, there is little relation between the amplitude ratio for /aw/ and intelligibility.

Similar results as those stated above were found when examining the relation between intelligibility and the average amplitude ratio of the vowel /i/ (Figure 2). Normal hearing speakers had an average amplitude ratio in the range of .8-1.1 with an intelligibility score of 100%. Again, many deaf speakers yielded similar ratios and high intelligibility scores. However, there is a significant number of deaf children who have normal/near-normal amplitude ratios in conjunction with very low intelligibility scores. There appears, therefore, to be no correlation ($r = 0.157$) between the average amplitude ratio of /i/ and the intelligibility of the speaker.

Intelligibility was also plotted versus the average formant ratios (F2/F1) for the vowels /aw/ and /i/. Since F1 and F2 are relatively close in frequency for a normal production of the vowel /aw/, one can expect that the average formant ratio will not be as great as that for the vowel /i/. In normal hearing speakers, the average formant ratio of the vowel /aw/ fell between 1.6 and 1.9 with an intelligibility score of 100%. Deaf speech, however, had a wide range in average formant ratio values (Figure 3). Some speakers had near normal ratios with near normal intelligibility, while others had near normal ratios with low intelligibility scores. As with the amplitude ratios, there is little correlation between the variables (r = -0.443).

For the vowel /i/, the "phonological space defined by the maximum and minimum frequencies of the first two formants" (Monsen, p. 206) is quite large; therefore, it is not surprising that the average formant ratio is much larger than that of the vowel /aw/. As seen in Figure 4, the ratio for the normal hearing speakers ranges from 6.3-9.4 with 100% intelligibility. The value of the formant ratio for deaf speech, however, is widely scattered. Again, some deaf children had similar ratio values as their normal hearing counterparts, however, intelligibility was not always the same. Upon examination of this data, we also find a much larger distribution of average formant ratio values. There does not appear to be a significant correlation between average formant ratios and intelligibility (r = 0.397).

Nasals: The final nasal metric, as well as the three individual factors, were studied for their correlation with intelligibility. As shown in Figure 5, which examines the relationship between the Factor D of the intended word-initial nasal and intelligibility, the data are widely scattered. Normal subjects' D values range from 1.6 to 2.0 with an intelligibility score of 100%. Many of the deaf subjects achieved a score in this range, indicating that the duration of the production is at or near 40 milliseconds. However, many of these subjects also had very low intelligibility scores, indicating that "correct nasal" duration alone is not sufficient to produce highly intelligible speech. But note, five talkers with D values less than 1.3 are all very poor talkers.

Factor L compared the energy of the intended nasal below 750 Hz to the overall energy in the production. Nasals are characterized by having the majority of their energy below 750 Hz; therefore, any spread of energy to the higher frequencies may reflect a distorted nasal production, or in some cases, the production of a completely different consonant. Using the equation previously stated, a Factor L value was found and plotted against intelligibility, as seen in Figure 6. The majority of the deaf subjects did achieve values at or near 2; yet, again, intelligibility varied greatly among the subjects. Therefore, it appears as though this factor in nasal production does not have much influence on the intelligibility of the production. Note, again, a value of Factor L less than 1.9 corresponds to very poor talkers.

Last, Factor R compared the energy of the intended nasal below 750 Hz to the energy below 750 Hz in 50 milliseconds of the succeeding vowel. As shown in Figure 7, this value is plotted against intelligibility. Again, a value at or near 2 is ideal and may indicate high intelligibility. Normal subjects have values in the range of 1.8 to 2.0 with 100% intelligibility. It is found that most deaf subjects who achieve scores at or near 2 also have high intelligibility. Similarly, those deaf subjects with lower Factor R scores tend to have lower intelligibility scores.

Once these three factors were calculated, the previously mentioned nasal metric equation was used to find the overall measure of nasal goodness. This score was then plotted versus intelligibility (Figure 8). A score of 1 is the highest possible score, indicating an ideal production. Normal values range from 0.8 to 1.0 with 100% intelligibility. Results from the deaf speakers' productions reveal a wide range of scores: from –0.5 to 1.0. The correlation of intelligibility to the nasal metric is 0.572. It is interesting to note, however, that all speakers with nasal metric scores of less than 0.2 also had very low intelligibility scores. As stated earlier, a score of 1.0 indicates an ideally produced nasal. However, intelligibility ranged from 50% to 100% among those speakers who achieved a nasal metric score of 1.0.

Finally, comparisons were made between the intelligibility of orally-educated deaf children and children in total communication programs. Data from the 1997 summer camp were analyzed for a possible relationship between intelligibility and communication mode. It was assumed that the normal hearing subjects had 100% intelligible speech. As seen in Figure 9, children enrolled in oral education programs had significantly more intelligible speech than children using the total communication mode. This difference in intelligibility may be attributed to the emphasis on speech and auditory cues offered in oral education programs; however, further research is needed.

4.) Conclusion

Vowels: Examining the amplitude and formant ratios of certain vowels may assist us in our understanding of the spectral information offered to cochlear implanted children in various educational settings. This study found that there appears to be no significant correlation between amplitude ratios and formant ratios of the vowels /aw/ and /i/. Many deaf speakers produced normal/near-normal amplitude and formant ratio values yet had very low intelligibility scores. Therefore, it does not appear that this information alone is a good predictor of speech intelligibility.

Nasals: The duration of the production (Factor D), the ratio of low frequency energy to total energy in the production (Factor L), and the comparison between the low frequency energy in the nasal to the low frequency energy in the succeeding vowel (Factor R) were examined and a nasal metric was calculated. These individual factor scores were also correlated with intelligibility. From this study, it does not appear that there is a strong correlation between the nasal metric and intelligibility. While it appears that a predictor score of less than 0.2 is associated with low intelligibility, there are many instances in which normal/near-normal nasal metric scores are also associated with low intelligibility. However, there are low nasal metric scores that correlate with a high intelligibility score. Therefore, it again appears that examining these nasal characteristics are not a good predictor of intelligibility.

It is important to realize that the normal hearing speakers were given an arbitrary score of 100% intelligibility. Perhaps in future studies, the normal hearing speakers should also produce the McGarr sentences and receive a more accurate intelligibility score based on the total key words correctly perceived by inexperienced listeners.

REFERENCES

Borden, G., Harris, K., & Raphael, L. *Speech Science Primer*. Williams & Wilkins, 1994.

Central Institute for the Deaf. Research Progress Report. September, 1999.

Goldhor, R. The Perceptual and Acoustic Assessment of the Speech of Hearing-Impaired Talkers. *Applied Speech Technology*, 1995, 521-525.

Kent, R., & Read, C. *The Acoustic Analysis of Speech*. Singular Publishing Group, 1992.

McGarr, N. The Intelligibility of Deaf Speech to Experienced and Inexperienced Listeners. *Journal of Speech and Hearing Research*, 1983, 26, 451-458.

Monsen, R. Toward Measuring How Well Hearing Impaired Children Speak. *Journal of Speech and Hearing Research*, 1978, 21, 197-219.

Appendix A:

1.) I am **tall**.
2.) *He* has a blue pen.
3.) Chuck *seems* thirsty after the race.
4.) She <u>*needs*</u> **strawberry** jam on her toast.
5.) Did you like the zoo this spring?
6.) Daddy took his <u>new</u> shoes.
7.) Show <u>*me*</u> the little duck.
8.) Sit at the table please.
9.) Sue's friend **bought** a toy ship.
10.) Feel the **soft** dog.
11.) <u>May</u> I *see* that rock?

**Note: Italicized words indicate samples measured acoustically for /i/ data; bold faced words indicate samples measured acoustically for /aw/ data. Underlined words indicate samples used in the nasal portion of the study.

Appendix B

McGarr (1983) Sentences

High Context

3 Syllables

Keep quiet.
Read the book.
Come *with* me.
Comb your *hair*.
That's no *good*.
The *dog* barks.

5 Syllables

The *cat* chased the mouse.
My *name* is Nancy.
Get your *coat* and hat.
Get your *ball* and bat.
Did you brush your *teeth*?
Is there no *more* milk?

7 Syllables

That *man* is not my father.
I *wish* I had a pony.
We have *food* for the picnic.
The flag is *red*, white and blue.
May I have a *piece* of cake?
Can you dive in *deep* water?

Low Context

3 Syllables

Feed the dog.
Have a lot.
You *did* it.
I *need* it.
Get the *cake*.
This is *his*.

5 Syllables

They *will* come again.
Is *that* the tall one?
Mother *has* the car.
Who wants *this* ice cream?
It's easy to *hear* her.
He said he *could* go.

7 Syllables

The *book* is on the table.
What *was* the name of that boy?
If it's *cool* I cannot go.
Is the *fat* baby crying?
It is nice on a *fall* day.
We will go to the *beach* today.

Words in italics indicate the test word.

Figure 1

Intelligibility vs. Average Amplitude Ratio (A2/A1) /aw/

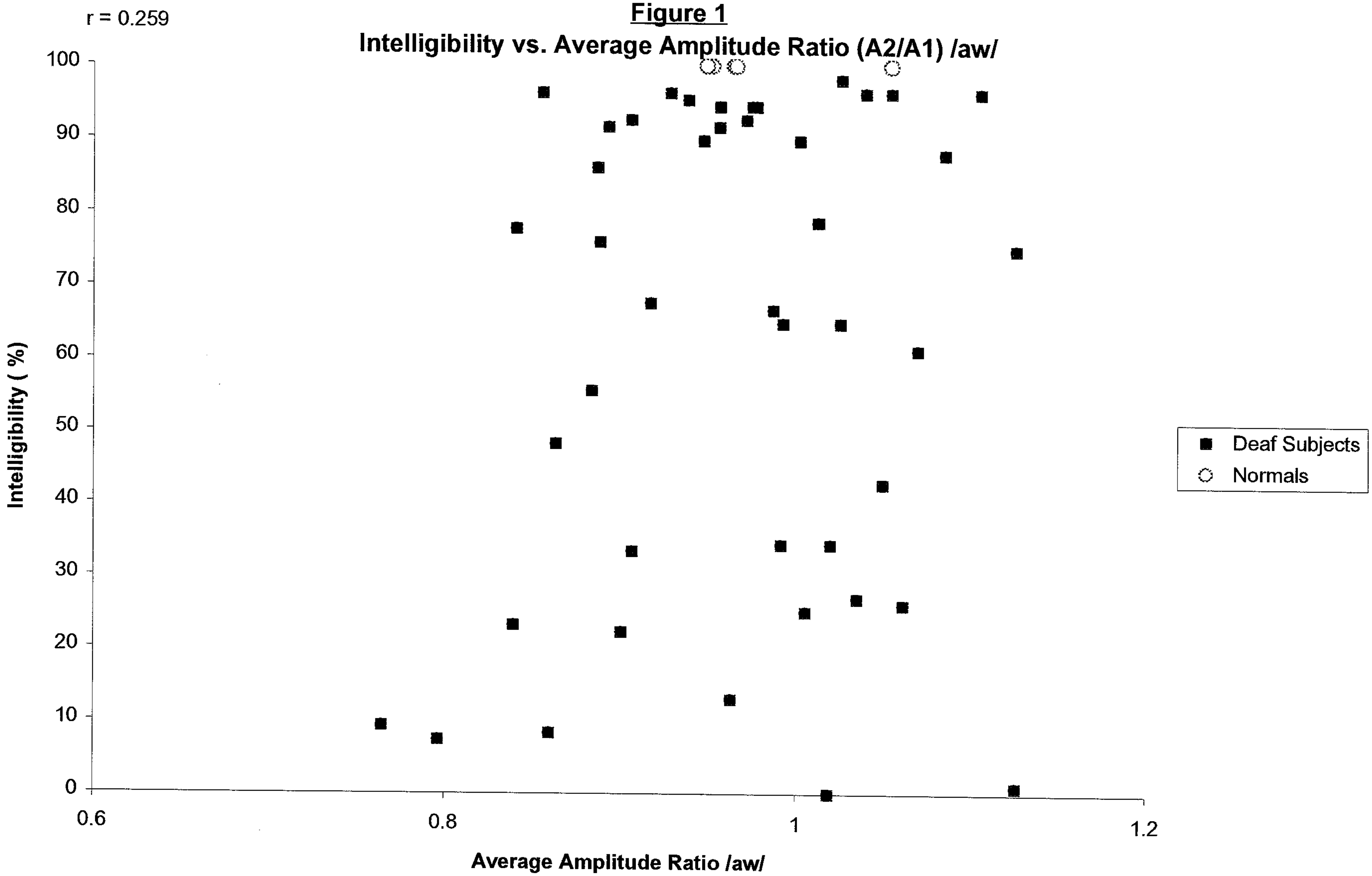
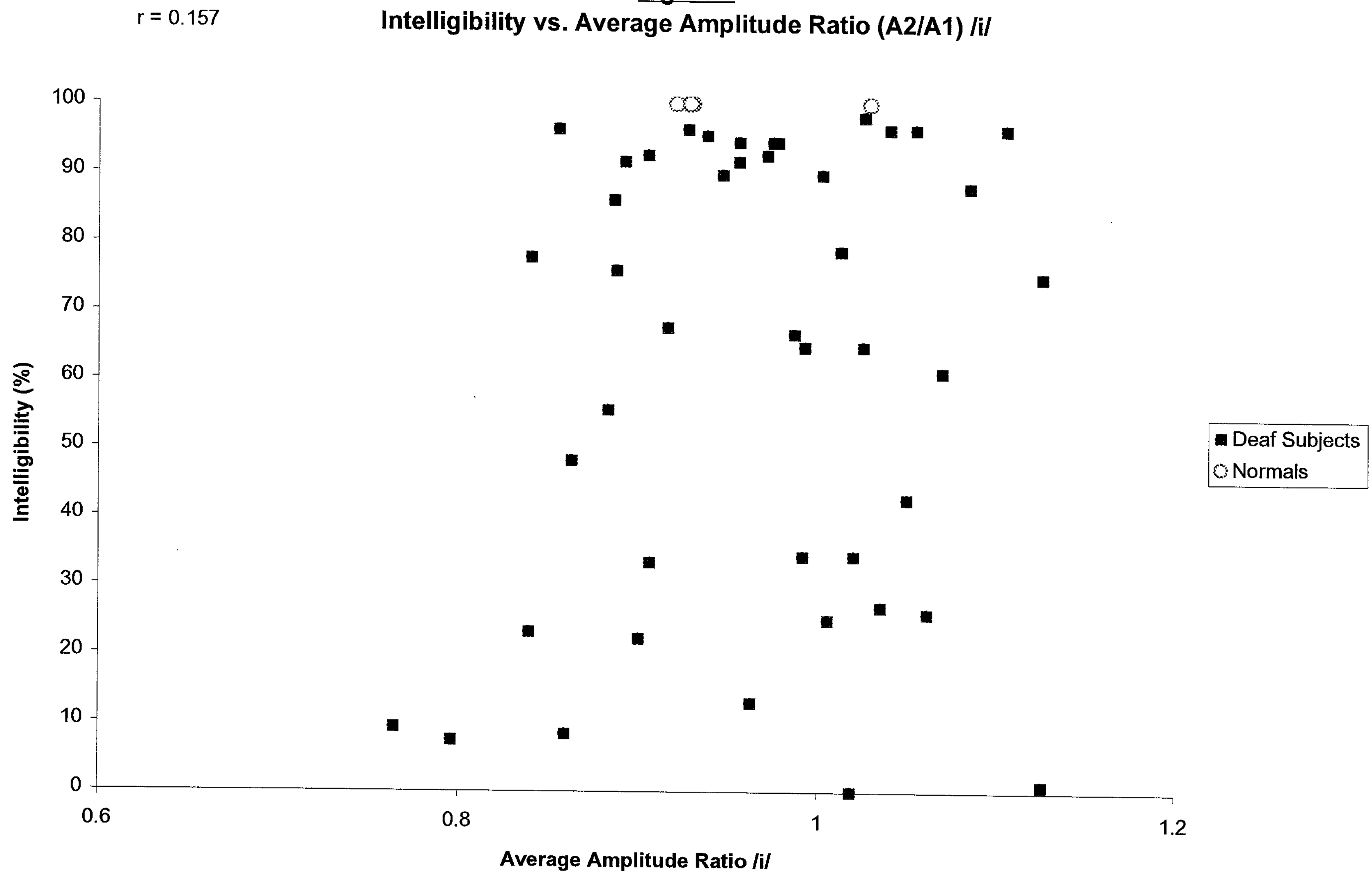


Figure 2

Intelligibility vs. Average Amplitude Ratio (A2/A1) /i/



r = -0.443

Figure 3

Intelligibility vs. Average Formant Ratio (F2/F1) /aw/

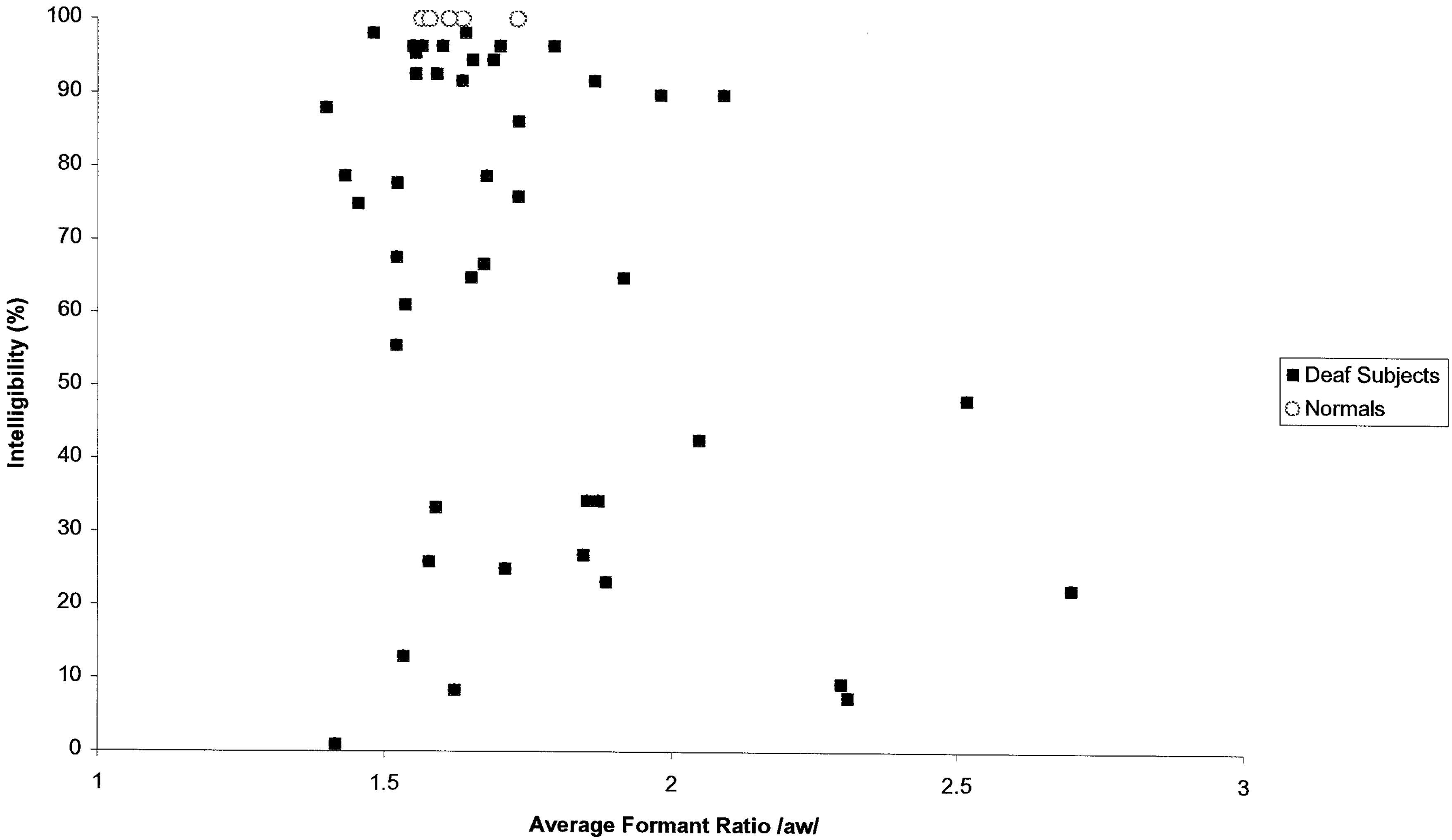


Figure 4

Intelligibility vs. Average Formant Ratio (F2/F1) /i/

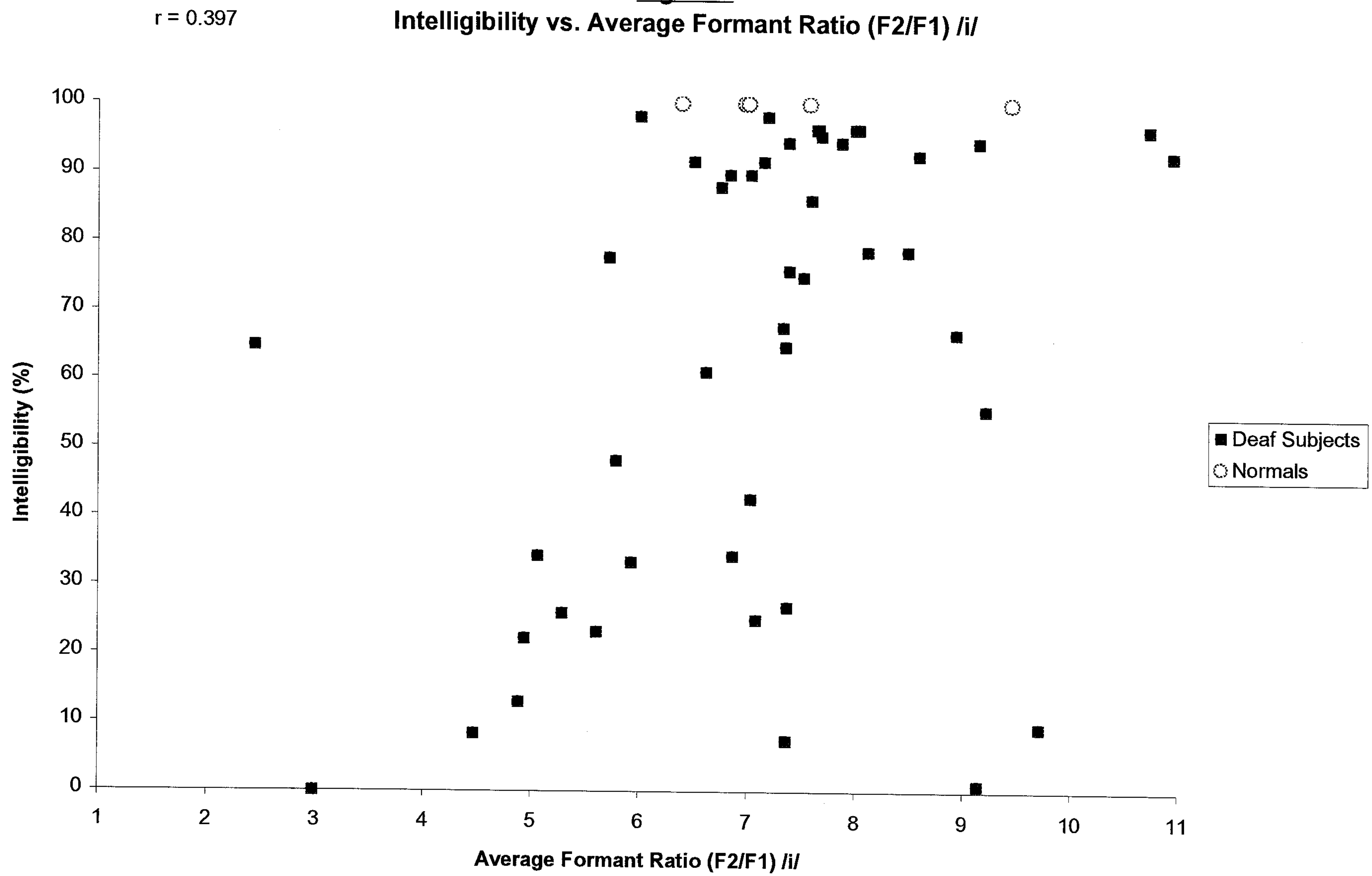


Figure 5

Intelligibility vs. Factor D

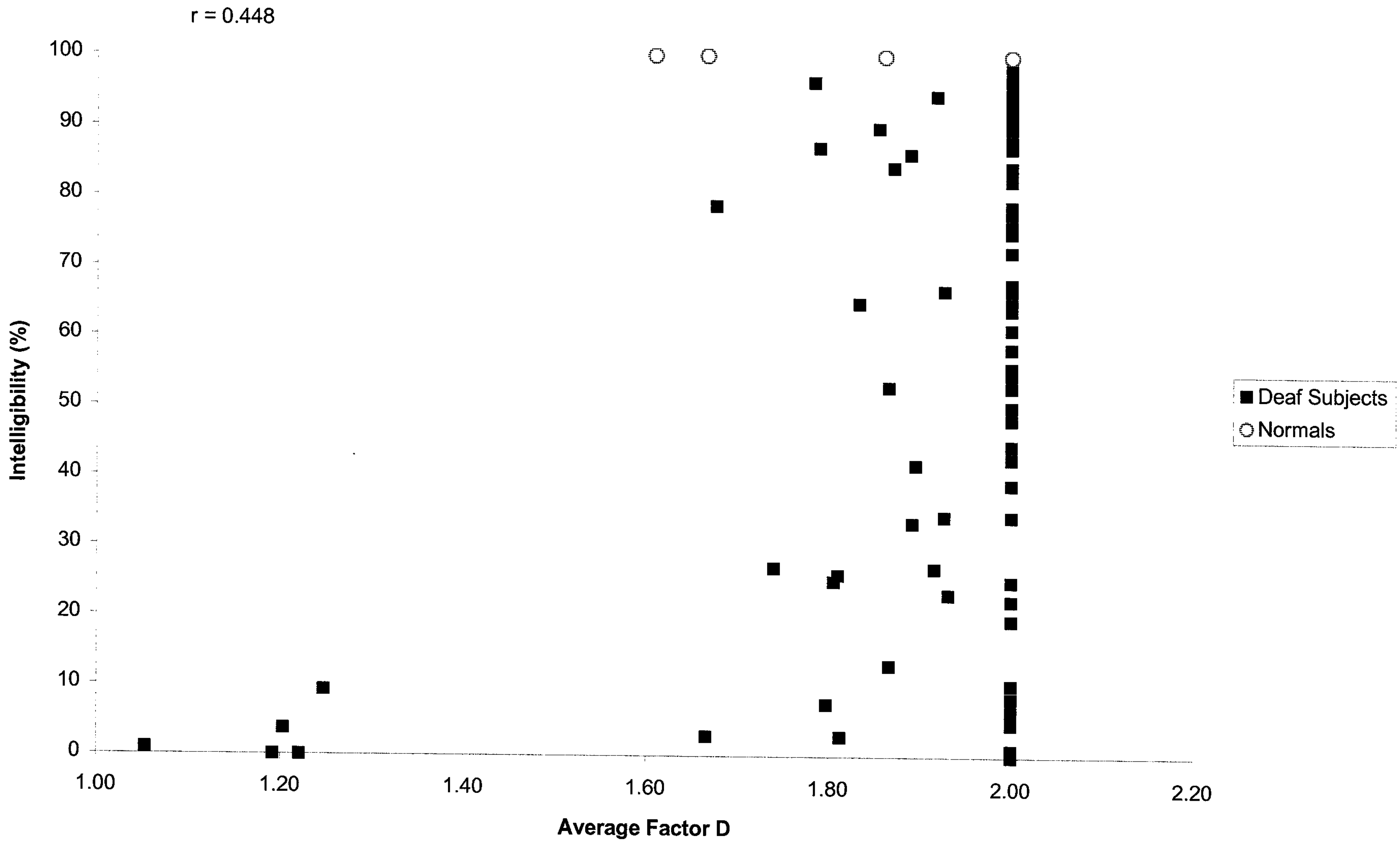


Figure 6
Intelligibility vs. Factor L

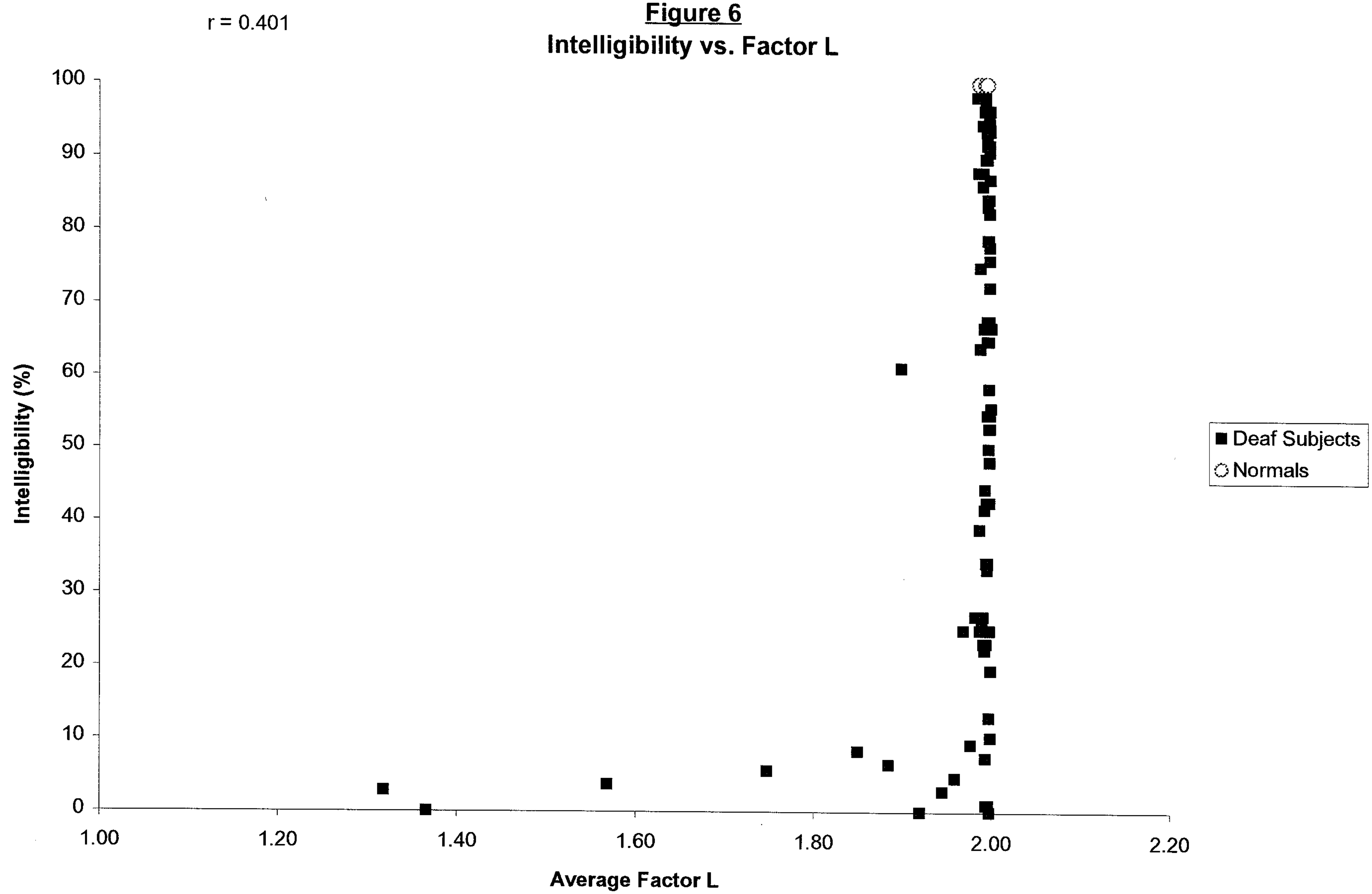


Figure 7

Intelligibility vs. Factor R

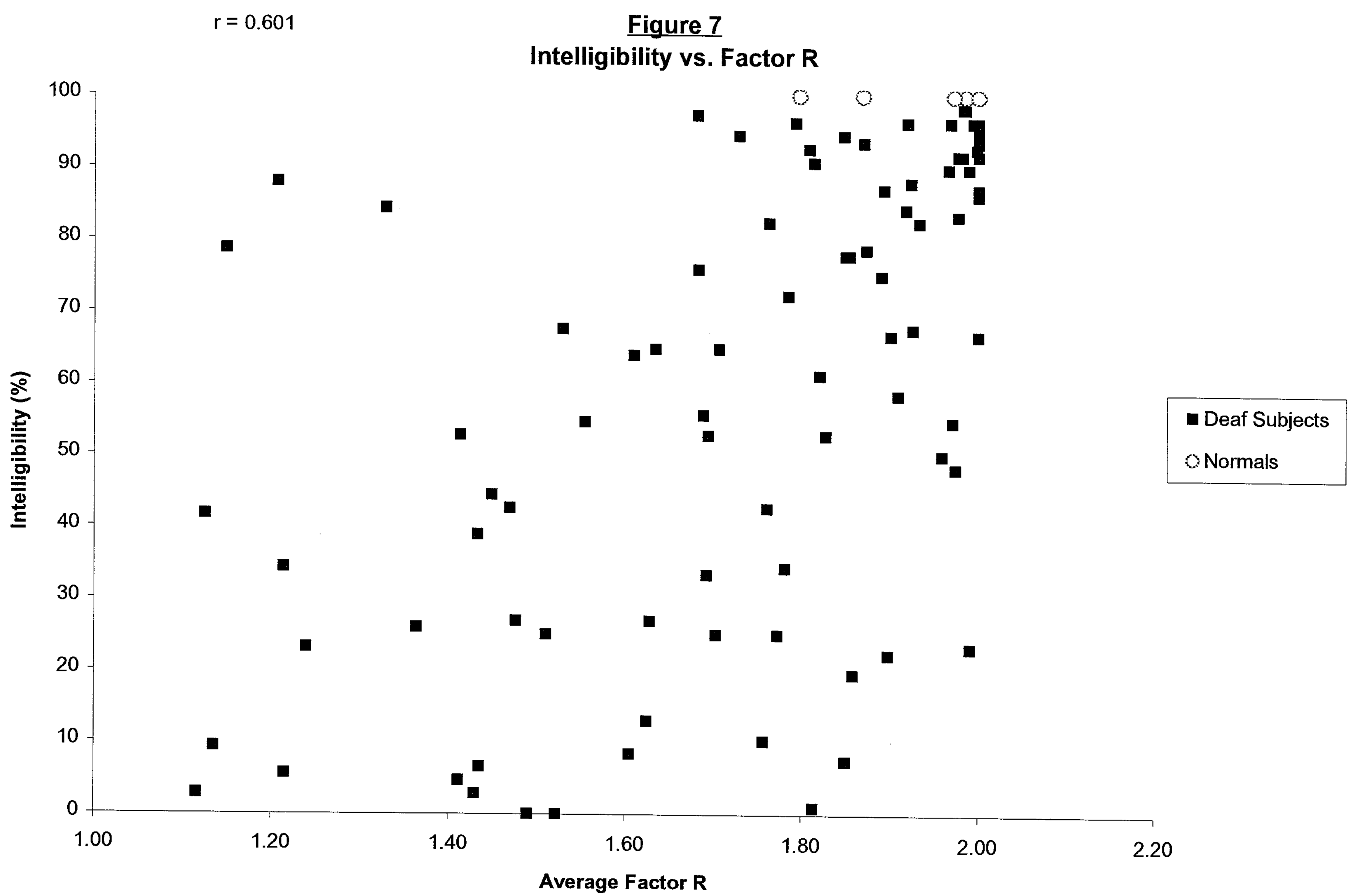


Figure 8

Intelligibility vs. Nasal Metric

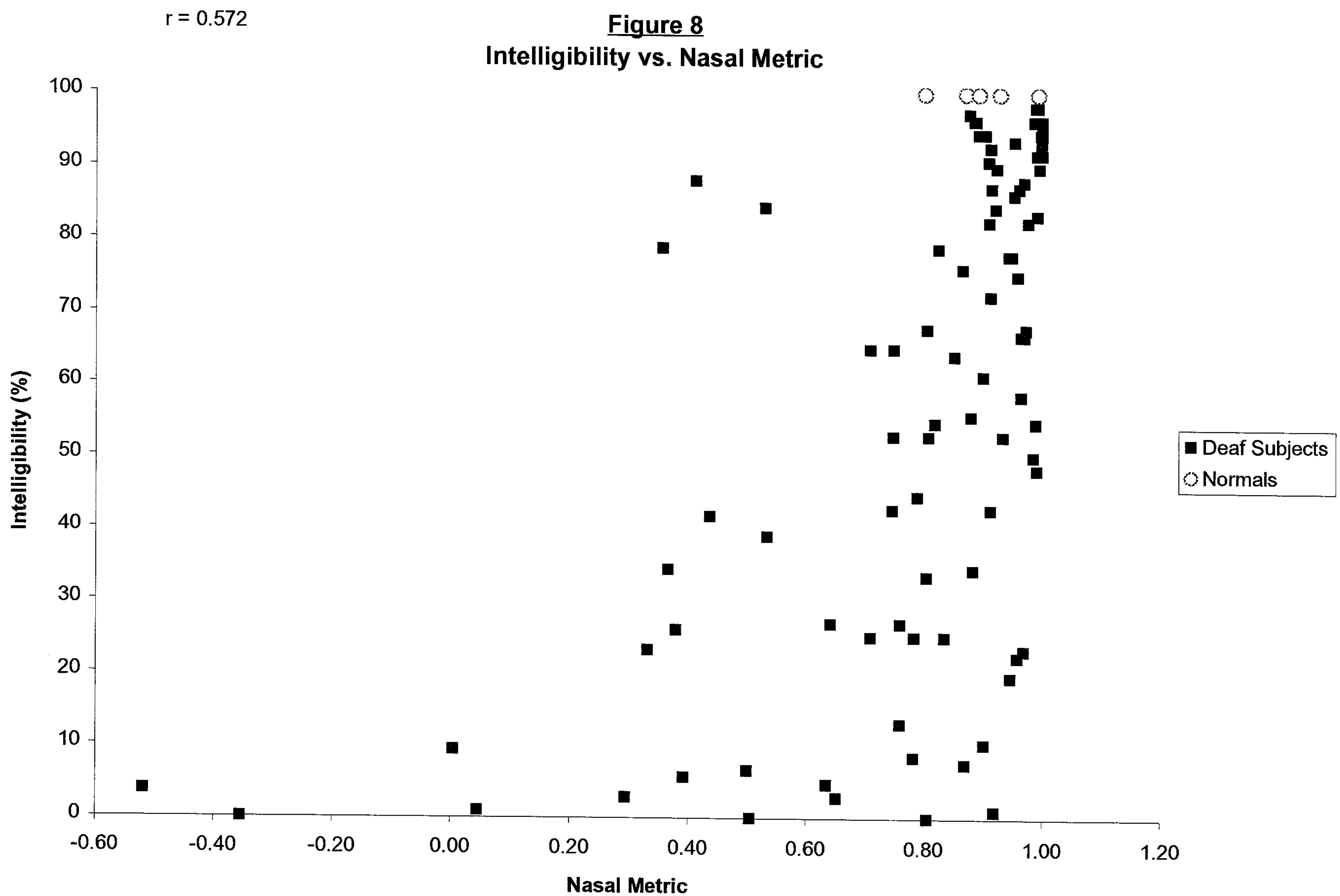


Figure 9

Average Overall Intelligibility vs. Communication Mode