JSLHR, Volume 40, 349-360, April 1997

Effect of Speech Dialect on Speech Naturalness Ratings: A Systematic Replication of Martin, Haroldson, and Triden (1984)

Linda S. Mackey
Claremore Regional Hospital
Home Health, OK

Patrick Finn
University of New Mexico
Albuquerque

Roger J. Ingham
University of California, Santa Barbara

This study investigated the effect of speech dialect on listeners' speech naturalness ratings by systematically replicating Martin, Haroldson, and Triden's (1984) study using three groups of speaker samples. Two groups consisted of speakers with General American dialect—one with persons who stutter and the other with persons who do not stutter. The third group also consisted of speakers who do not stutter but who spoke non-General American dialect. The results showed that speech naturalness ratings distinguished among the three speaker groups. The variables that appeared to influence speech naturalness ratings were type of dialect, speech fluency, and speaking rate, though they differed across speaker groups. The findings also suggested that strength of speech dialect may be a scaleable dimension that judges can rate with acceptable levels of reliability. Dialect may also be an important factor that needs to be incorporated or controlled within systems designed to train speech naturalness ratings. It may also be an important factor in determining the extent to which stuttering treatment produces natural sounding speech.

KEY WORDS: speech naturalness, stuttering, reliability, multicultural, assessment

Considering current demographic trends, it is clear that there has been a significant increase in cultural diversity in the United States (Taylor, 1993). One consequence of this change for speech-language pathologists is that, regardless of their own cultural background, they should expect to provide clinical services to clients with backgrounds different from their own (Battle, 1993). Numerous researchers (Cooper & Cooper, 1993; Leith, 1986; Shames, 1989; Watson & Kayser, 1994) have discussed the implications of this difference for the disorder of stuttering and have reached the conclusion that clinicians who treat persons who stutter from cross-cultural backgrounds should be responsive to the effect of these differences on the clinical relationship. Surprisingly, there has been little discussion of the impact these differences may have on the reliability and validity of current measures for stuttering treatment. This assumes that current speech measures are suitable for clients from different cultural backgrounds.

There are several reasons for believing that regardless of the cultural background of the person who stutters, measures of stuttering frequency, speech rate, and speech quality will continue to be mainstays in stuttering treatment evaluation. First, there is convincing evidence

1092-4388/97/4002-0349

that speech characteristics that typically describe stuttering are evident across cultures (Bloodstein, 1995; Van Riper, 1982). Second, in a recent unpublished survey, bilingual speech-language pathologists generally agreed that current stuttering treatment procedures and goals are suitable for the range of cultural groups found in America (Cooper & Cooper, 1993). Third, researchers in other countries have now begun to employ these measures to evaluate their stuttering treatment programs. Nevertheless, there is little empirical data on the reliability of these measures when the clinician and client do not share the same cultural background. Such information would seem especially important in the case of speech quality measurement. Conceivably, a clinician's speech naturalness ratings may be influenced by the extent to which a client's cultural dialect departs from the clinician's cultural dialect.

The most widely used measure of speech quality in stuttering research is the speech naturalness rating scale described by Martin, Haroldson, and Triden (1984). (For a comprehensive review see Schiavetti and Metz, 1997.) This 9-point scale was introduced for the purpose of assessing speech quality because of problems with speech naturalness emerging from treatments based on prolonged speech (Martin, 1981; Onslow & Ingham, 1987). Considerable research has since demonstrated that listeners are able to use this scale with satisfactory levels of intra- and interjudge agreement (Ingham, Gow, & Costello, 1985; Martin & Haroldson, 1992; Martin et al., 1984; Schiavetti, Martin, Haroldson, & Metz, 1994), although findings from some studies have not been as convincing (Finn & Ingham, 1994; Kalinowski, Noble, Armson, & Stuart, 1994; Onslow, Adams, & Ingham, 1992).

The clinical validity of the speech naturalness scale is also well documented. Research has shown that naturalness ratings can reliably distinguish between the speech of persons who stutter, persons treated for stuttering, and persons who do not stutter (Ingham, Gow, & Costello, 1985; Ingham & Onslow, 1985; Martin et al., 1984; Onslow, Hays, Hutchins, & Newman, 1992; Schiavetti et al., 1994). More importantly, clinicians' naturalness ratings can be used not only to evaluate treatment (Onslow, Costa, Andrews, Harrison, & Packman, 1996), but also to improve the quality of speech resulting from treatment (Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Ingham & Onslow, 1985).

In light of these findings, it is important to learn more about variables that might influence listeners' speech naturalness ratings. In speech-language pathology, there is considerable evidence that speech rate and speech fluency scores are correlated with naturalness ratings (see Finn & Ingham, 1994; Martin & Haroldson, 1992; Martin et al., 1984; Onslow, Hays, et al., 1992), although these correlations have varied (see Ingham, Gow, & Costello, 1985). Voice onset time, sentence duration (Metz, Schiavetti, & Sacco, 1990), and phonation duration (Gow & Ingham, 1992) may be related to naturalness judgments as well. However, there may be other general speech behavior factors that affect the rating scale's usefulness in routine clinical conditions. Among the most relevant might be variations in clients' speech dialect.[1]

In the area of applied linguistics, there are numerous studies focusing on the influence of a speaker's dialect (and its various components) on ratings and judgments of the speaker's communicative proficiency, ascribed speaker status, and numerous social and psychological variables or traits (e.g., Brennan & Brennan, 1981; Giles, Hewstone, Ryan, & Johnson, 1987; Giles, Williams, Mackie, & Rosselli, 1995; Munro & Derwing, 1994; Ryan & Carranza, 1975; Ryan & Giles, 1982; Taylor, 1973). Many studies have shown that phonetic norms exist for various properties of native speech, such as voice onset time, vowel duration before stops, vowel formant frequencies, and stress placement (Crowther & Mann, 1992; Flege, 1987; Flege & Bohn, 1989; Flege & Hillenbrand, 1984; Juffs, 1990; Port & Mitleb, 1983). It has also been shown that native listeners can translate variation away from these properties into degrees of perceived accentedness (Anderson-Hsieh, Johnson, & Koehler, 1992; Munro, 1993).

Once such perceptual distinctions are made, there is a tendency for native-speaker listeners to downgrade non-native speakers on ratings of competence and social acceptability in terms of these perceived degrees of accentedness (Fayer & Krasinski, 1987; Kalin & Rayko, 1978; Koster & Koet, 1993; Ryan & Carranza, 1975). The tendency to recognize accents and then attach evaluative judgments to them has also been noted with ethnic and regional variations in the same language (e.g., Gallois & Callan, 1989; Giles & Sassoon, 1983; Luhman, 1990; Nesdale & Rooney, 1990; Pittam, 1987; Taylor, 1973). Given the significant influence of accent or dialect on various kinds of evaluative judgments by native listeners, it is likely that speech dialect will also have an impact on

[1]In the present study no distinction is made between the terms *dialect* and *accent*. For our purposes, it is assumed that *dialect* is a generic term that encompasses accent. Several points support this assumption. First, because we were using student judges who were linguistically naïve (e.g., they would not view language technically as linguists), it was determined that a more general terminology consistent with their expectations should be employed. For this purpose, the term *dialect* is generally accepted as the nontechnical equivalent of *accent*. This statement aligns with dictionary definitions of the terms (see Simpson & Winer, 1989) and with general accepted usage (Edwards, 1989; Gardner, 1985; Lippi-Green, 1994). Second, the use of the terms *accent* and *dialect* as equivalent terms is often found in technical articles in applied linguistics (e.g., Cargile, Giles, Ryan, & Bradac, 1994; Giles et al., 1995). Third, the term *dialect* was used and defined as the cover term of *accent* within the instructions (see *Methods*) to the judges in the present study. Finally, our assumptions were verified as evidenced by the experimental differences obtained. As reported in the *Results* section, our judges were able to distinguish differences in accentedness in the *D* group (see *Method*) while using the nontechnical term *dialect*.

the ratings of speech naturalness as measured by the speech naturalness scale. Signs of this effect were noted within a study by Onslow, Adams, and Ingham (1992). They examined a number of variables that might influence the reliability of listeners' ratings during stuttering treatment. A post hoc analysis revealed that speaker dialect may have had an unanticipated, negative impact on listener reliability. It was this finding, in particular, that suggested the importance of a systematic investigation of the effect of dialect on speech naturalness ratings.

The general purpose of the present study was to conduct a systematic replication of Martin et al.'s (1984) investigation of speech naturalness ratings of persons who stutter and do not stutter by determining the effect of speech dialect on such ratings. The specific aims were (a) to determine whether speakers who do not stutter but who were judged to have relatively strong dialects receive speech naturalness ratings that are significantly different from the ratings of persons who either stutter or do not stutter, yet are judged to have no evidence of dialect; (b) to assess the relationship between speech naturalness ratings and the dimensions of dialect, disfluency, and rate; and (c) to assess the effect of speech dialect on the reliability of speech naturalness ratings.

Method

Participants

Three groups of participants, with 10 participants in each group, were assembled and labeled as follows:

1. Ten mild-to-severe speakers who stutter (9 males, 1 female), ranging in age from 17 to 38 years, who spoke with a General American English dialect. This group of nondialect speakers was labeled Group *S*. The basis for describing these speakers as nondialect is described later.
2. Ten speakers who do not stutter (5 males, 5 females), ranging in age from 20 to 26 years, who also spoke with a General American English dialect. This group of nondialect speakers was labeled Group *N*.
3. Ten speakers who do not stutter (5 males, 5 females), ranging in age from 19 to 30 years, who spoke with dialects other than General American English. English was either the first or second language of these speakers who were selected to provide a heterogeneous array of samples.[2] Four participants spoke English as their first language and presented dialects characterizing their birthplace, culture, and education—an Australian, British, Canadian, and New Zealand dialects. Six speakers spoke English as their second language and presented with dialects characteristic of their first language. Three had learned Chinese as their first language, two had learned Vietnamese, and one had learned Dutch. This group of different dialect speakers was labeled Group *D*. For this study, types of dialect were allowed to vary as long as they met the operational definition described below.

[2]Selection of a heterogeneous array of speakers is consistent with a similar strategy of including a wide range of language backgrounds in many attitude studies in applied linguistics (e.g., Anderson-Hsieh et al., 1992). It is believed that such heterogeneity will ensure greater authentic variability in degrees of accentedness and in the number and frequency of dialectical features present. The variability is needed for the judges to establish their perceptual norms upon which they will base their evaluation or ratings.

Speech Samples: Selection and Measures

Speakers

Participants were audio-recorded in a sound-treated room while speaking spontaneously on self-selected topics for 5 to 20 minutes. The speech samples from the 10 persons who stutter were originally recorded for a Stuttering Measurement Training Program (Ingham & Ingham, 1986). The first author then selected a one-minute speech sample from each speaker's recording; the main requirement being that the sample did not contain excessive nontalking time.

The three speaker groups' speech samples were randomized within groups and dubbed to a master tape with a 7-s interval between samples. Second and third tapes were also prepared with the running order of the speaker groups alternated. Three practice samples were placed at the beginning of all tapes and were not included in the data analysis.

Dialect Sample Evaluation

Twelve undergraduate speech and hearing students, born and raised in California (age range: 19 to 24 years), rated all 30 samples for strength of speech dialect relative to General American English. All 12 students reported General American English as their dialect and English as their first language. They rated each sample using a 9-point Speech Dialect Rating Scale, which was a modified version of the Speech Naturalness Rating Scale (see Martin et al., 1984). The instructions were as follows:

> You will hear a number of 1-minute speech samples….Your task is to rate the speech production dialect for each speech sample in relation to a General American English dialect. (For your information, *Webster's Dictionary* defines General American English as "the native speech of natives of the United States whose speech is not that of the South or of the r-dropping Northeast."[3]) If no dialect is noted in the sample, circle

[3]Gove, 1986, p. 944.

the 1 on the scale. If a very strong dialect is noted in the sample, circle the 9 on the scale. If the sample sounds somewhere between "no dialect" and a "very strong dialect," circle the appropriate number on the scale. Do not hesitate to use the ends of the scale (1 or 9) when appropriate.

Given this operational definition, speech samples rated close to 1 on the scale suggested that judges perceived a dialect similar to their own (General American English). Thus, these samples could be described as being obtained from Californian American or nondialect speakers. In contrast, samples rated close to 9 suggested a dialect different from the listeners' dialect. Thus, these samples could be described as being obtained from dialect speakers.

The student judges rated the 33 1-minute samples; the first 3 were practice samples. At least one week later, they rerated the same samples in a different randomized order. When their ratings were averaged across these two judgment occasions there was essentially no within-judge overlap between the rating given to the *D* group and to the other groups. It was found that the 10 normal speakers with non-General American English dialect (Group *D*) were assigned speech dialect ratings that exceeded the highest rating given by any one rater to any of the other 20 speaker samples.

Speech Rate

Using transcripts of the speech samples and real time measures, two speech and hearing graduate students independently counted syllables spoken in each speaker's sample. Agreement estimates were based on a frequency ratio in which the lower syllable count was divided by the higher syllable count for each sample. Average interrater agreement was 95%.

Frequency of Stuttering

As mentioned above, the samples of the persons who stutter were obtained from a judgment training program. In preparing this program five experienced (minimum 5 years) clinicians identified syllables stuttered on the basis of a perceptual threshold definition (Martin & Haroldson, 1981). The clinicians were provided with the sample transcripts and repeatedly viewed each sample in 5-s units. If 4 out of 5 clinicians agreed a syllable was stuttered, then it was counted as a moment of stuttering. Therefore, by definition, the minimum level of syllable-by-syllable interrater agreement was 80%. The mean percentage of syllables stuttered (%SS) was 9.5 %SS, with a range of 2.6 to 23.6 %SS (see Appendix).

Two speech and hearing graduate students rated the nonstuttering speakers' samples for moments of stuttering using the same definition of stuttering as above. These students had been previously trained to judge stuttering on the Stuttering Measurement Training Program (Ingham & Ingham, 1986) mentioned above. They completed this program by achieving agreement with the instances and, hence, frequencies of stuttering previously recorded for the *S* group. These two judges recorded no stuttering events in any of the nonstuttering speakers' samples.

Frequency of Normal Disfluencies

The same two graduate students also independently counted the number of normal disfluencies for all 30 speaker samples. Normal disfluencies were defined as normal disruptions in speech flow that were not perceived as moments of stuttering. They completed a rerating of all samples at least one week later. In terms of frequency ratios, average intrarater agreement was close to 80% (S = 75%, D = 81%, N = 83%). Overall interrater agreement was 77% (S = 88%, D = 62%, N = 81%). As mentioned below (see *Discussion*), these relatively low agreement scores may lower confidence in the disfluent syllable data, but they were retained because they still yielded interpretable findings.

Experimental Procedure

Thirty undergraduate speech and hearing students (age range: 19 to 24 years) served as judges for the speech naturalness rating task. All reported General American English as their dialect and English as their first language. No judge participated in any other part of this study. These judges rated the same 30 speech samples (minus the 3 practice samples) on two occasions with 1 to 3 weeks between rating occasions. The sample order was randomized across occasions as described earlier.

For both occasions, the judges were instructed to rate the naturalness of each speech sample on a 9-point speech naturalness scale, where 1 represented *sounds highly natural* and 9 represented *sounds highly unnatural*. Their instructions were identical to those used by Martin et al. (1984).

Results

Speech Naturalness Ratings

The mean, range, and standard deviation for each speaker group are displayed in Table 1. Mean naturalness ratings were 5.59 for the S, 3.43 for the D, and 1.71 for the N. An ANOVA showed a significant main effect for speaker group (F = 51.23, df = 2,81, $p < .01$), but a nonsignificant effect for listening order of speech samples and for any interaction between speaker group and listening order. A Newman-Keuls test revealed that the S mean was significantly ($p < .01$) higher than the other

Table 1. Mean, range, standard deviation, and total number of speech naturalness ratings assigned by 30 judges to 10 speech samples each from the following groups: Nondialect Speakers Who Stutter (*S*), Dialect Speakers Who Do Not Stutter (*D*), and Nondialect Speakers Who Do Not Stutter (*N*).

				Speech Naturalness Rating Scores								
Speech group	*M*	Range	*SD*	1	2	3	4	5	6	7	8	9
S	5.59	1–9	2.46	3	38	47	25	31	31	34	44	47
D	3.43	1–8	1.92	54	63	59	30	42	29	18	5	0
N	1.72	1–6	0.94	161	90	32	12	4	1	0	0	0

speaker groups and the *D* mean was significantly ($p < .01$) higher than the *N* mean. Eta squared, the proportion of variance in speech naturalness ratings accounted for by group membership (Young, 1994), was 64.65%, and the proportion of variance misclassified was 2.5%.

Acceptable levels of intrarater agreement were defined as two ratings obtained on separate occasions of the same sample by the same judge that were identical or differed no more than plus or minus one (±1.0) rating score (Martin et al., 1984). For each group of speaker samples, 300 comparisons were possible (10 samples and 30 judges). Table 2 shows that intrarater agreement (within ±1.0) was 76% for the *S* group, 80% for the *D*, and 98% for the *N*. Overall intrarater agreement (within ±1.0) was 85%.

For interrater agreement, each judge's rating of a speech sample was compared with the ratings of the same sample by the other 29 judges. A total of 4,350 comparisons were possible for each speaker group, (30 x 29)/2 x 10. As summarized in Table 2, interrater agreement (within ±1.0) was 59% for the *S*, 56% for the *D*, and 80% for the *N*. Overall interrater agreement (within ±1.0) was 65%.

The effect of gender differences between the *S* group (9 males; 1 female) and the *N* and *D* groups (5 males; 5 females) was investigated as a possible source of influence on the speech naturalness ratings. A Mann-Whitney (Siegel, 1956) test of differences between the naturalness ratings given to the 10 male and 10 female speakers in the *N* and *D* groups found no significant differences ($U = 48$, $p = ns$). In other words, there was no indication that the different gender ratios across the groups influenced the pattern of speech naturalness ratings.

Table 2. Cumulative number and percentage of intrarater and interrater agreements of speech naturalness rating scores for the Nondialect Speakers Who Stutter (*S*), Dialect Speakers Who Do Not Stutter (*D*), and Nondialect Speakers Who Do Not Stutter (*N*).

	Agreement (in rating scores)								
Speech group	±0.0	±1.0	±2.0	±3.0	±4.0	±5.0	±6.0	±7.0	±8.0
				Intrarater Agreement					
S	113	227	271	289	297	300	300	300	300
	38%	76%	90%	96%	99%	100%	100%	100%	100%
D	126	239	277	296	300	300	300	300	300
	42%	80%	92%	99%	100%	100%	100%	100%	100%
N	192	293	298	300	300	300	300	300	300
	64%	98%	99%	100%	100%	100%	100%	100%	100%
				Interrater Agreement					
S	1102	2554	3301	3845	4188	4298	4331	4350	4350
	25%	59%	76%	89%	96%	99%	100%	100%	100%
D	1005	2383	3280	3805	4144	4304	4350	4350	4350
	23%	56%	75%	87%	95%	99%	100%	100%	100%
N	1790	3499	4070	4285	4339	4350	4350	4350	4350
	41%	80%	94%	99%	100%	100%	100%	100%	100%

Dialect Ratings

As summarized in Table 3, the mean dialect rating for Groups *N* and *S* was, respectively, 1.95 and 1.80, whereas for Group *D* it was 7.26. An ANOVA showed a significant main effect for speaker group ($F = 155.61$, $df = 2,29$, $p < .01$). A Newman-Keuls test revealed that the *D* mean was significantly ($p < .05$) higher than the other speaker groups and there was no significant difference between the *S* and *N* means. At the same time, Table 3 also shows that there were some speakers among the *S* and *N* samples who were also judged to have relatively strong dialects.

As shown in Table 4, intrarater agreement was based on 12 judges rating 10 samples (120 ratings) for each of the three groups. Following the procedure used by Martin et al. (1984), acceptable levels of agreement were defined as two ratings obtained on separate occasions of the same sample by the same judge that were identical or differed no more than plus or minus one (±1.0) rating score. Using this criterion, intrarater agreement was 87% for the *S*, 93% for the *D*, and 83% for the *N*. Overall intrarater agreement was 87.7%—essentially identical to the level of intrarater agreement (88%) for rating speech naturalness reported by Martin et al.

For interrater agreement, each judge's rating of a speech sample was compared with the ratings of the same sample by the other 11 judges (see Martin et al., 1984). For each speaker group, a total of 660 paired comparisons was possible, (12 x 11)/2 x 10. Acceptable levels of agreement were again defined as two ratings of the same sample by different judges that were identical or within ±1.0 rating score. Table 4 summarizes the frequency of agreement for each group of speech samples.

Table 3. Mean, range, standard deviation, and total number of speech dialect ratings assigned by 12 judges to 10 speech samples each from the following groups: Nondialect Speakers Who Stutter (*S*), Dialect Speakers Who Do Not Stutter (*D*), and Nondialect Speakers Who Do Not Stutter (*N*).

				Speech Dialect Rating Scores								
Speech group	*M*	Range	*SD*	1	2	3	4	5	6	7	8	9
S	1.95	1–9	1.40	43	49	16	5	3	1	1	1	1
D	7.26	2–9	1.70	0	1	4	5	11	13	18	32	36
N	1.80	1–6	1.10	61	42	10	6	0	1	0	0	0

Table 4. Cumulative number and percentage of intrarater and interrater agreements of speech dialect rating scores for the Nondialect Speakers Who Stutter (*S*), Dialect Speakers Who Do Not Stutter (*D*), and Nondialect Speakers Who Do Not Stutter (*N*).

	Agreement (in rating scores)								
Speech group	±0.0	±1.0	±2.0	±3.0	±4.0	±5.0	±6.0	±7.0	±8.0
				Intrarater Agreement					
S	55	104	110	115	116	117	118	120	120
	46%	87%	92%	96%	97%	98%	98%	100%	100%
D	66	111	120	120	120	120	120	120	120
	55%	93%	100%	100%	100%	100%	100%	100%	100%
N	55	100	112	112	118	119	120	120	120
	46%	83%	93%	93%	99%	100%	100%	100%	100%
				Interrater Agreement					
S	234	509	583	617	627	635	643	652	660
	35%	77%	88%	93%	95%	96%	97%	99%	100%
D	164	399	528	602	632	651	655	660	660
	25%	60%	80%	91%	96%	99%	99%	100%	100%
N	316	582	644	655	660	660	660	660	660
	48%	88%	98%	99%	100%	100%	100%	100%	100%

Interrater agreement (within ±1.0) was 77% for the *S*, 60% for the *D*, and 88% for the *N* group of speakers. Overall, interrater agreement was 75%—identical to interrater agreement (75%) for rating speech naturalness reported by Martin et al.—but it is noteworthy that agreement was lowest for the *D* group.

Inspection of the listeners' judgments shows that there was a bimodal trend across the listeners' ratings: 3 of 12 judges tended to judge that some speakers did not have a very strong dialect, whereas all other judges consistently rated the same speakers as having a very strong dialect. When those 3 judges' ratings were omitted, interrater agreement rose to 77.2% for the *D* group—virtually identical to the level achieved for the *S* group.

Speech Rate

Average speech rate in syllables per minute (SPM) was 177 SPM (range: 72 to 236 SPM) for *S*, 214 SPM (range: 180 to 244 SPM) for *D*, and 254 SPM (range: 211 to 310 SPM) for *N* (see Appendix for individual scores). A one-way analysis of variance (ANOVA, Winer, Brown, & Michels, 1991) identified a significant difference among speaker groups ($F = 83.79$, $df = 2,27$, $p < .01$). Newman-Keuls (Winer et al., 1991) post hoc comparisons revealed that the *S*'s mean SPM was significantly slower ($p < .01$) than the other two speaker groups, and the *D*'s mean SPM was significantly slower than the *N* group mean ($p < .01$).

Normal Disfluencies

Mean percentage of disfluent syllables (%disf) was 18.8 %disf (range: 4.7 to 41.7 %disf) for the *S*, 4.4 %disf (range: 1.2 to 7.8 %disf) for the *D*, and 3.1 %disf (range: 0.7 to 5.5 %disf) for the *N* (see Appendix). A one-way ANOVA revealed a significant difference between means ($F = 17.36$, $df = 2,27$, $p < .01$). A Newman-Keuls test showed that the *S* group mean was significantly larger than the other two speaker group means. The means for the two groups of nonstuttering speakers (*D* and *N*), however, were not significantly different.

Correlations Among Speech Measures and Speech Naturalness Ratings

Pearson correlation coefficients were calculated between the speech behavior measures and speech naturalness ratings. The results are summarized in Table 5.

Significant negative correlations were found between naturalness ratings and speech rate for *N* ($r = -.86$, $p < .01$) and *S* ($r = -.60$, $p < .05$), but were nonsignificant for *D* ($r = -.59$). Speech naturalness ratings were correlated with dialect ratings for *N* ($r = .91$, $p < .01$), but nonsignificant for the other two groups. There were significant positive correlations ($p < .01$) between naturalness ratings and %disf (syllables disfluent) for *S* ($r = .83$) and *D* ($r = .83$), as well as between naturalness ratings and %SS (syllables stuttered) for *S* ($r = .81$).

Table 5. Summary of correlations among speech naturalness ratings, dialect ratings, speech rate (SPM), percentage of syllables stuttered (%SS), and percentage of disfluent syllables (%disf) for the three speaker groups: Nondialect Speakers Who Stutter (*S*), Dialect Speakers Who Do Not Stutter (*D*), and Nondialect Speakers Who Do Not Stutter (*N*).

Group	Variable	SPM	Dialect	%disf	%SS
S	Naturalness	−.60*	.15	.83**	.81**
	SPM		.47	−.77**	−.81**
	Dialect			−.18	.16
	%disf				.89**
DS	Naturalness	−.58*	.28	.83**	
	SPM		.00	−.68*	
	Dialect			.46	
N	Naturalness	−.86*	.91**	.22	
	SPM		−.08	−.08	
	Dialect			.34	

*$p < .05$; **$p < .01$

Significant correlations were found among the speech measures. Speech rates were not significantly correlated with dialect ratings for any of the speaker groups, although they were negatively correlated with %disf scores for *S* ($r = -.77$, $p < .01$) and *D* ($r = -.68$, $p < .05$), but not for group *N* ($r = -.09$). Predictably, speech rate was negatively correlated with the *S*'s %SS scores ($r = -.82$, $p < .01$). Dialect ratings were not significantly correlated with %SS for *S* or with %disf for any of the speaker groups.

Dialect Speakers: Additional Data Analyses

The *D* group consisted of speakers with various dialects. These speakers were divided into two subgroups in terms of whether English was their first language ($n = 4$) or their second language ($n = 6$). Comparisons were made between the subgroups' dialect ratings, speech naturalness ratings, and %disf scores.

A Mann-Whitney U Test (Siegel, 1956) was used to examine differences between the ranking of scores shown in Table 6 for the two subgroups. No significant difference between subgroups was found for dialect ratings, but there were significant differences between their speech naturalness ratings ($U = 0$, $p < .01$) and their %disf scores ($U = 0$, $p < .01$).

Table 6. Individual scores for dialect ratings, speech naturalness ratings, and percent of disfluencies (%disf) for Dialect Speakers (*D*). Speakers are divided into two groups: (1) English as first language and (2) English as second language. Place of dialect origin is included.

Language	Place of origin	Dialect rating	Speech naturalness rating	%disf
English as first language	Australia	7.2	1.9	3.8
	Britain	6.8	1.6	2.0
	Canada	6.7	3.0	1.2
	New Zealand	7.5	2.1	2.2
M		7.1	2.2	2.3
English as second language	China	7.9	4.0	4.4
	China	8.3	5.3	7.8
	China	7.7	4.0	5.8
	Holland	5.0	4.0	4.1
	Vietnam	8.3	4.5	6.8
	Vietnam	6.8	3.9	6.0
M		7.3	4.3	5.8

Discussion

The findings from this study are consistent with past research showing that listeners' speech naturalness ratings can distinguish between persons who stutter and persons who do not stutter (see Schiavetti & Metz, 1997). The present study also demonstrated that listeners' naturalness ratings distinguished between nonstuttering speakers with and without a speech dialect. In general, persons who stutter were judged as sounding the most unnatural and nondialect speakers who do not stutter as sounding the most natural, with the ratings of the dialect speakers who do not stutter located in between. The dialect speakers were judged as sounding significantly less natural than the nondialect speakers who do not stutter, but significantly more natural than the nondialect speakers who stutter. Hence, the findings suggest that listeners judge a speaker with a different dialect from theirs as sounding unnatural, though not as unnatural as a nondialect speaker who stutters.

In considering these findings, it is important to note that the speech naturalness ratings for the speakers with a dialect were *not* significantly correlated with their dialect ratings. However, it was found that the subgroup of speakers in the dialect group who had learned English as a first language were judged as sounding natural (2.2) even though they received ratings of strong dialect (7.1). Whereas, the speakers who had learned English as a second language were judged as sounding relatively unnatural (4.3), yet they received comparable dialect ratings (7.3). Therefore, it does appear that the type rather than the strength of a dialect will influence speech naturalness ratings. This is also consistent with the finding that speech naturalness ratings for the nondialect speakers who do not stutter were highly correlated with their dialect ratings. In other words, highly natural sounding speech was influenced by the type of dialect—General American English—or, more specifically, the absence of dialect as defined in this study.

Another factor that may have influenced listeners' speech naturalness ratings was the level of speech fluency. The most telling evidence was the significant difference found between the frequency of normal disfluencies in the speech of the two subgroups of dialect speakers. The dialect speakers who learned English as a second language exhibited more normal disfluencies and were judged more unnatural sounding than the dialect speakers who learned English as a first language. There are two possible reasons for this. First, this could mean that disfluency rather than type of dialect influenced listener's naturalness ratings. Second, it is possible that both speech fluency and degree of dialect had an impact on the speech naturalness ratings. It has been repeatedly demonstrated that listeners combine various linguistic and even descriptive information in constructing their judgments of various types of speaker attributes and status. As demonstrated by the work of Bradac and Wisegarver (1984), Brown, Giles, and Thackerar (1985), and others (e.g., Ludwig, 1982; Munro, 1993; Ryan & Bulik, 1982), there appears to be a combinatorial rather than a singular system at work when determinants of perceived status and listener evaluation are processed. This could mean that disfluency as well as type of dialect influences listeners' naturalness ratings.[4]

Less confidence should be ascribed to the findings based on disfluency scores because of the relatively lower intra- and interjudge reliability of disfluency judgments (see *Results*). Nevertheless the abovementioned conclusion is fortified by other evidence. For instance, the four dialect participants, who had English as a first language as well as low disfluency counts, were assigned naturalness ratings (2.2) similar to those given to nondialect normal speakers (1.72), who also exhibited low disfluency counts. This occurred even though these four speakers were given dialect ratings that were almost identical to those given to speakers for whom English was a second language. In addition, an examination of the two disfluency judges' levels of intrarater agreement shows that the more consistent (reliable) judge (81% vs. 72%) completely distinguished between the two subgroups, whereas the less reliable judge did

[4]One additional possibility is that word-finding problems among the *D* group speakers may have influenced the naturalness ratings that they received. In all likelihood this feature would be more common in this group. This variable merits investigation in future research in this area.

not. And finally, a significant correlation was found between the frequency of normal disfluencies and speech naturalness ratings for the persons who stutter (also between their speech naturalness ratings and frequency of stuttering). The finding was consistent with those for the dialect group. Cumulatively, this evidence would suggest that level of speech fluency did influence listeners' naturalness ratings.

Speech rate may have also influenced the listeners' speech naturalness ratings. Significant correlations were found between naturalness ratings and speech rate for the nondialect speakers who stutter and do not stutter. The group of speakers who stutter had the slowest rate of speech and the most unnatural sounding ratings, whereas the nondialect speakers who do not stutter had the fastest rate of speech and the most natural sounding ratings—a pattern consistent with findings reported by Martin et al. (1984). However, for the present study, the correlation between speech rate and speech naturalness ratings was not significant for the dialect group. In other words, speech rate may influence speech naturalness ratings only when dialect is not prominent.

Another important finding is that listeners were able to rate speaker dialect using the 9-point dialect scale with satisfactory levels of intrarater agreement. The raters were able to distinguish clearly between speaker samples that had a relatively strong dialect and those that did not. Moreover, their ratings did not appear to be influenced by stuttering, normal disfluencies, or speech rate. This suggests that strength of dialect could be a scaleable and independent dimension that can be rated reliably. This is not surprising given the numbers of rating scales and evaluative instruments that have been used successfully in applied linguistics for this same task (e.g., Fayer & Krasinski, 1987; Koster & Koet, 1993; Ludwig, 1982; Pittam, 1987; Ryan & Giles, 1982). Munro and Derwing (1994) found that a very similar kind of 9-point scale was effective for eliciting judgments of nonnative speech. Indeed, a comparison with levels of rater agreement reported by Martin et al. (1984) demonstrates that overall the dialect rating scale can be used at levels identical to those originally reported on the 9-point speech naturalness scale. Therefore, this rating scale may have potential utility in future research that explores dialect and other speech behavior or language-related variables. That promise is only partially tempered by the relatively low interrater agreement for the dialect group. The fact that judges show high consistency in their individual dialect ratings means that the scale might be very useful when assessing the effect of training designed to alter dialect. The scale's scores might also be employed to alter dialect directly, perhaps in the same way that speech naturalness scores have been used in stuttering treatment (Ingham & Onslow, 1985).

The reliability of speech naturalness ratings in the present study adds to the growing pattern of results obtained in studies on this scale. In the present study the overall level of intrarater agreement (85%) appeared to be satisfactory and comparable to levels reported in past research (e.g., Martin & Haroldson, 1992; Martin et al., 1984). However, listeners were more consistent in their naturalness ratings for the nondialect (98%) and dialect (80%) speakers who do not stutter than they were when rating the speech of nondialect speakers who stutter (76%). A similar low level of intrarater agreement for speech samples of persons who stutter (72.3%) was also reported by Onslow et al. (1992).

Equally consistent with previous studies was the relatively lower overall interrater agreement (65%) for speech naturalness ratings. The listeners notably disagreed with each other on their ratings for the *D* (56%) and *S* (59%) groups, yet they showed satisfactory levels of agreement when they rated *N* (80%) speakers. In this respect the present study failed to replicate the relatively tolerable levels of interrater agreement (at least 75%) reported in previous studies (e.g., Martin & Haroldson, 1992; Martin et al., 1984; Schiavetti et al., 1994). Rather, the present findings are consistent with the interrater agreement levels of 64% and 62% reported, respectively, by Kalinowski et al. (1994) and Onslow, Adams, and Ingham (1992). Taken together, though, these findings underscore the need for clinicians to be trained to a certain level of reliability before using the speech naturalness scale, especially in treatment settings. There is now general agreement that improved intra- and interrater agreement in the use of the scale is most likely to emerge from the development of training exemplars of naturalness scores or anchor stimuli (Gerratt, Kreiman, Antonanzas-Barroso, & Berke, 1993; Schiavetti & Metz, 1997)—a feature that seems to be urgently required for many commonly employed measures in speech pathology (Cordes, 1994; Kent, 1996). The present findings suggest that dialect may be an appropriate factor to be considered when preparing such stimuli.

Finally, it is especially interesting that some speech naturalness judges in the present study (4/30) when making their judgments seemingly ignored the dialects of the speakers who learned English as a second language. They reliably rated these speakers as highly natural sounding (3 or less) despite their relatively different sounding speech; yet, they also rated stuttered speech as unnatural sounding. In other words, some judges are able to judge speakers with a very strong dialect as very natural sounding, but they are then able to judge some speakers who stutter as not very natural sounding. This finding does suggest that judges do have the capacity to ignore dialects that differ from their own when judging the naturalness of clients with disordered

speech patterns. At the same time, the overall findings suggest that in a clinical setting a speaker with a dialect differing from the clinician's dialect may not be judged as producing natural sounding speech even though the speaker is neither stuttering nor displaying a treatment-induced speech pattern. This finding may have important implications for evaluating the outcome of stuttering treatment.

In summary, the present study did find that judges do judge the speech naturalness of speakers with a dialect as different from that of persons who stutter and normal speakers who do not evidence a prominent dialect. Strong correlations were found between speech naturalness ratings and speech rate, speech fluency and dialect, but they differed across speaker groups. Intrarater agreement for speech naturalness ratings given to speakers with a dialect was satisfactory. It was also satisfactory for ratings of dialect, which suggest that the dialect scale employed in this study may have promise for research in this area. However, the level of interrater agreement for speech naturalness rating was generally poor for all speakers; this highlights the urgent need for systems designed to train speech naturalness ratings.

Acknowledgments

Authorship of this paper is equal. The authors wish to thank Jack Damico, PhD, for his invaluable suggestions and contributions. Preparation of this paper was supported (in part) by research grant number 5 R01 DC 00060-05 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529–555.

Battle, D. E. (1993). Introduction. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. xv–xxiv). Boston, MA.: Andover Medical Publishers.

Bloodstein, O. (1995). *The handbook of stuttering* (5th ed.). San Diego, CA: Singular Press.

Bradac, J. J., & Wisegarver, R. (1984). Ascribed status, lexical diversity, and accent: Determinants of perceived status, solidarity, and control of speech style. *Journal of Language and Social Psychology, 3,* 239–255.

Brennan, E. M., & Brennan, J. S. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research, 10*, 487–501.

Brown, B., Giles, H., & Thackerar, J. (1985). Speaker evaluations as a function of speech rate, accent, and context. *Language and Communication, 5*, 207–220.

Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language and Communication, 14,* 211–236.

Cooper, E. B., & Cooper, C. S. (1993). Fluency disorders. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. 189–211). Boston, MA.: Andover Medical Publishers.

Cordes, A. K. (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research, 37,* 264–279.

Crowther, C., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America, 92,* 711–722.

Edwards, J. R. (1989). *Language and disadvantage.* London: Cole and Whurr.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37,* 313–326.

Finn, P., & Ingham, R. J. (1994). Stutterers' self-ratings of how natural speech sounds and feels. *Journal of Speech and Hearing Research, 37,* 326–340.

Flege, J. E. (1987). The production of new and similar phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15,* 47–65.

Flege, J. E., & Bohn, O. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition, 11*, 35–62.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76,* 708–721.

Gallois, C., & Callan, V. J. (1989). Attitudes to spoken Australian English: Judgments of ingroup and ethnic outgroup speakers. *Australian Journal of Linguistics, 9,* 149–160.

Gardner, R. C. (1985). *Social psychology and second language learning.* London: Arnold.

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research, 36,* 14–20.

Giles, H., Hewstone, M., Ryan, E. B., & Johnson, P. (1987). Research in language attitudes. In U. Ammon, N., Dittmar, & K. J. Mattheier (Eds.), *Sociolinguistics: Vol. 1. An interdisciplinary handbook of the science of language* (pp. 585–597). Berlin: DeGruyter.

Giles, H., & Sassoon, C. (1983). The effect of speaker's accent, social class background and message style on British listeners' social judgments. *Language and Communication, 3,* 305–313.

Giles, H., Williams, A., Mackie, D. M., & Rosselli, F. (1995). Reactions to Anglo- and Hispanic-American-accented speakers: Affect, identity, persuasion, and the English-only controversy. *Language and Communication, 15,* 107–120.

Gove, P. B., Ed. (1986). *Webster's third new international dictionary of the English language.* Springfield, MA.: Merriam-Webster.

Gow, M., & Ingham, R. J. (1992). Modifying electroglottograph-identified intervals of phonation: The effect on stuttering. *Journal of Speech and Hearing Research, 35,* 495–511.

Ingham, J. M., & Ingham, R. J. (1986). *Stuttering measurement training program.* Santa Barbara: Department of Speech and Hearing Sciences, University of California, Santa Barbara, CA.

Ingham, R. J., Gow, M., & Costello, J. M. (1985). Stuttering and speech naturalness: Some additional data. *Journal of Speech and Hearing Disorders, 50,* 217–219.

Ingham, R. J., Martin, R. R., Haroldson, S. K., Onslow, M., & Leney, M. (1985). Modification of listener-judged naturalness in the speech of stutterers. *Journal of Speech and Hearing Research, 28,* 495–504.

Ingham, R. J., & Onslow, M. (1985). Measurement and modification of speech naturalness during stuttering therapy. *Journal of Speech and Hearing Disorders, 50,* 261–281.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics in Language Teaching, 28,* 99–117.

Kalin, R., & Rayko, D. S. (1978). Discrimination in evaluative judgments against foreign-accented job candidates. *Psychological Reports, 43,* 1203–1209.

Kalinowski, J., Noble, S., Armson, J., & Stuart, A. (1994). Pretreatment and posttreatment speech naturalness ratings of adults with mild and severe stuttering. *American Journal of Speech-Language Pathology, 3,* 61–66.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5,* 7–23.

Koster, C. J., & Koet, T. (1993). The evaluation of accent in the English of Dutchmen. *Language Learning, 43,* 69–92.

Leith, W. R. (1986). Treating the stutterer with atypical cultural influences. In K. St. Louis (Ed.), *The atypical stutterer* (pp. 9–34). San Diego, CA: Academic Press.

Lippi-Green, R. (1994). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society, 23,* 163–198.

Ludwig, J. (1982). Native speaker judgments of second learners' efforts at communication: A review. *The Modern Language Journal, 66,* 274–283.

Luhman, R. (1990). Appalachian English stereotypes: Language attitudes in Kentucky. *Language in Society, 19,* 331–348.

Martin, R. R. (1981). Introduction and perspective: Review of published research. In E. Boberg (Ed.), *Maintenance of fluency* (pp. 1–30). New York: Elsevier.

Martin, R. R., & Haroldson, S. K. (1981). Stuttering identification: Standard definition and moment of stuttering. *Journal of Speech and Hearing Research, 46,* 59–63.

Martin, R. R., & Haroldson, S. K. (1992). Stuttering and speech naturalness: Audio and audiovisual judgments. *Journal of Speech and Hearing Research, 35,* 521–528.

Martin, R. R., Haroldson, S. K., & Triden K. A. (1984). Stuttering and speech naturalness. *Journal of Speech and Hearing Disorders, 49,* 53–58.

Metz, D. E., Schiavetti, N., & Sacco, P. R. (1990). Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. *Journal of Speech and Hearing Disorders, 55,* 516–525.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech, 36,* 39–66.

Munro, M. J., & Derwing, T. M. (1994). Evaluation of foreign accent in extemporaneous and read material. *Language Testing, 11,* 253–266.

Nesdale, A., & Rooney, R. (1990). Effect of children's ethnic accents of adults' evaluations and stereotyping. *Australian Journal of Psychology, 42,* 309–319.

Onslow, M., Adams, R., & Ingham, R. J. (1992). Reliability of speech naturalness ratings of stuttered speech during treatment. *Journal of Speech and Hearing Research, 35,* 994–1001.

Onslow, M., Costa, L., Andrews, C., Harrison, E., & Packman, A. (1996). Speech outcomes of a prolonged speech treatment for stuttering. *Journal of Speech and Hearing Research, 39,* 734–749.

Onslow, M., Hays, B., Hutchins, L., & Newman, D. (1992). Speech naturalness and prolonged-speech treatments for stuttering: Further variables and data. *Journal of Speech and Hearing Research, 35,* 274–282.

Onslow, M., & Ingham, R. J. (1987). Speech quality measurement and the management of stuttering. *Journal of Speech and Hearing Disorders, 52,* 2–17.

Pittam, J. (1987). Listeners' evaluations of voice quality in Australian English speakers. *Language and Speech, 30,* 99–113.

Port, R., & Mitleb, F. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics, 11,* 219–229.

Ryan, E. B., & Bulik, C. M. (1982). Evaluations of middle class speakers of standard American and German-accented English. *Journal of Language and Social Psychology, 1,* 51–61.

Ryan, E. B., & Carranza, M. (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican-American accented English. *Journal of Personality and Social Psychology, 31,* 855–863.

Ryan, E. B., & Giles, H. (Eds.) (1982). *Attitudes toward language variation: Social and applied contexts.* London: Arnold.

Schiavetti, N., Martin, R. R., Haroldson, S. K., & Metz, D. E. (1994). Psychophysical analysis of audiovisual judgments of speech naturalness of nonstutterers and stutterers. *Journal of Speech and Hearing Research, 37,* 46–52.

Schiavetti, N., & Metz, D. E. (1997). Stuttering and the measurement of speech naturalness. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed., pp. 398–412). Boston: Allyn and Bacon.

Shames, G. H. (1989). Stuttering: An RFP for a cultural perspective. *Journal of Fluency Disorders, 14*, 67–77.

Siegel, S. S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Simpson, J. A., & Winer, E. S. C. (1989). *The Oxford English dictionary* (2nd ed.). Oxford: Clarendon Press.

Taylor, O. (1973). Teacher's attitudes toward Black and nonstandard English as measured by the language attitudes scale. In R. Shuy & R. Fasold (Eds.), *Language attitudes: Current trends and prospects* (pp. 174–210). Washington, DC: Center for Applied Linguistics.

Taylor, O. (1993). Foreword. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. xii–xiii). Boston, MA.: Andover Medical Publishers.

Van Riper, C. (1982). *The nature of stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Watson, J. B., & Kayser, H. (1994). Assessment of bilingual/bicultural children and adults who stutter. *Seminars in Speech and Language, 15,* 149–163.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.) New York: McGraw-Hill.

Young, M. A. (1994). Evaluating differences between stuttering and nonstuttering speakers: The group difference design. *Journal of Speech and Hearing Research, 37,* 522–534.

Received May 20, 1996

Accepted August 30, 1996

Contact author: Patrick Finn, PhD, Department of Communicative Disorders, University of New Mexico, Albuquerque, NM 87131. Email: pfinn@unm.edu

Appendix. Individual scores for each speaker, plus average syllables per minute (SPM) and percentage of disfluencies (%disf) for each speaker group: Nondialect Speakers Who Stutter (S), Dialect Speakers Who Do Not Stutter (D), and Nondialect Speakers Who Do Not Stutter (N). Percent syllables stuttered (%SS) is listed only for speakers who stutter.

S *n* = 10			*D* *n* = 10		*N* *n* = 10	
SPM	**%disf**	**%SS**	**SPM**	**%disf**	**SPM**	**%disf**
170	26.5	14.7	212	3.8	211	3.8
158	16.5	12.0	183	4.4	241	2.1
171	12.8	7.6	184	6.0	245	3.7
193	17.6	7.9	222	6.8	250	1.6
174	30.5	9.9	180	7.8	310	3.2
167	8.5	3.6	241	2.0	298	0.7
230	4.7	2.6	226	5.8	254	2.0
72	41.7	23.6	219	4.1	242	5.0
236	12.3	8.1	226	2.2	273	3.2
195	17.0	5.0	244	1.2	218	5.5
177	18.8	9.5	214	4.4	254	3.1