

RATINGS OF SPEECH NATURALNESS OF CHILDREN AGES 8–16 YEARS¹

SUZANNE COUGHLIN-WOODS AND MARK E. LEHMAN

Central Michigan University

PAUL A. COOKE

Michigan State University

Summary.—The focus of this cross-sectional study was the investigation of Speech Naturalness (speech that sounds normal or natural to the listener) of 60 normal speaking children and adolescents between the ages of 8 and 16 years. 26 naive adult listeners rated the naturalness of videotaped and computer-presented speech samples, using a 9-point-Likert rating scale (1: highly natural sounding speech and 9: highly unnatural sounding speech). The children and adolescents who participated as speakers were distributed among 5 age groups (8, 10, 12, 14, and 16 yr.) with 6 boys and 6 girls in each group. Each child demonstrated normal articulation, language, voice, and speech fluency skills. Age and sex comparisons indicated boys' and girls' speech was rated comparably; however, 8-yr.-olds' speech was rated as significantly less natural than those of 12-, 14-, and 16-yr.-olds. Preliminary ratings of Speech Naturalness for normal speaking children were presented. Suggestions for the clinical application of the findings as a target criterion in treatment programs with communicatively impaired children were suggested. Replication with a larger and more representative sample is in order.

The term "Speech Naturalness" has appeared in the professional literature for more than fifty years. Speech Naturalness is a term defined from the listener's perspective as speech that "sounds normal or natural" and allows the listener's attention to focus on the message rather than the speech pattern (Parrish, 1951; Nichols, 1966; Martin, Haroldson, & Triden, 1984; Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Onslow, Hayes, Hutchins, & Newman, 1992). Speech-language pathologists have been primarily concerned with ratings of Speech Naturalness of adults who stutter before and after treatment because research has suggested that the stutter-free speech of posttreatment stutterers' remains perceptually distinguishable from the speech of normally fluent counterparts (Love & Jeffress, 1971; Runyan & Adams, 1979; Finn, 1997; Dayalu & Kalinowski, 2002). In fact, Martin, *et al.* (1984), using a 1–9 interval scale (1 = highly natural to 9 = highly unnatural), found mean values of Speech Naturalness for adult nonstutterers ranged

¹Address correspondence to Suzanne Coughlin-Woods, Ph.D., Communication Disorders, Central Michigan University, 2171 Health Professions Building, Mt. Pleasant, MI 48859 or e-mail (Cough1ss@cmich.edu).

from 2.1 to 3.6 and those of the fluent speech of posttherapy stutterers ranged from 4.26 to 5.92. This interval scale noted by Martin, *et al.* was constructed through psychological comparison by magnitude estimation as an appropriate measure of rating speech naturalness (Metz, Schiavetti, & Sacco, 1990). Studies of Speech Naturalness with adult stutterers (using the same scale) also showed faster, fluent speech is rated as more natural than slower speech or speech containing disfluencies (Martin, *et al.*, 1984; Ingham, *et al.*, 1985; Ingham & Onslow, 1985).

In contrast to ratings available of naturalness of adults' speech, limited data have been published about children, normal speaking or otherwise. Ingham and Onslow (1985) concluded a study aimed at the measurement and modification of speech with five adolescents who stuttered (one age 10 years, two age 13 years, and two age 14 years), who were enrolled in a fluency-shaping program reliant on procedures to control speech rate. Ingham and Onslow (1985) found rated Speech Naturalness valuable in application for clinical treatment because subjects improved their ratings when given only numerical feedback of ratings. Ingham, Sato, Finn, and Belknap (2001), when replicating Ingham and Onslow's 1985 study, supported clients' knowledge of their Speech Naturalness ratings as a tool in treatment of the rhythm of speech. Ingham and Riley (1998) suggested that such measurements be included in treatment for young children who stutter, underscoring its importance as a measure of the treatment's efficacy.

Although there are limited data on Speech Naturalness of normal speaking children, some differences in the development of normal speech rate with typical disfluencies, variables known to affect naturalness ratings in adults' speech, have been noted as children age. Dawson (1929) and later Kowal, O'Connell, and Sabin (1975), concluded that speech rate varies by age and sex throughout development. Dawson, while studying 200 children in 12 grades, concluded that speech rate increases predominately in Grades 1 to 3, followed by smaller, yet steady, increases as children approach Grade 12. Dawson also stated that development of speech rate varied by sex, noting that girls spoke faster than boys until approximately age 12. Then, between ages 12 and 19, minor and inconsistent changes in speech rate in both sexes were noted. Kowal, *et al.* (1975) found that the frequency of normal nonfluency, duration of unfilled pauses, and length of utterance significantly increased from Kindergarten through Grade 4 in comparison with later grades. Starkweather (1985) concurred that, normally throughout childhood, rate and fluency patterns differ among age groups and between boys and girls. Girls tend to speak at faster rates than boys until age 12 when a change occurs and boys' speech rate becomes more rapid than girls. Between ages 12 and 19, rate of speech between girls and boys varies. Speech rate and number of disfluencies significantly affect listeners' ratings of Speech Natur-

alness with adults (Martin, *et al.*, 1984; Ingham, *et al.*, 1985). Given these findings, it is reasonable to expect that children's Speech Naturalness ratings would be variable as they age.

As suggested by Ingham and Onslow (1985) and Ingham and Riley (1998), knowing the Speech Naturalness ratings of normally fluent children would assist speech and language clinicians by providing a target criterion for children with all types of communication impairments in their attempts to promote the most natural sounding speech possible (Kalinowski, Noble, Armson, & Stuart, 1994; Onslow, Costa, Andrews, Harrison, & Packman, 1996). To date, the target value for Speech Naturalness employed when rating of adolescents' speech has been the mean naturalness value assigned to the speech of normally speaking adults (2.2 on 1 [highly natural] to 9 [highly unnatural], Ingham & Riley, 1998).

The purpose of this study was to identify and compare ratings of Speech Naturalness for normally speaking children and adolescents. The particular focus was to establish preliminary norms or target criteria of Speech Naturalness ratings of normally speaking children between the ages of 8 and 16. It is hypothesized that (1) normally speaking children show differences in Speech Naturalness ratings and these vary across ages. Specifically, it was expected that the speech of older children would be rated more natural than the speech of younger children, and (2) differences in Speech Naturalness ratings for normally speaking children also vary across sex.

METHOD

Participants

Sixty normally speaking children and adolescents, 30 boys and 30 girls, participated in this study. Twelve speakers at each ages 8, 10, 12, 14, and 16 were chosen (six boys and six girls), and each was within 4 months from the mid-year of a birthday. For example, for inclusion in the group of 8-yr.-olds, a participant's age was between 8 yr., 2 mo. and 8 yr., 10 mo. Sixty speakers were assessed by the researchers to be a sufficient number to represent the range of variability typically found in speech (Johnson, 1987; Curlee, 1993). After providing a 50-utterance speech and language sample, each speaker was judged by two certified speech-language pathologists as having normal articulation, language, voice, and fluency skills.

Speech Naturalness has been judged using normal and disordered speakers in listening and rating tasks (Martin, *et al.*, 1984; Ingham & Onslow, 1985; Martin & Haroldson, 1992). Although focusing on Speech Naturalness ratings of normally speaking children, the current study also included disordered speech samples so that listeners' responses to the rating tasks would correspond to the normal versus abnormal speaker paradigm usually used in studies of naturalness. In contrast, if the design of this study did not include

disordered samples, the perceptual anchors and comparisons could be altered, then it would be possible that normal speakers could be rated more variably. That would challenge the relation of these data to previously published results. Therefore, approximating the national statistic that as many as 1 in every 10 persons has a communication impairment (Van Riper & Emerick, 1990), a group of 10 children between the ages of 8 and 16 with varying types of communication disorders represented this portion of the population. Each child in the impaired communication group was diagnosed as having a moderate to severe impairment by two certified speech-language pathologists not otherwise involved in this study. The communication disorder group was comprised of seven boys and three girls with the following communication disorders: two 8-yr.-old boys having severe articulation disorders, two 10-yr.-old boys with voice disorders; one 12-yr.-old girl, and one 12-yr.-old boy with language disorders, two 14-yr.-old girls with hearing impairments, and two 16-yr.-old boys who stuttered. In total, 70 speech samples were included. Sixty were normally speaking children, the focus of this study, and 10 were children with moderate to severe communication disorders.

Speakers were videotaped in color, so that their upper bodies and faces were clearly visible and were instructed to tell the examiners what "kids like to spend their time doing when they are not in school." Six written topic suggestions were provided to assist the speaker, e.g., "Tell me about your family pet." Each topic suggestion was read by the examiner from 5- by 8-in. index cards, which were then placed within the speaker's view. A minimum of 3 min. of monologue or conversation was elicited from each speaker. Speakers were allowed to select more than one topic, if needed, to facilitate 3 min. of recorded speech.

All videotaped samples were reviewed, and a 30-sec. continuous segment of speech was chosen from each sample to serve as a stimulus in this study. Fourteen speech samples, 20% of the 70 speech segments to be rated, were chosen to serve as duplicate presentations for intrarater reliability. Of the 14 samples, 12 were from the normally speaking group and 2 were from the communication disorder group. Consequently, a total of 84 30-sec. speech samples (70 original + 14 duplicate presentations) were prepared to be rated by each listener.

All speech samples were ordered using a quasi-random distribution. The 60 normative samples were randomly placed into five groups of 12 each. Two speech samples from the 10 communication disorder children were randomly selected and included in each of the five subgroups of normally speaking children. This gave a total of 14 speech samples in each subgroup. The 14 speech samples in each of the five subgroups were randomly ordered. Finally, the order of the five subgroups was randomized to complete

the ordering of stimuli to be rated by listeners. For purposes of intrarater reliability, the 14 speech samples chosen to serve as duplicate presentations were randomly ordered in a subgroup and presented, following the 70 randomly ordered speech stimuli in a manner similar to Martin and Haroldson's procedure (1992). Consequently, a total of 84 videotaped samples were presented each listener, with the first 70 being unduplicated (from the 60 normal and 10 communication disorder speakers), and the final 14 segments representing the samples chosen as duplicate presentations for estimating intrarater reliability analysis.

A CD-ROM of all speech samples was made by professional videographers for raters to view. Each 30-sec. video sample was transferred to digital video using *MEDIA 100* Software. Each digitized segment was edited and stored as a computerized file using *QUICKTIME* and transferred to an audiovisual authoring station for scripting and navigation, using the software program *DIRECTOR* prior to being transferred to CD-ROM using the software program *TOAST*.

Listeners and Listening Task

A number of researchers have suggested that unsophisticated listeners, who have no specialized knowledge of speech-language pathology and who demonstrate high reliability, would best represent the judgments of the general listening public and that their evaluations of naturalness should be the primary concern of investigating researchers and speech-language pathologists (Runyan & Adams, 1979; Onslow, *et al.*, 1992; Curlee, 1993). Therefore, 39 native English-speaking adults without specialized knowledge of communication disorders volunteered as raters in response to local advertisements and were paid \$10.00 for participation. The listener group of 19 men and 20 women were between ages of 18 and 56 ($M=25.4$; $SD=7.1$). Each listener passed a puretone hearing screening at 20 dB HL (bilaterally) in a sound-treated booth for the frequencies of 500, 1000, 2000, 4000, and 8000 Hz and were judged by a certified speech-language pathologist as having normal speech.

For the listening task, each listener was seated individually in a 10-ft. by 10-ft. quiet room and rated each speech sample provided using the 9-point interval scale previously used by many researchers (Martin, *et al.*, 1984; Ingham & Onslow, 1985; Martin & Haroldson, 1992). No other verbal descriptions appeared on the scale. The researcher remained in the room with each listener. The CD-ROM containing the speech samples was inserted into a computer and viewed using a 14-in. color monitor. Each listener was given instructions to operate the CD-ROM program to view speech samples at an independent, comfortable pace. Each listener wore a headset to maximize audio reception of the speech samples and reduce ambient noise. Directions,

adapted from those employed by Martin and Haroldson (1992) were read and presented in writing to each listener along with a packet of 84 rating scales.

Three practice ratings were provided listeners before the samples to be rated and were followed by an opportunity to ask procedural questions. If no questions were asked, the researcher instructed the listener to begin rating the speech samples.

Listeners' Reliability

Intralistener and interlistener reliabilities were estimated prior to data analysis. To judge intralistener's rating reliability, responses on replicated samples were analyzed in a procedure similar to that used by Ingham, *et al.* (1985), which defined listeners as reliable with themselves if 75% of their ratings when exposed to the same sample were within one interval of the original rating. For example, if the original presentation was rated 4, then a rerating of 3, 4, or 5, would be considered reliable. Given that 75% of the 14 duplicated reliability samples equaled 10.5, the criterion established for intrarater reliability in this study was increased to 11 of 14 samples (79%). Listeners' initial ratings were compared with their ratings of Speech Naturalness on each of the 14 duplicated samples and listeners whose ratings were within one interval for 11 samples or greater were considered reliable. As a result, 32 of the original 39 listeners met this criterion, and the data from the remaining 7 listeners' ratings (1 woman and 6 men) were eliminated from further analysis. The 32 reliable listeners (19 women and 13 men) ranged in age from 18 to 56 years, with an average age of 22 yr., 7 mo. for the women ($SD=9.6$) and 29 yr., 4 mo. for the men ($SD=10.6$).

Interlistener reliability was estimated for the remaining 32 listeners by calculating the percentage of time that each listener's rating agreed within 1 rating point of the other 31 listeners for all 60 stimuli (Ratcliff, Coughlin, & Lehman, 2002). For a 9-point scale of the type used in this study, the percentage of agreement due to chance alone would be 30.9% (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). Moreover, since in practice many listeners avoid extreme ratings (Kreiman, *et al.*, 1993), viewing the scale as consisting of seven points results in a percentage of agreement due to chance of 38.8%. Therefore, an arbitrary level of 60% was set as the criterion for acceptable interlistener reliability. Six listeners did not reach that and were eliminated from further analyses. This left 26 listeners who were judged to show both acceptable intra- and interlistener reliabilities.

RESULTS

Table 1 shows the mean Speech Naturalness ratings for boys and girls at each of the five ages, as well as combined values by age.

TABLE 1
MEANS, STANDARD DEVIATIONS, AND RANGES OF RATINGS SPECIFIC TO
AGE AND SEX OF 60 NORMALLY SPEAKING CHILDREN

Age and Sex	<i>n</i>	<i>M</i> *	<i>SD</i>	Range
8-yr.-olds				
Girls	6	3.0	1.7	1.5-5.3
Boys	6	2.1	.4	1.6-2.5
Combined	12	2.6	1.3	1.5-5.3
10-yr.-olds				
Girls	6	1.8	.4	1.4-2.5
Boys	6	2.2	.6	1.7-3.3
Combined	12	2.0	.5	1.4-3.3
12-yr.-olds				
Girls	6	1.6	.3	1.3-1.9
Boys	6	1.9	.5	1.3-2.8
Combined	12	1.8	.4	1.3-2.8
14-yr.-olds				
Girls	6	1.6	.3	1.2-2.0
Boys	6	1.8	.3	1.5-2.2
Combined	12	1.7	.3	1.2-2.2
16-yr.-olds				
Girls	6	1.7	.7	1.2-3.1
Boys	6	1.6	.5	1.2-2.5
Combined	12	1.7	.6	1.2-3.1
All Age Groups				
Girls	6	1.9	1.0	1.2-5.3
Boys	6	1.9	.5	1.2-3.3
Combined	12	1.9	.8	1.2-5.3

*1 = Highly natural and 9 = Highly unnatural

Age and Sex Comparisons

Group means of the 60 normally speaking samples were compared to assess whether ratings related to sex or age were significantly different. A two-way analysis of variance indicated statistical significance ($F_{4,50} = 3.71$, $p < .01$) between Speech Naturalness ratings of age groups, and no significant differences in the Speech Naturalness ratings of male and female speakers ($F_{1,50} = .01$, $p > .05$). No interaction between age and sex was noted ($F_{4,50} = 2.01$, $p > .05$).

Tukey test *post hoc* analysis was applied to specify group comparisons accounting for statistical significance noted in speakers' ages. Significant differences were noted between the mean ratings for the following multiple comparisons: 8-yr.-olds and 12-yr.-olds (mean difference = .8, $p < .05$), 8-yr.-olds and 14-yr.-olds (mean difference = .90, $p < .05$), and 8-yr.-olds and 16-yr.-olds (mean difference = .91, $p < .05$). The higher mean Speech Naturalness ratings for the 8-yr.-old girls relative to the other groups may have contributed substantially to these differences. There were no significant differences

in the ratings between 8- and 10-yr.-olds or among ratings of the remaining age groups.

DISCUSSION

Speech Naturalness data from 8-, 10-, 12-, 14-, and 16-yr.-old speakers were compiled to assess whether Speech Naturalness ratings of these children were related to sex or age. Analysis showed that the ratings of boys and girls by adult untrained listeners were statistically comparable, an unexpected result given the previously cited differences in speech rate and fluency development of boys and girls during these ages (Dawson, 1929; Kowal, *et al.*, 1975; Starkweather, 1985). These findings suggest that developmental differences in speech patterns of speech rate, pause, and flow may not affect perceptions of Speech Naturalness or were difficult for untrained or unprompted listeners to detect, particularly in the absence of pretask instructions drawing attention to these parameters. Additional research is needed to address the relationship between Speech Naturalness ratings and developmental speech rate, pause, and fluency. Further research should also investigate whether developmental sex characteristics are more noticeable if listeners are instructed to pay attention to specific items such as speech rate, pause, or flow.

Results indicated that listeners perceived the Speech Naturalness of 8-yr.-olds' mean speech as significantly different from those of 12-, 14-, and 16-yr.-olds. Eight-yr.-old speakers were rated as having the largest range of mean values (2.1 to 3.0) and the largest standard deviation (1.3) of all normal speakers. These data suggest that perception of Speech Naturalness of the 8-yr.-olds was significantly different during this developmental period, and perhaps related to changing speech rate, elevated numbers of disfluencies, and stress pattern variances as suggested by Dawson (1929) and Kowal, *et al.* (1975). Further research correlating these naturalness ratings with speech rate, pause occurrence and duration, and number of typical disfluencies of the samples would add valuable information to these data.

Curlee (1993) suggested that Speech Naturalness ratings from a number of inexperienced raters would serve as the most valuable assessment of treatment efficacy both within as well as outside the clinic. Given Curlee's assertion, ratings from peer-age groups could become a pertinent and valuable feedback tool for children in speech-language therapy programs. In addition to investigating the Speech Naturalness ratings of children by their normally speaking peers, a comparison of naturalness ratings of normal speaking children by communication disorder children would verify how children with communication disorders perceive the speech of children that normally speaking adults and children consider natural.

Speech Naturalness ratings by multiple listeners provide valuable feed-

back to both clients and clinicians in decisions related to therapy transfer tasks, consideration of treatment dismissal, and posttreatment maintenance. Professional literature has indicated a need to establish normative data for Speech Naturalness that encompasses an increased number of normal speakers. The Speech Naturalness data compiled in this study incorporated 312 ratings by age for children ages 8 through 16. These data offer speech-language pathologists preliminary age-referenced Speech Naturalness rating targets that could be useful for children aiming to improve Speech Naturalness, irrespective of communication disorder diagnosis.

REFERENCES

- CURLEE, R. F. (1993) Evaluating treatment efficacy for adults: assessment of stuttering disability. *Journal of Fluency Disorders*, 18, 319-331.
- DAWSON, L. O. (1929) A study of the development of the rate of articulation. *Elementary School Journal*, 29, 610-615.
- DAYALU, V. N., & KALINOWSKI, J. (2002) Pseudofluency in adults who stutter: the illusory outcome of therapy. *Perceptual and Motor Skills*, 94, 87-96.
- FINN, P. (1997) Adults recovered from stuttering with formal treatment: perceptual assessment of speech normalcy. *Journal of Speech, Language, and Hearing Research*, 40, 821-831.
- INGHAM, J. C., & RILEY, G. (1998) Guidelines for documentation of treatment efficacy for young children who stutter. *Journal of Speech, Language, and Hearing Research*, 41, 753-770.
- INGHAM, R. J., MARTIN, R. R., HAROLDSON, S. K., ONSLOW, M., & LENNEY, 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.
- INGHAM, R. J., SATO, W., FINN, P., & BELKNAP, H. (2001) The modification of speech naturalness during rhythmic stimulation treatment of stuttering. *Journal of Speech Language and Hearing Research*, 44, 841-852.
- JOHNSON, L. (1987) Multidimensional scaling of speech naturalness in stutterers. Unpublished doctoral dissertation, Univer. of Minnesota.
- 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 and Language Pathology*, 3, 61-66.
- KOWAL, S., O'CONNELL, D. C., & SABIN, E. F. (1975) Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research*, 4, 195-207.
- KREIMAN, J., GERRATT, B., KEMPSTER, G., ERMAN, A., & BERKE, G. (1993) Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, 36, 21-40.
- LOVE, L. R., & JEFFRESS, L. A. (1971) Identification of brief pauses in the fluent speech of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, 14, 229-240.
- 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. C., 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.
- NICHOLS, A. C. (1966) Audience ratings of the "naturalness" of spoken and written sentences. *Speech Monographs*, 33, 156-159.
- 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., HAYES, 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.
- PARRISH, W. M. (1951) The concept of naturalness. *Quarterly Journal of Speech*, 37, 448-450.
- RATCLIFF, A., COUGHLIN, S., & LEHMAN, M. (2002) Factors influencing ratings of speech naturalness in augmentative and alternative communication. *Augmentative and Alternative Communication*, 18, 11-19.
- RUNYAN, C. M., & ADAMS, M. R. (1979) Unsophisticated judges' perceptual evaluations of the speech of "successfully treated" stutterers. *Journal of Fluency Disorders*, 4, 29-38.
- STARKWEATHER, C. W. (1985) The development of fluency in normal children. In H. H. Gregory (Ed.), *Stuttering therapy: prevention and intervention with children*. (Publication 20) Memphis, TN: Stuttering Foundation of America. Pp. 67-100.
- VAN ROPER, C., & EMERICK, L. (1990) *Speech correction: an introduction to speech pathology and audiology*. (2nd ed.) New York: Macmillan.

Accepted January 11, 2005.