ELSEVIER

Journal of Communication Disorders
35 (2002) 63–82

Journal of Communication Disorders

The effect of temporal manipulation on the perception of disfluencies as normal or stuttering

Ofer Amir[a,*], Ehud Yairi[b]

[a]*Department of Communication Disorders, Tel-Aviv University, Tel-Aviv, Israel*
[b]*Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, IL, USA*

Received 5 April 2001; received in revised form 10 August 2001; accepted 28 August 2001

Abstract

The purpose of this investigation was to study the effect of temporal features within repetition of speech segments on the perception of stuttering. Past research has provided evidence that certain temporal aspects of repetitions produced by people who stutter tend to be shorter than those produced by normally fluent speakers. The effect of these temporal factors on the perception of the disfluency as "stuttering" or "normal" has not yet been studied. Conversational speech of five children who stutter was recorded. Two short utterances, one containing part-word repetition (PWR) and one containing whole-word repetition (WWR), were identified in the speech of each child and then manipulated by the CSL and CSpeech computer softwares. Two selected elements within repetitions, namely the vowel of the repeated unit and the interval between the repeated units (e.g., but-but), were lengthened to simulate normal disfluency. Results indicated that both factors (interval duration and vowel duration) moderately affected listeners' perception. In general, repetitions with short vowel and interval durations were judged as more representative of stuttering, whereas repetitions with longer vowel and interval duration were judged as more representative of normal speech. Learning outcomes: As a result of this activity, the reader will learn about (1) various factors that influence the perception of disfluent segments as stuttering, (2) the special effect of duration of specific elements within repetitions on the perception of disfluency as stuttering, and (3) the possible implications

* Corresponding author. Department of Communication Disorders, The "Chaim Sheba" Medical Center, Tel-Hashomer, 52621, Israel.
E-mail address: oferamir@post.tau.ac.il (O. Amir).

PII: S0021-9924(01)00067-3

of the new information for therapeutic considerations.

Keywords: Stuttering; Interval duration; Vowel duration; Temporal manipulation; Perception

1. Introduction

Research concerning differences between disfluencies that are "normal" and those characterized as "stuttering" has taken on two major tracks: studies that focused on disfluency output and studies that focused on listener perception. The present investigation pertains to perception, an issue that was thrust into prominence after the early exposition of the diagnosogenic theory (Johnson, 1942). With its basic tenet that stuttering is caused by listeners' erroneous judgment of a child's normal disfluencies as being "stuttering," the theory has provided an impetus for many studies of factors that influence such perception.

Studies concerning overall conditions and circumstances that affect listener judgment of a speaker as "a stutterer," or specific disfluent events as "stuttering" or "normal," have shown the impact of experience, background, psychological set, and experimental instructions on listeners' shifting judgment (Cordes, Ingham, Frank, & Ingham, 1992; Curlee, 1981; Tuthill, 1940, 1946; Williams & Kent, 1958). Other perceptual studies have shown that certain types of disfluency, such as syllable repetitions and sound prolongations, are more likely to be judged as "stuttering" than other types (e.g., Boehmler, 1958; Young, 1961) and that the number of disfluencies in a given sample, regardless of type, is also likely to influence stuttering judgment (Curran & Hood, 1977; Hegde & Hartman, 1979a, 1979b). The number of repetition units per disfluent event was also found to influence listeners' perception. Sander (1963) found that it took fewer instances of double-unit repetition (e.g., bu-bu-but) than single-unit repetition (bu-but) to be judged by listeners as stuttering. Curran and Hood (1977) reported similar perceptual findings.

Although perceptual studies have identified basic trends about disfluency typically perceived as stuttering or as normal, inconsistent judgments, overlap, ambiguities, and contradictory findings still exist (Curlee, 1981). For example, although changing instructions caused listeners to judge the same disfluencies as both stuttering and normal (Williams & Kent, 1958), some disfluent events were judged as stuttering more consistently than other disfluent events. Recently, Cordes (2000) reported that some disfluent events classified as "stuttering-like disfluencies" (SLD) were judged as normal. Thus, what makes a specific disfluent event more likely to be perceived as "normal" or "stuttering," and what contributes to more stable judgment of a given disfluent event, is not yet understood.

One reason for the difficulty in achieving better appreciation of features that influence stuttering–normal judgment is the fact that traditional disfluency classification systems have been grounded in linguistic terminology such as "part-word repetition (PWR), word-phrase repetition, sound prolongation," etc. (Johnson et al., 1959). Although these descriptors are certainly useful for many purposes, they are apparently rather general and do not readily lend themselves to analysis of fine features. For example, it would seem reasonable to hypothesize that acoustic features, such as intonation, stress, etc., may help listeners differentiate between disfluencies produced by persons who stutter and fluent speakers and thus contribute to listeners' identification of stuttering (Schwartz, 1995; Van Riper, 1982).

Attention to acoustic correlates of disfluencies regarded as stuttering or normal has been limited. The parameter of duration is of particular interest to the present study. Zebrowski (1991) compared overall duration of disfluencies in the speech of preschoolers who stutter to a control group and found no statistically significant differences. More promising findings, however, were reported for duration of segments within disfluencies. Howell and Vause (1986) analyzed speech samples of eight children who stutter and found that the duration of vowels produced during stuttered PWR was significantly shorter than the duration of vowels produced during the fluent speech of each speaker. They concluded that this shortening of duration led the listeners to believe that the vowel was neutralized to schwa. Howell, Williams, and Vause (1987) also reported that vowel duration within stuttered PWR was shorter than vowels within the fluent speech of the same speakers. Additionally, stuttered vowels were lower in amplitude than fluent vowels. In a third study, Howell and Williams (1992) continued to study this question from a perceptual perspective. Although the results also revealed durational differences between vowels produced in stuttered and fluent repetitions produced by the same speakers, manipulation of vowel duration, by itself, was not sufficient for changing listeners' perception of an intended vowel to a schwa.

Another group of studies pursued a different temporal parameter of duration of subsegments within the disfluencies. Yairi and Hall (1993) reported a trend for the intervals between segments of a whole-word repetition (WWR) (e.g., the interval between "but-but") of young children who stutter to be shorter than those in the repetitions of children who did not stutter. In a follow-up investigation, Throneburg and Yairi (1994) confirmed this trend, reporting significant group differences in the interval duration between repetition units in PWR and WWR. For example, normally fluent children had, on the average, intervals between the consecutive segments in PWR that were approximately three times longer than intervals in PWR produced by stuttering children. That is, children who stutter repeated faster than children who do not stutter did. At present, the effect of the duration of the interval on listener judgment of disfluencies is not known.

In summary, it has been well documented that various conditions, instructions, type of disfluency, the amount of disfluency in the speech sample, and the number of repetition units influence listeners' perception of stuttering. It has been suggested that acoustic properties of disfluencies, such as duration, may differentiate between disfluencies produced by persons who stutter and fluent speakers.

More recently, several investigators reported that stuttered vowels are shorter than fluent vowels in the speech of the same stuttering speakers and that durational features within repetitions produced by children who stutter differ from repetitions produced by normally fluent speakers. To date, however, there are not sufficient data to indicate whether these durational characteristics, evident in the speech output, also contribute to the perception of stuttering in speech of young children. The present study is an initial attempt to investigate this question. Consonant duration within repetition was not reported previously to differentiate stuttered speech from fluent speech. Thus, it was not included as a variable in this study. In addition, the use of natural speech rather than synthetic speech was deem desirable for this study in light of past findings regarding the presence of additional acoustic features that might contribute to listener identification of a disfluency as stuttering. Utilizing synthetic speech could, on one hand, isolate the parameters tested but, on the other hand, might present biased results. Listeners might respond differently to synthetic and natural disfluency.

Both from a theoretical and clinical points of view, additional information regarding the relation between the physical features of early disfluency and listener's reactions to them should help clarify long-standing questions concerning the onset of stuttering. Are there factors beside the type and frequency of disfluency that register in listeners' minds as abnormal or stuttering? Is it possible that some apparently "simple, easy" repetitions contain additional elements that are not conveyed by their linguistic descriptions, making them different from similar disfluent events in the same class? Are some repetitions more "stuttering" than others are? In other words, what else is there that finally convinces parents and others that a child's repetition is a cause for concern? Additionally, data concerning the relation between physical characteristics of disfluent events and their perception may contribute to understanding of stuttering severity, provide clues for what might be important data to obtain in clinical evaluation, and might be useful in new development in the area of automatic recognition of stuttering. Information of this nature should also enhance formulating therapeutic objectives and, finally, help determine when and why children return to being normally fluent speakers.

The present investigation was an exploratory study designed to assess the possibility that certain temporal acoustic features of specific components of disfluent events affect the perception of disfluencies as being normal or stuttering. Inasmuch as information on this parameter is scarce, we were looking for trends that might provide clues for further research concerning the relation between acoustic features and the normal/stuttering distinction. If interval and vowel duration within repetitions affect the perception of stuttering, stuttered vowels should sound more like normal disfluency if they are lengthened. To test this assumption, vowel duration and interval duration taken from speech samples of stuttering children were systematically manipulated. The effect of these manipulations was then evaluated perceptually by listeners. To limit the scope of the study, it was decided to manipulate only two types of disfluency: one-unit PWR and one-unit monosyllabic WWR. Incorporating other types of disfluency into this study would have required examination of multiple temporal factors as well as other acoustic features, thus utilizing different manipulation procedures. PWR and WWR were selected as the target disfluencies in this study because they are more likely to be found in normally fluent speech as well as in the speech of stuttering speakers. In addition, these disfluencies can be judged by listeners as both stuttering and normal disfluency.

2. Method

2.1. Speakers

Five children who stutter were chosen to participate. They were selected from more than 150 children currently enrolled in the longitudinal Stuttering Research Project at the University of Illinois. The criteria for inclusion of children in the study were identical to those used previously in this center and were given in detail by Ambrose and Yairi (1999). In essence, each child was judged by two experienced speech language pathologists and by his parents as having a stuttering problem. In addition, a child had to exhibit three or more SLD (PWR, single-syllable WWR, and disrythmic phonations that mostly consist of sound prolongations).

The five children have stuttered for at least 24 months and were recorded close to the age of 5 years. Their mean age was 60.6 months (S.D. = 3.51) at the time of the recordings. Mean time postonset of stuttering was 30.80 months (S.D. = 8.35). All children were males to reduce variability between speakers.

2.2. Speech sample recording

Ten utterances were ultimately selected: two utterances per speaker, one containing single-syllable WWR and one containing PWR. Section 2.3 will describe the process by which these 10 utterances were chosen. All children were audio- and video-recorded in an IAC soundproof room using a Crown PPC-160 cardioid microphone connected to a Yamaha KM608 preamplifier. Each child sat in a chair with the microphone placed approximately 20 in. from his mouth. The audio signal was directed to a Tascam 122MKII cassette recorder and recorded onto Maxell II S-90 cassette tapes. Nevertheless, because movements of the child slightly varied the exact mouth-to-microphone distance, measurement of intensity for the different segments of the speech samples were not performed.

2.3. Token selection

To locate candidate utterances for this study, a recorded speech sample of approximately 1000 syllables of conversational speech was first collected and analyzed to identify and classify disfluencies. Speech was recorded during verbal interaction with one parent and also included several minutes of interaction with one investigator.

One experimenter (OA) listened to the complete recorded sample of the five selected speakers and digitized each segment that followed all these criteria: (a) contained one incidence of one-unit repetition (PWR or WWR), (b) did not contain any other type of disfluency except for that repetition, (c) included four to eight words, (d) was shorter or equal in duration to 4 s, (e) did not include the conversational partner's voice, (f) was representative of the child's normal voice quality in the complete speech sample, (g) included only minimal background (electrostatic or nonspeech) noise, and (h) visual identification of the different components of the repetition (based on a wide-band spectrogram) was possible.

Identification of disfluency type (PWR versus WWR) and number of repeated units within the repetition was first performed perceptually by an experienced listener who transcribed the speech sample and then independently confirmed by the first experimenter. Following that, an acoustic analysis was performed, in which visual inspection of a wide-band spectrogram accompanied by audio verification was carried out. Only instances in which both the transcriber and the experimenter agreed, and which were also identified as such by inspecting the spectrogram display, were included in the initial part of the study. Sixty-seven utterances were selected from the speech samples of the stuttering children.

To verify that the utterances collected from the stuttering children are indeed recognized as stuttered, all 67 utterances were presented to four listeners highly experienced in identification of disfluency and stuttering. These listeners had a mean of 9.5 years of research and clinical experience with stuttering (range: 7 – 14 years). Each listener listened individually to the speech samples via earphones. The experienced listeners were asked to decide, using a binary forced-choice evaluation form, whether each utterance contained "stuttering" or "normal disfluency." Only utterances on which at least three of the four experienced listeners agreed on the categorization, "stuttering," were included in the study. Additionally, an attempt was made to select utterances in which the disfluency occurred at the beginning of the utterance. Following this procedure, two utterances were selected from the speech sample of each child: one containing a monosyllabic WWR [e.g., (but-but)] and one containing a PWR [e.g., (bu-but)]. This resulted in 10 utterances taken from the stuttering children, regarded as the "original utterances."

2.4. Stimuli description

Table 1 presents the 10 original utterances selected from the speakers' speech samples. Additionally, the loci of the disfluency in the utterance and the number of words, syllables, and phonemes in each utterance are reported.

Table 1
Description of the 10 selected utterances, loci of disfluency, and number of words, syllables, and phonemes in each utterance

Speaker	Utterance	Word	Syllable	Phoneme
E08	[WWR] make-make a dinosaur out of it	6	8	17
	[PWR] A-I (a-ai) like that song	4	4	10
E73	[WWR]I-I just stay in the car	6	6	14
	[PWR]A-I (a-ai) like the blue guy	5	5	11
E23	[WWR]the-the wings are straight	4	4	13
	hit me [PWR]i-in the head with the bowl	8	8	20
E76	[WWR]he-he takes toys from me	5	5	15
	[PWR]A-I (a-ai) hit the target	4	5	12
E10	[WWR]she-she flew in a house	5	5	11
	[PWR]tha-that the good witch gave to her	7	7	18

The demonstration of the PWR of "I" as "A-I" is aimed to represent the fact that the first unit of the repetition consisted of the neutral vowel, schwa, or of the vowel /a/ and not of the diphthong /ai/ (which would be considered a WWR).

The mean number of words per utterance was 5.4 (S.D. = 1.35), mean number of syllables per utterance was 5.7 (S.D. = 1.49), and mean number of phonemes per utterance was 14.1 (S.D. = 3.35).

The duration of the vowel within the first spoken unit of the repetition and the duration of the interval between the two spoken units of the repetition were measured for all utterances. Mean vowel and interval duration for utterances containing PWR and WWR are presented in Table 2.

No significant difference was found between interval duration within WWR and PWR ($P = .091$), although mean interval duration was approximately 60 ms shorter in PWR than in WWR. In addition, no significant difference was found between vowel duration within WWR and PWR ($P = .264$). It should be noted that the durational values presented here fall within the range of values reported by Throneburg and Yairi (1994) for different groups of children.

2.5. Signal processing and segmentation

Each of the 10 utterances was low-pass filtered at 7.5 kHz, digitized at 20,000 samples per second (DT 2821 A/D converter), and stored on a computer disk.

Table 2
Interval and vowel duration in the original utterances produced by the stuttering speakers prior to durational manipulation

Type of disfluency	Interval duration [ms; mean (S.D.)]	Vowel duration [ms; mean (S.D.)]
PWR	33.63 (39.04)	146.09 (27.93)
WWR	96.67 (27.93)	139.73 (27.93)

Signals were again low-pass filtered at 7.5 kHz for playback during analysis. Acoustic analyses were performed from a fast Fourier transformation (FFT)-based spectrogram display and from the acoustic waveform displays, employing the CSpeech computer software (Milenkovic, 1987). Utilizing this procedure, the target disfluency event was located and its components (the vowel within the first spoken unit and the immediate following interval) were identified and measured. Cursors were placed at the initiation and termination of spectral energy that correspond to the following three points: (a) Beginning of the vowel in the first spoken unit of the repetition. This point was defined as the first zero crossing on the enhanced time wave display that corresponded to the beginning of the vowel formants, as displayed in the spectrogram. (b) The end of the vowel in the first spoken unit of the repetition. This point was defined as the last zero crossing on the enhanced time wave display that corresponded to the end of the vowel formants, as displayed in the spectrogram. (c) The beginning of the second spoken unit. This point was defined as the first burst of spectral energy after the interval within the repetition, as displayed on the time wave and the spectrogram. The interval was defined as the duration between points (b) and (c); that is, between the two spoken units of the repetition (for example: bu-but).

Fig. 1 illustrates an example of an enhanced spectrogram of the disfluent segment [I-I] from the utterance [I-I just stay in the car]. The positions of the points (a), (b), and (c) are marked with vertical lines.

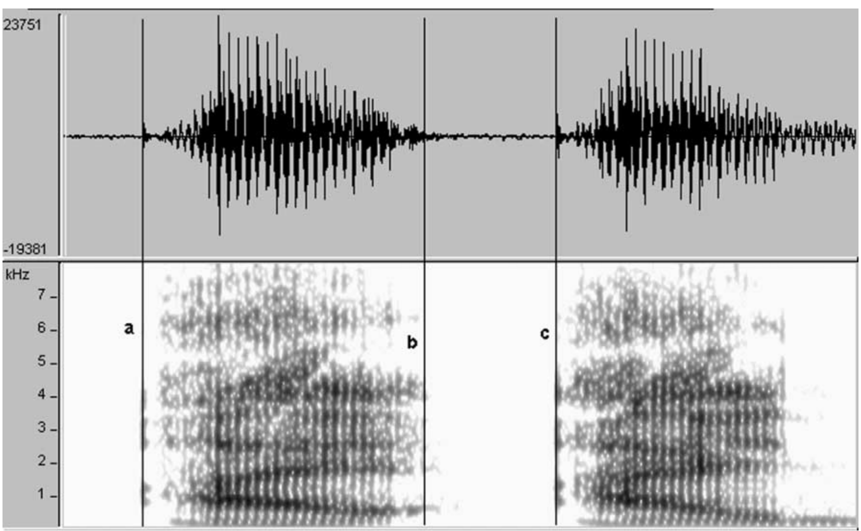

Fig. 1. A spectrographic enhanced display (wide band) combined with a time-intensity waveform display of the segment [I-I] from the utterance "[WWR]I just stay in the car." Location of the cursors for segmentation is marked by vertical lines: (a) beginning of first spoken unit within the repetition, (b) end of first spoken unit within the repetition, and (c) beginning of second spoken unit within the repetition.

2.6. Temporal manipulation procedure

Interval duration and vowel duration in the original utterances were digitally prolonged, using the CSpeech and the CSL commercial programs, to simulate normal disfluency. First, each original interval was prolonged to create three additional conditions of interval duration of approximately 100, 300, and 500 ms. In two utterances, where the original interval duration was longer than 100 ms, only the two longer conditions were created. These intervals were chosen in an attempt to cover both ranges of typical interval duration within repetitions produced by children who stutter as well as by children who do not stutter, based on the results reported by Throneburg and Yairi (1994). The manipulation of each interval was performed by first identifying it between the two spoken units as described above. Then, the central portion of the interval was taken out of its original position (without including format transition to or from the adjacent consonants that might have been present). Following, the isolated segment was duplicated several times as needed. Next, the modified segment was inserted into its original position between the two repetition units. This procedure maintained all background, static, and nonspeech noise throughout the manipulated interval identical to its original characteristics. The connecting points between the original signal and each added portion were enhanced for visual and perceptual examination. A smoothing procedure was used on each connection point, utilizing the CSpeech program (Milenkovic, 1987), to remove any distortions or acoustic artifacts that resulted from the duration manipulation or from the cutting-and-pasting process. The smoothing procedure included, first, visual inspection of the enhanced time wave display. In many utterances, the product of the manipulations was a visible peak in the sound wave, which resulted in an audible click sound. To account for that, a manual alteration of the sound wave, using the "edit" command of the CSpeech program, was performed, establishing a smooth transition between the two segments. Then, each spectrogram was reexamined, visually and perceptually by the experimenter and by a second listener, to verify that any abnormality or acoustic distortion was eliminated.

Using a similar procedure, each original vowel was artificially prolonged to create two additional conditions of vowel duration of approximately 150% and 200% of the original duration. This procedure was designed in light of the results presented by Howell and Vause (1986) that indicated longer vowels in fluent speech and shorter vowels in disfluent speech. Vowel duration manipulation was performed utilizing the CSL program. The steady state of the vowel within the repeated unit was taken out of the signal and manipulated separately.

The "Voice Period Marks" command of the CSL software was used to identify each subsequent voicing impulse in the waveform and place marks at the locations following positive-going-zero crossing that precede the first positive-going amplitude peaks of the voicing impulse. Then, the increase in duration was automatically performed by breaking the voiced signal into its individual pitch separations and operating on the waveform by repeating each voicing cycle. This procedure results in double duration of the selected segment, without affecting its fundamental frequency.

After completion of this procedure, the artificially prolonged segment was returned to its original position within the repeated unit. The 150% and the 200% conditions were selected in view of the technical limitation of the system to create other vowel durations and in an attempt to cover the range presented in the Howell and Vause (1986) study. The connecting points between the original signal and the manipulated vowel were enhanced and examined, and the same smoothing procedure described above was employed here again, removing any distortions or acoustic artifacts that resulted from the duration manipulation. Following interval and vowel manipulation, all combinations of interval duration condition (four conditions: original, 100, 300, and 500 ms) and vowel duration condition (three conditions: original, 150%, and 200%) were created. These procedures resulted in a total of 114 utterances.

It should be noted that although specific values were targeted for both interval and vowel duration, these values were eventually approximated though not exactly matched due to the acoustic features of the connecting points between the modified and the original segment. In utterances where the difference between the two segments was too large to enable a smooth transition, an alternative "connecting point" was selected. These adjustments resulted in slightly different durational values of the modified utterance than the intended fixed values. To account for this, all modified segments were remeasured after the completion of the manipulation, as will be described in the following sections.

All manipulated utterances were evaluated by two listeners to ensure that no acoustic distortion or artifacts were perceivable or could mark these utterances as artificial. In addition, a random sample of 12 utterances, six manipulated ones and six original ones, was selected and presented in a random order to a group of five graduate students in the Speech and Hearing Science Department at the University of Illinois. The five listeners were given a brief description of the listening task and were asked to mark, on a five-point scale, their responses to the question: "Is this sentence presented in its original form or in its computer manipulated form," where 1 represented "I am sure the sentence was presented in its original form" and 5 represented "I am sure the sentence was manipulated." The average score given to the original utterances by the listeners was 2.8 (S.D. = 0.40); the score for the manipulated utterances was 2.7 (S.D. = 0.86). A one-way ANOVA revealed no significant differences between the responses of the listeners to the two groups of utterances ($P = .80$).

In addition to the 114 test utterances, 10 utterances that qualified for the criteria for candidate utterances but were not selected for the study were recorded separately and served as practice items. The purpose of the practice items was to familiarize the listeners with the listening task, with the time constrains, and to enable them to adjust playback loudness level to a comfortable level.

Twenty-four of the test items were selected randomly, and were included in the listening task twice, for the purpose of test–retest reliability evaluation. Also, additional 48 utterances with similar disfluencies were recorded from 12 other children (four utterances from each child). Each utterance was recorded twice, creating a pool of 96 nontest utterances. The purpose of adding these utterances to the test items was to reduce the possible effect of familiarity with the voices of the five speakers and to reduce possible bias to a specific child's voice or a specific utterance. The nontest items were excluded from the final analysis of the results. The grand total number of all utterances was 234 (114 test items + 96 nontest items + 24 reliability items). These utterances were randomized and prepared for presentation to the listeners with a 5-s interval between consecutive utterances.

2.7. Listeners

Forty listeners were selected from a group of 50 young adults (undergraduate students). The selected 40 listeners were native English speakers, were naive about stuttering, had no friends or relatives who stutter, and had no history of speech or hearing problems. Fifteen listeners were males and 25 were females. Their mean age was 19.2 years (S.D. = 1.24; range = 17 – 22).

2.8. Listening task

Each listener was seated in a quiet room and completed the listening task individually. The instructions were presented to the listeners in writing and were also recorded on a tape and played for them as they read along. In essence, each listener was instructed to listen to one utterance at a time and, after the utterance was completed, to circle the corresponding number, according to his/her judgement, on a six-point scale. In this scale, 1 represented normal disfluency and 6 represented stuttering. The intermediate options on the scale were numbered but not labeled. The complete instruction sheet presented to the listeners is presented in Appendix A. Thirty-five listeners used the six-point scale in its full range. That is, they rated different tokens as 1 – 6 on the scale. Three listeners used five points on the six-point scale, and the other two listeners used only four points on the scale. This distribution of the listeners' responses can imply that the scale was a valid psychometric method for this specific listening task. The utterances were played to the listeners on a Denon DRM-700 cassette deck and were presented individually to each listener through a Koss Pro/4AAA set of headphones.

The 10 practice items were presented prior to the beginning of the listening task, and then the test items were presented with the test – retest items, all pooled and randomized. A 2-min break was given to the listeners every approximately 10 min. The complete listening task, including instructions, test items, and breaks between the sets, took approximately 60 min.

After completing the listening task, each listener filled out a questionnaire that included items regarding demographic information, language, and speech background and information regarding previous experience with stuttering. The listeners filled out the questionnaire after the completion of the listening task to reduce possible bias of the questions on the responses to the test items. As described above, based on the listeners' responses to the questionnaire, 10 of the initial group of 50 listeners were removed from the study and were not included in the analysis. These omitted listeners had either previous experience with stuttering or they did not meet one of the criteria listed above.

2.9. Reliability

Interlistener reliability was estimated using the Interclass Correlation (Shrout & Fleiss, 1979). As described previously, 24 utterances were selected randomly and presented again to the listeners for listener reliability evaluation. A two-way ANOVA in which "Listeners" was treated as a random variable, "Test–Retest" as a fixed variable, and Listener-by-Retest interaction was assumed to be zero and used as the estimate of error (Winer, 1962, pp. 124–132) was used to establish reliability of listener's judgement. The resultant reliability value was .94.

2.10. Data treatment

To examine differences in listeners' judgment of the different vowel and interval duration conditions, a mean value was calculated for each interval duration condition and for each vowel duration condition for utterances containing PWR and WWR separately. The mean score was derived from the ratings of the 40 listeners for each test item on the six-point scale.

Initial observation of the data revealed variability in the duration of the segments originally produced by the children. In addition, as described above, although an attempt was made to manipulate interval and vowel duration to several fixed levels, some variability in final segment duration was evident due to technical limitation of the system used. To account for this variability, all test items (original and manipulated) were measured again and then regrouped according to their absolute interval and vowel duration. Utterances were divided into four categories of absolute interval duration: (1) <50, (2) 70–250, (3) 270–450, and (4) >470 ms. These levels were constructed in light of the natural distribution of the data. The first interval duration category (<50 ms) was constructed to include a large group of utterances, in which interval duration was shorter than 50 ms. A gap of 20 ms was allowed between the different category to prevent overlap between them. The second and third categories (70–250 and 270–450 ms) were constructed due to the fact that many utterances (manipulated, as well as original) included interval durations that fall within these ranges. Thus, it was decided to create these two 180-ms time windows, with a 20-ms gap between them. The >470-ms category was constructed to include the remaining utterances that fall within this durational range.

The utterances were also divided into four categories, according to their absolute vowel duration, on an a priori basis. Each category was set as a 60-ms time window: (1) 90–150, (2) 150–210, (3) 210–270, and (4) $\geq$270 ms. The distribution of the vowel duration data did not allow a 20-ms gap between the different categories, as was designed for the interval duration categories.

The constructed durational reorganization described above yielded four categories of interval duration and four categories of vowel duration. Because all available combinations of these categories were created, 16 interval – vowel combinations were constructed. The mean number of tokens per interval – vowel combination was 16.4 (S.D. = 7.2) for the WWR utterances and 18.8 (S.D. = 9.7) for the PWR. The only exception was the combination of the 0 – 50-ms interval duration with the ≥ 270-ms vowel duration in the WWR category. As described above, all utterances in the 0 – 50-ms interval category were original utterances, because temporal manipulations were designed to create 100 ms or longer interval durations. No utterance in the WWR category was originally produced, with interval duration shorter than 50 ms, and at the same time vowel duration longer than 270 ms. Therefore, such combination did not exist.

3. Results

Table 3 presents mean ratings and standard deviations of the 40 listeners to the test items produced by the five stuttering speakers, grouped according to the time categories described above. As can be seen, data are given for all interval – vowel duration combinations, with the exception of the 0 – 50-ms interval duration with the ≥ 270-ms vowel duration in the WWR category, for which no tokens were available. Inspection of group means reveals that all mean ratings ranged between the values of 2.74 and 4.03.

An analysis of variance with repeated measures separating the effects of interval duration, vowel duration, and type of disfluency revealed a significant main effect for interval duration [$F(3,4529) = 4.148$, $P = .006$]. Post hoc analysis, using Tukey's HSD test, demonstrated, however, that the shortest interval duration

Table 3
Group means and standard deviations of the 40 listeners' ratings on a six-point scale of the WWR and PWR utterance variations

Disfluency type	Vowel duration	Interval duration [mean (S.D.)]			
		0 – 50 ms	70 – 250 ms	270 – 450 ms	470 ms and up
WWR	90 – 150 ms	3.70 (1.33)	3.59 (1.30)	3.24 (1.29)	3.28 (1.22)
	150 – 210 ms	3.58 (1.36)	3.22 (1.33)	3.44 (1.25)	3.35 (1.34)
	210 – 270 ms	3.44 (1.25)	3.31 (1.30)	3.30 (1.27)	3.14 (1.14)
	270 ms and up	N/A	2.99 (1.25)	2.74 (1.24)	2.75 (1.28)
PWR	90 – 150 ms	4.03 (1.41)	3.42 (1.55)	3.73 (1.47)	3.81 (1.43)
	150 – 210 ms	3.87 (1.36)	3.38 (1.54)	3.39 (1.42)	3.54 (1.39)
	210 – 270 ms	3.53 (1.43)	3.23 (1.35)	3.68 (1.35)	3.43 (1.40)
	270 ms and up	3.67 (1.44)	3.66 (1.42)	3.70 (1.50)	3.79 (1.35)

condition (0 – 50 ms) was the only condition to be rated significantly higher (more representative of stuttering) than the other three conditions (with individual comparison error rate of $P < .001$). A statistically significant main effect were also found for the vowel duration variable [$F(3,4529) = 9.791$, $P < .001$]. Post hoc analysis, using Tukey's HSD test, demonstrated a significant decrease in ratings, as vowel duration increased, among all pairs of vowel duration conditions that were not adjacent (with individual comparison error rate range of $P = .001 - .014$). In other words, an increase in vowel duration of more than 60 ms caused listener to judge the utterance as more representative of normal disfluency. In addition, utterances containing PWR were rated significantly higher on the six-point scale than utterances containing WWR [$F(1,4529) = 56.840$, $P < .001$].

Type of disfluency was found to interact with the vowel duration factor [$F(1,4529) = 15.908$, $P < .001$]. It was also found to interact with the interval duration factor [$F(1, 4529) = 3.278$, $P = .02$]. No other significant interactions were observed.

4. Discussion

It has been well documented in past literature that listeners' perception of stuttering is influenced by different variables, such as the amount and type of the disfluency in the speech sample, number of spoken units in repetitions, and duration of sound prolongations. As mentioned earlier, the present investigation was an initial exploration designed to assess the possibility that certain temporal acoustic features of specific components of disfluent events affect listeners' perception of disfluencies as being normal or stuttering. Specifically, we were interested to learn whether durational characteristics within repetition contribute to the normal/stuttering distinction in speech of preschool age children. In view of the limited information on this parameter, we were looking for trends that might provide clues for further research in this general domain. The results provided evidence that interval duration affects the perception of disfluency. Repetitions recorded from children who stutter were perceived more like stuttering when the interval between repeated units was shorter than 50 ms. When interval duration was longer than 70 ms, listeners did not react to durational manipulation.

Throneburg and Yairi (1994) reported that interval duration in repetitions produced by stuttering children was shorter than those produced by normally fluent children. The present study demonstrated that although differences in interval duration exist between stuttered and normal speech, it does not affect listeners' judgment through the entire range of duration. A perceptual change occurred only when interval duration was increased to be longer than 70 ms. Beyond that point, no perceptual change was observed. This finding may reflect a perceptual threshold with regard to interval duration. The effect of interval duration manipulation by itself, however, was not sufficient to induce a categorical shift in perception from stuttering to normal disfluency. Further investigation of this domain is needed to better explain the relation between the production of disfluency and its effect on the perception of stuttering.

Vowel duration was also found to affect the perception of stuttering by listeners. When vowel duration within the repetition was increased by more than 60 ms, listeners modified their judgment gradually toward the category of "normal disfluency." This trend was observed more clearly when interval duration was kept short (0–50 ms) than when it was prolonged. It seems that the duration of the vowel within the repetitions contributes to the identification of stuttering versus normal. It is interesting to note that listeners shifted their judgement gradually, as a reaction to vowel duration modification, whereas interval duration modifications induced a judgement pattern that was more typical of a threshold phenomenon.

The effect of vowel duration manipulation, within the range used in the present study, was not sufficient to induce a categorical shift in perception between stuttering and normal disfluency. This finding is consistent with previous descriptive findings presented earlier. Howell et al. (1986, 1987, 1992) reported shorter vowel durations in stuttered PWR than in fluent speech of the same speakers. In their study, however, shortening vowel duration alone was not sufficient to create perceptual change of the identity of the vowel. Throneburg and Yairi (1994) also reported that the total duration of repeated segments produced by stuttering children were not significantly shorter than similar segments produced by normally fluent children. It seems, then, that vowel duration within repetitions produced by stuttering children tends to be slightly shorter than in repetitions produced by normally fluent children. Thus, these small differences in vowel duration do not seem to have a marked impact on listeners' perception of the repetition as stuttering. To induce a significant change in listeners' judgement, a vowel prolongation of more than 60 ms was required.

Past research has suggested that fluent speech of people who stutter is different from that of normally fluent speakers in several acoustic features. Such acoustic parameters include, for example, duration of intervals between repeated units within repetitions produced in normally (dis)fluent speech (Love & Jeffress, 1971), formant transition (Montgomery & Cooke, 1976), consonant–vowel proportion within the syllable (Di Simoni, 1974a, 1974b), vowel intensity (Howell et al., 1987), and duration of the speech segment presented to the listener (Cordes & Ingham, 1994; Cordes et al., 1992). Wendahl and Cole (1961) suggested that there are probably additional differences in terms of vocal tension and stress, and Prins, Hubbard, and Krause (1991), as well as Wingate (1984), have shown that syllabic stress also coincides with stuttering. It is justified, therefore, to suspect that the utterances produced by the stuttering children in the present study included additional acoustic features (such as those listed above) that are relevant in the identification of stuttering. These features have probably influenced listeners' perception, thus somewhat "masking" the direct influence of the temporal manipulation performed here. The present study utilized speech produced by stuttering children in order to maintain naturalness of the speech segments that were later manipulated. It is possible that if the same perceptual paradigm would be used with synthetic speech samples, a better representation of the manipulation effect would be observed. Future study will be needed to verify this aspect of the results.

An additional implication of the differences in listeners' responses to WWR versus PWR applies to the debate of whether WWR should be regarded as stuttering or as normal (e.g., Wingate, 2001; Yairi, Watkins, Ambrose, & Paden, 2001). In the present study, utterances with PWR were rated between half a point to one point higher on the six-point scale than utterances with WWR, indicating that listeners are somewhat more inclined to perceive PWR as stuttering. This result is in accord with previous reports (e.g., Bloodstein, 1995; Boehmler, 1958; Curran & Hood, 1977; Young, 1961, 1984). The present data revealed that monosyllabic word are more sensitive than PWR to manipulation of the two durational parameters employed in this study. There was a significant interaction between type of disfluency and vowel duration. Monosyllabic WWR were judged as more representative of normal (dis)fluency when vowel duration was prolonged. This pattern was not observed as clearly for PWR. Nevertheless, individual variability existed and it is clear that WWR were also identified as stuttering. One of the five children received higher ratings for utterances containing WWR. Interestingly, this speaker (E10) was the only one who produced shorter interval in WWR than in PWR. Whereas previous interpretations (e.g., Wingate, 1984, 2001) emphasized the importance of the linguistic categorization (word or part word) in the identification of disfluency as either "stuttering" or "normal," our data suggest that disfluencies were judged by the listeners according to their durational attributes rather than by their linguistic class. On this basis, it appears to us that/when listeners evaluate nontypical disfluencies (PWR with longer interval than WWR), they might be reacting to the temporal characteristics of the disfluency rather than to the linguistic categorization (part and whole word). We note, however, that these findings are based on a small set of data. Replication on a larger set of data would be necessary before such conclusion can be solidified.

Finally, inasmuch as the present study deals, in part, with the question of what it takes to normalize stuttered repetitions, the findings provide hints for clinical implications. We found that for repetitions identified as stuttering, increasing vowel duration in steps longer than 60 ms influenced listeners to perceive them as less representative of stuttering. In addition, an increase in interval duration within the repetition beyond a specific point also influenced listeners to perceive the disfluencies as more normal. Thus, slight alterations toward slower tempo of repetitions (longer intervals) and perhaps longer vowels within the repetitions ("stretching") should work toward normalizing the disfluencies. It is difficult to ignore the similarities of our findings to ideas expressed in traditional therapeutic approaches that emphasized modification of the "moment of stuttering" (Gregory, 1979) such as Van Riper's (1973) concept of "easy stuttering" and Johnson's (1946) "bouncing" method that emphasized the use of easy repetitions. Such changes in the production of the disfluency can be also relatively easy to monitor automatically (Bakker, 1997, 1999a, 1999b). Although these techniques were applied primarily to adults, the past two decades has seen a tendency to employ various techniques of slow speech and modification of moments of stuttering in therapy programs for preschool ages children who stutter (Lincoln & Harrison, 1999, Shine,1980). Therefore, further research along the lines of the present study might provide useful information specifically geared to this age group.

Acknowledgments

This research reported here was performed while Dr. Amir was affiliated with the University of Illinois at Urbana-Champaign and was based on part of his doctoral dissertation that was directed by Ehud Yairi. This research was supported by grant #R01-DC00459 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. Principal Investigator: Ehud Yairi.

Appendix A. Instruction for the listening task

You are going to participate in a study regarding stuttering and normal disfluency. This is a listening task divided into five parts. Each part will take approximately 8 min. After the third part, you will be given a 5-min break, then continue with the additional two parts.

As you may know, the speech of most people contains incidences of interruption (disfluencies). Some of these interruptions are regarded as normal, whereas others may be regarded as stuttering. During the session, you will hear recordings of children saying one sentence at a time. Each child, however, was recorded many times, saying the same sentence in slightly different ways. All the sentences that will be presented to you include fluency disruptions of some kind. Neither stuttering nor normal disfluency will be defined for you.

You should be aware that sometimes, normal disruption may occur in the speech of a child who stutters and that stuttering can occur in the speech of a normally speaking child. In other words, either type of disruption (normal and stuttering) may be present in the speech of each child. The same child can sometimes stutter and sometimes have a normal disruption, even when repeating the same sentence.

Each sentence that will be presented will be followed by a 5-s pause. Your task is to determine the extent to which each sentence you hear contains stuttering or normal disruption. Mark your judgment during the 5-s pause following each sentence. You will have to mark your decision on a six-point scale, on which 1 means "this sounded like normal disruption" and 6 means "this sounded like stuttering." All the sentences that you will hear include repetitions of the same kind, for example: "I wa-wanna go home" or "who took-took my ball." You have to decide how sure you are that it was either stuttering or normal disruption, and circle the corresponding number on the answering form.

There are no "correct" answers for this task, as individuals differ in their perception. We are interested in your personal judgment about the sentences you hear.

Appendix B. Continuing education

1. Listeners in the present study identified disfluencies as more "normal" when:
   a. vowel duration was prolonged
   b. interval duration was prolonged
   c. both vowel duration and interval duration were prolonged
   d. both vowel duration and interval duration were shortened
   e. none of the above
2. Listeners' perception of stuttering was documented in the literature to be influenced by different variables. Among these variables is/are:
   a. Amount of disfluency
   b. Type of disfluency
   c. Number of spoken unit in repetition
   d. Duration of sound prolongation
   e. All of the above
3. In the present study, listeners judged one-unit repetitions as more representative of stuttering when interval duration within the repetition was:
   a. 0–50 ms
   b. 70–250 ms
   c. 270–450 ms
   d. $\geq$470 ms
   e. 150–210 ms
4. Manipulation of vowel duration affected listeners' perception of stuttering, in this study, when the increase in vowel duration was greater than:
   a. 30 ms
   b. 60 ms
   c. 90 ms
   d. 120 ms
   e. 240 ms
5. Listeners identify PWR differently from WWR. Which statement describes the results more accurately?
   a. WWR are identified as representative of stuttering as PWR
   b. WWR are identified as more representative of stuttering than PWR
   c. WWR are identified as less representative of stuttering than PWR
   d. All of the above are correct
   e. None of the above is correct

References

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, *42*, 859–909.

Bakker, K. (1997). Instrumentation for the assessment and treatment of stuttering. In: R. Curlee, & G. Siegel (Eds.), *Nature and treatment of stuttering: new directions* (2nd ed.), pp. 377 – 397, Boston: Allyn & Bacon.

Bakker, K. (1999a). Technical solutions for quantitative and qualitative assessments of speech fluency. *Seminars in Speech and Language*, *20*, 185–196.

Bakker, K. (1999b). Clinical technologies for the reduction of stuttering and enhancement of speech fluency. *Seminars in Speech and Language*, *20*, 271–280.

Bloodstein, O. (1995). *A handbook on stuttering* (5th ed.). San Diego: Singular Pub. Group, California.

Boehmler, R. M. (1958). Listener responses to non-fluencies. *Journal of Speech and Hearing Research*, *1*, 132–141.

Cordes, A. K. (2000). Individual and consensus judgements of disfluency types in the speech of persons who stutter. *Journal of Speech, Language, and Hearing Research*, *43*, 951–964.

Cordes, A. K., & Ingham, R. J. (1994). Time-interval measurement of stuttering: effect of interval duration. *Journal of Speech and Hearing Research*, *37*, 779–788.

Cordes, A. K., Ingham, R. J., Frank, P., & Ingham, J. C. (1992). Time-interval analysis of interjudge and intrajudge agreement for stuttering event judgment. *Journal of Speech and Hearing Research*, *35*, 383–394.

Curlee, R. F. (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research*, *24*, 595–600.

Curran, M. F., & Hood, S. B. (1977). Listener ratings of severity for specific disfluency types in children. *Journal of Fluency Disorders*, *2*, 87–97.

Di Simoni, F. G. (1974a). Preliminary study of certain timing relationships in the speech of stutterers. *Journal of the Acoustical Society of America*, *56*, 695–696.

Di Simoni, F. G. (1974b). Some preliminary observations on temporal compensation in the speech of children. *Journal of the Acoustical Society of America*, *56*, 697–699.

Gregory, H. H. (1979). *Controversies about stuttering therapy.* Baltimore: University Park Press.

Hegde, M. N., & Hartman, D. E. (1979a). Factors affecting judgments of fluency: I. Interjections. *Journal of Fluency Disorders*, *4*, 1–11.

Hegde, M. N., & Hartman, D. E. (1979b). Factors affecting judgments of fluency: II. Word repetitions. *Journal of Fluency Disorders*, *4*, 13–22.

Howell, P., & Vause, L. (1986). Acoustic analysis and perception of vowels in stuttered speech. *Journal of the Acoustical Society of America*, *79*, 1571–1579.

Howell, P., & Williams, M. (1992). Acoustic analysis and perception of vowels in children and teenagers stuttered speech. *Journal of the Acoustical Society of America*, *91*, 1697–1706.

Howell, P., Williams, M., & Vause, L., (1987). Acoustic analysis of repetitions in stutterers' speech. In: H.F.M. Peters, & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering*. pp. 371 – 380, New York: Springer-Verlag.

Johnson, W. (1942). A study of the onset and development of stuttering. *Journal of Speech Development*, *7*, 251–257.

Johnson, W. (1946). *People in quandaries*. New York: Harper and Row.

Johnson, W., Boehmler, R. M., Dahlstrom, W. G., Darley, F. L., Goodstein, L. D., Kools, J. A., Neelley, J. N., Prather, W. F., Sherman, D., Thurman, C. G., Trotter, W. D., Williams, D., Young, M. A. (1959). *The onset of stuttering*. Minneapolis: University of Minneapolis Press.

Lincoln, M., & Harrison, E. (1999). The Lidcombe program. In: M. Onslow, & A. Packman (Eds.), *Early stuttering intervention.*, pp. 113 – 127, San Diego: Singular Publishing Group.

Love, L. R., & Jeffress, L. A. (1971). Identification of brief pauses in the fluent speech of stutterers and non stutterers. *Journal of Speech and Hearing Research*, *14*, 229–240.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research*, *30*, 529–538.

Montgomery, A. A., & Cooke, P. A. (1976). Perceptual and acoustic analysis of repetitions in stuttered speech. *Journal of Communication Disorders*, *9*, 317–330.

Prins, D., Hubbard, C. P., & Krause, M. (1991). Syllabic stress and the occurrence of stuttering. *Journal of Speech and Hearing Research*, *34*, 1011–1016.

Sander, E. K. (1963). Frequency of syllable repetition and 'stutterer' judgment. *Journal of Speech and Hearing Disorders*, *28*, 19–30.

Schwartz, R. G. (1995). Effect of familiarity on word duration in children's speech: a preliminary investigation. *Journal of Speech and Hearing Research*, *38*, 76–84.

Shine, R. (1980). Direct management of the beginning stutterer. In: W. Perkins (Ed.), *Strategies in stuttering therapy.* pp. 104 – 127, New York: Thieme-Stratton.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Throneburg, R. N., & Yairi, E. (1994). Temporal dynamics of repetitions during the early stage of childhood stuttering: an acoustic study. *Journal of Speech and Hearing Research*, *37*, 1067–1075.

Tuthill, C. (1940). A quantitative study of external meaning with special reference to stuttering. *Journal of Speech Disorders*, *5*, 189–191.

Tuthill, C. E. (1946). A quantitative study of experimental meaning with special reference to stuttering. *Speech Monographs*, *13*, 81–98.

Van Riper, C. (1973). *The treatment of stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

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

Wendahl, R. W., & Cole, J. (1961). Identification of stuttering during relatively fluent speech. *Journal of Speech and Hearing Research*, *4*, 281–286.

Williams, D. E., & Kent, L. R. (1958). Listener evaluation of speech interruptions. *Journal of Speech and Hearing Research*, *1*, 124–131.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Wingate, M. (1984). Stutter events and linguistic stress. *Journal of Fluency Disorders*, *9*, 295–300.

Wingate, M. E. (2001). SLD is not stuttering. *Journal of Speech, Language, and Hearing Research*, *44*, 381–383.

Yairi, E., & Hall, K. D. (1993). Temporal relations within repetitions of preschool children near the onset of stuttering: a preliminary report. *Journal of Communication Disorders*, *26*, 231–244.

Yairi, E., Watkins, R., Ambrose, N., & Paden, E. (2001). What is stuttering? *Journal of Speech, Language, and Hearing Research*, *44*, 585–592.

Young, M. A. (1961). Predicting ratings of severity of stuttering. *Journal of Speech and Hearing Disorders, Monographs, Supplement*, *7*, 31–54.

Young, M.A. (1984). Identification of stuttering and stutterer. In: D. Curlee, & W. Perkins (Eds.), *Nature and treatment of stuttering*. pp. 13 – 30, San Diego: College Hill.

Zebrowski, P. M. (1991). Duration of the speech disfluencies of beginning stutterers. *Journal of Speech and Hearing Research*, *34*, 483–491.