Speech Effort Measurement and Stuttering: Investigating the Chorus Reading Effect

Roger J. Ingham
Allison Warner
Anne Byrd
John Cotton
University of California, Santa Barbara

Purpose: The purpose of this study was to investigate chorus reading's (CR's) effect on speech effort during oral reading by adult stuttering speakers and control participants. The effect of a speech effort measurement highlighting strategy was also investigated.
Method: Twelve persistent stuttering (PS) adults and 12 normally fluent control participants completed 1-min base rate readings (BR–nonchorus) and CRs within a BR/CR/BR/CR/BR experimental design. Participants self-rated speech effort using a 9-point scale after each reading trial. Stuttering frequency, speech rate, and speech naturalness measures were also obtained. Instructions highlighting speech effort ratings during BR and CR phases were introduced after the first CR.
Results: CR improved speech effort ratings for the PS group, but the control group showed a reverse trend. Both groups' effort ratings were not significantly different during CR phases but were significantly poorer than the control group's effort ratings during BR phases. The highlighting strategy did not significantly change effort ratings.
Conclusion: The findings show that CR will produce not only stutter-free and natural sounding speech but also reliable reductions in speech effort. However, these reductions do not reach effort levels equivalent to those achieved by normally fluent speakers, thereby conditioning its use as a gold standard of achievable normal fluency by PS speakers.

KEY WORDS: stuttering, speech effort, chorus reading, speech naturalness

The characteristics of normally fluent speech are not easily defined. Nevertheless, few people would dispute that the concept of normally fluent speech incorporates (a) relatively little disfluency, (b) a normal utterance rate, and (c) a natural sounding quality and (d) is produced with relatively little effort in most situations (see Finn & Ingham, 1989; Starkweather, 1987). This subdivision of the characteristics not only has been of some benefit in prescribing the goals of stuttering treatment, but it also has permitted the development of measures that can be used to program or shape normally fluent speech during therapy (see Ingham & Cordes, 1997). The first three features (a–c) have been operationally defined and can be quantified for therapeutic use (see Ingham & Cordes, 1997), but "speech effort" is still an elusive concept. According to Starkweather (1987), fluent speech can be conceptualized as being effortless in two ways: (a) it requires little cognitive preparation, and (b) it requires only a small amount of muscular exertion. Starkweather recognized the problems involved in trying to measure cognitive effort but argued that muscular effort might be somewhat easier to measure. Muscular effort may correspond to changes in an electromyography signal, although others have contended that speech effort corresponds to the "duration, spectral balance, and the intervocalic sound energy difference of consonants" (e.g., van Son & Pols, 1999, p. 125). However, the level of effort required for speech is still a privately perceived sensation that ultimately must be judged subjectively. Attempts have been made to quantify observer-judged articulatory effort (muscular effort in producing phonemes; Locke, 1972; Malecot, 1955; Parnell & Amerman, 1977; Young, 1981), but these endeavors have not been successful. Young (1981), reviewing the results of his investigation of the reliability of observer judgments of articulatory effort, concluded "that there is only a limited perceptual reality of articulatory effort as it is currently measured" (p. 231). In short, the ultimate referent for a valid measure of speech effort, such as pain, is likely to be based on speaker perception.

Speech effort may also be a problematic feature of poststuttering therapy speech quality, especially for many prolonged speech treatments (Ingham, 1984). Such treatments are often reported to achieve speech that is stutter free, but only at the cost of extensive and non-normal self-control (Perkins, 1985). Some attempts have been made to assess this characteristic. For instance, Perkins (1981) developed the Speech Performance Questionnaire, which asks (among other questions) whether the respondent is "Now able to speak normally without thinking about controlling speech." Boberg and Kully (1994) used the Speech Performance Questionnaire to assess the effects of a therapy program that incorporated prolonged speech features. This involved 30 persistent developmental stuttering speakers (PSs) who were assessed 12 to 24 months after their therapy ceased. It was reported that 50% of PSs responded to this item with "sometimes" or "never," and no participant answered "always." Nonetheless, there is encouraging evidence that some stuttering therapy programs using variants of prolonged speech or its components are now overcoming the problem of producing unnatural sounding speech. For instance, stuttering treatment studies reported by the senior author and his colleagues (Ingham et al., 2001; Ingham & Onslow, 1985) have demonstrated that programs using variants of prolonged speech are capable of producing highly natural sounding and stutter-free speech. However, it is not known whether this natural sounding speech is also acceptably effortless speech.

To date, no studies have examined the relationship between stuttering and speech effort. In fact, only a few studies have reported on attempts to measure speech effort within the nonstuttering literature (see Ingham & Cordes, 1997). However, studies conducted on the measurement of speech naturalness, especially self-ratings of speech naturalness, may have relevance. A study by Ingham, Ingham, Onslow, and Finn (1989) showed that adult-treated PSs were able, while speaking or listening, to rate the naturalness of their own speech on a 9-point scale (Martin, Haroldson, & Triden, 1984) with very high levels of reliability and accuracy. In some instances, PSs reliably perceived (and quantified) changes in their speech naturalness that were not perceived by independent judges. Finn and Ingham (1994) obtained very similar findings when adult PSs rated how natural their speech sounded while speaking with essentially no stuttering during various rhythmic stimulation conditions. Of more relevance though, was Finn and Ingham's (1994) finding that when PSs rated how natural their speech *felt* and how natural it *sounded,* there was a strong relationship between their 9-point scale ratings, although it was not a perfect relationship. Although the issue has yet to be investigated, it seems likely that when PSs rate how natural their speech feels they also take into account how effortful it is to produce. It is noteworthy, though, that Finn and Ingham's (1994) participants used descriptors that referred to either "cognitive effort" or "attention to speech" when they later tried to identify features they used when rating their speech during the different rhythmic stimulation conditions.

The present study extends the above-mentioned perceptual judgment studies on stuttered speech by investigating the level of self-rated speech effort during chorus reading (CR), a well-established fluency-inducing procedure (Bloodstein, 1995; Ingham, 1984). Research on the CR effect has suggested that this procedure may produce relatively natural sounding speech (oral reading) in stuttering speakers (Ingham, Ingham, Finn, & Fox, 2003; Ingham & Packman, 1979). Given that CR does produce perceptually fluent speech, it has recently become a source of considerable interest for brain imaging researchers seeking to elucidate the neural system(s) that PSs might use to support normal fluency (see Brown et al., 2005). However, it has yet to be established that the improved speech during CR is also experienced as effortless fluent speech. The aim of the present study, therefore, was to investigate the effect of CR on self-rated speech effort and experimenter-judged speech naturalness by adult stuttering speakers and normally fluent control participants. A supplementary aim was to investigate the effect of an effort measurement control strategy, highlighting the participant's effort ratings, on intraspeaker speech effort rating agreement during oral reading.

Method

Participants

Participants were 12 PSs (10 men and 2 women, mean age = 28 years) and 12 normally fluent controls (10 men and 2 women, mean age = 25 years). All PS participants reported a history of stuttering since childhood and had participated in a variety of therapy programs;

all reported that treatment had not been beneficial. All were paid volunteers for a brain imaging study on stuttering being conducted at the Research Imaging Center within the University of Texas Health Science Center, San Antonio. Prior to this particular experiment, each participant had completed a 3-min monologue. The PS group's percentage syllables stuttered (%SS) scores on this task ranged from 18.3 to 1.6 (M = 9.36%SS).

Control participants were paid volunteers from the faculty, staff, and students at the University of California, Santa Barbara. All controls were recorded in the Department of Speech and Hearing Sciences at the University of California, Santa Barbara under conditions identical to those at the University of Texas Health Science Center, San Antonio. All participants had hearing within normal limits and were assumed to be of normal cognitive ability; all were successful professionals or high school or university students. All control participants met most selection criteria established by Finn (1996) for an investigation using normally fluent speakers: they self-judged themselves to be normally fluent speakers, reported no history of a communication disorder, and were judged by the experimenter as normally fluent and free of any speech or language disorder. Finn's recommended additional criterion, corroboration by a friend, was not used.

Setting and Apparatus

During the experiment, participants sat alone in a sound-treated room and were monitored by the experimenter, who was located in the control room. All participants were audiovisually monitored by the experimenter via a TV screen and recorded using a digital video camera (Sony VCR HC20) with an external lapel microphone. Each recording file was stored on a computer hard drive for later analysis. A parallel computer terminal arrangement (linked screens in front of the participant and the rating judge) enabled the participant and the judge to be aware of the duration of the trial; the terminal screen also alerted the participant to make a speech effort rating when the trial ceased (1 = *very effortless speech*; 9 = *very effortful speech*). The Stuttering Measurement System (SMS) program (Ingham, Bakker, Ingham, Kilgo, & Moglia, 1999) was used for this study. The SMS is stored on the experimenter's computer. The SMS program also produces a 5-s tone at the end of the 1-min trial so as to alert the judge to make a speech naturalness rating.

Before the experiment, the research assistant and independent judges were trained on the SMS and the Stuttering Measurement and Assessment Training (Ingham, Cordes, Kilgo, & Moglia, 1998) programs. These programs provide standardized training to make real-time stuttered event judgments, syllable counts, and speech naturalness ratings.

The CR stimulus conditions were provided by an audio recording of a college-age female fluent speaker orally reading three different "garden passages" that differed from those that participants read aloud in base rate (BR) conditions. Her oral reading rates on these three original recordings were, in order, 206, 225, and 202 syllables per minute (SPM). These recordings were then redigitized at different speeds using GoldWave software so that 14 different accompanist reading rate recordings were available for each of the three CR passages; they ranged from 70% to 130% of the original recording speech rate.[1] The accompanist stimulus was transmitted unilaterally (preferred ear) via insert ear phones (E-A-R Tone front tubes; Model 3A) fitted with foam ear-tips.

Design and Procedure

The format of this experiment can be conceptualized as a time-series ABA experimental design (Kazdin, 1998) that was preceded by preliminary AB phases. The entire format required three 1-min trials within each of the following phases: BR1/CR1/BR2/CR2/BR3 (BR = base rate, CR = chorus reading conditions). The preliminary base rate (BR1) and chorus reading (CR1) phases were conditions that provided each participant with practice in making speech effort ratings under both experimental conditions. This practice was designed to help participants make their most accurate speech effort ratings during the subsequent BR2/CR2/BR3 phases (thus, in time series design terms, an A phase was a BR phase and a B phase was a CR phase).

During the experiment, each participant sat alone in a sound-treated room in front of a table housing a computer monitor and keyboard. The attending research assistant provided the participants with reading material. These were passages describing the construction and management of a domestic garden. Participants then completed three 1-min oral readings; they were prompted to start and stop by a computer tone. During each reading task, the research assistant recorded syllables spoken, syllables stuttered and, at the end of each 1-min trial, a speech naturalness rating on the 1-to-9 scale (1 = *highly natural sounding speech*; 9 = *highly unnatural sounding speech*; Martin et al., 1984); all are incorporated within the SMS program. The syllables-spoken and syllables-stuttered data were then converted into SS% and SPM scores by the SMS program. In addition, the SMS program recorded the number of stuttered 5-s intervals during a 1-min trial; by excluding such intervals, it is then possible to derive a measure of stutter-free syllables per minute (SFSPM).

Before the start of the experiment each participant read the following instructions:

[1]Contact Roger J. Ingham for copies of the passages and recordings.

> We are studying the amount of effort you feel when you speak. Sometimes you might feel as though you use a lot of physical effort, while at other times you might feel as though you can speak without much effort at all. On this occasion we want you to judge the amount of effort required while reading aloud different passages. During the first part of this experiment we want you to read aloud from each passage instructed for 1 min. We will tell you when to start.
>
> At the completion of each 1-min reading task you will be instructed on the computer screen to make an effort rating from 1 to 9, to indicate just how much effort it took to read aloud. If reading aloud felt very effortless, then we want you to press "1" on the keyboard. If it felt very effortful, then we want you to press "9" on the keyboard. If it felt somewhere between very effortless and very effortful, then press the appropriate number between 1 and 9 on the keyboard. We want you to make this rating independent of whether you are stuttering or not stuttering. Just concentrate on the amount of physical effort needed to speak, whether or not you were stuttering.
>
> Do you have any questions?

The reference to stuttering was omitted from the control group's instructions. To ensure that each participant fully understood the instructions, the experimenter removed the instruction sheet and asked the participant to explain precisely what was required.

It is important to note that at no point throughout the experiment was the research assistant aware of the participant's speech effort rating; neither was the participant aware of the experimenter's data scores. The participant's keyboard was not visible to the experimenter; the parallel computer screen arrangement indicated to the experimenter only when the participant had registered a speech effort rating.

Before the first CR1 trial, each participant's most comfortable CR rate and volume level were identified by arranging for the participant to practice orally reading for short intervals (~10 s) along with the accompanist at different speeds and signal intensities. This procedure was continued until the participant identified the fastest and most comfortable oral reading rate; that stimulus setting was then used in all CR trials for the participant. The instructions to the participant were as follows:

> Reading aloud the same passage as another reader—a procedure that is called chorus reading—often helps people reduce their stuttering. In another words, some people who stutter find it easier to speak more fluently when accompanied by someone else reading the same material. The research assistant will play recordings of a person reading the passage in front of you at different speeds. We want you to practice briefly reading the passage in front of you to the accompaniment of the recording until you find the speed that best fits your natural reading rate. As soon as you hear the recording we want you to try to read the same words AND then tell us if you want the volume adjusted. We will increase or decrease the volume until you find a level that allows you to hear the recorded words as you read them aloud and that you find the level to be comfortable.
>
> Now we want you to select one of the recording speeds that you feel most comfortable in reading along with. Take your time and feel free to ask us to replay different recording rates until you find one that is most suitable for you.
>
> When the experimental trials begin don't worry if you fall behind the accompanist. Just move on in the text and try to keep up. Do the best you can.
>
> Do you have any questions?

For control participants, the reason for reference to stuttering was explained in conjunction with the instructions.

After completing the first two experimental phases (BR1 and CR1), the experimenter then provided each participant with the following additional instructions in order to control and possibly improve rating accuracy:

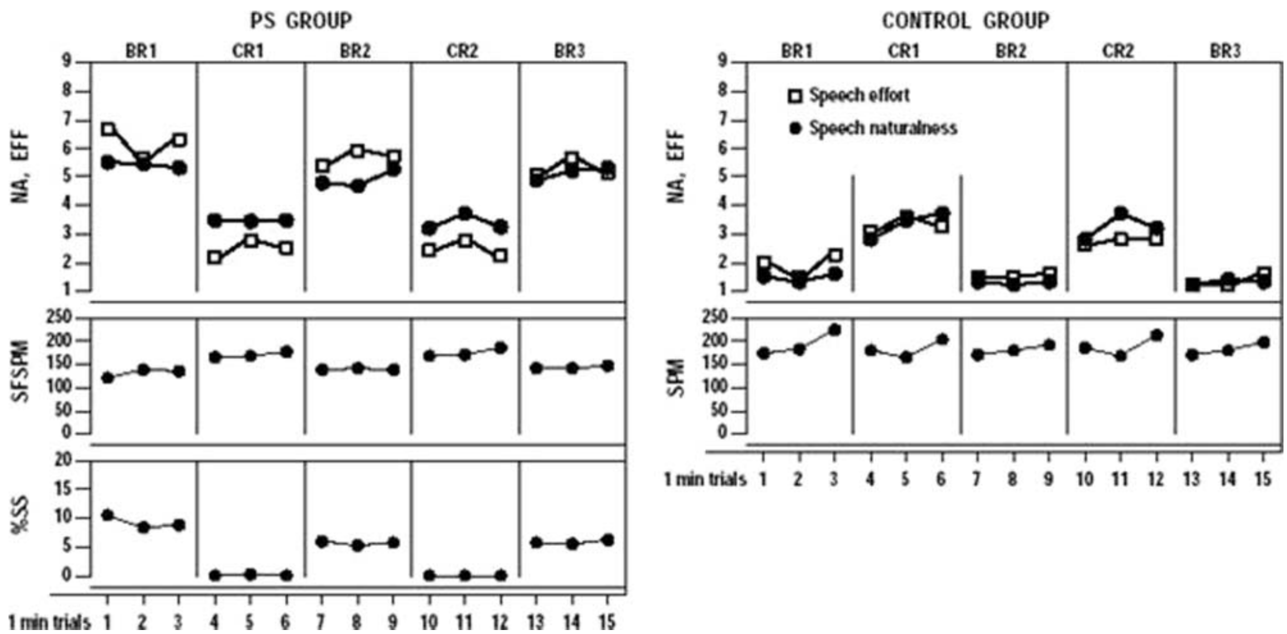
> At the completion of the previous speaking task you rated the amount of physical effort you experienced while reading aloud with the accompanist. You gave that speaking task a rating of X on the 1–9 Speech Effort Scale (1 = very effortless speech; 9 = very effortful speech). During the next speaking task you will be reading aloud a similar passage, but without hearing the accompanist. We want you to use your speech effort rating for the previous task as a guide for rating the level of speech effort you experienced during this task. It may have been more or less effortful. With that in mind, please make your rating for this task on the 1–9 scale.
>
> Do you have any questions?

Once again, after reading the instructions the participant was then asked to explain to the experimenter precisely what was required. The experimenter ensured that the participant understood the instructions before beginning the next 1-min reading task.

Reliability

Interjudge measurement reliability was determined by arranging for two independent judges, also trained on the SMS and Stuttering Measurement and Assessment Training programs, to rescore all session trial recordings from half of the PS and control group participants. Both

Figure 1. Mean 1-min trial scores for the persistent stuttering (PS) group (n = 12) and control group (n = 12) across the study's five experimental phases. During base rate (BR) phases, participants read aloud, and during chorus reading (CR) phases participants read aloud while a recorded accompanist was heard reading the same passage. Training instructions were provided between phases CR1 and BR2. %SS = percentage syllables stuttered; SFSPM = stutter-free syllables per minute; NA = naturalness; EFF = effort.



judges were unfamiliar with the design of the study. Six of the PS participants' trials (fifteen 1-min trials) were rescored by one judge, and six of the control participants' trials were rescored by the other judge. All trials for a participant were presented to judges in a randomized order. They rescored the trial recordings from a computer monitor using concurrently the Windows Media Player and the SMS program (Ingham et al., 1999). Reliability was established by making a graphic comparison between the experimenter's and independent judge's scores to establish the validity of the experimental effects achieved during the experiment. The findings are discussed in the Results section.

Results

Group Data

The mean trial scores for the PS and control groups across all experimental phases are summarized and shown graphically in Figure 1. The mean scores within each of the five experimental phases are shown in Table 1. For the most part, the findings from both groups are reasonably straightforward and are readily interpreted by inspecting Figure 1. It is apparent that introducing and withdrawing CR conditions produced substantial changes in most of the experimenter-judged and self-judged measures.

Stuttering Frequency

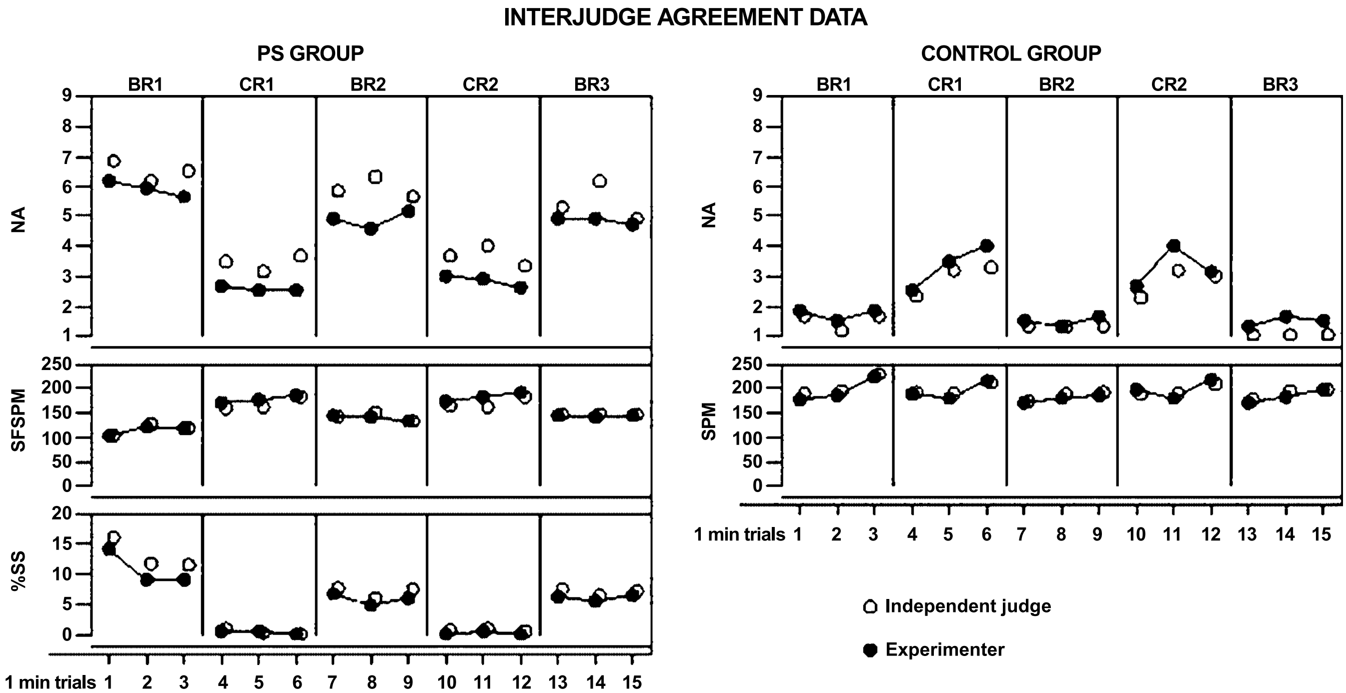
As expected, the PS group showed a dramatic and immediate reduction in stuttering frequency during the CR phases. The overall difference between the mean %SS scores for the three BR phases (7.15%SS) and two CR phases (0.26%SS) was confirmed as significant within a repeated measures analysis of variance (ANOVA; Winer, Brown, & Michels, 1991), $F(1, 154) = 149.8$, $p < .0001$, $d = 1.860$.[2]

Speech Rate

The effect of the CR conditions on the SFSPM scores for the PS group was also evident in the differences between the overall mean scores for the three BR phases (147.6 SFSPM) and the two CR phases (178.7 SFSPM). This difference was also confirmed as significant within a repeated measures ANOVA, $F(1, 154) = 56.32$, $p < .0001$, $d = 1.144$. For the control group, the effect of the CR condition on speech rate was assessed by reference to their SPM scores. It should be noted that

[2]Because the five phases of this experiment (BR1, CR1, BR2, CR2, and BR3) do not imply an orthogonal nesting of the trials within phases, expected mean squares based on the general linear model for ANOVA show odd hypotheses being tested with a standard analysis. Therefore, it is useful to think of the tests performed here as following from a cell means model (e.g., Kirk, 1995, pp. 240–244; Searle, 1987, especially chap. 3) in which the hypotheses tested are about equality of averaged appropriately weighted population cell means.

Figure 2. Experimenter's and independent judge's mean scores for 6 PS participants and 6 Control participants. The experimenter's mean scores are shown in closed circles, and the independent judge's scores are shown in open circles.



these SPM data are derived from the number of syllables spoken within a 1-min trial, whereas the PS group's SFSPM data were derived from the syllables spoken during 5-s intervals that were stutter free within a 1-min trial (as isolated by the SMS program). This meant that it was not possible to make a legitimate direct comparison between the PS group's and control group's speech rate data. An overall comparison between the control group's BR and CR phase means (186.3 vs. 185.4 SPM) shows that they were not significantly different, $F(1, 154) = 0.15$, $p = .6999$, $d = 0.016$. In short, the CR effect data in this study cannot be attributed to a reduction in oral reading rate.

Speech Quality

Inspection of Figure 1 shows that the effect of CR on each group's speech effort and speech naturalness ratings is reasonably obvious and unambiguous. It is also obvious, however, that the effect of CR on these qualitative measures differs in direction and magnitude across the groups. In the PS group, there was an improvement in these ratings each time CR was introduced, but the control participants' effort ratings increased during CR conditions. These trends were elucidated in the data analyses.

Speech naturalness ratings. A repeated measures ANOVA was again used separately for the speech naturalness and speech effort data. In the case of speech naturalness ratings, there was a significant group difference between ratings, $F(1, 22.25) = 38.79$, $p < .0001$, $d = 1.090$, but also a significant Group × Phase interaction effect, $F(1, 88) = 107.3$, $p < .0001$, $d = 1.426$. The source of these interaction effects emerged within the planned comparisons between the phase means for both groups. The speech naturalness planned comparison results for the PS and control groups are shown in the left column of Table 2. This table shows that for both groups the differences between the BR and CR phase means were always significant, but there were no significant differences between BR phase means (BR1 vs. BR2, etc.) or between CR phase means (CR1 vs. CR2).[3] Of special interest, though, were the comparisons between the PS and control groups for each phase (see PS × Control Group in Table 2); they show that the groups' BR phase means all differed significantly, but for the CR phases the differences between the groups' means were not significant.

Speech effort ratings. The repeated measures ANOVA for the speech effort data showed there was a significant group difference between ratings, $F(1, 22.3) = 27.83$, $p < .0001$, $d = 0.898$, but also a significant overall difference between the BR and CR phase means, $F(1, 88) = 18.56$, $p < .0001$, $d = 0.801$. However, this difference obscured an

[3]Three levels of significance ($p < .05$, $p < .01$, and $p < .001$) are reported here. Given the 45 possible comparisons of two cells, a Bonferroni correction divides .05 by 45 = .001+, yielding a familywise correction for one or more significant findings by chance.

Table 1. The experimental phase mean scores.

	PS group					Control group				
	BR1	CR1	BR2	CR2	BR3	BR1	CR1	BR2	CR2	BR3
Effort	6.22	2.47	5.70	2.47	5.31	1.94	3.31	1.53	2.72	1.30
NA	5.51	3.47	4.89	3.38	5.12	1.44	3.33	1.25	3.22	1.28
SFSPM/SPM	134	177	153	180	156	195	182	181	188	183
%SS	9.6	0.4	5.9	0.2	6.0					

Note. Data provided for the persistent stuttering (PS) group are mean effort ratings, naturalness (NA) ratings, stutter-free syllables per minute (SFSPM), and percentage syllables stuttered (%SS). Data provided for the control group are mean effort ratings, naturalness ratings, and syllables per minute (SPM). BR = base rate; CR = chorus reading.

obvious Group × Phase interaction effect, $F(1, 88) = 119.76$, $p < .0001$, $d = 2.884$.[4] As was true for the speech naturalness data, the source of these interaction effects for the speech effort ratings were revealed by the planned comparisons between the phase means for both groups. The speech effort planned comparison results for the PS and control groups are shown in the right column of Table 2. This table shows that, with one exception, the differences between the BR and CR phase means for both groups were always significant. The exception was between BR1 and CR2 for the control group, which did not reach significance, although the experimental effect was preserved by the comparison among the BR2/CR2/BR3 phase means.

[4]We calculated effect sizes ($\widehat{ES}$) for the ANOVA interactions within these data by using proportionate subclass frequencies. For a 2 × 2 interaction, a customary expression of estimated effect is proportional to the difference between means for columns in Row 1 less the difference between means for columns in Row 2:

$$\overline{Y}_{11} - \overline{Y}_{12} - \overline{Y}_{21} + \overline{Y}_{22}, \tag{1}$$

where the first subscript refers to a row and the second to a column. With equal cell numbers n, the quantity

$$n(\overline{Y}_{11} - \overline{Y}_{12} - \overline{Y}_{21} + \overline{Y}_{22})^2/4 \tag{2}$$

is algebraically equal to sum of squares interaction computed with more familiar raw score or deviation score formulas. A little algebra shows that the associated F test having 1 degree of freedom for its numerator implies that

$$t = \frac{\overline{Y}_{11} - \overline{Y}_{12} - \overline{Y}_{21} + \overline{Y}_{22}}{\sqrt{2}\sqrt{2MS_{Error}/n}} \tag{3}$$

with $\sqrt{MS_{Error}}$ estimating the population σ, and an estimated $\widehat{ES}$ to be discussed shortly.

In the present case, however, $n = 36$ (3 scores for each of 12 participants) for a group measured for baseline and $n = 24$ (2 scores for each of 12 participants) for that group when measured otherwise. Hence, the approach above must be modified for unequal but proportionate ns. Using Winer et al.'s (1991, pp. 117–118) report of use of harmonic means of sample sizes in a slightly related case, we computed a harmonic mean n of 28.8 for this example. The formulas just presented using the modified n now make our new sums of squares, Fs, and ts equivalent to those obtained with standard formulas and unequal but proportionate ns.

The foregoing t formula implies

$$\widehat{ES} = \frac{(1/\sqrt{2})(\overline{Y}_{11} - \overline{Y}_{12} - \overline{Y}_{21} + \overline{Y}_{22})}{\sqrt{MS_{Error}}}, \tag{4}$$

exactly the same formula as if all ns had been the same. For the current data, $\widehat{ES} = 1.426$ for the naturalness data and 2.884 for the effort data.

With that exception, there were also no significant differences between BR phase means (BR1 vs. BR2, etc.) or between CR phase means (CR1 vs. CR2). The comparisons between the PS and control groups for each phase were of additional interest. These comparisons showed that the groups' BR phase means all differed significantly, and this was true for all BR– CR comparisons. The PS group's CR ratings were not significantly poorer than the control group's CR ratings. In other words, CR produced effort ratings that did not differ significantly between the groups. However, the PS group's overall mean speech effort ratings during the CR phases were significantly poorer than the control group's mean BR phase ratings (2.47 vs. 1.59, $p = .0047$, $d = 0.676$). In other words, CR did produce speech that PS participants adjudged to be relatively less effortful than normal speech, but it was also rated as significantly more effortful than the control group's normally fluent oral reading.

Effects of Highlighting Speech Effort Ratings

Additional instructions were introduced at the completion of the CR1 phase for both groups (see Method section). These instructions highlighted the speech effort ratings that participants gave to their speech during BR and CR conditions to assist judgment accuracy. Therefore, comparisons were made to determine whether these instructions changed speech effort ratings made in subsequent experimental phases. The effects of the highlighting were expected to be shown by a significant change in speech effort ratings from CR1 to CR2, and from BR1 to BR2 and BR3. However, as previously noted, Table 2 shows that for both groups there was no significant change in speech effort ratings across these comparisons. In short, there was no evidence that the highlighting instructions influenced subsequent ratings.

Individual Differences

The resemblance between the overall trend of the speech effort and naturalness ratings, as shown in Figure 1, made it of interest to examine the levels of correlation between these ratings. Despite the relatively restricted range of the effort and naturalness ratings in

Table 2. Planned comparisons for identifying differences between all experimental phase means for speech naturalness and effort ratings.

	Naturalness ratings						Effort ratings				
	PS group						PS group				
	BR1	CR1	BR2	CR2			BR1	CR1	BR2	CR2	
CR1	***					CR1	***				
BR2	*ns*	**				BR2	*ns*	**			
CR2	***	*ns*	**			CR2	***	*ns*	***		
BR3	*ns*	***	*ns*	***		BR3	*ns*	***	*ns*	***	
	Control group						Control group				
CR1	***					CR1	***				
BR2	*ns*	***				BR2	*ns*	**			
CR2	***	*ns*	***			CR2	*ns*	*ns*	*		
BR3	*ns*	***	*ns*	***		BR3	*ns*	***	*ns*	***	
	PS × Control group						PS × Control group				
	BR1	CR1	BR2	CR2	BR3		BR1	CR1	BR2	CR2	BR3
BR1	***	***	***	***	***	BR1	***	***	***	***	***
CR1	***	*ns*	***	*ns*	***	CR1	*ns*	*ns*	*ns*	*ns*	*
BR2	***	***	***	***	***	BR2	***	***	***	***	***
CR2	***	*ns*	***	*ns*	***	CR2	*ns*	*ns*	*	*ns*	*
BR3	***	***	***	***	***	BR3	***	***	***	***	***

Note. Comparisons are shown for the PS and control groups, plus comparisons between groups (PS × Control Group) for each phase.
*$p < .05$. **$p < .01$. ***$p < .0001$.

the control group, the average Pearson product–moment correlation between each control participant's ratings was .62 ($p < .01$). The PS group's average participant correlation was very similar, at .57 ($p < .01$). However, this correlation score obscures a dichotomous pattern within the PS group's data. Inspection of Table 3 reveals that 7 of 12 PS participants had correlations between adjudged ratings that ranged between .76 and .97, but 2 of 12 actually had slightly negative correlations (–.12 and –.21). Overall, there is generally a strong relationship between the effort and naturalness ratings; 7 of 12 PS participants' ratings, and 6 of 12 control participants' ratings, showed correlations of .72 or more. Nonetheless, it is clear that speech effort and speech naturalness are independent dimensions of speech.

Interjudge Agreement

Figure 2 provides a graphic depiction of the experimenters' and independent judges' speech naturalness, SFSPM/SPM, and %SS scores. In general terms, the results are of interest because they show clearly that the overall experimental trends were also replicated when data from 50% of the participants in the PS and control groups were analyzed. At a more specific level, the results show that the experimenters' and independent judges' mean scores for both groups displayed almost identical score trends across the five experimental phases for the experimenter-judged scores. In fact, 87.2% (157 of 180) of the experimenters' and independent judges' naturalness ratings differed by either one rating unit or were identical. In short, the interjudge scores for this experiment show that the experimental effects were reliable.

Discussion

The present findings provide encouraging but qualified indications that a self-rating scale measure of speech effort could be a valid and clinically useful means of helping to evaluate the quality of a speaker's fluency. This would have applicability for stuttering therapy, where there is continuing concern about the nature and quality of fluency that emerges from many current treatments. More particularly, the method used in this study to measure speech effort may also form the basis of a viable strategy for determining whether treatment does produce improvements in fluency that approximate normally fluent speech. It is conceivable that the CR strategy—and perhaps other fluency-inducing strategies—might be used to give a speaker access to a perceptual "yardstick" for gauging the quality of his or her fluency. Indeed, if CR is consistently reported to achieve stutter-free, or almost stutter-free, speech, then speech effort rating data during this task could be normalized in much the same way as

Table 3. Pearson product–moment correlations for each PS and control (Cont.) participant's possible pairs of scores across the fifteen 1-min experimental trials.

Group	#	Effort NA	Effort SFSPM	NA SFSPM	Effort %SS	NA %SS	SFSPM %SS
PS	1	.94	−.88	−.90	.92	.92	−.92
PS	2	.97	−.94	−.91	.95	.92	−.97
PS	3	.50	−.88	−.57	.91	.49	−.85
PS	4	.78	−.83	−.74	.73	.84	−.86
PS	5	.90	−.89	−.98	.82	.96	−.97
PS	6	.76	−.60	−.23	.93	.87	−.51
PS	7	−.21	.54	.02	.20	−.09	.09
PS	8	.13	−.01	−.23	.02	−.76	−.05
PS	9	.83	−.64	−.60	.93	.91	−.67
PS	10	.41	−.12	−.39	.80	.77	−.33
PS	11	.91	−.94	−.93	.93	.87	−.96
PS	12	−.12	.04	−.32	−.13	.64	−.73
Cont.	1	.81	.13	−.19			
Cont.	2	.69	−.72	−.53			
Cont.	3	.42	.08	−.11			
Cont.	4	.58	.32	.36			
Cont.	5	.76	.45	.29			
Cont.	6	.36	.66	.29			
Cont.	7	.72	.45	.38			
Cont.	8	.72	−.19	−.44			
Cont.	9	.91	−.38	−.46			
Cont.	10	.46	−.15	.04			
Cont.	11	.81	−.11	−.51			
Cont.	12	.28	.27	−.02			

speech naturalness ratings are now being normalized (see Schiavetti & Metz, 1997). In summary, the current findings show that in addition to reduced stuttering, CR produces relative improvements in speech naturalness and speech effort. The levels of speech effort induced by CR, although virtually the same for stuttering and normally fluent speakers, appear to indicate slightly more speech effort than that produced by normally fluent speakers when orally reading without an accompanist.

The findings from the present study replicate those from previous studies that have shown that CR produces reduced stuttering and improved speech naturalness. However, the level of speech naturalness that the stuttering speakers achieved during this study was still significantly poorer than the level achieved by normally fluent speakers when orally reading without a chorus accompanist. In this regard, the findings fail to replicate those previously reported by Ingham et al. (2003). This might be related to the fact that the latter findings were obtained during very different conditions; participants orally read for a shorter period within a brain imaging scanner. The current findings are also supported by the equivalence between the speech naturalness ratings of the control and PS groups during CR conditions. Essentially identical conclusions apply to the speech effort and speech naturalness ratings.

One interesting "non-finding" was the absence of any evidence that the highlighting of the speaker's speech effort ratings (instituted partway through the experiment) produced any positive (or negative) effect on speech effort rating. In essence, the data suggested that simply highlighting or reminding speakers of the ratings they gave their speech during the two different speaking conditions is likely to be an innocuous procedure in this type of experiment. On a more positive note, though, the results did show that the speakers were surprisingly consistent in the ratings they gave for their speech effort during the repeated CR conditions and during the repeated nonchorus oral reading conditions. For the control participants and the PSs, 72.2% and 69.4%, respectively, of their speech effort ratings given prior to the rating instructions agreed (within ±1 rating unit) with their postinstruction ratings during CR.

A point of concern regarding the fidelity of the study might be whether participants were actually paying attention to the level of physical effort and not the cognitive demands of the speaking task when they were rating their speech effort. An informal postexperiment questioning of the PS participants and a sample of control participants confirmed that their ratings were directly related to the physical ease of production. This may have also been enhanced in this study by the relative simplicity of the oral reading material. In any event, the issue of precisely how stuttering speakers distribute their expenditure of effort during speaking tasks, such as those used in the present study, probably involves a complex distribution of neural executive systems, as has been suggested through studies by Gopher and colleagues (Gopher, 1996; Iani, Gopher, & Lavie, 2004) Perhaps, therefore, an understanding of the basis of such neural system arrangements that might support the concept of speech effort will be more appropriately investigated in the future using brain imaging technologies.

Even if brain imaging were involved in future research on speech effort, one issue that is still worth considering is whether speech effort might be more accurately determined by measuring a physical or physiological parameter of speech production. Some studies have reported attempts to use psychophysical measures of a "sense of effort" during speech production. Prosek and Montgomery (1969), for example, reported that the level of perceived effort involved in speaking loudly by normal speakers was functionally related to their levels of subglottic air pressure. Wright and Colton (1972) and Colton and Brown (1972) made similar measurements, with the latter reporting that normal speakers' sense of effort was highly correlated with frequency, intensity, and subglottic pressure during phonation. McCloskey and colleagues (see, e.g., McCloskey, 1981) have also conducted numerous studies showing the crucial role of a sense of effort in executing complex motor movements, and Somodi, Robin, and Luschei (1995) found that a sense of effort, as recorded by hand pressure, can be directly and accurately related to certain levels of tongue pressure.

The preceding studies suggest that speakers can accurately quantify perceived changes in the effort levels required for normal speech production, but the relationship between perceived effort and known deficits in speech physiology is, of course, even more relevant to stuttering. This issue was pursued in studies reviewed recently by Solomon, Robin, Lorell, Rodnitzky, and Luschei (1994) on the level of perceived effort required to maintain tongue pressure in normal and dysarthric speakers. Verdolini-Marston, Hoffman, and McCoy (1992) also reported that magnitude estimations of effort were sensitive to changes in vocal structure and function after adults with voice disorders had received voice therapy. Watterson, McFarlane, and Diamond (1993) similarly reported that patients with voice disorders, and some normal speaker control participants, could recognize differences in the degree of vocal effort required for the production of voiceless obstruents and, in turn, with sonorants and nasals. A notable feature of Watterson et al.'s study was that a group of experienced listeners did not recognize differences in vocal quality across samples of speech produced with differences in speaker-judged vocal effort, a finding quite similar to the speech naturalness results reported by Finn and Ingham (1994) and by Ingham et al. (1989).

The results of the present study may, at first glance, appear to be at odds with the above findings. The graphed mean data trends (see Figure 1) appear to show a very strong relationship between speaker-judged effort ratings and experimenter-judged speech naturalness ratings. In fact, however, the overall correlation between the mean speech effort and speech naturalness ratings was rather moderate (controls = .62; PSs = .57) and obscured considerable individual differences. For instance, 4 of 12 PS participants displayed either a negligible or slightly negative relationship between these ratings; this was also true for 3 of 12 control participants. On the other hand, a better than +.9 correlation occurred between these ratings for 4 of 12 PSs and 1 of 12 control participants, which suggests, not surprisingly, that speech naturalness and speech effort are functionally interrelated for some, but not all, speakers. In this respect, therefore, the findings align with those reported by Young (1981), who failed to find a consistent concordance between assumed changes in speech effort and observer judgments. The findings also align with those reported by Finn and Ingham (1994), who found a pattern of similarity between how natural a PS's speech felt and how natural it sounded (to the speaker and listeners) during different rhythmic stimulation conditions. However, it is noteworthy that, as mentioned earlier, Finn and Ingham (1994) also determined that their participants' ratings of how natural their speech felt and how natural it sounded were "based on different criteria" (p. 337).

CR provides an indirect method for modifying speech effort, but there may be other ways of doing so. One possible avenue is to use speech effort measures to assess the efficacy of different therapy strategies. It may not be possible to use speech effort as an independent variable in experimental manipulations as part of treatment—as is the case with speech naturalness ratings (Ingham & Onslow, 1985)—but it is surely possible to use speech effort as a dependent variable. In other words, the extent to which different therapy strategies actually aid or improve fluency might be determined through such measures. Also, of course, self-ratings of speech effort might be used to help evaluate the effects of different stuttering therapy strategies during different speaking situations or speaking tasks. In the meantime, there is ample reason to begin investigations that will determine whether speech effort ratings differ reliably across different fluency-inducing strategies (e.g., rhythmic stimulation, masking, etc.) and, more importantly, whether there is a reliable difference in ratings across these strategies. Such findings might assist in deriving a standardized method for measuring self-judged speech effort and adding to the repertoire of methods that clinicians and researchers have available for measuring fluency.

Acknowledgments

This study was made possible by a grant from the University of California, Santa Barbara, Research Committee. Special thanks are due to Kristin White, Cari Nicholson, Irene Seybold, and Danielle Brett for their contributions as research assistants in preliminary trials that contributed to the method and design of this study. Thanks are also due to Janis Ingham and Irene Seybold, who served as independent judges, and to Don Robin for thoughtful comments on our conclusions.

References

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

Boberg, E., & Kully, D. (1994). Long-term results of an intensive treatment program for adults and adolescents who stutter. *Journal of Speech and Hearing Research, 37,* 1050–1059.

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping, 25,* 105–117.

Colton, R. H., & Brown, W. S. (1972, November). *Some relationships between vocal effort and intra-oral air pressure.* Paper presented at the 84th meeting of the Acoustical Society of America, Miami, FL.

Finn, P. (1996). Establishing the validity of recovery from stuttering without formal treatment. *Journal of Speech and Hearing Research, 39,* 1171–1181.

Finn, P., & Ingham, R. J. (1989). The selection of "fluent" samples in research on stuttering: Conceptual and methodological considerations. *Journal of Speech and Hearing Research, 32,* 401–408

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

GoldWave. (2004). GoldWave digital audio editor [Computer software]. Retrieved from http://www.goldwave.com

Gopher, D. (1996). Attention control: Explorations of the work of an executive controller. *Cognitive Brain Research, 5,* 23–28

Iani, C., Gopher, D., & Lavie, P. (2004). Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology, 41,* 789–798

Ingham, R. J. (1984). *Stuttering and behavior therapy: Current status and empirical foundations*. San Diego, CA: College-Hill Press

Ingham, R. J., Bakker, K., Ingham, J.C., Kilgo, M., & Moglia, R. (1999). *Stuttering Measurement System* (software and manual). Santa Barbara: University of California.

Ingham, R. J., & Cordes, A. K. (1997). Self measurement and evaluating treatment efficacy. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (pp. 413–438). Boston: Allyn & Bacon

Ingham, R. J., Cordes, A. K., Kilgo, M., & Moglia, R. (1998). *Stuttering Measurement Assessment and Training* (software and videodisk). Santa Barbara: University of California.

Ingham, R. J., Ingham, J. C., Finn, P., & Fox, P. T. (2003). Towards a functional neural systems model of developmental stuttering. *Journal of Fluency Disorders, 28,* 297–318

Ingham, R. J., Ingham, J. C., Onslow, M., & Finn, P. (1989). Stutterers' self-ratings of speech naturalness: Assessing effects and reliability. *Journal of Speech and Hearing Research, 32,* 495–504

Ingham, R. J., Kilgo, M., Ingham, J. C., Moglia, R., Belknap, H., & Sanchez, T. (2001). Evaluation of a stuttering treatment based on reduction of short phonation intervals. *Journal of Speech, Language, and Hearing Research, 44,* 1229–1244

Ingham, R. J., & Onslow, M. (1985). Measurement and modification of speech naturalness during stuttering therapy. *Journal of Speech and Hearing Disorders, 50,* 261–268

Ingham, R. J., & Packman, A. (1979). A further evaluation of the speech of stutterers during chorus- and nonchorus-reading conditions. *Journal of Speech and Hearing Research, 22,* 784–793.

Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Boston: Allyn & Bacon.

Kirk, R. E. (1995). *Experimental design. Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Locke, J. (1972). Ease of articulation. *Journal of Speech and Hearing Research, 15,* 194–200.

Malecot, A. (1955). An experimental study of force of articulation. *Studia Linguistica, 9,* 35–44.

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

McCloskey, D. I. (1981). Corollary discharges: Motor commands and perception. In V. B. Brookes (Ed.), *Handbook of physiology: A critical comprehensive presentation of physiological knowledge and concepts* (pp. 1415–1448). Bethesda, MD: American Physiological Society.

Parnell, M., & Amerman, J. D. (1977). Subjective evaluation of articulatory effort. *Journal of Speech and Hearing Research, 20,* 644–652.

Perkins, W. (1981). Measurement and maintenance of fluency. In E. Boberg (Ed.), *Maintenance of fluency* (pp. 147–178). New York: Elsevier.

Perkins, W. H. (1985). Horizons and beyond: Confessions of a carpenter. *Seminars in Speech and Language, 6,* 233–244.

Prosek, R. A., & Montgomery, A. A. (1969, November). *Some physical correlates of vocal effort and loudness.* Paper presented at the annual convention of the American Speech and Hearing Association, Chicago.

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* (pp. 398–412). Boston: Allyn & Bacon.

Searle, S. R. (1987). *Linear models for unbalanced data*. New York: Wiley.

Solomon, N. P., Robin, D. A., Lorell, D. M., Rodnitzky, R. L., & Luschei, E. S. (1994). Tongue fatigue testing in Parkinson disease: Indications of fatigue. In J. Till, K. Yorkston, & D. Beukelman (Eds.), *Dysarthria and apraxia of speech* (pp. 147–160). Baltimore: Brookes.

Somodi, L. B., Robin, D. A., & Luschei, E. S. (1995). A model of "sense of effort" during maximal and submaximal contractions of the tongue. *Brain and Language, 51,* 371–382.

Starkweather, C. W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice Hall.

van Son, R. J. J. H., & Pols, L. C. W. (1999). An acoustic description of consonant reduction. *Speech Communication, 28,* 125–140.

Verdolini-Marston, K., Hoffman, H., & McCoy, S. (1992). Non-specific laryngeal granuloma: A literature review and case study involving a professional singer. *National Center for Voice and Speech Status and Progress Report, 3,* 157–174.

Watterson, T., McFarlane, S. C., & Diamond, K. L. (1993). Phoneme effects on vocal effort and vocal quality. *American Journal of Speech Language Pathology, 2,* 74–78.

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

Wright, H. N., & Colton, R. H. (1972, November). *Some parameters of autophonic level.* Paper presented at the Annual Convention of the American Speech and Hearing Association, San Francisco.

Young, M. A. (1981). Articulation effort: Transitivity and observer agreement. *Journal of Speech and Hearing Research, 24,* 224–232

Received August 12, 2005

Accepted November 8, 2005

DOI: 10.1044/1092-4388(2006/048)

Contact author: Roger J. Ingham, Department of Speech and Hearing Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106. E-mail: rjingham@speech.ucsb.edu