Frequency Discrimination Deficits in People With Specific Language Impairment: Reliability, Validity, and Linguistic Correlates

G. M. McArthur
D. V. M. Bishop
University of Oxford,
Oxford, England

The reliability and validity of a frequency discrimination (FD) task were tested in 16 people with specific language impairment (SLI) and 16 people with normal spoken language (controls). The FD thresholds of the 2 groups indicated that FD thresholds for 25-ms and 250-ms tones were remarkably stable across 18 months. The FD thresholds were lower for control listeners than for listeners with SLI for both duration conditions, and the FD thresholds for both groups of listeners were lower for 250-ms tones than for 25-ms tones. Moreover, the FD thresholds were influenced little by nonperceptual, task-related abilities (e.g., paired-associative learning, memory for temporal order, sustained attention, and control of attention) of the listener groups. The significant group difference between the mean FD thresholds of the SLI and control groups was explained by a subgroup of people with SLI who had particularly poor thresholds compared with those of controls and the majority of the SLI group. This subgroup did not differ from the remainder of the SLI sample in terms of age or nonverbal ability but was characterized by very poor reading that was associated with poor phonemic awareness.

KEY WORDS: specific language impairment (SLI), frequency discrimination, psychoacoustics, language, reading

The speed and facility with which children learn their native language is impressive. Nevertheless, some children are late to achieve language milestones and may persist in using simplified grammar and vocabulary well into the school years. These children, who are also at risk for developing reading problems (Catts, 1993), are said to have a specific language impairment (SLI). The nature of SLI is remarkable because there is no obvious factor in a child's history that can explain their selective problem with language. They have nonverbal cognitive abilities in the normal range, their hearing appears to be adequate when assessed with traditional techniques, and there are no physical or environmental abnormalities that could account for their difficulties.

One theory explains SLI as a consequence of a reduced ability to process sounds that occur briefly or rapidly in time (Tallal, 2000). This "rapid auditory processing deficit" may distort the perception of speech sounds, which requires listeners to discriminate and classify spectrally complex sounds that change rapidly in time. This distortion may interfere with the development of stable neural representations of the basic speech sounds in a language (phonemes), impairing both the

comprehension and production of speech (Tallal, 2000). Interestingly, poor rapid auditory processing has also been held accountable for specific reading disability (i.e., developmental dyslexia; Farmer & Klein, 1995). To read fluently, a child must learn to segment words into phonemic units, which may not be fully formed if a rapid auditory processing deficit has compromised the development of phonemic representations.

Although the rapid auditory processing deficit account of SLI and dyslexia has received support from many studies, it is not universally accepted because some studies have failed to find evidence for a rapid auditory processing deficit in people with SLI and dyslexia. After reviewing this literature, McArthur and Bishop (2001) suggested that one reason for these mixed findings may be that poor performance on rapid auditory processing tasks arises from an impairment other than rapid auditory processing. Furthermore, we proposed that studies using rapid auditory processing tasks that stressed this impairment were more likely to find a rapid auditory processing deficit in people with SLI than were studies that used tasks that did not stress this impairment. We also suggested that one likely candidate for this alternative impairment was a frequency discrimination (FD) deficit, which is an impaired ability to distinguish between sounds that differ in frequency. Support for the idea that people with SLI have an FD deficit rather than a rapid auditory processing deficit comes from four sources:

1. In common with other researchers (e.g., Helzer, Champlin, & Gillam, 1996; Rosen, 1999), we have not found reliable group differences between listeners with SLI and listeners with normal spoken language (controls) on tests of rapid auditory processing (Bishop, Carlyon, Deeks, & Bishop, 1999; McArthur & Bishop, 2004).

2. A number of studies have found that at least some people with SLI or dyslexia show deficits in FD even when the rapid auditory processing demands of the task are minimal (Ahissar, Protopapas, Reid, & Merzenich, 2000; Baldeweg et al., 1999; Cacace, McFarland, Ouimet, Schreiber, & Marro, 1999; De Weirdt, 1988; France et al., 2002; Hari, Saaskilahti, Helenius, & Uutala, 1999; Hill et al., 1999; Korpilahti, 1995; McAnally & Stein, 1996; McArthur & Bishop, 2004). Of particular note is a study by Wright et al. (1997), which has been widely cited as supporting the rapid auditory processing deficit account of SLI on the grounds that children with SLI showed a striking deficit on a test of auditory backward detection masking. However, much less attention has been given to the other finding of this study, which was that the poor thresholds of children with SLI on a backward masking task that used a spectrally "notched" masking noise reached control levels when the width of

the spectral notch was increased. As Wright et al. pointed out, this indicated that the children with SLI "did not display a general deficit in the perception of rapidly presented sounds" and that they were "less able than controls to take advantage of a frequency separation between the tone and noise to aid detection of the tone" (p. 177).

We came to a similar conclusion in a study that asked people with SLI and control listeners to make an FD judgment (which one of two tones was higher in pitch) when each tone was followed by a masking sound. The masker was presented after a silent interstimulus interval (ISI) that differed in each of four conditions (i.e., 20, 50, 150, or 300 ms; McArthur & Bishop, 2004). We predicted that people with SLI would do particularly poorly in the backward masking conditions that used a short ISI (i.e., 20 and 50 ms). In fact, we found that a subset of people with SLI did have poor FD thresholds at short ISIs. However, these people also had poor FD thresholds in a control condition that presented tones without a masker. This result strengthens the case for a deficit in FD rather than rapid auditory processing in at least a subset of people with SLI.

3. The classic task used to assess rapid auditory processing, the Rapid Perception Test, relies heavily on FD. In this task, the listener has to recreate the order of pairs of high and low tones (i.e., high-low, low-high, high-high, low-low) using appropriate key presses. The silent ISI that separates the tones in each pair defines the rate of stimulus presentation: Short ISIs mean that the tones are presented rapidly whereas long ISIs mean that the tones are presented slowly. A number of studies have found that children with SLI or dyslexia perform worse than controls when the tones are presented rapidly but not when the tones are presented slowly (e.g., Farmer & Klein, 1995; Reed, 1989; Tallal & Piercy, 1973; Tallal, Stark, Kallman, & Mellits, 1981). This result has been interpreted as evidence for a specific rapid auditory processing deficit in children with SLI or dyslexia. However, more recent studies have found that these children discriminate tones presented slowly just as poorly as tones presented rapidly (Bishop, Bishop, et al., 1999; Cacace et al., 1999; France et al., 2002; Lincoln, Dickstein, Courchesne, Elmasian, & Tallal, 1992; Marshall, Snowling, & Bailey, 2001; Nittrouer, 1999; Share, Jorm, MacLean, & Matthews, 2002; Waber et al., 2001). Furthermore, some children with SLI or dyslexia have been excluded from studies using the Rapid Perception Test because they were unable to make the appropriate response to tones even when these stimuli were presented individually in practice trials (Bishop, Bishop, et al., 1999; Breier, Gray, Fletcher, Foorman, & Klaas, 2002; Heath, Hogben, & Clark, 1999; Reed, 1989; Tallal et al., 1981). These findings suggest that, in some cases,

poor scores on the Rapid Perception Test may be a product of poor FD rather than a rapid auditory processing deficit (Heath et al., 1999; McArthur & Bishop, 2001).

4. Finally, there is a theoretical basis for an "FD deficit" account of SLI. The speech signal is composed of rapid changes in spectral information across time. If a person is less able to discriminate between the frequencies of sounds, then he or she may have problems discriminating the spectral patterns of certain speech sounds. This limitation could affect the development of neural representations for the different phonemes in a language. In turn, an FD deficit could impair the comprehension and production of phonemes, which is a characteristic of SLI. Skeptics may argue that an FD deficit alone could not impair speech perception, because adults are able to perceive speech even when many spectral cues are absent (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). However, adults have well-established neural representations of phonemes to use as "templates" to interpret the degraded speech signal. We cannot assume that children, who are still formulating the phonemic templates of their language, can cope as well as adults without all the available speech cues.

Given our interest in FD deficits in people with SLI, the first issue we addressed in this experiment was whether our previous finding of an FD deficit in a subgroup of people with SLI (McArthur & Bishop, 2004) was reliable over time. We tested this by reassessing most of the people from our previous study to see if they had the same FD thresholds 18 months later.

The second issue addressed in this study was the validity of our initial findings. One factor we needed to consider was whether the poor FD thresholds in our original study arose because we used relatively brief (25 ms) tones. Children with SLI have been found to perform particularly poorly on the Rapid Perception Test with very short stimuli (Tallal & Piercy, 1973). This result could arise if children with SLI had poor rapid auditory processing associated with a slower rate of perceptual sampling of sounds. Fewer samples of a stimulus would leave them with less information to detect or discriminate short sounds. Thus, we administered another FD task to the SLI and control groups that used 250-ms tones rather than 25-ms tones.

A further factor we evaluated was the impact that nonperceptual, task-related factors had on FD thresholds. In the traditional two-interval, two-alternative, forced-choice (2I-2AFC) task used by McArthur and Bishop (2004), the participant had to (a) attend to a "low" tone (600 Hz) and a "high" tone (higher than 600 Hz), (b) create a memory of the tones, (c) compare them, (d) recall the temporal order of the tones, and then (e) access the verbal concepts of "high" and "low" to report the interval that contained the higher tone. Any of these

steps could pose problems for people with SLI. In the present study, we tested the SLI and control groups for four task-related factors that could affect performance on FD tasks: paired-associative learning, memory for temporal order, sustained attention, and control of attention.

The third issue we addressed was the linguistic profile of people who have poor FD. McArthur and Bishop (2004) found that one third of the listeners with SLI had poor FD thresholds compared with controls, which suggested that only a subset of people with SLI have auditory difficulties. If these difficulties are causally linked to language or reading problems, then we would expect to see relations between FD thresholds and scores on measures of spoken language, reading, and phonological processing. Accordingly, we administered a battery of language and literacy tasks to our participants to explore these associations.

Method

All methods were approved by the Ethics Committee of the Department of Experimental Psychology at the University of Oxford. Informed consent was obtained from each listener and his or her parent or guardian to participate in the research.

Participants

Eleven people with SLI and 13 controls who participated in the study 18 months previously (McArthur & Bishop, in press) agreed to return for the present experiment. Their data were used to test the long-term reliability of FD thresholds. To boost the power of the analyses for the rest of the study, we recruited a further 5 people with SLI and 3 controls so that each group comprised 16 people.

The people with SLI (7 females) ranged in age from 12 to 21 years and were recruited from language development centers and support groups around England. The people with normal spoken language skills (9 females) were matched for age and nonverbal IQ to the people with SLI and were recruited from scout and guide groups, a college, and a high school in Oxfordshire. All participants had nonverbal IQ scores within the average range; had no reported auditory, physiological, or neurological problems; and had good hearing sensitivity for a 750-Hz tone (i.e., were able to detect its presence at 20 dB HL). The people with SLI scored more than 1 standard deviation below the level expected for their age on at least two of four key spoken language tests (see below). Control participants scored within the average range on at least three of the four spoken language tests. Group statistics are illustrated in Table 1.

Table 1. Mean age, Nonverbal IQ, and spoken language scores of the SLI and control groups.

Variable	SLI (n = 16)			Control (n = 16)			Comparison	$\eta^{2\,a}$
	M	SD	Range	M	SD	Range		
Age	15.82	2.70	12–20	15.02	2.58	12–21	$t(30) = 0.85, p = .40$	.02
Nonverbal IQ	91.87	13.36	75–113	97.50	9.70	75–113	$t(30) = 1.36, p = .18$	.06
BPVS	79.12	15.88	56–110	109.94	9.68	93–128	$t(30) = 6.63, p < .001^b$	.59
Figurative language	4.38	1.78	3–9	11.12	3.05	7–16	$t(30) = 7.64, p < .001^b$	.66
Recreating sentences	4.31	2.15	3–6	8.43	2.13	6–13	$t(30) = 5.45, p < .001^b$	.50
Recalling sentences	4.25	1.34	3–10	10.25	1.98	6–13	$t(30) = 10.02, p < .001^b$	.77

Note. SLI = specific language impairment; BPVS = British Picture Vocabulary Scale, Long Form.

[a]Effect size is represented by eta squared (η^2), which is calculated in the same way as r^2 but is reported as η^2 when one variable is nonlinear.
[b]$p < .008$ (Bonferroni correction for six analyses).

Nonverbal Intelligence Test

Nonverbal intelligence was assessed with the Standard Progressive Matrices (Raven, Raven, & Court, 1998). In each of 60 items, the participant was presented with a large pattern with a missing section. Below the large pattern were six sections with different patterns. The participant had to indicate which one of the six sections matched the missing section of the large pattern. Performance was expressed as age-standardized scores with a mean of 100 and a standard deviation of 15. Norms for young adults were not provided, so the scores of the people who were 17 years or older were calculated from the norms provided for the oldest age group (i.e., 16-year-olds).

Spoken Language Tests

We assessed spoken language with four standardized tests of receptive and expressive semantics and syntax that are widely used in the United Kingdom.

British Picture Vocabulary Scale (BPVS), Long Form (Dunn, Dunn, Whetton, & Pintilie, 1982)

In this test of receptive semantics, participants indicated which of four pictures matched a word spoken by the tester. Scores were expressed as age-standardized scores with a mean of 100 and a standard deviation of 15. The scores of people aged 18 years or older were based on the norms provided for 17-year-olds.

Recalling Sentences

Recalling Sentences is a test of expressive syntax and is a subtest of the Clinical Evaluation of Language Fundamentals–Revised (Semel, Wiig, & Secord, 1987). It uses 26 sentences of increasing difficulty, each one of which the participant repeats after presentation by the examiner. Age-standardized scores had a mean of 10 and a standard deviation of 3. The standard scores of people older than 17 years were calculated from the norms provided for 16-year-olds.

Recreating Sentences

Recreating Sentences is a test of expressive semantics and syntax, and is a subtest of the Test of Language Competence–Expanded Edition (TLC-EE; Wiig & Secord, 1989). It is composed of 13 items that presented a picture and three printed words. The participant's task is to create a sentence that was relevant to the picture and contained the three words in any order. Standard scores had a mean of 10 and a standard deviation of 3. The scores of people who were 19 years or older were taken from the norms provided for 18-year-olds.

Figurative Language

Another subtest of the TLC-EE, Figurative Language, is a test of receptive semantics, which is composed of 12 two-part items. In the first part, the participant interpreted a situation that had been described by the experimenter. In the second part, the participant selected one of four written expressions that best reflected the meaning of the same situation. Standard scores had a mean of 10 and standard deviation of 3. The scores of people older than 19 were taken from norms provided for 18-year-olds.

Reading Tests

Reading was not used as a criterion for inclusion in the study. However, we tested three types of reading in the SLI and control listeners, as previous findings suggested that reading skill may differentiate people with SLI with and without auditory processing deficits (McArthur & Bishop, 2004).

Nonword Reading

The ability to read nonwords (i.e., nonsense words such as *pag*) is an index of phonological decoding of unfamiliar words (i.e., the ability to transform the visual symbol of the nonword *pag* into the corresponding speech sounds of /p/, /a/, /g/). Nonword reading was assessed with the Martin and Pratt Nonword Reading Test (Martin & Pratt, 1999), which is composed of 54 increasingly difficult nonwords that were read aloud until eight consecutive incorrect responses were made. Standard scores had a mean of 100 and standard deviation of 15. Scores for people 17 years or older were calculated from the norms provided for the oldest age group (16;0 [years; months] to 16;11).

Regular Word Reading

The ability to read regular words (i.e., words that follow letter-to-sound conversion rules, such as *pat*) is an index of phonological decoding of familiar words (i.e., the ability to transform the visual symbol of the word *pat* into the corresponding speech sounds of /p/, /a/, /t/). We tested this using a list of 30 regular words that were modified by Coltheart and Leahy (1996) and by Edwards and Hogben (1999) from Castles (1993). The words were presented on cards that the participant shuffled and then read aloud at his or her own pace. Scores were the number of regular words read correctly out of 30.

Irregular Word Reading

The ability to read irregular words (i.e., words that do not follow letter-to-sound conversion rules, such as *yacht*) is an index of whole-word reading (e.g., the ability to transform the visual symbol of the word *yacht* into the corresponding speech sound that sounds like "yot"). We tested this using 30 irregular words, which were also modified by Coltheart and Leahy (1996) and Edwards and Hogben (1999) from Castles (1993) and were presented on cards that the participant shuffled and then read aloud at his or her own pace. Scores were the number of irregular words read correctly out of 30.

Phonological Processing Tests

People with SLI and controls were tested for their phonological processing skills because an FD deficit account of SLI predicts that impaired FD gives rise to SLI by impairing the comprehension and production of phonemes.

Phoneme Deletion

The Phoneme Deletion test (Bruck, 1990) taxes phonemic awareness, memory, and retrieval. This test is composed of 24 items in which the participant repeats a nonword after removing the first or last sound. For example, if instructed to remove the last sound of *bink*, the correct response would be *bin*. Scores were the number of items correct out of 24.

Phoneme Discrimination

Phoneme discrimination is the ability to discriminate between words or nonwords that differ in one (or more) of their component speech sounds (e.g., *cab* versus *cad*). This was tested with the Alliteration Test With Pictures, which is a subtest of the Phonological Assessment Battery (PhAB; Frederickson, Frith, & Reason, 1997). The task was split into two halves, each presenting two practice items and five test items. Each item was composed of three words that the tester read aloud while pointing to the appropriate picture on a card. The participant had to report, either verbally or by pointing to the pictures, which one of two words started with the same sound. Because this test only had norms for children from ages 6 to 9 years, raw scores (out of 10) were used to represent performance.

Naming Speed Test for Pictures

Semantic retrieval, phoneme retrieval, and phoneme production abilities were assessed with another subtest of the PhAB, the Naming Speed Test for Pictures. The participant named a series of 50 pictures printed on a card (5 pictures repeated 10 times in random order) as quickly as possible. Standardized scores (with a mean of 100 and standard deviation of 15) were calculated from the time taken to name all 50 items. Scores for participants older than 15 were taken from norms provided for the oldest age group (i.e., 14;11).

Naming Speed Test for Digits

The Naming Speed Test for Digits test assessed phonemic retrieval and production in a manner similar to the Naming Speed for Pictures, except that the participant had to name a series of single digits presented on a card (the numbers 1 to 9 repeated five or six times in random order). Because digits are overlearned stimuli, this task placed fewer demands on semantic retrieval than Naming Speed for Pictures, making it a purer test of phonemic retrieval and production.

Frequency Discrimination Tasks

Listeners completed three FD tasks. One task was the same 2I-2AFC task used by McArthur and Bishop (2004) as a control task. Another FD task was a three-interval, two-alternative forced-choice task, which, like the 2I-2AFC task, presented 25-ms tones (3I-2AFC-25). However, unlike the 2I-2AFC task, the 3I-3AFC-25 task

was designed to minimize the influence of verbal labeling and memory for the temporal order of sounds. The third FD task was the same as the 3I-2AFC-25 task except that it presented 250-ms tones (3I-2AFC-250 task). Longer tones were used to address our concern that performance on the 2I-2AFC and 3I-2AFC-25 tasks may be confounded by the brevity of the 25-ms tones, which may pressure rapid auditory processing by requiring the formation of an auditory representation in a short period of time. These tasks are described in more detail below.

The 2I-2AFC Task

Each trial of the 2I-2AFC task presented two sinusoidal tones that were separated by a 500-ms silent interval. The two stimulus intervals were represented visually on a computer monitor by a pair of square buttons that flashed when the respective tone was played (marked "1" for the first tone and "2" for the second tone). The standard tone had an intensity of 80-dB SPL, a duration of 25 ms (including 2.5-ms onset and offset), and a frequency of 600 Hz. The signal tone was the same except that it had a higher frequency. The listener's task was to use a computer mouse to click on the button on the monitor that represented the interval containing the higher-frequency signal tone. A correct response was rewarded by a colored "thumbs-up" sign on the monitor. An incorrect response prompted an uninteresting black cross.

The higher frequency of the signal tone was adjusted over 60 trials using a 3-down, 1-up tracking algorithm (Levitt, 1971) to measure 79% correct identification of that tone. The frequency of the signal tone was initially set at 700 Hz (the ceiling value was 800 Hz) and was adjusted in 25-Hz steps. These steps were reduced to 5 Hz after the first four reversals in response adjustment. The FD threshold estimate was based on the mean frequency of the signal tone calculated from the last even number of the small step-size reversals. This threshold estimate was considered valid only if the listener's performance was stable after the fourth reversal in response adjustment (i.e., the midpoints between the reversal points formed a flat line, and the final 4–6 reversal points were less than 20 Hz away from the estimated threshold point). Acceptable FD threshold values were transformed into logarithmic units to normalize the distribution for statistical analyses.

The 3I-2AFC-25 Task

FD thresholds for 25-ms tones were also measured in 3I-2AFC-25 trials, using the same standard tones as the 2I-2AFC task. Each trial was composed of three consecutive tones, each tone presented in an interval separated by a 500-ms period. For all 3I-2AFC-25 trials, the second interval, which was visually cued by a

"jumping" dinosaur on the computer monitor, always contained the standard tone. The standard tone was also presented with equal likelihood in either the first interval (cued by a yellow ball to the left of the dinosaur) or the third interval (cued by a red ball to the right of the dinosaur) of the trial. The remaining interval contained the signal tone with a frequency higher than 600 Hz. The listener used the computer mouse to click on the ball that represented the interval containing the signal tone. A correct response was rewarded with a small token on the computer screen that was stacked in a column. When the column reached the top of the screen, the listener was rewarded by the dinosaur and the two balls jumping to a short jingle.

The frequency of the higher tone was adjusted between trials using the parameter estimation by sequential tracking (PEST) procedure (Taylor & Creelman, 1967). This algorithm tracked 79% correct identification of the signal tone to achieve the same threshold point as that in the 2I-2AFC task. The higher frequency of the signal tone was initially set at 1000 Hz (the ceiling value) and was adjusted in 80-Hz steps. These steps were reduced to 2 Hz after the first six response reversals. The task was finished once the listener either made 10 response reversals or completed 60 trials, whichever came first. The threshold estimate was based on the mean frequency of the higher tone after the sixth response reversal. A threshold estimate was accepted if the listener's performance was stable after the sixth reversal (i.e., the midpoints between the reversal points formed a flat line, and the final 4–6 reversal points were close to the threshold point). FD threshold estimates were again transformed into logarithmic units to normalize their distribution for statistical analyses.

As mentioned above, the 3I-2AFC-25 task was designed to minimize the influence of verbal labeling and memory for the temporal order of sounds. In contrast to the 2I-2AFC task, the 3I-2AFC-25 task minimized verbal labeling by asking the listener to identify the interval that contained the tone that sounded different from the standard tone (i.e., the second tone). This task was easier than making a relational judgment of "high" or "low" that is required in the 2I-2AFC task (Bishop, Carlyon, et al., 1999; Maxon & Hochberg, 1982; Thompson, Cranford, & Hoyer, 1999). The 3I-2AFC-25 task also reduced the load on memory for temporal order by requiring relative judgments about the stimuli (i.e., were the two tones different?) rather than both relative and absolute judgments about the stimuli (i.e., were the tones different, and if so, then were they high then low, or low then high?). It could do this in two ways. First, the listener may compare the second tone to the first and judge whether they were the same or different and then compare the third tone to the second and judge whether they

were the same or different. Listeners could then decide which pair sounded more different and make their response accordingly. Alternatively, a listener might simply compare the second tone to the first and judge whether they were the same or different. Either approach would be less taxing of memory for the temporal order of the sounds than the 2I-2AFC task because the listener in the 3I-2AFC-25 task would not have to identify which interval contained the "high" tone. Instead, the listener just had to detect whether there was a difference. Another advantage of the 3I-2AFC-25 task was that it used a child-friendly, gamelike format that held the attention of the listeners better than the 2I-2AFC task.

The 3I-2AFC Task With 250-ms Tones (3I-2AFC-250)

The 3I-2AFC-250 task was the same as the 3I-2AFC-25 task, except that the duration of each tone was 250 ms rather than 25 ms. As mentioned above, we used longer tones to ascertain whether the 2I-2AFC and 3I-2AFC-25 threshold estimates of FD were confounded by using short stimuli (i.e., 25-ms), which could place extra pressure on rapid auditory processing ability.

Nonperceptual, Task-Related Abilities

To address our concern that poor FD estimates may stem from poor task-related abilities rather than a bona fide FD deficit, we tested listeners for four task-related abilities: memory for temporal order, paired-associated learning, sustained attention, and control of attention.

Auditory Temporal Order Task

Auditory temporal order judgment was measured with a task that tested how well a listener remembered the temporal order of two sounds that were easy to discriminate on the basis of many cues. Each trial was composed of two stimulus intervals and a probe interval. The two stimulus intervals were visually represented by a blue or green dinosaur that appeared on a computer screen. One dinosaur emitted a "woof-woof" sound, and another dinosaur emitted a "growl." The woof-woof and growl sounds were acoustically complex and long (the woof-woof was 400 ms, and the growl was 700 ms), making them easy to discriminate. The two sounds were followed by a probe interval that presented either the woof-woof or the growl sound alone (i.e., without the dinosaur). The listener had to use two buttons on the keyboard to indicate as quickly and accurately as possible whether the probe sound was emitted by the first or second dinosaur. The task comprised eight practice trials followed by 10 trials of each of four trial types (growl-woof with growl probe; growl-woof with woof

probe; woof-growl with growl probe; woof-growl with woof probe) presented in random order (i.e., random selection without replacement). Scores were number and mean reaction time of correct and incorrect trials.

Auditory Paired-Associative Learning Task

The ability to associate a stimulus with a response was measured with a choice reaction time paradigm. This paradigm involved two tasks: a simple reaction time task and a choice reaction time task. Both tasks comprised 20 practice trials and 50 experimental trials. Each trial presented an 80-dB SPL, 250-ms tone that had a frequency of 600 Hz (low) or 1000 Hz (high), with frequency varied randomly between trials. These tones were long enough and different enough to be easily discriminated and were comparable to the tones used in the FD tasks.

In the simple reaction time task, the listener held the mouse with their nonpreferred hand so that the two mouse keys were on the right-hand side (i.e., the mouse was rotated 90° to the right). They then used the index finger of their preferred hand to click on the top key as quickly and accurately as possible when they heard a tone, regardless of whether it was high or low. In the choice reaction time task, the listener used same index finger to press the top mouse button if a tone was high (1000 Hz) and the bottom button if a tone was low (600 Hz). Because performance on the choice reaction task relied on simple reaction time plus choice reaction time, choice reaction time alone was calculated by subtracting simple reaction time values from choice reaction time values (i.e., number and mean reaction time of correct and incorrect trials.)

Sustained Attention

Sustained attention to auditory stimuli was measured with the !Score subtest of the Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson, & Nimmo-Smith, 1999), which has norms for people aged 6 to 16 years. !Score comprised two practice trials and 10 experimental trials, each comprising 9 to 15 "laser gun" noises that were separated by long, irregular silences. The participant counted the number of noises heard. Because the task was protracted and uninteresting, it taxed the ability to sustain attention to the task. Standard scores had a mean of 10 and standard deviation of 3. The scores of people older than 16 were taken from the norms provided for the oldest age group (15;0 to 15;11).

Control of Attention

Another subtest of the TEA-Ch (Manly et al., 1999), Control of Attention was measured over two phases. The

Figure 1. Associations between new (present study) versus old (previous experiment; McArthur & Bishop, 2004) two-interval, two-alternative, forced-choice (2I-2AFC) frequency discrimination (FD) thresholds on log-log frequency coordinates. SLI = specific language impairment.
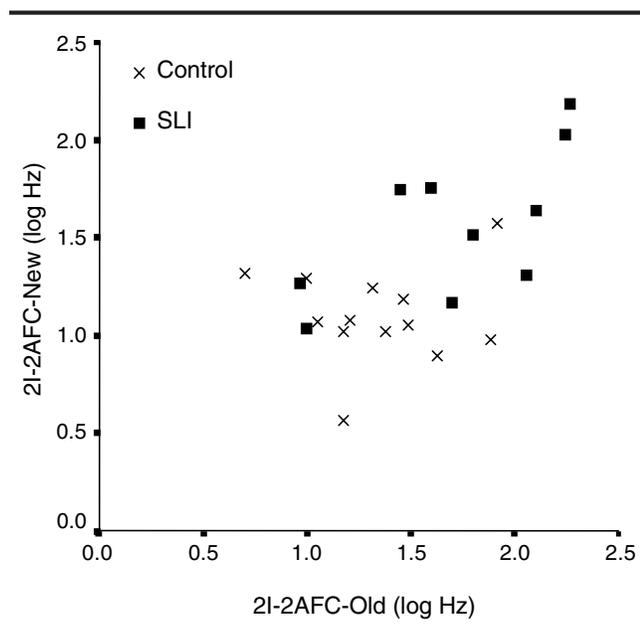
first part, called Same World, presented a series of boxes in a meandering path that were each labeled with either a "1" or "2." Standard scores, which had a mean of 10 and standard deviation of 3, were calculated from the time taken to name all the digits in order as fast as possible. Scores of people older than 16 were calculated from the norms provided for the oldest age group (15;0 to 15;11).

The second part of the test, called Opposite World, was the same as the Same World subtest, except that

the participants had to reverse their response (i.e., report "1" when they saw a 2 and report "2" when they saw a 1). Again, standard scores were calculated from the time taken to report all the numbers in the path. The authors of this test note that the "speed with which children can perform this cognitive reversal is the crucial measure" (Manly et al., 1999, p. 7). However, they do not provide standard scores for the amount of extra time taken to complete the Opposite World compared with the Same World subtests, so we subtracted each person's Same World standard score from their Opposite World standard score to get a measure of their control of attention required for the cognitive reversal.

Results

Reliability of the 2I-2AFC Thresholds

The individual 2I-2AFC FD thresholds of the 11 people with SLI and 13 controls in the present and previous studies are compared in Figure 1 (2I-2AFC-new as a function of 2I-2AFC-old, respectively). Higher thresholds represent poorer FD. The group mean thresholds are compared in Table 2. The SLI group was missing a data point because 1 listener had an invalid 2I-2AFC-old threshold (i.e., the midpoints between the reversal points did not form a flat line, and the final 4–6 reversal points were more than 20 Hz away from the estimated threshold point). A criterion of $p < .05$ was used in all analyses unless a Bonferroni correction was required.

The Pearson r correlation coefficient between the 2I-2AFC-new and 21-2AFC-old FD thresholds evaluated across the two groups was .61 ($p = .002$). This moderately strong correlation indicates that the 2I-2AFC

Table 2. 2I-2AFC-Old, 2I-2AFC-New, 3I-2AFC-25, 3I-2AFC-250, and mean FD thresholds (log Hz) of the SLI and control groups.

Variable	SLI			Control			Comparison	η^2 [a]
	M	SD	n	M	SD	n		
2I-2AFC-old	1.72	0.47	10	1.34	0.35	13	$t(21) = 2.22, p = .04$	.19
2I-2AFC-new	1.58	0.44	16	1.24	0.43	16	$t(30) = 2.23, p = .03$	.14
3I-2AFC-25	1.42	0.46	16	1.14	0.25	16	$t(30) = 2.17, p = .04$	.14
3I-2AFC-250	1.18	0.57	16	0.84	0.29	16	$t(30) = 2.20, p = .04$	.14
Composite mean FD	1.40	0.45	16	1.07	0.24	16	$t(30) = 2.54, p = .02$	.18
2I-2AFC-old vs. 2I-2AFC-new	$t(9) = 1.12, p = .29$			$t(12) = 1.15, p = .27$				
2I-2AFC-new vs. 3I-2AFC-25	$t(15) = 2.12, p = .05$			$t(15) = 1.06, p = .31$				
3I-2AFC-25 vs. 3I-2AFC-250	$t(15) = 3.35, p = .004$ [b]			$t(15) = 3.74, p = .002$ [b]				

Note. 2I-2AFC-old and 2I-2AFC-new = two-interval, two-alternative, forced-choice task in a previous study (McArthur & Bishop, 2003) and in the present study, respectively; 3I-2AFC-25 and 3I-2AFC-250 = three-interval 2AFC task with 25-ms and 250-ms tones, respectively; FD = frequency discrimination.

[a]Effect size is represented by eta squared (η^2).   [b]$p < .02$ (Bonferroni correction for three condition comparisons).

thresholds of the SLI and control groups were reasonably consistent over 1.5 years. In general, people with SLI who had poor FD thresholds in the previous experiment still had poor FD thresholds 18 months later.

The mean 2I-2AFC-old and mean 2I-2AFC-new FD thresholds of the SLI group were higher (1.72 and 1.58 log Hz, respectively) than those of the control group (1.34 log Hz and 1.24 log Hz, respectively; see Table 2), but these effects just failed to reach significance after the Bonferroni correction. Both the SLI and control groups produced similar thresholds on the 2I-2AFC-old and 2I-2AFC-new tasks (see Table 2), and these were not significantly different within the groups.

Validity of 2I-2AFC-New Thresholds

### 2I-2AFC-New Thresholds Versus 3I-2AFC-25 Thresholds

The Pearson r correlation coefficient between thresholds produced by all the people who participated in this study (i.e., 16 people with SLI and 16 controls) on the 2I-2AFC-new task and the 3I-2AFC-25 task (designed to minimize the influence of memory for temporal order and verbal labeling) was .67 ($p < .001$). This moderately strong, statistically significant, relationship suggests that 2I-2AFC-new thresholds may have been influenced somewhat by task-specific factors such as variations in memory for temporal order or verbal labeling but not enough to confound the validity of the 2I-2AFC-new thresholds as an acceptable index of FD. The general similarity of the two sets of thresholds, which were collected in the same session, also suggests that the different procedures of the 2I-2AFC-new and 3I-2AFC-25 tasks generally did not affect the pattern of results across the listeners.

Like 2I-2AFC-new thresholds, the mean 3I-2AFC-25 threshold of the SLI group was higher (i.e., 1.42 log Hz) than that of the control group (1.14 log Hz; see Table 2), but this difference failed to reach significance after the Bonferroni correction. There also was no significant difference between the mean 2I-2AFC-new and 3I-2AFC-25 thresholds for either the SLI or control groups (see Table 2).

3I-2AFC-25 Thresholds Versus 3I-2AFC-250 Thresholds

We compared individual FD thresholds on the 3I-2AFC-25 task with corresponding FD thresholds on the 3I-2AFC-250 task to determine the effect of stimulus duration on performance (see Figure 2). The Pearson r between 3I-2AFC-25 and 3I-2AFC-250 thresholds was .78 ($p < .001$). Both the SLI and control groups had significantly lower thresholds on the 3I-2AFC-250 task

Figure 2. Association between 25-ms versus 250-ms FD thresholds for the three-interval, two-alternative, forced-choice (3I-2AFC) tasks displayed on log-log frequency coordinates.
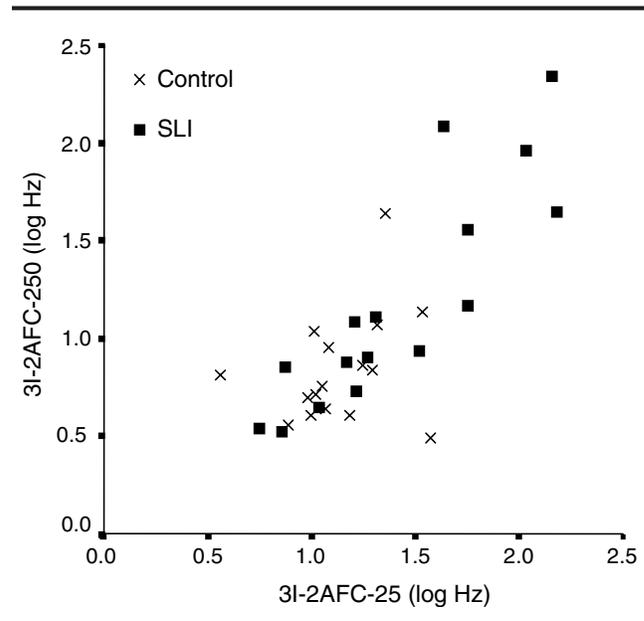

(1.18 and 0.84 log Hz, respectively) than on the 3I-2AFC-25 task (1.42 and 1.14 log Hz, respectively; see Table 2 for means). This finding indicates that the longer tones made the task easier for both groups. However, people with the poorest 3I-2AFC-25 thresholds also had the poorest 3I-2AFC-250 thresholds, which is evident in Figure 2. This indicates that poor 3I-2AFC-25 thresholds, although higher than for the 3I-2AFC-250 task, were not disproportionately influenced by the brevity of the 25-ms tones used in the task.

Like the 2I-2AFC-new and 3I-2AFC-25 thresholds, the 3I-2AFC-250 thresholds of the SLI group were higher (1.18 log Hz) than the corresponding thresholds of the control group (0.84 log Hz), but this difference just failed to reach statistical significance after the Bonferroni correction (see Table 2). Thus, increasing the length of the tones had a similar effect of reducing FD thresholds in both the SLI and control groups (i.e., there was no significant interaction).

Composite Mean FD Thresholds of the SLI and Control Groups

To obtain a more stable estimate of FD thresholds, we averaged the three FD threshold values for the 2I-2AFC-new, 3I-2AFC-25, and 3I-2AFC-250 tasks for each listener. This composite calculation seemed reasonable for two reasons: (a) the significant relations between 2I-2AFC-new and 3I-2AFC-25 thresholds ($r = .67$), 2I-2AFC-new and 3I-2AFC-250 thresholds ($r = .59$), and

Figure 3. Composite mean FD thresholds for individual listeners in the SLI and control groups.
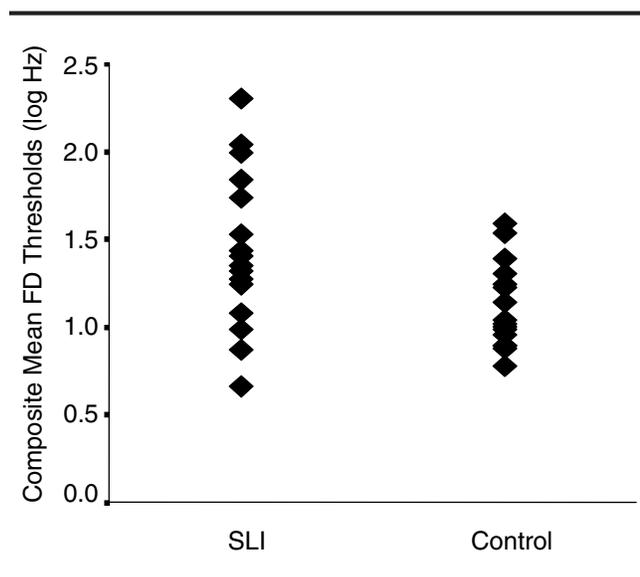
3I-2AFC-25 and 3I-2AFC-250 thresholds ($r = .78$) indicated similar response patterns for each task, and (b) we had no evidence to suggest that one task was a better measure of FD than any other. The resulting composite mean FD thresholds for the individual listeners in the SLI and control groups are shown in Figure 3.

There was a sizeable difference between the composite mean FD thresholds of the SLI (1.40 log Hz) and control groups (1.07 log Hz). The difference, which just failed to reach significance after a Bonferroni correction (see Table 2), is explained by 5 people with SLI who

had mean composite FD thresholds that fell above the 95% confidence limits of the control group's composite mean thresholds. Accordingly, we subdivided the SLI group into two: 5 people with SLI with poor FD (SLI-poor FD group; $M = 1.94$, $SD = 0.22$) and 11 people with SLI with normal FD (SLI-normal FD group; $M = 1.15$, $SD = 0.27$).

FD Thresholds, Age, Nonverbal IQ, and Task-Related Factors

To ascertain whether the poor composite mean FD thresholds of the SLI-poor FD group were a function of task-related factors, we compared the SLI-poor FD, SLI-good FD, and control groups for their age, nonverbal IQ scores, and performance on the auditory temporal order, auditory paired associative-learning, sustained attention, and control of attention tasks (see Table 3). There were no significant differences between the mean scores of the SLI-poor FD, SLI-normal FD, and control groups on the attention tests. There were only nonsignificant trends for group differences in age, nonverbal IQ, reaction time for correct responses on the temporal order and paired association tasks, and reaction time for errors on the paired association task. The mean reaction time for errors on the temporal order test was significantly greater (i.e., later) in the SLI-poor FD group, indicating that their incorrect answers were not due to impulsive responses. The SLI-poor FD group made significantly more errors on the temporal order and paired association tasks than the SLI-normal FD and control groups, although the error rate of the SLI-poor FD group was reassuringly low. Further analysis

Table 3. Age, Nonverbal IQ, task-related measures (temporal order, paired association, sustained attention, control of attention) in the SLI-Poor FD, SLI-Normal FD, and control groups.

Variable	SLI-Poor FD (n = 5)		SLI-Normal FD (n = 11)		Controls (n = 16)		Comparison	η^{2a}
	M	SD	M	SD	M	SD		
Age	13.99	1.43	16.64	2.78	15.02	2.59	$F(2, 29) = 2.30$, $p = .12$	.14
Nonverbal IQ	85.20	11.10	94.91	13.64	97.50	9.69	$F(2, 29) = 2.22$, $p = .13$	.13
Temporal order								
RT correct	1187.30	294.80	877.67	411.28	729.24	221.19	$F(2, 29) = 4.24$, $p = .02$	.23
RT errors	1665.62$_a$	426.18	743.33$_b$	322.47	624.61$_b$	390.40	$F(2, 20) = 13.80$, $p < .001^b$	.58
Errors	5.60$_a$	3.91	1.73$_b$	1.27	1.00$_b$	1.09	$F(2, 29) = 12.40$, $p < .001^b$	.46
Paired association								
RT correct	383.20	112.35	250.58	82.93	244.25	184.58	$F(2, 29) = 1.81$, $p = .18$	.11
RT errors	1429.69	310.13	1506.52	378.90	1342.83	280.30	$F(2, 24) = 0.70$, $p = .51$	.06
Errors	6.60$_a$	4.93	2.09$_b$	1.97	1.94$_b$	1.39	$F(2, 29) = 7.86$, $p = .002^b$	.35
Sustained attention	9.40	4.04	10.18	1.94	10.06	3.27	$F(2, 29) = 0.12$, $p = .88$	.01
Control of attention	−2.25	1.50	−0.18	2.99	0.50	2.13	$F(2, 29) = 2.07$, $p = .14$	.13

Note. Means with different subscripts differed significantly on a Scheffe test ($p < .05$). RT = reaction time; PA = paired association.

[a]Effect size is represented by eta squared (η^2).   [b]$p < .005$ (Bonferroni correction for 10 comparisons).

revealed that the number of errors on the temporal order task was significantly associated with nonverbal IQ (Pearson's $r = -.45$, $p = .01$) and performance on the Opposite World control of attention task ($r = -.57$, $p = .001$). Similarly, errors on the paired association task were significantly associated with control of attention ($r = -.45$, $p = .01$).

This pattern of results suggested that it was the temporal order and paired association tasks that were affected by task-related factors rather than the FD thresholds. We tested this idea by recomparing the SLI-poor FD, SLI-normal FD, and control groups for temporal order errors, paired association errors, and the mean FD thresholds while taking age, nonverbal IQ, and attention into account. Using a Bonferroni correction for four analyses ($p < .01$), we found a significant difference between the three groups for the mean FD thresholds, $F(2, 25) = 6.12$, $p = .007$, but not for the mean number of errors on the paired association and temporal order tasks: paired association, $F(2, 25) = 0.70$, $p = .51$; temporal order, $F(2, 25) = 5.09$, $p = .015$. We also found that the mean FD thresholds of the SLI-poor FD, SLI-normal FD, and control groups were still significantly different when the number of errors on both the temporal order and paired association tasks were accounted for simultaneously, $F(2, 27) = 6.43$, $p = .005$. Considered together, these findings indicate that, although temporal order and paired association may account for some variance in FD thresholds, this amount is too small to explain the poor FD thresholds of the SLI-poor FD group. In contrast, a significant amount of variance in temporal order and paired association scores may be explained by task-related factors such as nonverbal IQ and attention.

FD Thresholds and Reading, Phonological Processing, and Spoken Language Measures

The significant relationships between FD thresholds and the measures of reading and phonological processing across groups (see Table 4) indicated that poor FD thresholds were associated with poorer reading of nonwords, regular words, and irregular words, poorer phoneme discrimination, phoneme deletion, and rapid naming of pictures, and poorer spoken language ability. To ensure that these associations were not the product of the trends for group differences in age, nonverbal IQ, temporal order errors, and paired association errors, we recompared the reading, phonological processing, and spoken language scores of the SLI-poor FD, SLI-normal FD, and control groups while taking these factors into account (see Table 4).

Differences between the mean scores of the three groups were still significant for nonword reading, regular word reading, irregular word reading, phoneme discrimination, and spoken language. The differences between the mean phoneme discrimination and rapid digit naming scores of the groups were significant at $p < .05$, but they were not large enough to pass the stricter criterion of the Bonferroni correction. The dramatic reduction in the group effect on phoneme deletion

Table 4. Reading, phonological processing, and spoken language scores for the SLI-Poor FD, SLI-Normal FD, and control groups.

Variable	SLI-Poor FD (n = 5)		SLI-Normal FD (n = 11)		Controls (n = 16)		Comparison	η^{2a}
	M	SD	M	SD	M	SD		
Reading								
Nonword reading	70.60$_a$	15.06	87.73$_a$	17.67	110.38$_b$	11.07	$F(2, 29) = 17.98$, $p < .001^b$	.55
Regular word	19.80$_a$	5.72	27.54$_b$	2.91	30.00$_b$	0.00	$F(2, 29) = 26.66$, $p < .001^b$	.65
Irregular word	18.20$_a$	5.93	21.64$_a$	4.15	27.94$_b$	5.14	$F(2, 29) = 20.15$, $p < .001^b$	.58
Phonological processing								
Phoneme Deletion	12.80$_a$	7.98	20.27$_b$	3.69	22.50$_b$	1.63	$F(2, 29) = 12.06$, $p < .001^b$	.45
Phoneme Discrimination	7.60$_a$	1.82	9.64$_b$	0.92	10.00$_b$	0.00	$F(2, 29) = 14.84$, $p < .001^b$	.51
Naming Pictures	80.00$_a$	11.44	96.91$_{a,b}$	14.79	107.56$_b$	14.15	$F(2, 29) = 7.65$, $p = .002^b$	.34
Naming Digits	82.40	16.26	85.73	15.70	101.31	17.00	$F(2, 29) = 4.15$, $p = .03$	.22
Spoken language								
BPVS	82.00$_a$	7.50	77.65$_a$	18.65	109.94$_b$	9.68	$F(2, 29) = 21.77$, $p < .001^b$	.60
Figurative Language	4.20$_a$	1.79	4.45$_a$	1.86	11.12$_b$	3.05	$F(2, 29) = 28.24$, $p < .001^b$	.66
Recreating Sentences	3.80$_a$	1.30	4.55$_a$	2.46	8.44$_b$	2.13	$F(2, 29) = 14.77$, $p < .001^b$	.50
Recalling Sentences	4.20$_a$	2.68	5.00$_a$	1.67	10.88$_b$	2.45	$F(2, 29) = 29.91$, $p < .001^b$	.67

Note. Means with different subscripts differed significantly on a Scheffe test ($p < .05$).

[a]Effect size is represented by eta squared (η^2).   [b]$p < .004$ (Bonferroni correction for 11 comparisons).

is noteworthy. This outcome was due primarily to the strong association between phoneme deletion and temporal order error scores ($r = -.76$, $p < .001$), suggesting that the phoneme deletion task was linked to the ability to recall the order sounds.

The profiles of the SLI-poor FD and SLI-normal FD groups on the various tests differed on post hoc tests (see Table 5). Although both groups were equally impaired on the four spoken language tests relative to controls, the SLI-poor FD group was significantly worse than the SLI-normal FD group at reading nonwords and regular words and had poorer phoneme discrimination. Thus, the SLI-poor FD group was characterized by the poorest reading and phonological processing skills. Although the SLI-normal FD group had significantly lower mean reading scores than the control group, their nonword and regular-word reading scores were more similar to those of the control group than to the scores of the SLI-poor FD group and fell within the normal range (see Table 4). The SLI-normal FD group also had similar phoneme discrimination scores to those of the control group. However, like the SLI-poor FD group, the SLI-normal FD group did have poor irregular-word reading scores and poor rapid naming scores for digits, compared with controls.

Discussion

In a previous study, we found that 31% of people with SLI produced poor FD thresholds compared with those of controls and that the remaining 69% of the SLI group had normal FD scores (McArthur & Bishop, in press). Most of these people returned to be retested on the same FD task 1.5 years later and they produced remarkably similar FD thresholds. This finding indicates that there is a distinct subset of people within the SLI population who have poor FD thresholds. Moreover, their poor FD persists across time.

After recruiting new people to boost the size of the SLI and control groups, we compared FD thresholds on the original task (2I-2AFC-new) to a new FD task that reduced the influence of memory for temporal order and verbal labeling (the 3I-2AFC-25 task). There was no significant difference between the 2I-2AFC-new and 3I-2AFC-25 FD thresholds. This outcome, combined with a significant association between an individual's 2I-2AFC-new and 3I-2AFC-25 FD thresholds, suggested that both measures were valid reflections of FD.

We went on to compare thresholds on the 3I-2AFC-25 FD task to scores on the 3I-2AFC-250 FD task, which used tones of 250 ms rather than 25 ms. This comparison allowed us to evaluate whether poor FD thresholds for 25-ms tones could be explained by diminished processing due to fewer samples of the shorter tones. We found that people with SLI and control listeners both performed better with the long tones than with the short tones, presumably because they had more opportunities to sample the longer sounds. There was also a significant association between the FD thresholds for the 25-ms and 250-ms tones, so that people with poor

Table 5. Comparison of reading, phonological processing, and spoken scores of the SLI-Poor FD, SLI-Normal FD, and control groups controlling for age, Nonverbal IQ, temporal order, and paired association.

Measure	Comparison	SLI-Poor FD	SLI-Normal FD	Control	$\eta^{2\,a}$
Reading					
Nonword Reading[b]	$F(2, 26) = 11.67$, $p < .001$[c]	a	b	c	.47
Regular Word	$F(2, 25) = 14.23$, $p < .001$[c]	a	b	c	.53
Irregular Word	$F(2, 25) = 14.97$, $p < .001$[c]	a	a	b	.54
Phonological processing					
Phoneme Deletion	$F(2, 25) = 0.46$, $p = .64$				.04
Phoneme Discrimination	$F(2, 25) = 4.93$, $p = .02$	a	b	b	.28
Naming Pictures[b]	$F(2, 26) = 3.02$, $p = .07$				.19
Naming Digits[b]	$F(2, 26) = 3.49$, $p = .05$	a,b	a	b	.21
Spoken language					
BPVS[b]	$F(2, 26) = 28.43$, $p < .001$[c]	a	a	b	.69
Figurative Language[b]	$F(2, 26) = 23.39$, $p < .001$[c]	a	a	b	.64
Recreating Sentences[b]	$F(2, 26) = 10.93$, $p < .001$[c]	a	a	b	.46
Recalling Sentences[b]	$F(2, 26) = 18.78$, $p < .001$[c]	a	a	b	.59

Note. Groups with different letters differed significantly ($p < .05$).

[a]Effect size is represented by eta squared (η^2). [b]Age was not included as a covariate for these variables as they used normative scores that already accounted for age differences. [c]$p < .004$ (Bonferroni correction for 11 comparisons).

thresholds for 25-ms tones also tended to have poor thresholds for 250-ms tones. The poor FD thresholds of the SLI-poor FD group for both 25-ms and 250-ms tones suggested that they had impaired FD rather than a slower sampling rate.

We then compared each listener's composite mean threshold across the three FD tasks to their age, nonverbal IQ, and performances on four nonperceptual, task-related abilities (paired associative learning, memory for temporal order, sustained attention, and control of attention). Poor FD thresholds were not significantly associated with age, nonverbal IQ, or attention, but were associated with the number of errors made on the temporal order and paired association tasks. However, temporal order and paired association error scores themselves were associated with task-related factors (nonverbal IQ and attention). When these factors were controlled, indices for temporal order and paired association explained little of the variance in the FD thresholds. These findings are in line with results from Ollson, Bjorkman, Haag, and Juslin (1998) who found that performance on temporal order tasks was significantly associated with nonverbal IQ. The findings are also consistent with the data from Amitay, Ben-Yehudah, Banai, and Ahissar (2002), who reported that adults with dyslexia have larger deficits in discriminating sound frequency than temporal order. Thus, it would appear that the poor FD thresholds of our SLI-poor FD group were valid reflections of poor FD rather than problems with nonperceptual, task-related factors that have the potential to affect performance on psychoacoustic tasks.

On establishing the reliability and validity of our FD thresholds, we switched our attention to the linguistic profile of people who have poor FD. Our previous experiment (McArthur & Bishop, 2004) indicated that people with SLI and poor FD do not necessarily have the most severe SLI, but they do tend to have the poorest nonword reading. The present study supported these findings. The people with both SLI and poor FD had a similar degree of language impairment to the people with SLI with normal FD. However, the SLI-poor FD group had reading skills that were significantly weaker than those within the SLI-normal FD group. One route by which poor FD may affect literacy would be by compromising the development of phonemic categories. In line with this, we found significant associations between FD thresholds within our SLI sample and measures of phonemic awareness and literacy. However, it should be noted that the reciprocal relations between phonemic awareness and reading complicate the interpretation of these results. That is, although phonemic awareness is important for learning to read, learning to read triggers improvements in phonemic awareness (Bentin & Leshem, 1993). Given this inextricable connection, particularly

in people who have been exposed to print for some time, we tentatively offer the conclusion that poor FD affects literacy development by impairing development of phonological awareness.

An unexpected finding of this study was the distinctive linguistic profile of the people with SLI who had normal FD. Although their spoken language skills were as impaired as those of individuals with SLI and poor FD, the SLI-normal FD group had neither the same degree of phonemic awareness nor the phonological decoding problems. In fact, their only linguistic weakness, apart from impaired spoken language, appeared to be irregular-word reading. The profile of this group highlights the large individual differences in the linguistic skills of the SLI population and challenges the general assumption that most people with SLI have phonemic processing and phonological decoding deficits. More information is needed about people with SLI and normal phonological processing, who have occasionally been noted in the literature (e.g., Briscoe, Bishop, & Norbury, 2001; Catts, 1993; McArthur, Hogben, Edwards, Heath, & Mengler, 2000) but have not received much attention. In this vein, an intriguing group to examine would be persons who are "poor comprehenders," who also have normal phonological decoding despite certain spoken language difficulties (Nation, 2001).

More information is also needed about how specific the auditory processing deficit in people with SLI actually is (Cacace et al., 1999). Our previous finding that a subgroup of people with SLI performed significantly worse than did controls on a FD task, but not on auditory backward recognition masking tasks, suggests that the SLI deficit relates to spectral processing rather than to rapid temporal processing (McArthur & Bishop, in press). However, it is possible that people with SLI have more than one auditory processing deficit (e.g., a subgroup of adults with SRD, described by Amitay, Ahissar, and Nelken [2002], who reported poor performances on a variety of auditory tasks). Alternatively, people with SLI might have different types of auditory processing deficit. Either possibility could explain why people with SLI and dyslexia have been found to perform poorly on a variety of nonverbal auditory processing tasks.

Another issue that needs resolution is whether the deficit is limited to the auditory modality or whether it is a more general problem that occurs in other sensory modalities (Cacace et al., 1999; Cacace & McFarland, 1998; McFarland & Cacace, 1995). Because this study measured the reliability and validity of FD thresholds for tones alone, we do not know whether a discrimination deficit in people with SLI only affects their audition or whether it is a global sensory problem. Resolving this issue is imperative for understanding the nature

of auditory processing deficits in SLI and should be given high priority in future research.

As a final point, we must stress that poor FD was found in only one third of people with SLI. Our results emphasize the heterogeneity within SLI and suggest that, for progress to be made, we need to examine individual differences within SLI instead of relying on traditional methods of analysis that emphasize differences in group means.

Acknowledgments

References

Ahissar, M., Protopapas, A., Reid, M., & Merzenich, M. M. (2000). Auditory processing parallels reading abilities in adults. *Proceedings of the National Academy of Sciences, 97,* 6832–6837.

Amitay, S., Ahissar, M., & Nelken, I. (2002). Auditory processing deficit in reading disabled adults. *Journal of the Association for Research in Otolaryngology, 3*, 302–320.

Amitay, S., Ben-Yehudah, G., Banai, K., & Ahissar, M. (2002). Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. *Brain, 125,* 2272–2285.

Baldeweg, T., Richardson, A., Watkins, S., Foale, C., & Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology, 45*, 495–503.

Bentin, S., & Leshem, H. (1993). On the interaction between phonological awareness and reading acquisition: It's a two-way street. *Annals of Dyslexia, 43,* 125–148.

Bishop, D. V. M., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech, Language, and Hearing Research, 42,* 155–168.

Bishop, D. V. M., Carlyon, R. P., Deeks, J. M., & Bishop, S. J. (1999). Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech, Language, and Hearing Research*, *42,* 1295–1310.

Breier, J. I., Gray, L. C., Fletcher, J. M., Foorman, B., & Klaas, P. (2002) Perception of speech and nonspeech stimuli by children with and without reading disability and attention deficit hyperactivity disorder. *Journal of Experimental Child Psychology, 82,* 226–250.

Briscoe, J., Bishop, D. V. M., & Norbury, C. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 329–340.

Bruck, M. (1990). Word recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology, 26,* 439–454.

Cacace, A. T., & McFarland, D. J. (1998). Central auditory processing disorder in school-aged children: A critical review. *Journal of Speech, Language, and Hearing Research, 41,* 355–373.

Cacace, A. T., McFarland, D. J., Ouimet, J. R., Schreiber, E. J., & Marro, P. (1999). Temporal processing deficits in remediation-resistant reading-impaired children. *Audiology Neuro-otology, 5,* 83–97.

Castles, A. (1993). *Varieties of developmental dyslexia*. Unpublished doctoral dissertation, Macquarie University, Sydney, Australia.

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36,* 948–958.

Coltheart, M., & Leahy, J. (1996). Assessment of lexical and non-lexical reading abilities in children: Some normative data. *Australian Journal of Psychology, 48,* 136–140.

De Weirdt, W. (1988). Speech perception and frequency discrimination in good and poor readers. *Applied Psycholinguistics, 9,* 163–183.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Scale*. Windsor, U.K.: NFER-Nelson.

Edwards V. T., & Hogben, J. H. (1999). New norms for comparing children's lexical and nonlexical reading: A further look at subtyping dyslexia. *Australian Journal of Psychology, 51,* 37–49.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin and Review, 2,* 460–493.

France, S. J., Rosner, B. S., Hansen, P. C., Calvin, C., Talcott, J., Richardson, A. J., & Stein, J. F. (2002). Auditory frequency discrimination in adult developmental dyslexics. *Perception and Psychophysics, 64,* 169–179.

Frederickson, N., Frith, U., & Reason, R. (1997). *Phonological Assessment Battery*. Windsor, U.K.: NFER-Nelson.

Hari, R., Sääskilahti, A., Helenius, P., & Uutela, K. (1999). Non-impaired auditory phase locking in dyslexic adults. *NeuroReport, 10,* 2347–2348.

Heath, S. M., Hogben, J. H., & Clark, C. D. (1999). Auditory temporal processing in disabled readers with and without oral language delay. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 637–647.

Helzer, J. R., Champlin, C. A., & Gillam, R. B. (1996). Auditory temporal resolution in specifically language-impaired and age-matched children. *Perceptual and Motor Skills, 3,* 1171–1181.

Hill, N. I., Bailey, P. J., Griffiths, Y. M., & Snowling, M. J. (1999). Frequency acuity and binaural masking release in dyslexic listeners [Letter]. *Journal of the Acoustical Society of America, 106,* L53–L58.

Korpilahti, P. (1995). Auditory discrimination and memory

function in SLI children: A comprehensive study with neurophysiological and behavioural methods. *Scandinavian Journal of Logopedics and Phoniatrics, 20,* 131–139.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 73,* 663–667.

Lincoln, A. J., Dickstein, P., Courchesne, E., Elmasian, R., & Tallal, P. (1992). Auditory processing abilities in non-retarded adolescents and young adults with developmental receptive language disorder and autism. *Brain and Language, 43,* 613–622.

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The Test of Everyday Attention for Children.* Suffolk, U.K.: Thames Valley Test Co.

Martin, F., & Pratt, C. (1999). *Martin and Pratt Nonword Reading Test.* Melbourne, Australia: ACER.

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in dyslexic and normal readers. *Journal of Speech, Language, and Hearing Research, 44,* 925–940.

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behaviour: Sensitivity and discrimination. *Ear and Hearing, 3,* 301–308.

McAnally, K. I., & Stein, J. F. (1996). Auditory temporal coding in dyslexia. *Proceedings of the National Academy of Sciences, 263,* 961–965.

McArthur, G. M., & Bishop, D. V. M. (2001). Auditory perceptual processing in people with reading and oral language impairments: Current issues and recommendations. *Dyslexia, 7,* 150–170.

McArthur, G. M., & Bishop, D. V. M. (2004). Which people with specific language impairment have auditory processing deficits? *Cognitive Neuropsychology, 21,* 79–94.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the 'specifics' of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41,* 869–874.

McFarland, D. J., & Cacace, A. T. (1995). Modality specificity as a criterion for diagnosing central auditory processing disorders. *American Journal of Audiology, 4,* 36–48.

Nation, K. (2001). Reading and language in children: Exposing hidden deficits. *Psychologist, 14,* 238–242.

Nittrouer, S. (1999). Do temporal processing deficits cause phonological processing problems? *Journal of Speech, Language, and Hearing Research, 42,* 925–942.

Ollson, H., Bjorkman, C., Haag, K., & Juslin, P. (1998). Auditory inspection time: On the importance of selecting the appropriate sensory continuum. *Personality and Individual Differences, 25,* 627–634.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Standard Progressive Matrices.* Oxford, England: Oxford Psychologists Press.

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology, 48,* 270–292.

Rosen, S. (1999). Language disorders: A problem with auditory processing? *Current Biology, 9,* R698–R700.

Semel, E., Wiig, E. H., & Secord, W. (1987). *Clinical Evaluation of Language Fundamentals–Revised.* San Antonio, TX: The Psychological Corporation.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303–304.

Share, D. L., Jorm, A. F., MacLean, R., & Matthews, R. (2002). Temporal processing and reading disability. *Reading and Writing: An Interdisciplinary Journal, 15,* 151–178.

Tallal, P. (2000). Experimental studies of language learning impairments: From research to remediation. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 131–155). Hove, U.K.: Psychology Press.

Tallal, P., & Piercy, M. (1973). Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia, 11,* 389–398.

Tallal, P., Stark, R., Kallman, C., & Mellits, D. (1981). A re-examination of some nonverbal perceptual abilities of language-impaired and normal children as a function of age and sensory modality. *Journal of Speech and Hearing Research, 24,* 351–357.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America, 41,* 782–787.

Thompson, N. C., Cranford, J. L., & Hoyer, E. (1999). Brief-tone frequency discrimination in children. *Journal of Speech, Language, and Hearing Research, 42,* 1061–1068.

Waber, D. P., Weiler, M. D., Wolff, P. H., Bellinger, D., Marcus, D. J., Ariel, R., et al. (2001). Processing of rapid auditory stimuli in school-age children referred for evaluation of learning disorders. *Child Development, 72,* 37–49.

Wiig, E. H., & Secord, W. (1989). *Test of Language Competence–Expanded Edition.* San Antonio, TX: The Psychological Corporation.

Wright, B. A., Lombardino, L. D., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature, 387,* 176–178.