

The gating paradigm: Effects of presentation format on spoken word recognition by children and adults

AMANDA C. WALLEY, VICTORIA L. MICHELA, and DAPHNE R. WOOD
University of Alabama, Birmingham, Alabama

This study focused on the impact of stimulus presentation format in the gating paradigm with age. Two presentation formats were employed—the standard, successive format and a duration-blocked one, in which gates from word onset were blocked by duration (i.e., gates for the same word were not temporally adjacent). In Experiment 1, the effect of presentation format on adults' recognition was assessed as a function of response format (written vs. oral). In Experiment 2, the effect of presentation format on kindergarteners', first graders', and adults' recognition was assessed with an oral response format only. Performance was typically poorer for the successive format than for the duration-blocked one. The role of response perseveration and negative feedback in producing this effect is considered, as is the effect of word frequency and cohort size on recognition. Although the successive format yields a conservative picture of recognition, presentation format did not have a markedly different effect across the three age levels studied. Thus, the gating paradigm would seem to be an appropriate one for making developmental comparisons of spoken word recognition.

The aim of this study was to assess the appropriateness of the gating paradigm for making developmental comparisons of spoken word recognition—that is, to ascertain whether there are characteristics of the paradigm itself that differentially influence children's and adults' performance and thus limit general conclusions about the growth of recognition.

In the gating paradigm as it was introduced by Grosjean (1980), listeners are presented with increasing amounts of acoustic–phonetic input from word onset and asked to identify the target after each gate. This paradigm is consistent with the widespread theoretical view of recognition as a discriminative process (the view that words must be discriminated from various lexical alternatives; see Luce, 1986) and has proved useful in the illumination of key theoretical issues regarding this process in adults (e.g., the extent to which it is sequential, with greater perceptual weighting of word-initial vs. non-initial input; cf. Grosjean, 1985; Salasoo & Pisoni, 1985). Less attention has been paid to recognition in childhood, in part because of a shortage of tasks that are amenable for the study of younger listeners. Without such tasks, we cannot assess theoretical issues regarding

development, including the proposal that recognition becomes an increasingly discriminative process with vocabulary growth (see, e.g., Jusczyk, 1992; Walley, 1993). The gating paradigm is promising in this respect.

As it is currently implemented, the gating paradigm constitutes an off-line task: listeners attempt to identify a target after stimulus presentation in the absence of any response deadline. The off-line character of this task makes it well suited for studying young children. Specifically, it does not require a speeded or very sophisticated response; children can simply guess aloud the target at their own pace and need not be aware of the segmental structure of speech, in contrast with what may be required in tasks such as speech shadowing, phoneme or word monitoring, and mispronunciation detection.¹ Increasingly, then, the gating task is being used in developmental research (Craig, Kim, Rhyner, & Chirillo, 1993; Elliott, Hammer, & Evan, 1987; Walley, 1988) and in research with special populations, such as children with language delays and reading problems, adolescents with Down's Syndrome, and adults with Alzheimer's dementia or aphasia (Elliott, Scholl, Grant, & Hammer, 1990; Marcell & Cohen, 1992; Marshall, Duke, & Walley, 1992; Metsala, 1993; Wingfield, Goodglass, & Smith, 1990).

Given this increasing use, an important question arises. Is gating an appropriate task for making *developmental* comparisons of recognition, or comparisons across different age levels? There is no research addressing this question, and thus no evidence of the sort adduced to support the validity of this task for studying adult recognition. In his original study with adults, Grosjean (1980) replicated several findings (word-frequency, word-length, and context effects) that obtain in a range

This research was supported by the Department of Psychology at the University of Alabama and by NICHD Grant HD30398. We thank Steve Smith for software development and Paul Luce for computational analysis of stimulus characteristics. James Flege and Linda Smith made helpful comments on this paper, as did Joanne Miller and three anonymous reviewers. Aisha Holmes, Cindy Marshall, and Lauren Randazza assisted in data collection and scoring. We are also grateful to the principal, Jack Allison, the staff, and the students of Hall-Kent Elementary School for their participation. Address correspondence to A. C. Walley, Department of Psychology, University of Alabama at Birmingham, Birmingham, AL 35294 (e-mail: sbsf080@uabdp0.bitnet).

of other tasks and found that the amount of word-initial input needed to isolate targets was, in absolute terms, similar to estimates of recognition time from on-line tasks, such as speech shadowing and word monitoring. Subsequent research has shown that adults' responses in the gating task are not markedly influenced by certain procedural variations. For example, performance in the standard, successive presentation format is not artifactually enhanced relative to an individual format, in which different subgroups of listeners hear the various gates for a given word—whether the stimuli are presented in isolation or in meaningful sentences (Bard, Shillock, & Altmann, 1988; Cotton & Grosjean, 1984; Salasoo & Pisoni, 1985).

There is little empirical evidence to contradict these findings. Even so, some retain misgivings about the gating task, which center on the possibility that it is subject to the influence of postperceptual processes and thus inappropriate for studying core aspects of recognition (e.g., Wayland, Wingfield, & Goodglass, 1989; cf. Craig et al., 1993). This criticism warrants scrutiny.

The gating task may well tap postperceptual effects, in addition to more basic perceptual ones (see Samuel, 1986). However, even as perceptual effects may not be restricted to on-line tasks, such tasks themselves are not necessarily pure measures of perceptual processing (see Balota & Chumbley, 1985). Furthermore, there is no theoretical consensus that recognition itself is a solely perceptual, bottom-up phenomenon, impervious to the influence of higher level sources of knowledge (see McClelland, 1991). Finally, postperceptual effects are not isomorphic with uninteresting, artifactual ones. According to Luce (1986), word frequency constitutes higher level knowledge that biases recognition, rather than directly affecting early stimulus encoding (cf. Marslen-Wilson, 1987); nevertheless, he views frequency biases as a fundamental aspect of recognition that helps to optimize lexical processing (e.g., under conditions of stimulus degradation). Differing stances on the locus of frequency effects underscore the importance of understanding all processes that affect recognition, whatever their level of operation.

Still, from our perspective, there remains a serious problem with the possible operation of postperceptual processes in the gating task—namely, that the nature of these processes varies across age level, rendering the task unsuitable for making developmental comparisons. More generally, a limitation of previous validation studies is that they have been conducted exclusively with adults. Thus, although the gating task yields results that converge with more on-line tasks, and although variations in presentation format have little impact on performance, we do not know that this is the case for children.

Recently, however, Walley and Ancimer (1990) examined 5-year-olds' recognition of words presented in either the successive or the individual format. More children identified the targets as stimulus duration increased for the individual as opposed to the successive format. Children were also more confident about their responses for the individual format than for the successive one and

proposed a smaller number of preliminary word candidates. Walley and Ancimer suggested that the successive format implicitly provides children with negative feedback or leads them to believe that they are on the "wrong track," since, even after early identification of a target, they hear additional gates for it.

In another recent study, Craig and Kim (1990) found that adults identified successively gated words with less accuracy and confidence, even after word offset, than they identified words presented only once in their entirety. One explanation offered was listener fatigue, due to the greater number of trials in the successive format than in the whole-word one. Another explanation was response perseveration, due to the repetition of stimuli corresponding to the same word (cf. Walley & Ancimer, 1990). Yet another explanation was that because word offset is not very salient in the gating task, listeners anticipate additional input that is not forthcoming; ultimately performance is impeded—especially for short words, the recognition of which is substantially influenced by subsequent context (see, e.g., Bard et al., 1988; Grosjean, 1985).

These recent studies indicate that the standard, successive format of the gating task may actually have an inhibiting effect on recognition. Clearly, we need a better understanding of why this effect is sometimes found, especially if this task is to be used in studies of nonadult and atypical populations, and in comparisons across different listener groups. Our approach was to assess directly whether variations in presentation format have a differential impact on children's and adults' recognition. Specifically, we compared kindergarteners', first graders' and adults' performance for the successive format with a "duration-blocked" one. (Children of similar ages have been included in previous research on spoken word recognition.) Each format involved the repeated presentation of increasingly longer gates of words to the same subjects and the total number of gates/trials heard across formats was identical. The crucial difference was that for the duration-blocked format, gates were blocked by stimulus duration, not by word; thus, gates for the same word were not successive or temporally adjacent. (Subjects heard all the 100-msec gates for the words, then the 150-msec gates, the 200-msec gates, etc.)

The duration-blocked format is more similar to the successive one than is the individual format used in previous studies (Cotton & Grosjean, 1984; Salasoo & Pisoni, 1985). In the latter format, different subgroups of subjects hear the various gates for a given word. Thus, subjects do not hear repetitions of stimulus input for the same word, they do not hear gates for the same word in temporal succession, and they do not hear increasingly longer word fragments. Furthermore, the total number of gates heard is less than that in the successive format. Because these differences also apply to Craig and Kim's (1990) comparison of the successive and whole-word formats, it is difficult to pinpoint why effects of presentation format have or have not been found in previous research.

Our format comparison should yield a more definitive assessment of the specific effect of successive stimulus

presentations. Successive presentations might, when unconfounded with differences in stimulus repetitions for a given word and total number of trials (differences that might induce listener fatigue; Craig & Kim, 1990), have a facilitating effect—for example, by enhancing stimulus encoding and/or helping to maintain partial input in memory while a lexical match is being sought. Alternatively, successive presentations might impede performance by providing implicit negative feedback (Walley & Ancimer, 1990). If so, the duration-blocked format might disrupt such feedback—for example, by disrupting memory for previous responses to gates for a given word. The successive format might also impede performance by promoting response perseveration (Craig & Kim, 1990), a possibility that we can evaluate by examining the variety of word candidates proposed prior to target recognition.

Our primary question was whether children's and adults' performance is affected similarly by successive stimulus presentations, or whether variations in this task characteristic have a differential impact depending upon age. The answer should provide much-needed information about the validity of the gating task for making developmental comparisons, and preliminary evidence about the level of processing issue. That is, to the extent that postperceptual processes are strategic/conscious or due to extensive higher level knowledge, we should find little evidence of them in children's performance. Young children do not spontaneously employ verbal strategies for mnemonic purposes (see Flavell, 1985), and lexical knowledge in early to middle childhood still differs substantially from that in adulthood (see Walley, 1993). Postperceptual processes might allow adults, but not children, to perform better in one format than in the other. Alternatively, adults' recognition from word beginnings might be so good (in part because of postperceptual processes) that task variations have little impact, whereas factors such as negative feedback and response perseveration might affect children's recognition.

One ancillary goal was to assess the impact of presentation format on adults' recognition as a function of response format (written vs. oral). Most gating studies with adults, including those assessing task validity, have used a written response format. Since young preliterate children are limited to oral responses and our ultimate goal was to compare children's and adults' performance across presentation format, we wanted to determine how important any confounding of response format with age might be (see Craig et al., 1993; Walley, 1988). Again, adults' recognition might be so good that performance for the two presentation formats does not vary qualitatively with response format. However, the opportunity to read and inspect responses to previous gates for a given word in the successive format might promote response perseveration and impede recognition (Craig & Kim, 1990); alternatively, such an opportunity might help to offset the negative feedback that Walley and Ancimer (1990) suggest is a salient aspect of the successive format for children.

Another ancillary goal was to replicate the word-frequency effect observed by Grosjean (1980). By this effect, high-frequency words should be recognized from less input than low-frequency ones. Word familiarity has seldom, by any operational measure, been varied in previous studies involving children, regardless of the paradigm employed. Thus, there is little *developmental* evidence bearing on the growth of word familiarity (see Walley, 1993). One exception is a gating study by Fox and Koenigsnecht (1989), who found that both children's and adults' recognition of low-frequency words was better when the targets came from a small as opposed to a large cohort (i.e., overlapped with few rather than many words in the first two or three phonetic segments); recognition of high-frequency words was not greatly influenced by cohort size. We also sought to replicate this word frequency \times cohort size interaction.

In summary, our primary aim was to assess the validity of the gating task from a developmental perspective by comparing children's and adults' recognition of words presented in either the standard successive or the duration-blocked format (Experiment 2). A direct age comparison of the effect of presentation format has not been made before. Also, our presentation format comparison will yield information not available from previous adult studies about the specific effect of successive stimulus presentations. The effect of response format on adult recognition was addressed in a preliminary experiment (Experiment 1); the effects of word frequency and cohort size were studied in both experiments.

METHOD

The method was the same across experiments with the exception of subject age and response format (see below).

Subjects

In Experiment 1, 48 introductory psychology students were tested in successive or duration-blocked presentation formats of a written response condition (M age = 22.65 and 21.0 years; SD = 4.25, 3.37) or an oral response condition (M = 22.27 and 20.55; SD = 4.28, 2.00). In Experiment 2, 24 kindergarteners were tested in successive or duration-blocked formats (M = 6.03 and 6.22; SD = .36, .38), as were 24 first graders (M = 7.32 and 7.08; SD = .34, .36). The children were tested in an after-school program. They made only oral responses, which were compared with those of the adults in the oral condition of Experiment 1. All subjects were native English speakers with no known speech, hearing, or (early) reading disorder, and no child had a recent ear infection, according to parental report. (It would, of course, have been desirable to administer a hearing screening directly to all subjects, but this was not possible because of time constraints and a lack of equipment. Therefore, we had to rely on subject/parent report data in assessing hearing status.)

Stimuli

Sixteen concrete nouns formed a 2 (word frequency) \times 2 (cohort size) stimulus design (see Table 1). Eight were of high frequency, eight of low frequency (median = 114 and 4), according to Kucera and Francis (1967). Half the high- and low-frequency words came from large cohorts, half from small cohorts (median = 32.50 and 11); cohort size was defined by the number of words with similar #CV or #CCV structure in an on-line version of Web-

ster's dictionary (1967) with about 20,000 entries.² The words were balanced, across cells, in number of syllables and similarity of #CV or #CCV structure. For example, with respect to the latter factor, *prison*, *priest*, *brother*, and *pudding* all begin with a labial stop consonant; however, these words are different enough to permit variation in cohort size, or the number of candidates that may be activated from word-initial input (Marslen-Wilson, 1987). Word-initial consonant clusters were also balanced across cells. The small number of words per cell (4) was limited by the total number of gates/trials that we could present in the one session during which subjects were available for testing and by our selection criteria.

A female speaker read each word aloud in a neutral sentence, "Now I will say the word ____." Sentences were recorded with a Sony cassette recorder and Nakamichi microphone in an IAC booth, bandpass filtered (60 Hz–10 kHz) and digitized at 20 kHz with 12-bit amplitude resolution. The words were excised from the sentences, normalized for peak amplitude, and stored on disk.

Two eight-word lists were created because presenting many short stimuli in the duration-blocked format (sixteen 100-msec gates, sixteen 150-msec gates, etc.) might be especially frustrating for children. The lists had equal numbers of one- and two-syllable words (one of each length from the four cells). Similarity of #CV or #CCV structure was preserved within and across lists. List A and B words were similar in duration ($M = 607$ and 628 msec; $SD = 86.19$, 75.33) and number of gates ($M = 11.12$ and 11.50); $SD = 1.64$, 1.60 ; total = 89 and 92).

For the successive format, the first stimulus in each sequence of gates for a word consisted of the first 100 msec of the entire waveform (see Elliott et al., 1987), with subsequent stimuli incremented in 50-msec steps. For the duration-blocked format, the first block of trials consisted of the 100-msec gates for the eight words in a list, then the 150-msec gates for the same words, and so

Table 1
Characteristics of the Test Words

Cohort Size	KF	CS	Kol	Dur	No. Gates
High Frequency					
Large					
<i>sun</i> (A)	278	80	76	697	13
<i>grass</i> (B)	53	31	100	697	13
<i>river</i> (A)	165	54	59	557	10
<i>prison</i> (B)	42	22	7	597	11
Low Frequency					
Large					
<i>stocking</i> (B)	1	28	25	732	14
<i>garbage</i> (A)	7	34	22	700	13
<i>robe</i> (B)	6	35	.5	621	11
<i>priest</i> (A)	16	18	.5	607	11
High Frequency					
Small					
<i>story</i> (B)	153	16	205	675	12
<i>girl</i> (A)	220	8	350	486	9
<i>rock</i> (B)	75	18	43	635	12
<i>brother</i> (A)	73	9	124	494	9
Low Frequency					
Small					
<i>snail</i> (A)	1	1	.5	670	12
<i>glue</i> (B)	8	4	8	497	9
<i>robin</i> (A)	2	18	.5	644	12
<i>pudding</i> (B)	.5	9	10	567	10

Note—KF, frequency based on Kučera and Francis (1967); CS, cohort size; Kol, frequency based on Kolson (1961); Dur, total duration in milliseconds; No. Gates, number of gates/trials per word. A and B refer to list membership. Words not found in a frequency count were assigned a value of .5.

forth. Order of gates for different words in each block was random. Since the words in a list varied in duration, shorter words dropped out of longer blocks, and, after Block 11 (600 msec), gates for only a few words remained. These were combined into one final block per list, with shorter gates for a given word preceding longer ones; thus, the number of stimuli across blocks was similar. If the difference between the "last" gate for a word and the entire waveform was less than 25 msec (half a gate), this gate was not used. For example, the 550-msec gate was omitted in the sequence for *river*, which ended with the entire waveform (see Table 2). This gate was also eliminated in the duration-blocked condition. All gated stimuli were ramped off over the final 10 msec. Digitized stimuli were recorded on audiotape with 1.5 sec between stimuli within each sequence/block and 3 sec between different sequences/blocks. Two practice sequences or blocks preceded the test words in each list.

Procedure

Lists A and B were both presented to each subject within each presentation/response format condition (Experiment 1) and each presentation format/age condition (Experiment 2), with list order counterbalanced across subjects. Half the subjects in each condition were administered Form L, the other half, Form M of the *Peabody Picture Vocabulary Test-Revised* (PPVT-R; Dunn & Dunn, 1981).

Subjects were tested individually in a 1-h session. They were told they would first hear only parts of words, then more and more of them. For the oral response format, subjects were asked to guess the target aloud after each gate and to give a confidence rating by pointing to a scale, which contained the numbers 1–7 with a straight-mouthed face and three question marks at the 1 end, and a happy face at the 7 end. For the written response format (Experiment 1 only), adults wrote their identification after each gate on a blank line of a booklet and circled a number for their confidence rating on the scale beside each line. A separate page was allotted for each sequence/block, so that the adults could not see their previous responses; each page had 20 blank lines, so they would not be aware of the total number of gates for a given sequence/block.

The stimuli were presented over matched and calibrated Sennheiser headphones at a comfortable listening level via a portable Uher tape recorder. After each trial in a sequence/block, the assistant stopped the tape and, for the oral response format, recorded the subject's response. For the written response format, the assistant waited for the subject to write his/her response. If the subject did not respond after about 30 sec, a null identification and a confidence rating of 1 were used. No specific feedback was given.

Data Analyses

We examined isolation points (IP), confidence ratings at the IP (CIP), and total acceptance points (TAP). The IP was the stimulus duration at which a subject first correctly identified a target; the TAP was the duration at which a subject identified the target and first gave a confidence rating of 6 or 7, without changing his/her identification or giving a lower confidence rating. The IP and TAP provide some index of the process by which listeners narrow in on a word, as more information about it accumulates (Grosjean, 1985). IPs and TAPs were converted to percentages of a target's total duration to adjust for differences in absolute duration across words.

When a target was not identified, even after all of it had been presented, its total duration plus 50 msec (one gate) was used for the IP and a value of 1 was used for the CIP; when subjects did not maintain a correct identification, or did not reach/maintain a high level of confidence (6 or 7), the target's total duration plus 50 msec was used for the TAP (see also Elliott et al., 1987). This scoring was used because of the small number of words per stimulus cell; we did not want to exclude any items from analysis. However, we conducted both subject and item analyses (F_1 and F_2) of the IP,

Table 2
Illustration of the Successive and Duration-Blocked
Presentation Formats (List A)

Format					
Successive			Duration Blocked		
Trial	Gate	Target	Trial	Gate	Target
Sequence 1			Block 1		
1	100	<i>river</i>	1	100	<i>sun</i>
2	150		2	100	<i>priest</i>
3	200		3	100	<i>brother</i>
4	250		4	100	<i>snail</i>
5	300		5	100	<i>river</i>
6	350		6	100	<i>girl</i>
7	400		7	100	<i>garbage</i>
8	450		8	100	<i>robin</i>
9	500				
10	557				
Sequence 2			Block 2		
11	100	<i>girl</i>	9	150	<i>garbage</i>
12	150		10	150	<i>sun</i>
13	200		11	150	<i>robin</i>
14	250		12	150	<i>river</i>
15	300		13	150	<i>priest</i>
16	350		14	150	<i>brother</i>
17	400		15	150	<i>snail</i>
18	450		16	150	<i>girl</i>
19	486				
Sequence 8			Block 12		
81	100	<i>brother</i>	84	650	<i>sun</i>
82	150		85	644	<i>robin</i>
83	200		86	670	<i>snail</i>
84	250		87	650	<i>garbage</i>
85	300		88	697	<i>sun</i>
86	350		89	700	<i>garbage</i>
87	400				
88	450				
89	494				

Note—All gates were from word onset and are given in milliseconds. Total number of trials/gates = 89.

CIP, and TAP data to ensure that the results were not due to idiosyncratic properties of the test words and to guard against spurious higher order interactions. Unless otherwise stated, only the main effects and interactions that were significant in both analyses at or beyond $\alpha = .05$ are reported.

We also examined the proportion of pre-IP trials, across all targets, on which a different word candidate was offered by individual subjects to assess whether or not they reproduced previous responses.

RESULTS

Experiment 1

A 2 (presentation format) \times 2 (response format) analysis of variance (ANOVA) of adults' standard *PPVT-R* scores revealed a main effect of response format; adults who made oral responses had higher receptive vocabulary scores than did those who made written responses [$M = 101.96$ vs. 94.5 , $SD = 9.06$, 11.48 ; $F(1,44) = 6.04$, $p < .025$]. Therefore, *PPVT-R* scores served as a covariate in analyses of the word candidate, IP, CIP, and TAP data. Each analysis involved a 2 (pre-

sentation format) \times 2 (response format) \times 2 (word frequency) \times 2 (cohort size) analysis of covariance (ANCOVA) for a mixed design, with a corresponding item ANOVA. For word candidates, only a 2 (presentation format) \times 2 (response format) ANCOVA was appropriate.

Word candidates. A main effect of presentation format [$F_1(1,43) = 6.55$, $p < .025$] was found; fewer different word candidates were proposed prior to target isolation for the successive format than for the duration-blocked one [M proportion = $.41$ vs. $.46$; $SD = .05$, $.08$].

Isolation points and confidence ratings.³ Analysis of the IP data revealed a main effect of presentation format [$F_1(1,43) = 14.25$, $p < .001$; $F_2(1,12) = 11.68$, $p < .001$]; adults needed to hear more of the successively-gated targets than the duration-blocked ones to first isolate them, even with low confidence ($M\% = 59.04$ vs. 53.13 ; $SD = 12.24$, 10.07). Given a mean duration of 617 msec for the 16 test words, these percentages correspond to an absolute difference of 36 msec (three-quarters of a gate in our study, a full gate in previous ones—e.g., Grosjean, 1980). As is shown in the left panel of Figure 1, this difference between presentation formats was found, regardless of whether adults made oral or written responses; that is, there was no main effect of response format, and this factor did not interact with presentation format at a significant level.

Main effects of word frequency and cohort size, together with a word frequency \times cohort size interaction and a presentation format \times word frequency \times cohort size interaction, were also found in the subject analysis of the IP data [respectively, $F_1(1,44) = 126.81$, $p < .001$; $F_1(1,44) = 14.92$, $p < .001$; $F_1(1,44) = 6.54$, $p < .025$; $F_1(1,44) = 5.27$, $p < .05$], but not in the item analysis [$F_2(1,12) = 3.36$, $.31$, $.16$, and 2.44 , $p > .09$, in each case]. Still, the nature of these effects was consistent with the finding that the isolation of low-frequency words is enhanced by membership in a small cohort (Fox & Koenigsnecht, 1989), especially in the duration-blocked format.

Analysis of the CIP data revealed a main effect of response format [$F_1(1,43) = 8.43$, $p < .01$; $F_2(1,12) = 34.16$, $p < .001$]; adults who spoke, rather than wrote their responses were more confident at the IP (M CIP = 4.14 vs. 3.41 ; $SD = 1.15$, 1.28). A word frequency \times cohort size interaction was found in the subject, but not the item analysis [$F_1(1,43) = 4.12$, $p < .05$; $F_2(1,12) = 0.81$, $p > .30$]; adults were most confident at the IP for low-frequency words from large cohorts, perhaps because their identifications tended to be based on more input.

Total acceptance points. Analysis of the TAP data revealed a main effect of presentation format [$F_1(1,43) = 14.65$, $p < .001$; $F_2(1,12) = 10.93$, $p < .01$]; more input was needed to identify words with high confidence for the successive format than for the duration-blocked one ($M\% = 81.42$ vs. 74.18 ; $SD = 8.79$, 8.33). A presentation format \times response format \times cohort size interaction was also found [$F_1(1,44) = 4.15$, $p < .05$; $F_2(1,12) = 5.21$, $p < .05$]. Newman-Keuls' comparisons indicated that TAPs were always longer for successively gated words than for

duration-blocked ones, regardless of whether the targets came from large or small cohorts, and regardless of whether adults made oral or written responses (see the right panel of Figure 1, which shows TAPs as a function of presentation and response format). However, TAPs were especially long when adults wrote their responses for successively gated targets from large cohorts; comparison of these responses with written responses for duration-blocked targets from large cohorts (85.58 vs. 72.66; $SD = 10.98, 8.88$) yielded the largest absolute difference—about 80 msec.

In summary, adults proposed fewer word candidates, or were more likely to reproduce past responses, and needed more acoustic–phonetic input to first isolate successively gated targets than they did for duration-blocked ones. Adults also needed more input to maintain correct identification responses with high confidence for the successive presentation format—particularly when the targets came from a large cohort and thus had to be discriminated from numerous alternatives, and when adults could read their various responses on previous trials. These results indicate that the successive presentation format of the gating paradigm (perhaps especially together with a written response format) yields a conservative portrait of adults' spoken word recognition ability. We turn next to our central question—namely, whether or not children's recognition performance is affected similarly by presentation format. To address this question, we compared kindergarteners' and first graders' responses for successively gated and duration-blocked targets with those of adults in the oral response condition of Experiment 1.

Experiment 2

Mean standard *PPVT-R* scores in the successive and duration-blocked formats were 102.17 and 104.50 for kindergarteners ($SD = 13.93, 10.73$), 106.33 and 98.92 for first graders ($SD = 11.73, 10.32$), and 100.58 and 103.33 for adults (8.93, 9.36). A 3 (age) \times 2 (presentation format) ANOVA yielded no significant effects at or beyond $\alpha = .05$. Thus, our subject groups were similarly

close to the norm for their age in overall receptive vocabulary knowledge, and each analysis of the IP, CIP, or TAP data involved a 3 (age) \times 2 (presentation format) \times 2 (word frequency) \times 2 (cohort size) ANOVA for a mixed design, with an item ANOVA. The word candidate analysis employed a 3 (age) \times 2 (presentation format) ANOVA.

Word candidates. Analysis of the word candidate data revealed a main effect of age [$F_1(2,66) = 9.41, p < .001$] and an age \times presentation format interaction [$F_1(2,66) = 4.49, p < .025$]. Mean proportions of different word responses prior to target isolation by kindergarteners, first graders, and adults were .57, .49, and .40 in the successive format ($SD = .12, .05, .05$) and .47, .52, and .44 ($SD = .12, .08, .07$) in the duration-blocked format. The only significant post-hoc comparison indicated that kindergarteners gave more varied responses for the successive format than for the duration-blocked one.

Isolation points and confidence ratings.⁴ A main effect of age was found for the IP data [$F_1(2,66) = 11.02, p < .001$; $F_2(2,24) = 14.98, p < .001$]; kindergarteners' IPs were longer than first graders', which were longer than adults' ($M\% = 65.80$ vs. 62.68 vs. 57.30 ; $SD = 13.69, 12.52, 11.44$). (The difference in kindergarteners' and adults' IPs was about 52 msec.) Similarly, a main effect of age was found for the CIP data [$F_1(2,66) = 28.33, p < .001$; $F_2(2,24) = 91.31, p < .001$]; both kindergarteners and first graders were more confident than adults [$M = 5.56$ and 5.90 vs. 4.15 ; $SD = 1.25, 1.14, 1.15$], perhaps because their IPs were based on more input.

Analysis of the IP data also revealed a main effect of word frequency [$F_1(1,66) = 190.67, p < .001$; $F_2(1,12) = 5.23, p < .05$] and a marginal presentation format \times word frequency \times cohort size interaction [$F_1(1,66) = 3.95, p = .051$; $F_2(1,12) = 4.83, p < .05$]. IPs were shorter (by about 89 msec) for high- as opposed to low-frequency targets ($M\% = 54.72$ vs. 69.13 ; $SD = 10.45, 11.25$). Post hoc tests confirmed that IPs were always shorter for high- as opposed to low-frequency targets, regardless of whether they came from large or small co-

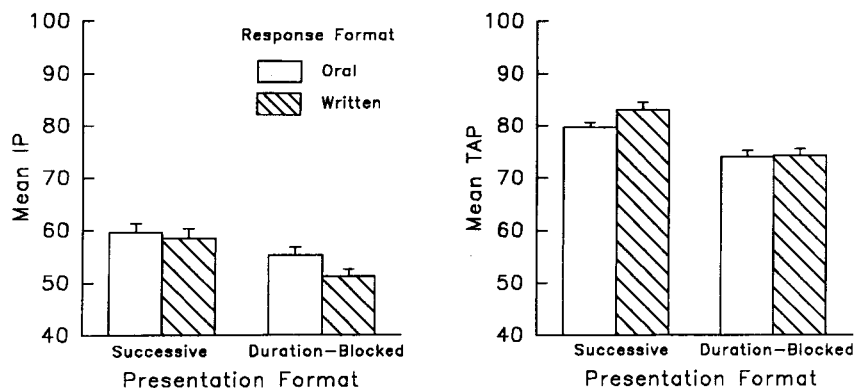


Figure 1. Mean isolation points (IPs) and total acceptance points (TAPs) as a function of presentation format and response format in Experiment 1. IPs and TAPs are expressed as percentages of the test words' total duration.

horts, and regardless of presentation format; however, IPs for high-frequency words from large cohorts were longer when they were presented in the successive, as opposed to the duration-blocked, format (58.73 vs. 53.57; $SD = 10.39, 9.80$ —an absolute difference of about 32 msec). Still, the effect of presentation format (like the effects of word frequency and cohort size) did not vary as a function of age (see left panel of Figure 2).

Total acceptance points. Main effects of age [$F_1(2,66) = 6.94, p < .001; F_2(2,24) = 22.18, p < .001$] and presentation format [$F_1(1,66) = 11.41, p < .001; F_2(1,12) = 29.45, p < .001$] were found in the analysis of the TAP data. Kindergarteners' TAPs were longer than both first graders' and adults' ($M\% = 84.30$ vs. 74.56 and 76.88; $SD = 15.22, 13.28, 7.95$; the difference between kindergarteners' and first graders' TAPs was 60 msec.) More input (47 msec) was needed to reach/maintain high confidence for targets in the successive than in the duration-blocked format (82.34 vs. 74.81; $SD = 12.08, 13.17$). Again, this effect of presentation format did not vary with age (see right panel of Figure 2).

In summary, high-frequency words were first isolated from less input than were low-frequency words; this advantage was less pronounced for high-frequency words that came from a large cohort and were presented in the successive, as opposed to the duration-blocked, format. More input was needed to reach/maintain a high level of confidence about the identity of all targets when they were successively gated than when they were duration blocked. These effects of presentation format held across age; that is, although IPs and TAPs generally increased with age, age did not interact with presentation format for these dependent measures. Only for the word candidate data was an age \times presentation format interaction observed; kindergarteners proposed a greater variety of word candidates prior to target isolation in the successive format than in the duration-blocked one, whereas the number of different word candidates offered by first graders and adults did not vary as a function of presentation format.

DISCUSSION

Two clear-cut results emerged. First, recognition was impeded by successive presentations of partial stimulus input for the same word, when such presentations were not confounded with differences in stimulus repetitions and total number of trials (cf. Cotton & Grosjean, 1984; Craig & Kim, 1990; Salasoo & Pisoni, 1985). In Experiment 1, adults' IPs were longer for words presented in the successive format as opposed to the duration-blocked one; this impeding effect was also observed for TAPs and was most pronounced for words from large cohorts in the written response format (when the targets had to be discriminated from many alternatives and adults could read their various responses on previous trials). In Experiment 2, the impeding influence of the successive format was initially more restricted—affecting only children's and adults' IPs for high-frequency words from large cohorts (words that have a "high profile" in the mental lexicon, but must be discriminated from many alternatives, the existence of which may be emphasized by the successive format); however, TAPs were longer for all words in the successive, as opposed to the duration-blocked, format. Second, age did not interact with stimulus presentation format, except in the analysis of word candidates. Thus, the successive format would seem to yield a *similarly* conservative estimate of children's and adults' recognition ability.

Why do successive stimulus presentations impede recognition? One factor implicated by our results is response perseveration (see also Craig & Kim, 1990). In Experiment 1, adults offered a smaller variety of word candidates in the successive format than in the duration-blocked one (i.e., they tended to reproduce past responses), and the isolation of all targets was delayed. Those who wrote rather than spoke their responses were less confident about early identifications of the targets, and their TAPs were longer for successively gated words from large, as opposed to small, cohorts, indicating that this response format may encourage perseveration.

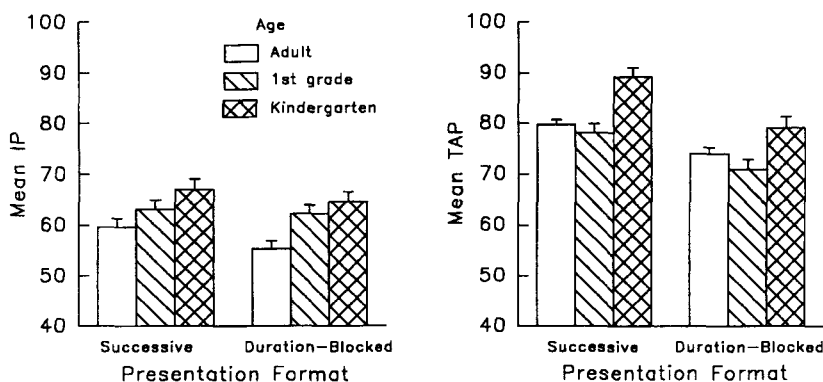


Figure 2. Mean isolation points (IPs) and total acceptance points (TAPs) as a function of age and presentation format. IPs and TAPs are expressed as percentages of the test words' total duration.

However, response perseveration cannot be the only relevant factor. In Experiment 2 also, the performance of both children and adults (who made only oral responses) was poorer in the successive format than in the duration-blocked one; yet the variety of word candidates proposed by first graders and adults did not differ across presentation format, and kindergarteners actually offered more candidates in the successive format. These results suggest that some sort of negative feedback may be in operation (see also Walley & Ancimer, 1990). The successive presentation of additional input for the same word may lead young children especially to infer that their responses on early trials are incorrect and thus to often change them, ultimately impeding recognition. The duration-blocked format may serve to disrupt such negative feedback (e.g., by disrupting memory for previous responses to a given target).⁵

In any event, the lack of marked age-related differences across different presentation formats suggests that the gating task is appropriate for making developmental comparisons of spoken word recognition. Few such tasks are available, since young children cannot make the speeded and sophisticated responses required in the on-line tasks typically used to study adult recognition. Still, our child and adult subjects might have performed similarly for different reasons. For example, adults, and perhaps first graders, may be more fully aware than kindergarteners that they are hearing overlapping repetitions for the same words in the successive format and may be better able to ignore any associated negative feedback or to implement strategies to counteract such feedback. More research is needed, then, if we are to fully understand this and other tasks that are of potential use in developmental studies (see Walley, 1993). From a theoretical perspective too, it will be important to scrutinize the extent to which various biases play a role in these recognition tasks, including information about age-related changes or constancies in these biases.

Our results also suggest that the successive presentation format of the gating task, especially in conjunction with written responses, yields a conservative portrait of adults' recognition ability. However, we did not observe pervasive effects of word frequency and cohort size, particularly in terms of their interaction with each other (cf. Fox & Koenigsnecht, 1989). To our knowledge, no other studies have assessed the effect of word frequency in combination with cohort size/membership as a measure of the acoustic-phonetic similarity relations among different words. The effect of word frequency and neighborhood size on recognition has been examined in several studies, however. In contrast to cohort size, which is defined in terms of word-initial overlap in the first two or three phonetic segments, neighborhood size is typically defined by single segment overlap in any word position. Close inspection of the literature suggests that an interaction between word frequency and the latter measure of acoustic-phonetic similarity may not be all that robust. In particular, studies of adults' recognition performance in a variety of tasks have often failed to result

in any interaction between these stimulus factors, or they have resulted in an interaction in subject, but not in item, analyses (e.g., Andrews, 1989; Goldinger, Luce, & Pisoni, 1989; Luce, 1986), despite the use of larger subject groups and many more test words than in our study. More recently, Metsala (1993) compared 7-year-olds', 9-year-olds', 11-year-olds', and adults' performance in the gating task and found a word frequency \times neighborhood size interaction in both subject and item analyses of adults' IPs, but only in the subject analyses of children's IPs. Together with our own results, this last finding suggests that more research is needed in order to track the *development* of acoustic-phonetic similarity effects on recognition in combination with word frequency. Such research might accord greater attention to the developmental appropriateness of various measures of word familiarity and acoustic-phonetic overlap (see Logan, 1992; Walley, 1993).⁶

Although the standard, successive format of the gating paradigm yields a conservative portrait of spoken word recognition ability, several final points should be emphasized. First, performance at three age levels (and of two groups of adults) was affected similarly by presentation format; thus, either the successive or the duration-blocked format would seem appropriate for making developmental comparisons of spoken word recognition. Second, from a purely design perspective, both presentation formats are simpler and more economical than the individual presentation format, in that a smaller number of subjects can be tested (see also Cotton & Grosjean, 1984). Third, the magnitude of the differences in performance for the successive and duration-blocked formats was not trivial; specifically, if one wishes to establish optimal recognition ability for certain listener groups, the observed difference in performance across formats of roughly 50 msec (which is, for example, sufficient to cue place of initial stop consonant articulation; see Kewley-Port & Luce, 1984) might be quite important. Finally, the duration-blocked format may simulate the demands of real-time speech processing better than the successive format to the extent that one typically hears "bits and pieces" of different words in a sentence, rather than the beginnings of the same word "over and over" again. Therefore, in some cases, the duration-blocked format might be of greater utility to researchers or clinicians than the standard, successive presentation format of the gating paradigm.

REFERENCES

- ANDREWS, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 802-814.
- BALOTA, D. A., & CHUMBLEY, J. I. (1985). The locus of word frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language*, **24**, 89-106.
- BARD, E. G., SHILLOCK, R. C., & ALTMANN, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics*, **44**, 395-408.
- COTTON, S., & GROSJEAN, F. (1984). The gating paradigm: A compar-

- ison of successive and individual presentation formats. *Perception & Psychophysics*, **35**, 41-48.
- CRAIG, C. H., & KIM, B. W. (1990). Effects of time gating and word length on isolated word-recognition performance. *Journal of Speech & Hearing Research*, **33**, 808-815.
- CRAIG, C. H., KIM, B. W., RHYNER, P. M. P., & CHIRILLO, T. K. B. (1993). Effects of word predictability, child development, and aging on time-gated speech recognition performance. *Journal of Speech & Hearing Research*, **36**, 832-841.
- DUNN, L. M., & DUNN, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- ELLIOTT, L. L., HAMMER, M. A., & EVAN, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics*, **42**, 150-157.
- ELLIOTT, L. L., SCHOLL, M. E., GRANT, J. O., & HAMMER, M. A. (1990). Perception of gated, highly familiar spoken monosyllabic nouns by children with and without learning disabilities. *Journal of Learning Disabilities*, **23**, 248-259.
- FLAVELL, J. H. (1985). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.
- FOX, R. A., & KOENIGSKNECHT, R. (1989, November). *The effect of cohort size on gated word recognition*. Paper presented at the American Speech-Language-Hearing Association meeting, St. Louis.
- GOLDINGER, S. D., LUCE, P. A., & PISONI, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language*, **28**, 501-518.
- GROSJEAN, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, **28**, 267-283.
- GROSJEAN, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, **38**, 99-310.
- JUSCZYK, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 17-64). Parkton, MD: York Press.
- KEWLEY-PORT, D., & LUCE, P. A. (1984). Time-varying features of initial stop consonants in auditory running spectra: A first report. *Perception & Psychophysics*, **35**, 353-360.
- KOLSON, C. (1961). *The vocabularies of kindergarten children*. Pittsburgh, PA: University of Pittsburgh Press.
- KUČERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LOGAN, J. S. (1992). A computational analysis of young children's lexicons. In *Research on spoken language processing* (Tech. Rep. No. 8). Bloomington, IN: University of Indiana, Department of Psychology, Speech Research Laboratory.
- LUCE, P. A. (1986). Neighborhoods of words in the mental lexicon. In *Research on speech perception* (Tech. Rep. No. 6). Bloomington, IN: University of Indiana, Department of Psychology, Speech Research Laboratory.
- MARCELL, M. M., & COHEN, S. (1992). Hearing abilities of Down syndrome and other mentally handicapped adolescents. *Research in Developmental Disabilities*, **13**, 533-551.
- MARSHALL, N. B., DUKE, L. W., & WALLEY, A. C. (1992, April). *The effects of priming on the speech recognition skills of normal young, normal elderly, and listeners with Alzheimer's disease*. Paper presented at the Fourth Cognitive Aging Conference, Atlanta.
- MARSLÉN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71-102). Cambridge, MA: MIT Press.
- MCCLELLAND, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, **23**, 1-44.
- METSALA, J. L. (1993). *An examination of speech perception in reading disabled and normally achieving children using the gating paradigm*. Unpublished doctoral dissertation, University of Toronto.
- MONTGOMERY, J. W., SCUDDER, R. R., & MOORE, C. A. (1990). Language-impaired children's real-time comprehension of spoken language. *Applied Psycholinguistics*, **11**, 273-290.
- SALASOO, A., & PISONI, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory & Language*, **24**, 210-231.
- SAMUEL, A. G. (1986). The role of the lexicon in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines* (pp. 89-112). New York: Academic Press.
- TOPPINO, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition*, **19**, 159-167.
- WALLEY, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development*, **3**, 137-165.
- WALLEY, A. C. (1993). The role of vocabulary growth in children's spoken word recognition and segmentation ability. *Developmental Review*, **13**, 286-350.
- WALLEY, A. C., & ANCIMER, C. (1990, March). *Spoken word recognition by young children in individual and successive presentation formats of the gating paradigm*. Paper presented at the Conference on Human Development, Richmond, VA.
- WALLEY, A. C., & METSALA, J. L. (1990). The growth of lexical constraints on spoken word recognition. *Perception & Psychophysics*, **47**, 267-280.
- WAYLAND, S. C., WINGFIELD, A., & GOODGLASS, H. (1989). Recognition of isolated words: The dynamics of cohort reduction. *Applied Psycholinguistics*, **10**, 475-487.
- WEBSTER'S SEVENTH COLLEGIATE DICTIONARY (1967). Los Angeles, CA: Library Reproduction Service.
- WINGFIELD, A., GOODGLASS, H., & SMITH, K. L. (1990). Effects of word-onset cueing on picture naming in aphasia: A reconsideration. *Brain & Language*, **39**, 373-390.

NOTES

1. Mispronunciation detection has been used with preschoolers (see, e.g., Walley & Metsala, 1990), but not as an on-line task. Word monitoring has been used with children as young as 5½ years of age (Montgomery, Scudder, & Moore, 1990), but considerable training was required.

2. Since the prefix [pri] is very common, *priest* likely has more neighbors than was identified by the computational analysis. Also, we tried to select words that were consistent with the adult frequency count in Kolson's (1961) juvenile production count (see Table 1). *Prison* violated this rule but was retained as a high-frequency item because of the difficulty in finding words that met all our selection criteria. Young children likely know this word, and indeed all the low-frequency ones, even if they do not produce them very often.

3. Adults rarely failed to identify the targets, even with low confidence: for the oral successive and duration-blocked formats, M proportion = .02 and .01 ($SD = .06, .05$); for the written successive and duration-blocked formats, $M = .03$ and .01 ($SD = .08, .04$).

4. A four-way ANOVA of arcsin-transformed proportions of failures to identify the targets revealed only a significant main effect of age [$F_1(2,66) = 4.00, p = .023$; $F_2(2,24) = 4.68, p = .019$]; the kindergarteners' errors were higher than first graders', which were higher than adults' (M untransformed proportion = .06, .03, .01, respectively; $SD = .14, .09, .06$). Still, the kindergarteners were able to identify most of the words.

5. Our presentation format effect resembles the well-established finding that learning and memory are generally superior for spaced or distributed, as opposed to massed, trials; that is, recognition was always better when gates for the same word were spaced (duration blocked) rather than massed (successively presented). Spaced presentations may be processed more fully and/or in more varied fashion, so that even words from large cohorts are better discriminated from their many neighbors than are massed ones. The size of this effect was similar across age, which is consistent with Toppino's (1991) finding that even 3- and 4-year-olds exhibit a (memory) spacing effect of the same magnitude as do older subjects, and thus his claim that automatic/fundamental memory mechanisms, rather than controlled processes or acquired strategies, are sufficient to produce the effect.

6. We found that kindergarteners' IPs were longer than first graders', which were longer than adults' (cf. Fox & Koenigsknecht, 1989); kindergarteners' TAPs were longer than first graders', whereas first graders' were similar to adults'. These slightly different age trends indicate that there are subtle, but detectable, advances in spoken word recognition during childhood.