Effects of Lexical Stress on Phonetic Categorization

Cynthia M. Connine[a], Charles Clifton[b], Jr., Anne Cutler[c]

[a] Department of Psychology, Indiana University, Bloomington, Ind., USA;
[b] University of Massachusetts at Amherst, Mass., USA;
[c] MRC Applied Psychology Unit, Cambridge, U.K.

Abstract. Three experiments investigated the use of lexical stress in auditory word recognition. Speech voicing continua were created in which lexical status resulted in one endpoint constituting a real word and the other end a nonword (e.g. diGRESS-tiGRESS, in which *digress* is a real word, and DIgress-TIgress, in which *tigress* is a real word). Subjects' identification of the initial segment of these items was biased in the midrange of the continua in that they were more likely to report a segment that resulted in a real word than one that resulted in a nonword. Alternative acoustically based explanations are discounted in favor of a lexically based account of the data. Possible mechanisms underlying the effect of lexical stress on speech perception are discussed.

In order to comprehend a spoken utterance, one of the tasks a listener must perform is the identification of the words in the utterance. The phonetic information available in spoken words is both segmental and suprasegmental. While segmental variation across words is language-universal, languages differ in whether and how they use suprasegmental variation to distinguish words from one another. The present studies concern one variety of lexically significant suprasegmental variation, namely lexical stress: the pattern of strong and weak syllables associated with a word's canonical pronunciation. In lexical stress languages such as English, stress does not occur in fixed position with respect to word boundaries, and hence is potentially available to distinguish between words, even between words whose pronunciation is otherwise (i.e. in the segmental domain) identical. Moreover, evidence from tip of the tongue experiences suggests that canonical lexical stress patterns form part of the lexical representations of words independent from the segmental representation. Speakers are often able to guess the stress pattern of a target item they are unable to retrieve [Brown and McNeill, 1966].

What has been more difficult to demonstrate is a role for lexical stress during auditory word recognition and recent studies have produced a conflicting pattern of results. Cutler [1986] has reported evidence

that lexical stress does not contribute to early deactivation of an inconsistent lexical entry. Using a cross-modal priming task, Cutler [1986] found that pairs of words with segmentally identical but suprasegmentally distinct pronunciations (such as *FORbear* and *for BEAR)* show the priming patterns characteristic of homophones - presentation of either results in both lexical representations being activated. This suggests that initial activation of a lexical representation is based on segmental information alone. Similarly, Cutler and Clifton [1984] found that prior knowledge of stress pattern did not facilitate lexical identification; nor did conformity to canonical noun-verb stress patterns facilitate grammatical class judgements. This pattern of results suggests that stress information is not used to restrict the set of potential word candidates and thereby speed word recognition. While preknowledge of lexical stress information does not facilitate word recognition, mis-stressing of words can significantly delay recognition [Cutler and Clifton, 1984]. Similar results have been reported by Bond and Small [1983] and Slowiaczek [1986]. This strongly suggests that stress patterns are indeed known to listeners, and are part of the lexical representation accessed in word recognition. A mismatch between stored and perceived information may inhibit responses in word recognition tasks.

The research reported here uses a paradigm developed by Ganong [1980; see also Connine and Clifton, 1987] to investigate the role of lexical stress in determining the identity of a word. Ganong [1980] presented listeners with synthesized stimuli from two types of voicing series (dash-tash, dask-task). In one series, the voiced endpoint formed a word (e.g. dash-tash): in a second

series (e.g. dask-task) the voiceless endpoint formed a word. Listeners tended to label stimuli in the midrange of the continua such that a word was formed. In the experiments reported here, this paradigm is utilized to determine whether lexical stress can similarly be used as a source of information to establish that a phonetic sequence is a real word and thus to influence identification of speech.

The logic is very similar to Ganong's except that lexical status is determined by the lexical stress pattern of the sequence. In experiment 1, two words were selected that differed phonemically only in the voicing feature of the initial stop consonant (/d/ vs. /t/) and in terms of their canonical stress pattern: diGRESS and Tigress (strong syllables are indicated by capital letters). We synthesized two voicing continua, diGRESS-tiGRESS and DIgress-TIgress, that differed from one another only in their stress pattern. In one series, diGRESS-tiGRESS, the voiced endpoint formed a word; conversely, in DIgress-TIgress, the voiceless endpoint formed a word. If lexical stress information is activated in lexical access and if (following Ganong) listeners' identifications of individual segments are biased by the lexical status of the items containing them, then we expect that the diGRESS-tiGRESS continuum, in which the /d/ endpoint is a word, would result in more /d/ responses than the DIgress-TIgress continuum.

One additional consideration is relevant. Acoustically, stress is typically signalled in the waveform by a complex pattern of information that includes duration, intensity and fundamental frequency [Gay, 1978]. Stressed syllables are typically longer in duration and have increased intensity and

higher fundamental frequency. In order to accomplish the contrast between a weak-strong sequence (WS, second-syllable stress) and a strong-weak sequence (SW, first-syllable stress) in the stimuli used in experiment 1, we manipulated the shape of the fundamental frequency contour and amplitude of voicing. In the SW stimuli, fundamental frequency reached its highest peak early in the first syllable, and was relatively low throughout the second syllable, By comparison, fundamental frequency of the WS stimuli was somewhat lower throughout the first syllable, but had its highest peak at the start of the second syllable. A number of researchers have demonstrated that fundamental frequency is in fact a cue to voicing [Haggard et al., 1970, Massaro and Cohen, 1976]. For example, Haggard et al. [1970] found that a high falling fundamental frequency produced more voiceless responses to a target ambiguous between /p/ and /b/ compared to the same stimulus with a low rising fundamental frequency contour. Similarly, Massaro and Cohen [1976] found that a falling fundamental frequency contour produced more voiceless responses than a rising fundamental frequency contour in a stimulus continuum ranging from /zi/ to /si/. Massaro and Cohen [1976] further demonstrated that the value of the fundamental frequency at syllable onset is the effective cue, not the fundamental frequency contour per se.

Thus, a shift in identification responses such that more /d/ (word) responses are obtained for the diGRESS-tiGRESS series compared to the Tigress-Digress series could potentially be explained by an acoustic bias due to fundamental frequency. In the first experiment, we attempted to control for this acoustically based explanation

of an obtained shift in identification functions by including an additional condition. Specifically, we presented listeners with just the initial syllable of each token in the SW and WS series. In this way, lexical information is no longer available and any effect found on identification responses in the one-syllable series may be attributed to fundamental frequency as a cue to voicing. If, for example, the WS dig-tig produced more /d/ responses than the SW DIG-TIG series, then this would suggest a role of fundamental frequency independent from lexical status. Alternatively, if comparable identification functions for the WS dig-tig and the SW DIG-TIG series are found, then any identification function shift in the two-syllable series is attributable to a true lexical effect.

Experiment 1

Materials

Eight tokens of each of the words *Tigress* and *diGRESS* (stress indicated by capitalization) were synthesized using the Klatt [1980] software synthesizer at the Laboratory of Experimental Psychology at Sussex University. The items with first-syllable stress will be referred to as the SW items and the items with second-syllable stress will be referred to as the WS items. A 10-kHz sampling rate was used (the rate used in all experiments to be reported), and the synthesized items were modeled on natural productions by a native speaker of British English. The base words *Tigress* and *diGRESS* were 670 ms long. The tokens along the SW and WS continua differed (apart from differences in voicing) only in their stress patterns, manipulated by varying fundamental frequency and amplitude of voicing parameters. Formant centers and other parameters were changed in a linear fashion between the inflection points shown in table I.

Voicing was manipulated by varying a complex acoustic parameter of the natural voiced/voiceless dimension, voice onset time (VOT), by 5-ms incre-

Table I. Parameters of stimuli: experiment 1

Time since start of word ms	Acoustic parameter													
	SW Tigress							WS diGRESS						
	AV	AF	AH	Fo	F1	F2	F3	AV	AF	AH	Fo	F1	F2	F3
0	0	0	0	146	380	1,700	3,000	0	0	0	138	380	1,700	3,000
15	50		52	150				54		52	138			
30	62							60			138			
50			57		580					57	123	580		
65						1,200	1,850						1,200	1,850
90					680							680		
155	60							59			118			
165				130										
170			55	135		2,000	2,200			55			2,000	2,200
185						2,100					110		2,100	
195	35					1,900		30					1,900	
200	0		32		380	1.170		0		32		380	1,170	
205			0							0				
250		0		135		1,170			0		110		1,170	
255		54			380	1,400	2,200		54		140	380	1,400	2,200
260					400							400		
275	0	50		112	400			0	50			400		
280	53	0	0					55	0	0				
295	60		0			1,200		62		0			1,200	
305			58				1,800			58				1,800
315						1,200	1,800						1,200	1,800
325											138			
340			5	108						5				
380					600							600		
405	56							58						
420					550	1,720	2,200					550	1,720	2,200
440		0							0					
445	30	35		110				30	35					
450	0			108				0						
620		32							32					
670	0	0	0	108	550	1,720	2,200	0	0	0	108	550	1,720	2,200

The changes in AV indicated as occurring at 15 and 30 ms actually occurred immediately after the end of the initial segment that was replaced with the natural /t/ sound. In addition to the parameters shown in the table, 52 dB of A6 were added from 445 ms to the end to create the /s/ sound, and 45-53 dB of A2-A5 were added from 255 to 275 ms to create the medial /g/ aspiration.

ments from 15 to 50 ms. The synthesized words included an initial acoustic segment 15-50 ms long. A waveform editor was used to replace the acoustic segment with an equal length of a digitized token of the t-burst and aspiration from the natural produc- tion of 'tigress', reduced 50% in amplitude. The natural and synthesized segments were joined at zero crossings. Immediately after the end of the initial segment, amplitude of voicing increased abruptly from 0 to 54 dB, and then increased over
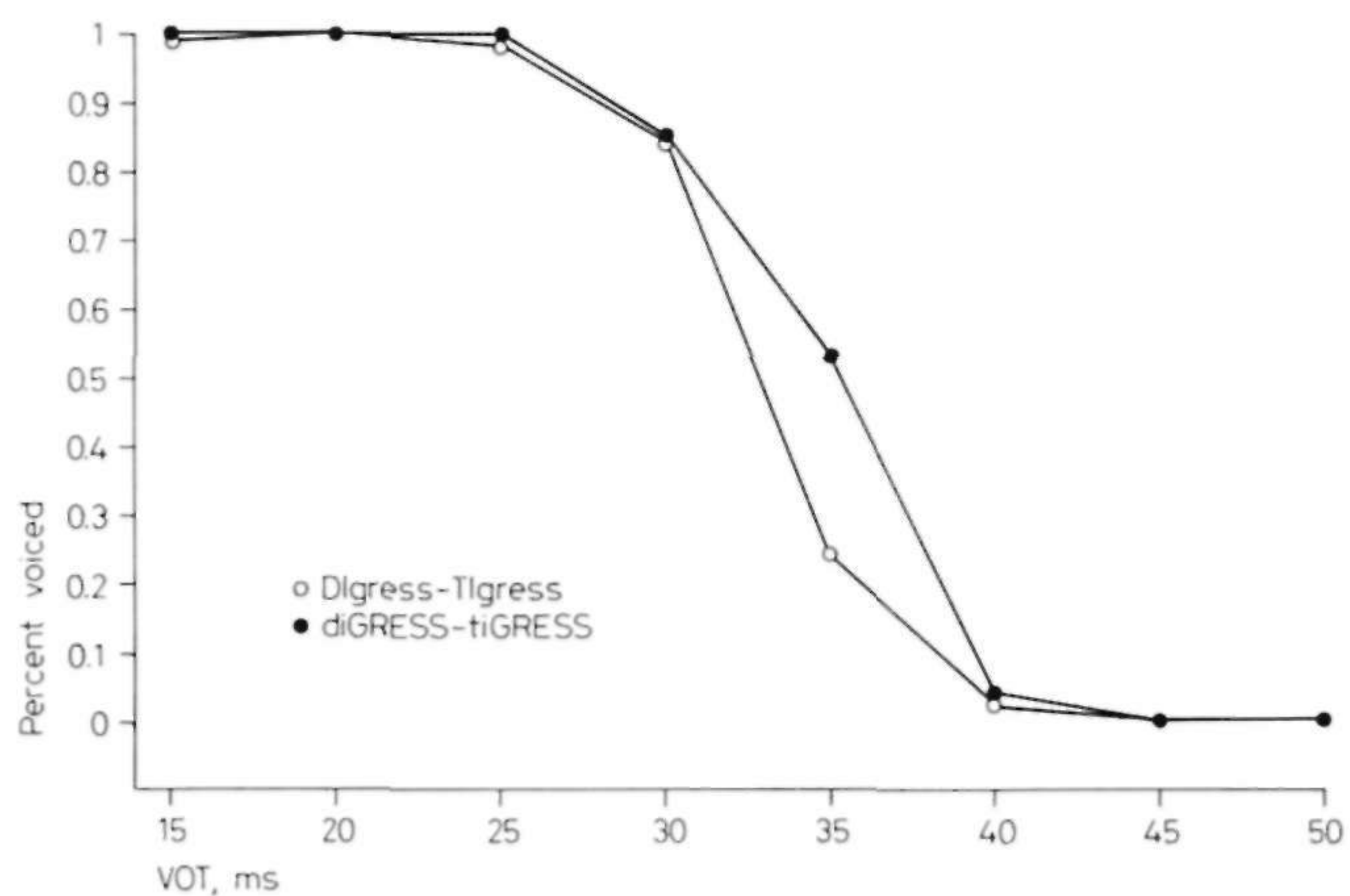
Fig. 1. Percent voiced (/d/) responses as a function of stimulus VOT (ms) for the two-syllable word stimuli. Experiment 1: di-GRESS/TIgress.

the next 20 ms to 62 dB (for SW stimuli) or 60 dB (for the WS stimuli). The 16 items created in this manner will be referred to as the WS and SW two-syllable word stimuli.

Each of the 16 items was then turned into a nonword by deleting the last 370 ms of each item. This resulted in two continua (the SW-TIG and WS-dig continua) of single-syllable nonwords tig-dig, one continuum having the amplitude of voicing and fundamental frequency characteristics of the first syllable of a SW item, and one continuum having the characteristics of the first syllable of a WS word. The 16 items created in this fashion will be referred to as the WS and SW one-syllable nonword stimuli. An additional two series of stimuli, tiger-diger and tiver-diver, were constructed and used in the experiment, but will not be discussed further.

Subjects and Procedure

Twelve members of the Sussex University community, all speakers of British English, were tested. They heard one practice block of 36 trials followed by 12 blocks of 32 trials each. On the practice block, they heard the six most extreme tokens of each of the six continua, in randomized order. On each following block, they heard four instances of each item from a single continuum. The first two of these blocks always presented the SW-dig and the WS-tig continua, counterbalanced in order across subjects. The other four continua were then presented, in a counterbalanced order, in the next four blocks. All six continua were presented again, each in its own block of trials, in the second half of the experiment.

Subjects heard the stimuli over headphones in a sound-damped chamber. A DEC VAX 780 computer presented the stimuli at a comfortable listening level, and controlled the timing of the experiment. Subjects were instructed to indicate whether each item began with a /t/ or a /d/ by pressing the appropriate key on a standard terminal keyboard. They were told to respond quickly and intuitively. A token was presented 1-2 s after the response to the previous stimulus was made.

Results and Discussion

The percentages of 'd' responses are plotted in figure 1 for the two-syllable word stimuli and in figure 2 for the one-syllable nonword stimuli. As can be clearly seen in figure 1, the identification function for the WS stimuli shows more 'd' responses than for the SW stimuli. Subjects more frequently made 'd' responses to the stimuli with second-syllable stress, thus forming a real word (diGRESS), than with first-syllable stress. No such displacement was apparent for one-syllable nonwords, as shown in figure 2.

To quantify the observed shift in identification responses, the VOT that corresponded to the point where the percentage of 'd' responses equalled 50% was deter-
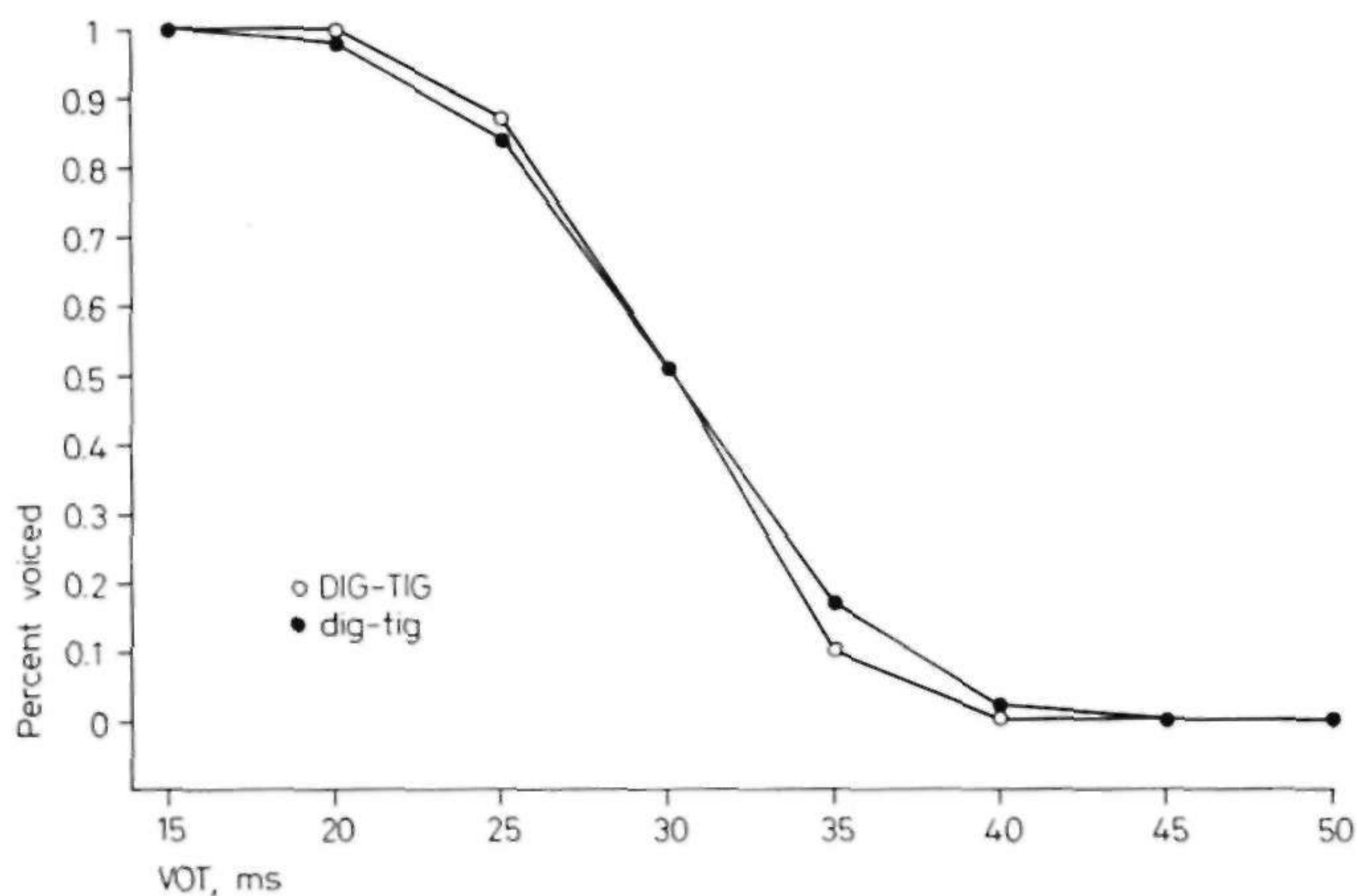
Table II. VOT values of the identification function midpoints: experiment 1

Stress	Continua	
	words (TIgress, diGRESS)	nonwords (SW TIG, WS dig)
SW	33.04	29.79
WS	34.76	30.03

mined for each subject's identification function. A straight line was fitted to the midrange of each function from the last point on the function that was greater than or equal to 85 % 'd' responses to the first point that was less than or equal to 15 % 'd' responses. The value of VOT at which this straight line function would result in 50% 'd' responses was determined.

The means of these midpoint values appear in table II. An analysis of variance indicated a significant effect (0.98 ms) of stress pattern [F(1,1 1) = 7.6, p<0.02] and a significant effect (3.99) of one vs. two syllables [F(1,11) = 17.0, p<0.01]. Subjects were more biased to report 'd' for the WS than for the SW stimuli, and for the two-syl-

lable than for the one-syllable stimuli. Although the interaction between number of syllables and stress pattern was nonsignificant [F(1,11) = 2.79, p>0.10], the effect of stress was significant only for the two-syllable words [1.72 ms, F(1,11) = 6.9, p<0.03] and not for the one-syllable nonwords [0.24 ms, F(1,11)<1].

It appears that fundamental frequency as an acoustic cue to voicing cannot alone account for the shift in the two-syllable series. It should be pointed out that while the data from the one-syllable nonword series appear on the surface to be at odds with the results of Massaro and Cohen [1976], it may be the case that the difference in fundamental frequency at syllable onset in our SW dig-tig and WS dig-tig stimuli may not have been extreme enough (a difference never greater than 12 Hz in the region that contains the consonant voicing information) to produce effects on perceived voicing. In the stimuli of Massaro and Cohen [1976], a difference of 30 Hz in starting value of fundamental frequency at the onset of voicing produced effects on identification.

On the basis of experiment 1, we tentatively conclude that lexical stress information can be used as a source of information to establish lexical status and can be used to influence identification of speech. The one-syllable control included in experiment 1 appears to rule out a simple acoustic cue-based explanation of the obtained shift n identification responses in the two-syllable series. However, the data do not unambiguously support this conclusion. Although simple effects tests revealed a significant effect only in the two-syllable series, the lack of a statistically significant interaction between one and two syllables in the two-way ANOVA is problematic. Our design, in which a given subject served in both one- and two-syllable conditions, may have inadvertently introduced additional error variability.

In addition, an alternative explanation, perceived stress effects that are unmediated by lexical effects, could account for the data. One version of a perceived stress effect is an effect due to perceived rate of speech. As previously mentioned, the contrast between strong and weak syllables is typically signalled acoustically by changes in fundamental frequency as well as in duration. In order to maintain durational equivalence across the SW and WS series, we did not manipulate the duration parameter as a cue. While the SW and WS series had the same objective first-syllable length, they may have differed in subjective length. The strong syllable in each series was in fact shorter than otherwise would be expected given that other cues (fundamental frequency and amplitude) were designed to signal a strong syllable. It is possible that subjects may have expected the strong syllable of the SW series to be longer than it

actually was and as a consequence perceived it as shorter, or as having been produced with a faster rate of speech. A number of researchers have demonstrated that the actual or perceived rate at which a syllable is uttered influences identification of voicing continua such that faster perceived utterance rates result in more voiceless responses [Summerfield, 1981; cf. Miller, 1981].

A perceived stress effect for our stimuli would result in relatively more /d/ (word) responses in the WS diGRESS-tiGRESS series compared with the SW Digress-Tigress series. Effects of perceived stress and lexical effects both account for the identification response shift obtained in the two-syllable series in experiment 1. In order to discriminate between these explanations, a second experiment was conducted that included a condition in which stress information was maintained but had no relevance to lexical status. If both factors are in operation in our lexical condition then we expect that the combination of lexical information and perceived stress will produce a larger effect on responses than in conditions where only perceived stress effects operate.

Experiment 2

Method

Eight tokens of the words Tigress and diGRESS, varying along the VOT continuum, were created just as in experiment 1, except that (1) the AH parameter of the Klatt synthesizer shown in table I was set to 0 throughout the length of the stimuli, (2) very minor adjustments of no more than 2 dB were made in the AV parameter, and (3) the final rise to 138 Hz in the F_0 parameter of the WS diGRESS stimuli was eliminated. Each of these 16 words was then turned into a nonword by deleting the last 255 ms (the en-
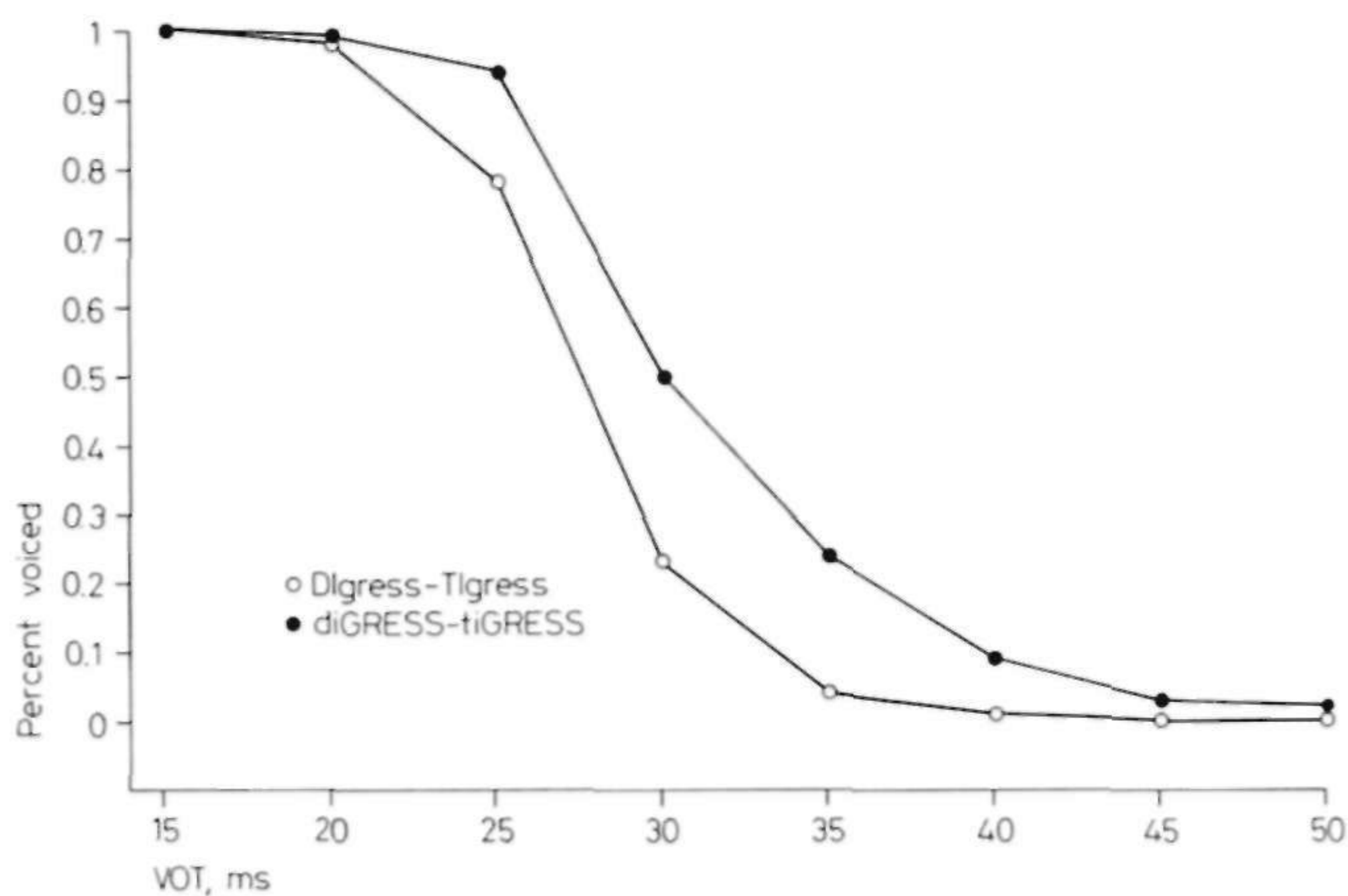
Fig. 3. Percent voiced (/d/) responses as a function of stimulus VOT (ms) for the real word stimuli. Experiment 2: diGRESS/TIgress.

tire length of the /s/) using a waveform editor. The resulting nonwords sounded like *tigre* and *digre* and were not identifiable as real words.

Two tapes, one of the SW Tigress and WS diGRESS words and one of the SW TIgre and WS diGRE nonwords, were recorded. Each tape included 20 instances of each of the 16 stimuli, in random order. Two practice tapes were also recorded. one containing two tokens of each of the endpoints of the word continua and one containing the corresponding tokens from the nonword continua.

Twenty-nine undergraduates at the University of Massachusetts were tested, 15 on the word tape (SW TIgress-WS diGRESS) and 14 on the nonword tape (SW TIgre-WS diGRE). One subject in the former group who did not label the stimuli consistently was discarded. Each subject was tested individually in a single 45-min session. The session began with the subject hearing the appropriate practice tape over headphones at a comfortable listening level and classifying each item as beginning with a /t/ or with a /d/ by pressing one of two buttons with the index finger of the right hand. The tape recorder was controlled by a microcomputer which also recorded responses and reaction times. At the end of the practice list, each subject was asked to say what items he or she had heard and to indicate whether each item was a real word or not. All subjects identified the SW Tigress and WS diGRESS endpoint stimuli as real words and the SW TIgre or WS diGRE endpoint stimuli as nonwords. Further, each subject

classified most or all of the endpoint stimuli on the practice tape correctly-

After practice, the subject heard the appropriate experimental tape and classified each item just as in practice. Subjects were instructed to respond as quickly as they found comfortable. They were told to base their responses on their intuitive reactions to the initial segment of the items, not to reflect upon what they had heard. No special emphasis was given to the word vs. nonword nature of the stimuli. Subjects were simply told that they would hear some computer-generated words and nonwords and should attend to the initial sound.

Results and Discussion

The mean percentages of 'd' responses appear in figure 3 for the real word stimuli and in figure 4 for the nonword stimuli. The curves for the WS stimuli are clearly displaced toward the /t/ endpoint of the scale relative to the SW stimuli, indicating that subjects more frequently made **'d'** responses to the stimuli with second-syllable stress than to stimuli with first-syllable stress.

The midpoints of the identification functions were determined as in experiment 1 and appear in table III. An analysis of vari-
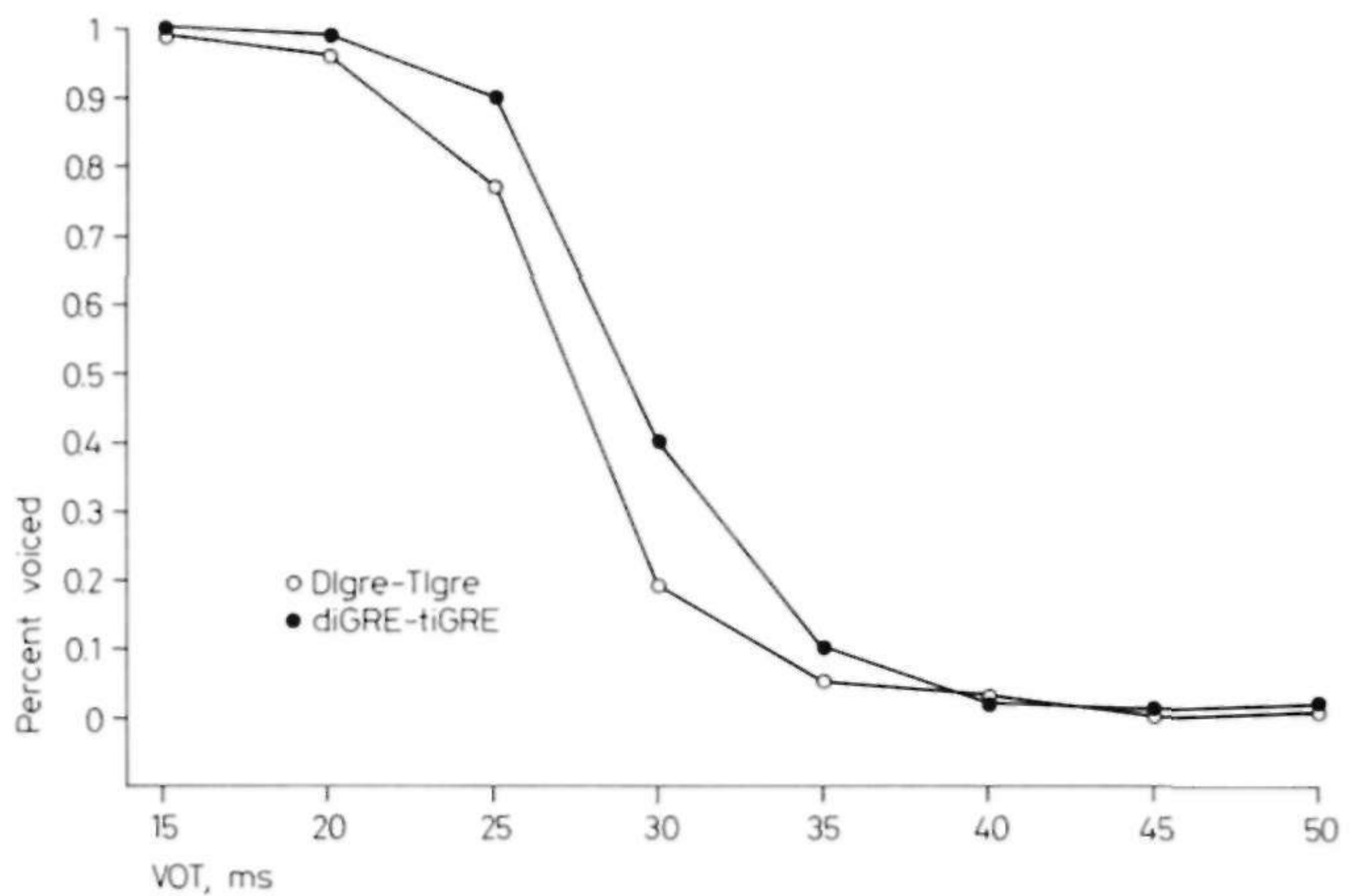
Fig. 4. Percent voice (/d/) responses as a function of stimulus VOT (ms) for the nonword stimuli. Experiment 2: diGRE/TIgre.

ance indicated a significant effect of stress pattern [2.85 ms, $F(1,26) = 63.37$, $p<0.001$] and a significant interaction between stress pattern and word/nonword continuum [$F(1,26) = 5.04$, $p<0.04$]. The difference between SW and WS was significant by a t test for both the word condition (3.65 ms) and the nonword (2.05) condition ($p<0.01$). The (between-groups) effect of word vs. nonword was nonsignificant [$F(1,26) = 1.91$, $p < 0.10$].

Subjects were more willing to produce /d/ identifications for WS stimuli than for SW stimuli, in that they persisted in reporting WS stimuli as beginning with a /d/ at longer VOT values. The presence of a shift in identification functions in the nonword condition indicates that a simple perceptual adjustment effect is influencing responses in both the word and nonword SW and WS stimuli. Importantly, however, the identification function shift was 1.6 ms greater in the word condition compared with the nonword condition. That is, effects on identification responses were larger when a /d/ response made the WS stimuli result in a real

Table III. VOT values of the identification function midpoints: experiment 2

Stress	Continua	
	words (Tigress, diGRESS)	nonwords (TIgre. diGRE)
SW	27.76	27.28
WS	31.41	29.33

word than when only nonwords were involved. This pattern of data suggests that two factors were involved in the word condition: the lexical effect and possibly a perceived stress effect.

Experiment 3

We have argued that lexical stress influences identification responses over and above perceived stress effects. An alternative way in which to test this hypothesis is to conduct an experiment in which lexical status is established in stimuli in which lexical effects are in opposition to effects of

Table IV. Parameters of stimuli: experiment 3

| Time since start of word ms	SW briefer						WS prefer					
	AV	AF	F_0	F_1	F_2	F_3	AV	AF	F_0	F_1	F_2	F_3
0	0	62	149	350	1,050	1,400	0	62	138	350	1,050	1,400
15	54	0					54	0				
25	62						59		138			
50									123			
115					1,800	2,600					1,800	2,600
140	60						54					
160									118			
170			130						110			
185					1,800						1,800	
200	35	0		360			35	0		360		
205	0	56			1,200	2,080	0	56			1,200	2,080
225					1,100						1,100	
235						2,080						2,080
240		60						60	110			
250									140			
260		130										
275	0											
280	53	60	112		1,270	1,540	54	60			1,270	1,540
285		0						0				
300		0					62					
325				470	1,270						1,270	
360			108									
400	56											
450	0		110									
490			108	420	1,310	1,540	50		108	420	1,310	1,540

perceived stress. In our third experiment, a pair of words that differed in the voicing value of the initial stop consonant and in stress pattern was selected (BRIEFer and preFER). This pair of stimuli differed from the diGRESS/TIgress pair used in experiments 1 and 2 in that the real word member of the pair with initial syllable stress began with a voiced consonant (/b/). Thus, lexical status information, a tendency for word or voiced responses in a SW environment, conflicts with perceived stress effects, a tendency for voiceless responses in a SW environment. If the combined effects of per-

ceived stress and lexical information account for the pattern of data found in experiment 2, then under conditions when these two sources of information conflict the net effect on identification responses will be reduced.

Method

Materials. In addition to the eight tokens of the SW Tigress series and the WS diGRESS series used in experiment 1, eight tokens of the SW word BRIEFer and eight of the WS word preFER were synthesized using the Klatt software formant synthesizer at Northeastern University. The synthesizing procedures used were the same as were used for

Fig. 5. Percent voiced (/d/) responses as a function of stimulus VOT (ms) for the diGRESS/TIgress stimuli. Experiment 3: diGRESS/TIgress.
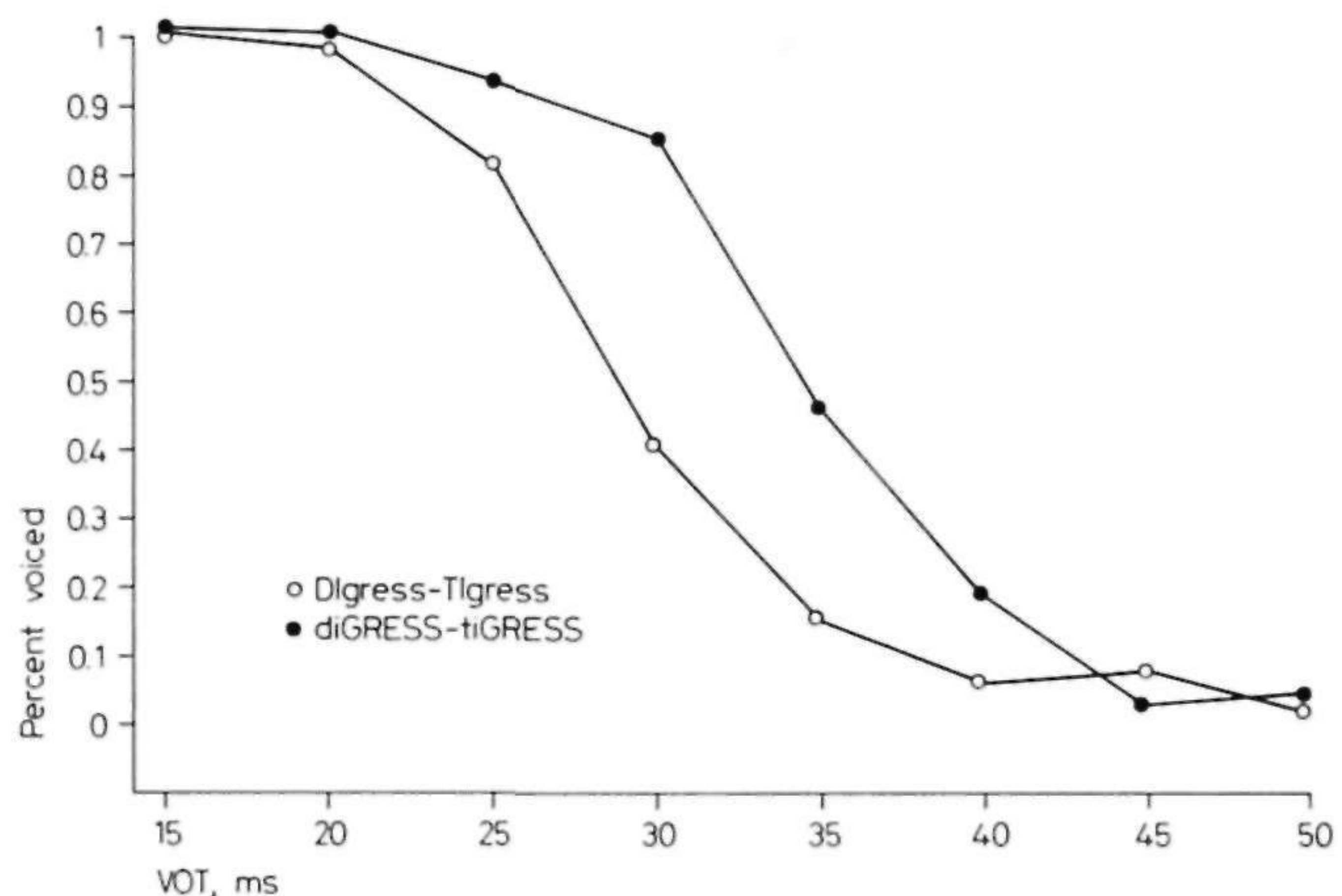
the stimuli of experiment 1. The inflection points of the synthesis parameters appear in table IV, which illustrates the stimuli with 15-ms VOT. As in experiment 1, tokens were made at 5-ms intervals from 15- to 50-ms VOT to create a continuum whose initial segments varied from a clear /b/ to a clear /p/. Two tapes were recorded containing ten randomly ordered occurrences of each of the 32 tokens. In addition, a pretest tape was recorded that contained four occurrences of each of the 32 tokens. Each stimulus was preceded by 400 ms by a brief, soft warning tone, and there were 2,000 ms between stimuli.

Subjects and Procedure. Seventeen undergraduates and graduate students at the University of Massachusetts were pretested using the procedures in experiment 2 except that they were told to press one button for either 'd' or [4]'b' and the other button for either Y or 'p\ Five subjects were rejected because they failed to show regular identification functions (e.g. they failed to consistently identify the endpoints of a continuum). Half of the remaining 12 subjects were tested on each of the two experimental tapes, again using the procedures of experiment 1.

Results

The mean percentages of voiced responses ['d' or 'b') appear in figure 5 for the TIgress-diGRESS stimuli and in figure 6 for the BRIEFer-preFER stimuli. The 50°/o

points of the functions for the individual subjects were estimated as in experiment 1. The means of these midpoint values appear in table V, separated by the contrast between /t/-/d/ and /b/-/p/ stimuli and the contrast between SW and WS stimuli. An analysis of variance with these factors indicated a significant interaction between the two factors [F(1,11) = 34.64, p<0.01] together with significantly longer VOT for the /t/-/d/ than the /b/-/p/ stimuli (31.91 vs. 27.65, respectively) [F(1,1 1) = 7.98, p<0.02] and a nearly significant effect of SW vs. WS (29.06 vs. 30.5 ms, respectively) [F(1,11) = 4.32, p<0.06]. A separate analysis conducted on the /t/-/d/ stimuli indicated that a significantly longer VOT was required for the WS than for the SW stimuli [3.53 ms, F(1,11) = 16.53, p<0.01]. A comparable analysis for the /b/-/p/ stimuli yielded an F< 1 (a difference of 0.65 ms).

The stress manipulation resulted in a significant shift in identification functions for the tigress/digress stimuli but not for the briefer/prefer stimuli. These data replicate the original finding in experiment 1 and experiment 2 and are consistent with the hy-
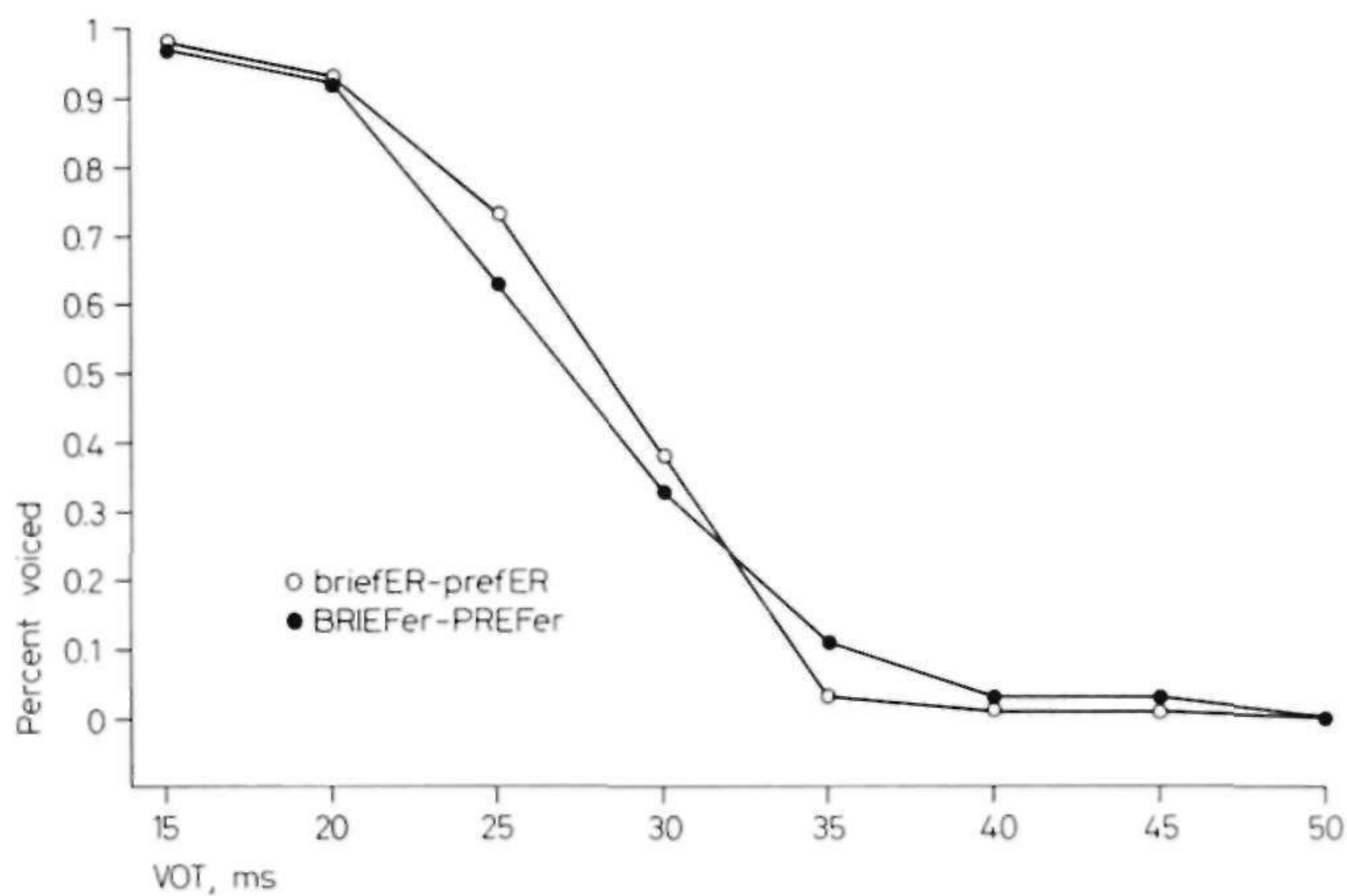
Fig. 6. Percent voiced (/b/) responses as a function of stimulus VOT (ms) for the BRIEFer/ prefER stimuli. Experiment 3: BRIEFer/prefER.

Table V. VOT values of the identification function midpoints

Stress	Continua	
	/t/–/d/ (TIgress, diGRESS)	/b/–/p/ (preFER, BRIEFer)
SW	30.15	27.98
WS	33.68	27.33

pothesis that two sources of information (perceived stress and lexical information) affected responses in experiment 2. In one case (experiment 3), lexical information combined with perceived stress effects counteracted one another resulting in an overall null effect on identification responses. Under conditions when both kinds of information are in accord with one another (experiment 2), the overall effect was larger than given one source of information alone (perceived stress).

General Discussion

The experiments presented here demonstrate that lexical stress information is lexically present and can be used as an additional source of information in auditory word recognition. Further, effects of lexical status are not simply due to acoustic biases either of fundamental frequency or of perceived stress. Experiment 1 showed no significant effect of a potential fundamental frequency-based bias. Experiment 2 demonstrated that the combined effects of the potential perceived stress factor (unmediated by lexical stress) and lexical status are larger than effects of perceived stress alone. Experiment 3 showed that under conditions in which perceived stress and lexical status conflict, neither effect proves to prevail over the other.

Clearly, these experiments support the involvement of lexical stress in auditory

word recognition. What remains at issue is the process that is influenced by lexical stress information. A number of mechanisms are possible to account for the effect of lexical stress on identification of speech, One possibility is that lexical stress information is used very early in speech perception to directly influence perceptual processing. In this characterization of lexical stress influences, activation of a lexical entry and presumably its canonical stress pattern is used to feed back to developing segmental hypotheses. The additional evidence supplied by the lexical level for word-initial ambiguous acoustic information (in the category boundary region) is used interactively to contribute to processes at a perceptual level of processing. Connine and Clifton [1987] have argued for an interactive account of lexical status effects on speech perception [Connine, 1987]. It was suggested that perceptual processes are directly influenced by prestored information.

This hypothesis contrasts with the demomstration by Cutler [1986] described in the introduction that stress information is not used prelexically. Cutler further argues that there is little premium in computing lexical stress prelexically, for two reasons. First, segmentally identical but suprasegmentally distinct pairs are rare in language and segmental information will nearly always be a sufficient determinant of a word's identity. Second, computation of lexical stress in the recognition of continuous speech may be costly, in that a word's stress pattern cannot be determined without knowing where the word begins and ends, and location of word boundaries in continuous speech is difficult. The notion that lexical stress does not contribute to the computation of the prelexical access code implies that lexical stress is used after segmental hypotheses have been established. Given perceptually ambiguous information, lexical stress information can be used to resolve the ambiguity in favor of a word.

The data presented here do not crucially distinguish between these two classes of mechanisms as explanations for the role of lexical stress in phonetic categorization. Further research is needed to determine the precise mechanism underlying effects of linguistic knowledge on perceptual processes. Such data will provide further insight into the basic architecture and organization of language processing. However, the present results do allow the firm conclusion that stress information *must* be represented in lexical memory. It is not possible to explain the results without the assumption that lexical stress information is directly relevant to the establishment of lexical status in perception.

Acknowledgements

References

Bond, Z. S.,; Small, L. H.: Voicing, vowel and stress mispronunciations in continuous speech. Perception Psychophysics *34:* 470-474 (1983).

Brown, R.; McNeill, D.: The ᵏtip of the tongue' phenomenon. J. verbal Learn, verbal Behavior *5:* 325-337 (1966).

Connine, CM.: Constraints on interactive processes in auditory word recognition: the role of sentence context. J. Memory Lang, (in press, 1987).

Connine, CM.; Clifton, C, Jr.: Interactive use of lexical status in speech perception. J. exp. Psychol. hum. Perception Performance (in press, 1987).

Cutler, A.: Forbear is a homophone: lexical prosody does not constrain lexical access. Lang. Speech 29:201-220(1986).

Cutler, A.; Clifton, C, Jr.: The use of prosodic information in word recognition; in Bouma, Bouwhuis, Attention and performance, vol. X, pp. 183-196 (Erlbaum, Hillsdale 1984).

Ganong, W. F.: Phonetic categorization in auditory word recognition. J. exp. Psychol, hum. Perception Performance *6:* 110-125 (1980).

Gay, T.: Physiological and acoustic correlates of perceived stress. Lang. Speech *21:* 347-353 (1978).

Haggard, M.; Ambler, S.; Callow, M.: Pitch as a voicing cue. J. acoust. Soc. Am. *47:* 613-617 (1970).

Klatt, D.H.: Software for a cascade/parallel formant synthesizer. J. acoust. Soc. Am. *67:* 971-995 (1980).

Massaro, D.M.; Cohen, M.M.: The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. J. acoust. Soc. Am. *60:* 704-707 (1976).

Miller, J.L.: Effects of speaking rate on segmental distinctions; in Eimas, Miller, Perspectives on the study of speech, pp. 39-74 (Erlbaum, Hillsdale 1981).

Slowiaczek, L. M.: Effects of lexical stress placement on auditory word recognition. Acoust. Soc. Am. Meet., Anaheim 1986).

Summerfield, Q.: Articulatory rate and perceptual constancy in phonetic perception. J. exp. Psychol. hum. Perception Performance 7: 1074-1095 (1981).

Cynthia M. Connine, Ph.D.
Department of Psychology
University Center at Binghamton
State University of New York
Binghamton, N.Y. 13901 (USA)