Published in final edited form as:

J Acoust Soc Am. 2006 April ; 119(4): EL55–EL59.

Polling the effective neighborhoods of spoken words with the verbal transformation effect

James A. Bashford Jr., Richard M. Warren, and Peter W. Lenz

Department of Psychology, University of Wisconsin—Milwaukee, PO Box 413, Milwaukee, Wisconsin 53201-0413

Abstract

Studies of the effects of lexical neighbors upon the recognition of spoken words have generally assumed that the most salient competitors differ by a single phoneme. The present study employs a procedure that induces the listeners to perceive and call out the salient competitors. By presenting a recording of a monosyllable repeated over and over, perceptual adaptation is produced, and perception of the stimulus is replaced by perception of a competitor. Reports from groups of subjects were obtained for monosyllables that vary in their frequency-weighted neighborhood density. The findings are compared with predictions based upon the neighborhood activation model.

1. Introduction

Most modern theories of spoken word recognition [TRACE (McClelland and Elman 1986), SHORTLIST (Norris, 1994), MERGE (Norris *et al.*, 2000), NAM (Luce and Pisoni, 1998), and PARSYN (Luce *et al.*, 2000)] assume that acoustic-phonetic input activates a set of structurally similar verbal representations in memory, which compete with one another in the process of speech perception. These activation-competition models of speech perception are supported by evidence from a variety of paradigms (e.g., shadowing, perceptual identification, and lexical decision), which have shown that processing speed and accuracy for a verbal stimulus are influenced by both its neighborhood density (i.e., the number of structurally similar, lexical neighbors) and by its neighborhood frequency (the sum of the word frequencies of its lexical neighbors). Words having large numbers of high-frequency lexical neighbors are generally processed more slowly and less accurately than words having only a few, low-frequency lexical neighbors (Goldinger *et al.*, 1989; Cluff and Luce, 1990; Luce *et al.*, 1990; Luce and Pisoni, 1998; Vitevitch and Luce, 1998, 1999).

The present study employs a novel procedure for directly accessing the neighborhoods of spoken words. Studies dealing with neighborhood effects typically employ a computational procedure that considers the effective neighbors of a stimulus to be those words in the lexicon that differ from the stimulus by the addition, deletion, or substitution of a single phoneme (Landauer and Streeter, 1973; Luce and Pisoni, 1998). Although this *a priori* method of identifying competitors has proven capable of predicting stimulus differences in word recognition, it likely provides only an approximation to the effective neighborhood of a stimulus (Luce and Pisoni, 1998, p. 16) and hence may benefit from further experimental refinement. The present study pursues this goal, using an auditory illusion known as the verbal transformation effect (Warren, 1961), which can be used to induce listeners to perceive and call out the effective neighbors of the stimulus word or syllable.

When listeners are presented with a clear recording of a syllable or word that is repeated over and over without change, they typically hear an abrupt and compelling illusory change to a

different word or syllable (Warren and Gregory, 1958; Warren, 1961). Verbal transformation (VT) appears to result from the operation of two concurrent processes (Warren, 1976): (1) a repetition-induced adaptation effect that progressively lowers the activation level of the initially dominant neural representation corresponding to veridical perception and (2) a repetition-induced verbal summation¹ that progressively increases the activation level of the most salient of the neural representations that are structurally similar to the stimulus. VT is considered to occur when the diminishing activation level of the perceived stimulus representation is exceeded by that of the next-most highly activated representation. Hence, it appears that VT may offer a means of directly accessing the sets of representations that compete in the process of spoken word recognition. The present study provides a preliminary test of this hypothesis. In order to maximize neighborhood effects upon transformations, neighborhood density and word frequency were manipulated conjointly, with half the stimuli having large numbers of high-frequency neighbors and half having only a few, low-frequency neighbors. The focus of analysis in this initial study is upon listeners' initial transformations reported for each stimulus, categorized in terms of lexicality, word frequency, and structural similarity to the stimulus.

2. Method

The 144 listeners in this study (three groups of 48) were undergraduate student volunteers from the University of Wisconsin—Milwaukee who were paid for their participation. All listeners were native monolingual English speakers who reported having no hearing problems and had normal bilateral hearing, as measured by pure tone thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz.

The test stimuli were 12 monosyllables that were selected to provide three replications of a 2×2 factorial crossing of stimulus lexicality and frequency-weighted neighborhood density (FWND), calculated using the procedure of Newman *et al.* (1997). This selection yielded three exemplar stimuli in each of four stimulus conditions: (1) the high-density lexical monosyllables (i.e., words with many neighbors), “boat,” “cane,” and “let,” having log word frequencies [$\log_{10}(\text{frequency})+1$] of 2.86, 2.15, and 3.58, respectively, and having neighborhood density scores of 32, 33, and 30, and FWND scores of 67.7, 67.2, and 72.1, respectively; (2) the high-density nonlexical monosyllables, “dake,” “leet,” and “nane,” having density scores of 25, 40, and 25, and FWND scores of 53.3, 82.4, and 53.2, respectively; (3) the low-density lexical monosyllables, “watch,” “gouge,” and “jibe,” having word frequency scores of 1, 1, and 2.91, respectively, and having neighborhood density scores of 5, 3, and 4, and FWND scores of 11.5, 5.7, and 6.4, respectively; and (4) the low-density nonlexical monosyllables, “powsh,” “chibe,” and “gouk,” having density scores of 5, 3, and 1, and FWND scores of 9.4, 4.2, and 1.8, respectively. An additional monosyllabic word, “lean,” was produced for use as a warm-up stimulus that preceded the experimental stimuli.

The stimuli were digitally recorded (44.1-kHz sampling, 16-bit quantization) by a male speaker having an average voicing frequency of approximately 100 Hz and no obvious regional accent, who produced the 500-ms monosyllabic stimuli. An additional 250-ms segment of digital silence was then appended to each capture to emphasize syllable boundaries and minimize any tendency for perceptual resegmentation of the stimuli (for example, when the stimulus “ace” repeats without an appreciable silent gap between repetitions, it is readily reparsed perceptually

¹Verbal summation was described and named by Skinner (1936), who employed a recording of repeating faint and indistinct speech sounds, which became organized into illusory words with continued listening. A more recent example of verbal summation, based upon clear speech sounds was reported by Warren *et al.* (1990), who employed a recording of repeating loud and clear sequences of three or more brief (30–100ms) isochronous steady-state vowels and reported that listeners could not identify the brief vowels, but after several repetitions of the sequences reported hearing syllables and words. Subsequently, Warren *et al.* (1996) reported that spectrograms showed a structural similarity between the vowel sequences and listeners' synchronous productions of the verbal forms being heard.

to yield “say”). Although this is an interesting phenomenon in its own right, this type of transformation could lead to the activation of neighborhoods other than those intended for the present study. The 500-ms stimulus, together with the added 250-ms silent gap, was digitally iterated to produce a 4-min test stimulus that provided 320 repetitions of the monosyllable.

The 144 listeners were randomly divided into three groups of 48, and each group of listeners received five VT stimuli, which included the initial, warm-up stimulus “lean” and one stimulus drawn from each of the four experimental conditions described above. The assignment of specific experimental stimuli to groups followed the order listed in the stimulus section: That is, the first group of listeners received the experimental stimuli boat, dake, watch, and powsh; the second group received cane, leet, gouge, and chibe; and the third group received the experimental stimuli let, nane, jibe, and gouk. The order of presentation for the experimental stimuli was pseudorandom, with the restriction that each stimulus was presented an equal number of times in each serial position across listeners in each group.

Testing was performed in a sound-attenuating chamber, with the VT stimuli delivered diotically through Sennheiser HD 250 Linear II Headphones at a slow-rms peak level of 70 dBA SPL, as measured using a Brüel & Kjaer model 2230 precision integrating sound level meter. For each VT stimulus presented, listeners were instructed to call out what the voice was saying at the stimulus onset, and then to call out what the voice was saying any time a change was heard. Listener’s responses were transcribed by the experimenter and also audiotaped for subsequent verification.

3. Results

Out of 576 experimental trials (144 listeners $\times$ 4 experimental stimuli), a total of 519 trials, or 90.1%, yielded at least one verbal transformation, and hence provided initial illusory forms data for use in this study (median time to first change was about 30 s). The percentages of trials producing at least one change for the four experimental conditions ranged from 86.8% to 92.4% and did not differ significantly by Z tests, $Z \leq 1.52$, $p \geq 0.129$, for the significance of a difference between proportions (Bruning and Kintz, 1977). All transforms reported by our monolingual English listeners were phonotactically legal. For the primary data analysis, transformations were categorized as one of four types: (1) lexical neighbors (words differing from the stimulus by a single phoneme), (2) lexical nonneighbors (words differing by two or more phonemes), (3) nonlexical neighbors, and (4) nonlexical nonneighbors. Table I presents the percentages of these four types of forms reported for the stimuli in each of the experimental conditions. The average log transformed word-frequencies (Kucera and Francis, 1967) for lexical forms reported in each condition are also included in parentheses.

The most frequently reported initial transformations were lexical neighbors of the stimuli (constituting about 45% of responses overall). However, it can be seen in Table I that this percentage varied substantially as a function of stimulus frequency-weighted neighborhood density (FWND). Lexical neighbors comprised about 75% of the transformations evoked by high-FWND stimuli, but comprised only about 14% of those reported for the low-FWND stimuli. This difference in percentages was significant for both the lexical stimuli, $Z=8.69$, $p<0.0001$, and the nonlexical stimuli, $Z=11.24$, $p<0.0001$. In contrast, stimuli that were low rather than high in FWND produced a much larger percentage of lexical nonneighbors, with overall averages of about 47% vs. 7%. This difference in percentages also was significant for both the lexical stimuli, $Z=6.16$, $p<0.0001$, and nonlexical stimuli, $Z=8.57$, $p<0.0001$. It is also of interest that the low- and high-FWND stimuli did not differ significantly in the time required to evoke VT, $F(1,143)=0.12$, $p>0.7$. This suggests that the lexical nonneighbors reported predominantly for low-FWND stimuli were equivalent in their salience to that of the lexical neighbors reported for the high-FWND stimuli. Finally, about 19% of reported VT forms were

nonlexical neighbors of the stimuli. The majority of these forms (72%) involved the external addition of a vowel or consonant in initial or final position, and thus preserved the component phonemes of the stimulus. About 22% involved the substitution of a consonant or vowel, and these forms were most frequently reported for nonlexical stimuli. Deletion of an individual consonant did occur, but rarely (6%), and then only with nonlexical stimuli.

4. DISCUSSION

The predominance of lexical neighbors obtained as transforms for the high-FWND stimuli is consistent with the neighborhood activation model (NAM), which currently considers that words differing from the stimulus by a single phoneme are its competitors during word recognition. However, the large percentage of lexical nonneighbors obtained as transforms for the low-FWND stimuli (typically differing from the stimuli by two phonemic changes) suggests that the “minimum-phoneme-distance rule” may be too restrictive, especially for stimuli having only a few, low-frequency neighbors. The present results also suggest that the dominance of lexical nonneighbors as transforms for the low-FWND stimuli is due to another factor conventionally considered critical in word recognition: competitor word frequency. Not surprisingly, the log word-frequency scores for lexical neighbors reported as transforms for the low-FWND stimuli were correspondingly low, with an average score of 1.37 (standard error=0.11). In contrast, the corresponding score for lexical nonneighbors was 2.93 (standard error=0.13). Lexical nonneighbors reported for the high-FWND lexical stimuli, though constituting a smaller percentage of VTs, showed a similarly high-word-frequency score, with an average value of 3.11 (standard error=0.29). Thus, it appears that word frequency may, under some conditions, override structural similarity in determining the effective neighborhood of a stimulus.

The results of this initial study suggest that the VT effect may indeed serve as a useful tool for directly accessing and identifying the salient competitors for lexical and nonlexical stimuli. In particular, the present study suggests that the number of high-frequency “remote neighbors” of a stimulus (i.e., those differing by two phonemes) may also contribute to measurable differences in processing speed and accuracy, especially for stimuli having sparse, low-frequency lexical neighborhoods, as conventionally calculated.

Acknowledgements

This work was supported by Grant No. DC 000208 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References and Links

- Bruning, J.L.; Kintz, B.L. Computational Handbook of Statistics. Scott Foresman; Glenview, IL: 1977.
- Cluff MS, Luce PA. Similarity Neighborhoods of Spoken Two-Syllable Words: Retroactive Effects on Multiple Activation. *J Exp Psychol Hum Percept Perform* 1990;16:551–563. [PubMed: 2144570]
- Goldinger SD, Luce PA, Pisoni DB. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *J Mem Lang* 1989;28:501–518.
- Kucera, H.; Francis, W.N. Computational Analysis of Present-Day American English. Brown U. P.; Providence, RI: 1967.
- Landauer T, Streeter LA. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *J Verbal Learn Verbal Behav* 1973;12:119–131.
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear Hear* 1998;19:1–36. [PubMed: 9504270]
- Luce, PA.; Pisoni, DB.; Goldinger, SD. Similarity neighborhoods of spoken words. In: Altmann, GTM., editor. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. MIT; Cambridge, MA: 1990.

- Luce PA, Goldinger SD, Auer ETJ, Vitevitch MS. Phonetic priming, neighborhood activation, and PARSYN. *Percept Psychophys* 2000;62:615–662. [PubMed: 10909252]
- McClelland JL, Elman JL. The TRACE model of speech perception. *Cogn Psychol* 1986;18:1–86.
- Newman RS, Sawusch JR, Luce PA. Lexical neighborhood effects in phonetic processing. *J Exp Psychol Hum Percept Perform* 1997;23:873–889. [PubMed: 9180048]
- Norris D. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 1994;52:189–234.
- Norris D, McQueen JM, Cutler A. Merging information in speech recognition: Feedback is never necessary. *Behav Brain Sci* 2000;23:299–325. [PubMed: 11301575]
- Skinner BF. The verbal summator and a method for the study of latent speech. *J Psychol* 1936;2:71–107.
- Vitevitch MS, Luce PA. When words compete: Levels of processing in perception of spoken words. *Psychol Sci* 1998;9:325–329.
- Vitevitch MS, Luce PA. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J Mem Lang* 1999;40:374–408.
- Warren RM. Illusory changes of distinct speech upon repetition—the verbal transformation effect. *Br J Psychol* 1961;52:249–258. [PubMed: 13783239]
- Warren, RM. Auditory illusions and perceptual processes. In: Lass, NJ., editor. *Contemporary Issues in Experimental Phonetics*. Academic; New York: 1976.
- Warren RM, Gregory RL. An auditory analogue of the visual reversible figure. *Am J Psychol* 1958;71:612–613. [PubMed: 13571475]
- Warren RM, Bashford JA Jr, Gardner DA. Tweaking the lexicon: Organization of vowel sequences into words. *Percept Psychophys* 1990;47:423–432. [PubMed: 2349054]
- Warren RM, Healy EW, Chalikia MH. The vowel-sequence illusion Intrasubject stability and intersubject agreement of syllabic forms. *J Acoust Soc Am* 1996;100:2452–2461. [PubMed: 8865650]

TABLE I

Types of initial verbal transformations reported for lexical and nonlexical stimuli having high- versus low-frequency-weighted neighborhood density (FWND). Percentages are given for the four categories of transformations reported in each experimental condition (see text). Values in parentheses are mean log transformed word-frequency scores for lexical forms reported [$\log_{10}(\text{frequency})+1$]. The values of N in the table are the numbers of trials, out of 144, producing at least one verbal transformation for a given type of experimental stimulus.

	Lexical neighbor	Lexical nonneighbor	Nonlexical neighbor	Nonlexical nonneighbor
Words				
High FWND $N=129$	61.2% (3.1)	13.2% (3.1)	15.5%	10.6%
Low FWND $N=132$	9.9% (1.5)	48.5% (3.6)	25.0%	16.7%
Nonwords				
High FWND $N=125$	88.0% (2.6)	0.0% (–)	4.8%	7.2%
Low FWND $N=133$	18.0% (1.3)	45.1% (2.4)	28.6%	8.3%