LANGUAGE AND COGNITIVE PROCESSES, 2001, *16* (5/6), 715–721

The source of a lexical bias in the Verbal Transformation Effect

Mark A. Pitt and Lisa Shoaf
Ohio State University, Columbus, USA

Comparisons of responses to words and pseudowords figure prominently in our efforts to understand how spoken words are recognised. This was apparent in many of the presentations at the SWAP meeting (e.g., Fowler & Brancazio, 2000; Frauenfelder & Content, 2000). Differences in responding to words and pseudowords are assumed to demonstrate the influence of lexical memory in the task at hand, from which we infer the structural and functional characteristics of the processing system.

Each task we use in this enterprise is often best suited to address a subset of theoretical issues. For example, phoneme detection and identification are used most often to examine issues about prelexical processing, such as whether there is lexical feedback and what the structure of phonetic categories is. Lexical decision and other word-based tasks tend to be used to explore the properties of lexical memory. So when a lexical effect is identified in an unfamiliar task, it is important to determine what we can learn about word processing from it. This has been the goal of our work into the origin of word-pseudoword differences in the Verbal Transformation Effect (VTE).

THE LEXICAL EFFECT

The VTE is an auditory illusion in which listeners report hearing illusory utterances after listening to a word repeat over and over at a rapid rate. Listeners respond on the fly, repeating illusory percepts into a microphone.

Requests for reprints should be addressed to Mark Pitt, Department of Psychology, 1885 Neil Avenue, Columbus, OH 43210-0222. E-mail: pitt.2@osu.edu

We thank Lyn Canterbury and Kalyani Subramaniam for help in scoring data. Rochelle Newman and Jean Vroomen provided constructive feedback on an earlier version of the manuscript. Sample stimuli are available at http://lpl.psy.ohio-state.edu.

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000056

Transformations are reported throughout a trial, with new ones being introduced and others being rereported, including the veridical (i.e., intended) word itself.

The first series of experiments (Shoaf & Pitt, submitted) focused on establishing the validity and reliability of the lexical effect in the VTE, and then probing the data further to identify its cause. Natsoulas (1965) was the first to study word-pseudoword differences in the VTE in depth. He measured the number of unique transformations (different verbal reports, irrespective of their frequency) that listeners reported, and found that pseudowords yielded more unique transformations than words.

We began by replicating and extending this finding to a wider range of stimuli under more controlled experimental conditions. Monosyllabic and disyllabic words and pseudowords were phonetically matched across conditions as closely as possible so that differences in performance could be attributable to differences in lexical status rather than to other properties of the stimuli. For the monosyllables, one or both consonants of an initial cluster were substituted (e.g., ''skunk''–''swunk''). For the disyllables, this involved swapping the order of the two syllables (e.g.,''center''–''tercen''). Stimuli were presented binaurally over headphones for 350 repetitions with an interstimulus interval (ISI) of 250 ms. Listeners (31) reported transformations into a microphone that were recorded onto cassette tape for later analysis.

For both the monosyllables and disyllables, listeners reported reliably more unique pseudoword transformations than word transformations. This effect was not due to a difference in the frequency with which transformations are reported to the two types of stimuli, as the total number of transformations was not consistently greater for pseudowords than words. Further analysis of the data provided insight into the cause of the lexical effect. Transformations were categorised as a function of whether they were reports of the veridical percept itself (i.e., the original utterance) or nonveridical percepts (i.e., word or pseudoword illusory transformations). When the repeating stimulus was a word (monosyllable or disyllable), it transformed back into the veridical stimulus reliably more often than when it was a pseudoword. However, nonveridical reports showed no such lexical bias, with illusory pseudoword transformations being more frequent than illusory word transformations. This finding, in addition to accounting for why fewer unique transformations are reported to words, suggests that lexical memory facilitates the veridical perception of the recycling stimulus. To understand the processes that underlie this lexical effect, it is first necessary to develop a better understanding of what causes verbal transformations. This was the objective of the following work.

CAUSES OF VERBAL TRANSFORMATIONS

One clue as to a cause of the VTE came from inspection of the transformations listeners reported to monosyllables that began with /s/ clusters, like ''skunk'' and ''swunk''. The /s/ was omitted in some of the transformations, such as ''gunk'' and ''wunk''. We hypothesised that these transformations were caused by the perceptual regrouping (i.e., streaming) of the acoustic elements that made up the word (Bregman, 1990). Repetitive presentation of the word caused the high-frequency frication corresponding to /s/ to split off from the remainder of the utterance and form its own perceptual stream. For example, /g/-initial transformations (e.g., ''gunk'') occurred 43% of the time, while there were no reports of /k/-initial transformations (e.g., ''kunk''), which suggests that the phonetic significance of the closure interval between /s/ and the remainder of the word was eliminated upon regrouping.

If streaming causes some verbal transformations, then listeners' reports should exhibit properties typical of regrouping percepts. Furthermore, which elements of an utterance split off should depend on the acoustic properties of the stimuli. These ideas were tested in two experiments using similar stimuli but different tasks (Pitt & Shoaf, in press). The stimuli were CVC pseudowords in which the strength with which the consonants bind with the vowel was manipulated. In the Intact condition, the consonants were continuants and nasals, which should be most resistant to streaming given their acoustic similarity (e.g., both are periodic signals and occupy similar frequency regions; /lom/, /wɛm/). In two other conditions, consonants were used that were expected to be most susceptible to streaming (e.g., fricatives, affricates, stops). These consonants occurred only syllable-finally in the Final condition (e.g., /lodʒ/, /wɛtʃ/), and in both consonant locations in the I+F (Initial plus Final) condition (e.g., /podʒ/, /pɛtʃ/).

In the first experiment, instructions to participants were modified slightly to extract more information from listeners about what they were hearing in order to test the regrouping account. Listeners (13) were trained to report the transformation, then the number of perceptual streams they heard at that time, and finally the contents of the background stream if one was present. There were 250 repetitions and a 0 ms ISI. Listeners' reports clearly suggested that streaming is one cause of the VTE. Sixty per cent of the transformations included reporting the presence of multiple streams. Almost without exception, the transformation formed the foreground stream and included one consonant and the vowel. The other consonant was reported as belonging to the background stream. The frequency of such ''streaming'' transformations varied across conditions, being 38% more prevalent in the Final and I+F conditions than in the Intact

condition. Across the three conditions, the consonants split off as a function of their relative cohesiveness with the vowel. In the Final condition, only the final consonant streamed off. When both consonants were not tied tightly to the vowel, as in the I+F condition, only one consonant tended to split off at a time (more often the final consonant), but there were reports in which both split off simultaneously, leaving the vowel as the only segment in the foreground stream.

Further support for perceptual regrouping being one cause of the VTE was obtained in the second experiment, which used a task that tested the streaming account indirectly, but which also enabled us to identify the portion of the stimulus that corresponded to the streaming transformation. Listeners (18) reported the first transformation, after which they isolated the portion of the stimulus that corresponded to the transformation by pressing buttons on a keypad that moved two cursors anywhere within the speech file. The stimulus continued to repeat at the same rate while the cursors were moved, but only the portion of the signal between the cursors was audible. Those portions outside of the cursors were no longer audible. Listeners had 200 repetitions during which to hear a transformation and isolate it. The only visual cues that were provided were numbers on a computer screen indicating how many cursor positions towards the centre of the file each cursor had been moved.

There was a high degree of consistency among listeners in the transformations that were reported and in where the cursors were placed. Two representative stimuli along with final cursor positions are shown in Figure 1. For stimuli in the Final condition (e.g., /lodʒ/), listeners moved the endpoint cursor towards the centre of the file past the frication corresponding to /dʒ/. For stimuli in the I+F condition (e.g., /pɛtʃ/), listeners did the same, but some also moved the initial cursor inward, suggesting that the stop burst and some of the aspiration split off. Cursor placement in the Intact condition was far less so, with much more variability in cursor placement and few listeners completely eliminating the initial or final consonant. Across all conditions, the segments that were eliminated were the same ones that were reported in the background stream in the immediately preceding experiment. The data from both tasks provide strong converging evidence that suggests some types of verbal transformations are due to perceptual regrouping.

More recent experiments have asked what besides perceptual regrouping causes verbal transformations. They were designed to minimise the frequency of streaming transformations so that other regularities might emerge in the data that would provide clues as to the identity and characteristics of other perceptual processes that are responsible for the illusion. Streaming was minimised by using Intact stimuli (e.g., /nal/, /jim/) and slowing stimulus presentation rate (ISI of

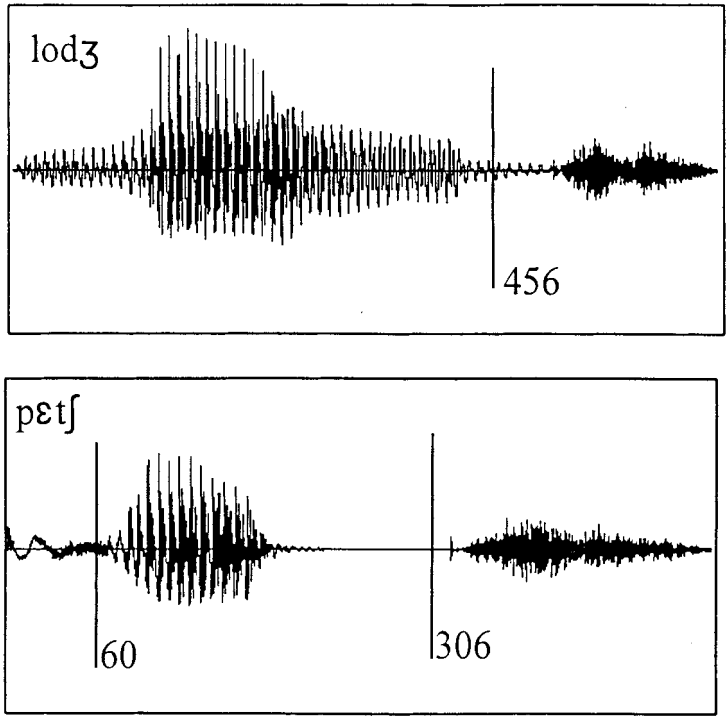

Figure 1. Waveforms of two stimuli. The vertical lines represent mean cursor positions measured in milliseconds from stimulus onset.

200 ms; 16 listeners) so that portions of each stimulus would be less likely to split off (Bregman, 1990).

This experiment succeeded in reducing the number of streaming transformations to 10%. The most frequent type of transformation (54%) was characterised by phoneme substitutions, with one or more veridical phonemes being replaced by a phonetically similar one. The remaining 36% were transformations back to the veridical percept. There was a great deal of regularity in the substitutions for both consonants and vowels. For the consonants, substitutions tended to change in place of articulation (e.g., /d/ for /b/ and /n/ for /m/). For vowels, front vowels tended to become lower and more backed (e.g., /I/ for /i/) and back vowels tended to be raised (e.g., /u/ for /U/). Satiation or fatigue of segmental representations may underlie this type of transformation. Repetitive presentation of an utterance may cause satiation, which leads to a slight shift in phoneme identity. Warren and Meyer (1988) and MacKay, Wulf, Yin, and Abrams (1993) have made similar proposals.

IMPLICATIONS FOR WORD PERCEPTION

Illusions can serve as windows into the processes that operate during veridical perception. By probing the causes of a verbal illusion like the VTE, one should be able to identify some of the processes involved in spoken word perception. The present findings suggest that perceptual regrouping and segmental satiation are two causes of the VTE. Both have been implicated in speech perception, so the present data reinforce these prior results (see Remez, Rubin, Berns, Pardo, & Lang, 1994; Samuel & Kat, 1996). That multiple processes appear to underlie the illusion makes the VTE a useful phenomenon for studying their simultaneous operation. Analysis of transformations across a range of testing conditions can identify these processes and demonstrate their interaction or independence. For example, the lexical effect could be driven by lexical processes influencing segmental satiation. Samuel's (2000) data showing lexical influences in selective adaptation, a paradigm which differs from the VTE only in *when* the effects of repeated presentation are measured, provide support for such a proposal. That regrouping and satiation act independently is suggested by the fact that streaming transformations have been found with and without phoneme substitutions.

The results showing that perceptual regrouping is one cause of the VTE suggest that lexical memory might also influence the strength with which the auditory elements of a word cohere. Nygaard (1993) found that lexical status influences the perceptual fusion of dichotically presented fragments of an utterance. It may well be that words which recycle transform back into the veridical percept more than pseudowords for the same reason: The lexical representation of the word may directly influence the perceptual organisation of speech and thereby serve as an anchor to stabilise perception. Put another way, lexical processes counter the effects of repetitive presentation that lead to streaming and help bind the acoustic elements of a word together. If this is the case, then words should yield fewer streaming transformations than pseudowords. Plans are underway to test this proposal. Finding such an effect would suggest that word perception involves more than processing the phonetic segments that make up a word. It also includes processes that assist in binding the acoustic elements of those segments together. Recent findings of Fowler and Brancazio (2000) demonstrating cross-modal lexical effects extend and reinforce this claim, so it is likely that word perception consists of multiple processes working in concert to match the acoustic signal with its representation in memory.

REFERENCES

Bregman, A.S. (1990). *Auditory Scene Analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Fowler, C.A., & Brancazio, L. (2000). Feedback in audiovisual speech perception. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes* (pp. 87–90). Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.

Frauenfelder, U., & Content, A. (2000). Activation flow in models of spoken word recognition. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes* (pp. 79–82). Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.

MacKay, D.G., Wulf, G., Yin, C., & Abrams, L. (1993). Relations between word perception and production: New theory and data on the verbal transformation effect. *Journal of Memory and Language*, *32*, 624–646.

Natsoulas, T. (1965). A study of the verbal-transformation effect. *American Journal of Psychology*, *78*, 257–263.

Nygaard, L. (1993). Phonetic coherence in duplex perception: Effects of acoustic differences and lexical status. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 268–286.

Pitt, M.A., & Shoaf, L.S. (in press). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human Perception and Performance.*

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., & Lang, J.M. (1994). On the perceptual organization of speech. *Psychological Review*, *101(1)*, 129–156.

Samuel, A.G. (2000). Some empirical tests of Merge's architecture. In A. Cutler, J.M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes* (pp. 51–54). Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.

Samuel, A.G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 676-694.

Shoaf, L.S., & Pitt, M.A. (submitted). A test of node stability in Node Structure Theory. Manuscript submitted for publication.

Warren, R.M., & Meyer, M.D. (1987). Effects of listening to repeated syllables: Category boundary shifts versus verbal transformations. *Journal of Phonetics*, *15*, 169–181.