BEYOND TRADITIONAL MEASURES OF LEXICAL INFLUENCES ON PERCEPTION

Mark A. Pitt, Lisa Shoaf
Ohio State University, Columbus, OH

ABSTRACT
Phoneme detection and identification tasks, and their more recent hybrids, have been the primary tools used to explore whether and how lexical memory influences earlier perceptual processes, with differences between words and pseudowords suggestive of such influences. Processing differences between words and pseudowords are also found in the Verbal Transformation Effect, which is a perceptual phenomenon in which continuous, rapid repetition of a word causes listeners to hear the word transform into other utterances. In this talk, the nature of this lexical effect will be described and two mechanisms responsible for verbal transformations will be discussed. The results tie in with recent work exploring lexical feedback and suggest that the influences of lexical memory extend to perceptual processes beyond those involved in phoneme perception.

1. INTRODUCTION
Research in language processing often centers around a phenomenon, such as priming, that is studied in depth to learn about the perceptual and cognitive processes that underlie it, and therefore language processing. Differences in responding to words and pseudowords (i.e., a lexical effect) have figured prominently in investigations of the architecture of the word recognition system.

Differences in responding to words and pseudowords have been reported in studies investigating the Verbal Transformation Effect (VTE), which is a perceptual phenomenon in which listeners report hearing illusory utterances after listening to a word repeat over and over [1]. Performance differences have been found along a couple of measures [2], differences that hold up when a meta-analysis is conducted on stimuli across studies. The reliability of this finding prompted us to explore its origins. This report is an overview of what we have found thus far and what it might tell us about how lexical memory influences perception.

2. EXPERIMENTS

2.1. Experiment 1
Experiment 1 was conducted to replicate the lexical effect reported in the literature.

2.1.1. Method
2.1.1.1. Stimuli
There were 24 stimuli, half monosyllables and half bisyllables, with an equal number of words and pseudowords in each category. Stimuli were matched phonetically across conditions as closely as possible so that differences in performance would be attributable to differences in lexical status rather than to other properties of the stimuli. For the bisyllables, this involved swapping the order of the two syllables. Example stimulus pairs are “skunk-swunk” and “center - tercen.”

2.1.1.2. Procedure
The experiment consisted of a lengthy instruction session, two practice trials, and then four test trials, in which one stimulus was randomly presented from each of the four conditions. Each stimulus was presented for 350 repetitions with an ISI of about 100ms. Stimuli were presented binaurally over headphones at a comfortable listening level. Listeners (31) reported transformations into a microphone that was fed to a cassette deck.

Only a small number of trials is presented because the experiment is extremely fatiguing, requiring a high level of concentration and alertness during each 4-minute trial. When given more than six trials, transformation reports can drop off noticeably.

2.1.2. Results
The most frequent measures of performance have been the number of verbal transformations reported and the number of forms (i.e., unique transformations) reported. Across experiments, we have found the former to yield few differences as a function of lexical status, whereas the latter has been much more reliable.

Data in the first column of Table 1 show that for both monosyllables and bisyllables, listeners reported fewer forms when hearing a word than a pseudoword, F(1,29)=11.51, p<.05. Data in the second column are from an analysis of the transformations back to the veridical percept (i.e., reporting "bliss" if the stimulus "bliss" is recycling). The results show that when the stimulus was a word, it transformed back into the veridical stimulus far more often than when it was a pseudoword. This was true for monosyllables and bisyllables.

These data indicate that listeners’ reports were influenced by the lexical status of the stimulus. They suggest that a lexical representation may serve as a perceptual anchor that prevents words from undergoing as many different transformations and instead transform back into the veridical form.

To learn about the origin of this lexical effect, it is necessary to learn what causes verbal transformations (VTs). The following experiments were an initial step in this direction.
2.2. Experiment 2

A number of the forms reported to the /s/-initial monosyllables did not contain the fricative (e.g., "swunk" -> "wunk"). We hypothesized that hearing the stimuli at

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean Number of Forms</th>
<th>Mean Proportion of Veridical Transformations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monosyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words</td>
<td>7.67</td>
<td>.39</td>
</tr>
<tr>
<td>Pseudowords</td>
<td>11.00</td>
<td>.25</td>
</tr>
<tr>
<td>Bisyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words</td>
<td>8.53</td>
<td>.41</td>
</tr>
<tr>
<td>Pseudowords</td>
<td>11.03</td>
<td>.18</td>
</tr>
</tbody>
</table>

a fairly high presentation rate caused the s/ split off perceptually from the remainder of the word and form its own perceptual stream, leaving the remainder to be heard as a new transformation (e.g., "wunk") in a separate stream. Experiment 2 tested this idea.

If streaming causes VTs, then listeners should report that two streams are heard and be able to identify the contents of each. Furthermore, based on what is known about the properties of sound and speech that promote streaming (e.g., frequency, temporal proximity and continuity)[3], we should be able to exert some control over the likelihood that a particular utterance will split into multiple streams and what portions of the utterance (e.g., initial or final consonants) are likely to do so.

2.2.1. Method
2.2.1.1. Stimuli

Six CVC pseudowords served as stimuli. In the Intact condition, the consonants were continuants and nasals, which should be most resistant to streaming given their close frequency proximity to the vowel and smooth formant transitions in and out of the vowel. In two other conditions, consonants were used that were expected to be most susceptible to streaming (affricates and voiceless fricatives and stops). These consonants occurred only at the end of the word in the Final condition, and in both consonant locations in the Initial plus Final (I+F) condition.

2.2.1.2. Procedure

Listeners (13) were familiarized with the phenomenon of streaming using tone sequences and then the VTE. There was then an extensive practice session, followed by the test session. They were instructed to report the transformation heard, determine how many streams were heard, and then identify the contents of the stream(s). Stimuli were presented for 250 repetitions. To promote streaming, no pause occurred between presentations.

2.2.2. Results

Listeners had no difficulty identifying transformations that streamed and those that did not. A sizeable proportion of streaming responses were reported in all three stimulus conditions, as the data in the first column of Table 2 indicate. Those stimuli created to stream (Final and I+F) did so reliably more often than those in the Intact condition, X²(2) = 10.02, p<.05. Data in the next two columns show the frequency with which the initial and final consonant streamed. As expected, only the final consonant streamed in the Final condition. Both initial and final consonants streamed in the I+F condition. The data in the fourth column show that even though the Intact items streamed, they did so less readily than the Final and I+F stimuli, requiring almost twice as many repetitions before the first streaming transformation was reported, F(2,22)=6.58, p<.05. Other measures yielded similar outcomes.

An example of a frequent transformation in each condition is shown on the right side of Table 2. Almost without exception, the longer segment (e.g., /lo/l/) is reported as the transformation, giving the impression that the remaining portion (e.g., /dʒ/) vanished. Reports in which an initial or final consonant has been deleted are a reliable marker that streaming caused the transformation.

Further insight into these results was obtained in experiments in which we used a methodology for determining the portion of the signal that corresponded to the percept in each stream. Using cursors that were linked to the beginning and end of the sound file, listeners isolated the portion of the repeating stimulus that corresponded to the transformation by moving the cursors back and forth. The vertical lines in the waveforms in Figure 1 divide the signal into the foreground (larger) and
background (smaller) streams for two stimuli. Note that the
lines divide the periodic portion from the aperiodic and
higher-frequency portion.

Figure 1

2.3. Experiment 3

The extent to which streaming causes VTs will depend on
the stimulus properties and methodological details in an
experiment. In some setups (slow repetition rate and no
initial or final voiceless consonants), few streaming
transformations are likely to be reported, in which case the
VTE should be due to other processes. Experiment 3 was
designed to increase the visibility of these other processes
by minimizing the probability of streaming
transformations.

2.3.1. Method
Seven CVC pseudowords were modeled after the Intact
items of Experiment 2 and the recycling rate was slowed
by inserting a 200ms pause between presentations. The
experiment was similar to Experiment 1 in other respects.
There were 250 repetitions.

2.3.2. Results
The transformations that predominated are qualitatively
very different from those in the preceding experiments.
Listeners reported few transformations in which a
consonant was deleted (10%). Instead, the majority (54%)
involved one or more phoneme substitutions (24% for
consonants and 30% for vowels). Furthermore, there is an
impressive regularity with which phonemes were
substituted. 93% of the substituted consonants differed
from the veridical consonant by a single phonetic feature
(e.g., /d/ for /b/, /s/ for /z/, /n/ for /m/). 68% of the time a
single vowel was substituted for the veridical vowel (e.g.,
/æ/ for /i/), and 80% of the time the substituted vowel
shared the same tenseness as the veridical vowel.

A brief shift in phoneme identity (i.e., substitution)
appears to be the immediate perceptual consequence of
repeated stimulus presentation. This finding suggests that
the segmental representation is being fatigued or satiated
[4,5]. That such a process would be identified as causing
VTs is not surprising given that the VT paradigm is very
similar to that of selective adaptation, in which the
operation of a fatigue mechanism has been extensively
explored. What differentiates the two is the point in time
when the effects of repeated presentation are measured. In
adaptation, the lingering effects of repeated stimulation are
measured in a labeling task that takes place after recycling
stops. In VT experiments, the immediate effects are
measured, as listeners freely report what they hear when
they hear it.

3. GENERAL DISCUSSION

The results of Experiments 2 and 3 suggest that streaming
and representational fatigue are two causes of the VTE.
Are the processes that cause streaming and fatigue related
to the lexical effect found in Experiment 1? In the case of
streaming, perceptual grouping processes have been shown
to be influenced by the lexical status of a word [6]. The
acoustic-phonetic elements that make up a dichotically
presented utterance cohere better when the utterance is a
word than a pseudoword. We are currently looking for
evidence of this in the data of Experiment 1.

Evidence connecting lexical effects with
representational fatigue is a bit more speculative. However,
the similarities of the VT and adaptation paradigms makes
it reasonable to suppose that similar stages of processing
are being affected in both cases. Some of these appear to
be prelexical based on adaptation data. The perceptual
substitution of phonemes may therefore occur prelexically.
If this is the case, the lexical effect found in Experiment 1
could come about through lexical feedback to a prelexical
stage, which has been demonstrated using the adaptation
paradigm [7].

The present results suggest that the VTE is a complex
illusion. It may in fact reflect the complexity of word
recognition itself. Careful study of the verbal reports that
listeners generate provides a glimpse of this complexity,
revealing some of the mental processes that must work in
concert to match the acoustic signal with its representation
in memory.

4. ACKNOWLEDGMENTS

Mark Pitt, Department of Psychology, 1885 Neil Avenue,
Ohio State University, Columbus, OH, 43210. Email:
pitt.2@osu.edu.
5. REFERENCES