Echoes of Echoes? An Episodic Theory of Lexical Access

Stephen D. Goldinger
Arizona State University

In this article the author proposes an episodic theory of spoken word representation, perception, and production. By most theories, idiosyncratic aspects of speech (voice details, ambient noise, etc.) are considered noise and are filtered in perception. However, episodic theories suggest that perceptual details are stored in memory and are integral to later perception. In this research the author tested an episodic model (MINERVA 2; D. L. Hintzman, 1986) against speech production data from a word-shadowing task. The model predicted the shadowing-response-time patterns, and it correctly predicted a tendency for shadowers to spontaneously imitate the acoustic patterns of words and nonwords. It also correctly predicted imitation strength as a function of "abstract" stimulus properties, such as word frequency. Taken together, the data and theory suggest that detailed episodes constitute the basic substrate of the mental lexicon.

Early in the 20th century, Semon (1909/1923) described a memory theory that anticipated many aspects of contemporary theories (Schacter, Eich, & Tulving, 1978). In modern parlance, this was an episodic (or exemplar) theory, which assumes that every experience, such as perceiving a spoken word, leaves a unique memory trace. On presentation of a new word, all stored traces are activated, each according to its similarity to the stimulus. The most activated traces connect the new word to stored knowledge, the essence of recognition. The multiple-trace assumption allowed Semon's theory to explain the apparent permanence of specific memories; the challenge was also to create abstraction from a collection of idiosyncratic traces. A resolution came from Galton (1883), who found that blending faces in a photographic composite creates the image of a "generic" face. Galton applied this as a memory metaphor: "Whenever a single cause throws different groups of brain elements simultaneously into excitement, the result must be a blended memory" (Galton, 1883, p. 229). Semon borrowed this idea, assuming that abstraction occurs during retrieval as countless partially redundant traces respond to an input.

For a variety of reasons (Schacter et al., 1978), Semon's (1909/1923) theory vanished from mainstream psychology. When cognitive science later resurfaced, its theories emphasized minimal, symbolic representations. Perception was theorized to entail information reduction, such that processing stages generate progressively more abstract representations of analog inputs (Posner, 1964). Whereas Semon's theory emphasized a proliferation of traces, later theories emphasized economy. Especially in psycholinguistic theories, the recoding of specific episodes (tokens) into canonical representations (types) remains a basic assumption. For example, models of spoken word perception generally assume a collection of canonical representations that are somehow accessed by variable, noisy signals (Goldinger, Pisoni, & Luce, 1996; Klatt, 1989).

In this article I propose a return to the episodic view, with specific application to the mental lexicon. Although the lexicon is theoretically involved in many linguistic behaviors, the present focus is limited to spoken word perception, production, and memory. To anticipate, I begin this article with a literature review on speaker normalization, focusing on memory for words and voices. This review suggests that many perceptual and memorial data are best understood in terms of episodic representations. After this, a specific model (MINERVA 2; Hintzman, 1986) is described and is applied to prior data (Goldinger, 1996). Three new shadowing experiments are then reported, along with MINERVA 2 simulations. The data and simulations support the basic ideas of episodic representation and access. In the General Discussion, the episodic view is considered in the context of other prominent theories, and several potential problems are addressed.

Speaker Normalization

In theories of speech perception, the assumption of an abstract lexicon is motivated by extreme signal variability. Speech acoustics are affected by many factors, including phonetic context, prosody, speaking rate, and speakers. Decades of research have revealed few invariant speech patterns that recognition systems can reliably identify (although see Cole & Scott, 1974; Stevens & Blumstein, 1981). Thus, speech variability is typically considered a perceptual "problem" solved by listeners, as it must be solved in recognition systems (Gerstman, 1968). Consider speaker variability: Speakers differ in vocal tracts (Peterson & Barney, 1952), glottal waves (Monsen & Engebretson, 1977), articulatory dynamics (Ladefoged, 1980), and native
diseases. Thus, great acoustic variability arises in nominally identical words across speakers. Nevertheless, listeners typically understand new speakers instantly.

Most theories of word perception assume that special processes match variable stimuli to canonical representations in memory (McClelland & Elman, 1986; Morton, 1969; Studdert-Kennedy, 1976; see Tenpenny, 1995). This is achieved by speaker normalization—"phonetically irrelevant" voice information is filtered in perception (Joos, 1948). Speaker normalization presumably allows listeners to follow the lexical–semantic content of speech; superficial details are exploited by the perceptual machinery, then discarded (Krueke, Tondo, & Wightman, 1983). For example, Halle (1985) wrote that when we learn a new word, we practically never remember most of the salient acoustic properties that must have been present in the signal that struck our ears. For example, we do not remember the voice quality, speed of utterance, and other properties directly linked to the unique circumstances surrounding every utterance. (p. 101)

Unfortunately, the speaker normalization hypothesis may be unfalsifiable, at least by perceptual tests. For example, Mullennix, Pisoni, and Martin (1989) compared listeners' responses to word sets spoken in 1 or 10 voices. Speaker variations reduced identification of words in noise and slowed shadowing of words in the clear, which led Mullennix et al. to suggest a capacity-demanding normalization process that usurps resources needed for primary task performance (see also Nusbaum & Morin, 1992). However, when researchers find no effects of speaker (or font) variation, they often conclude that automatic normalization occurs early in perception (Brown & Carr, 1993; Jackson & Morton, 1984; Krueke et al., 1983). Apparently, both positive and null effects reflect normalization. This reasoning seems to occur because normalization is required by the assumption of an abstract lexicon. If a theory presumes that variable speech signals are matched to ideal templates or prototypes, successful perception always implies normalization.

Given their basic representational assumptions, most theories of word perception are forced to assume normalization. However, in a lexicon containing myriad and detailed episodes, new words could be compared directly with prior traces. By this view, speaker normalization becomes a testable hypothesis, rather than an assumed process, equally evidenced by positive or null effects. As it happens, many contemporary models resemble Semon's (1909/1923) theory, positing parallel access to stored traces (Eich, 1982; Gillund & Shiffrin, 1984; Hintzman, 1986, 1988; Medin & Schaffer, 1978; Nosofsky, 1984, 1986; Underwood, 1969). Such theories are partly motivated by common findings of memory for "surface" details of experience. Outstanding memory for detail has been reported for many nonlinguistic stimuli, including faces (Bahrick, Bahrick, & Wittlinger, 1975; Bruce, 1988), pictures (Roediger & Srinivas, 1992; Shepard, 1967; Snodgrass, Hirshman, & Fan, 1996; Standing, Conezio, & Haber, 1970), musical pitch and tempo (Halpern, 1989; Levitin & Cook, 1996), social interactions (Lewicki, 1986), and physical dynamics (Cutting & Kozlowski, 1977). Indeed, Smith and Zarate (1992) developed a theory of social judgment based on MINERVA 2, and Logan (1988, 1990) developed an episodic model of attentional automaticity. Similarly, Jusczyk's (1993) developmental model of speech perception incorporates episodic storage and on-line abstraction, as in Semon's theory.

Contrary to many views, linguistic processes often create lasting, detailed memories. People spontaneously remember the presentation modalities of words (Hintzman, Block, & Insko, 1972; Hintzman, Block, & Summers, 1973; Kirsner, 1974; Lehman, 1982; Light, Stanisbury, Rubin, & Linde, 1973), the spatial location of information in text (Lovelace & Southall, 1983; Rothkopf, 1971), and the exact wording of sentences (Begg, 1971; Keenan, MacWhinney, & Mayhew, 1977). Experiments on transformed text show the persistence of font details in memory after reading (Kolers, 1976; Kolers & Ostry, 1974), and similar findings occur with isolated printed words (Hintzman & Summers, 1973; Kirsner, 1973; Roediger & Blaxton, 1987; Tenpenny, 1995). Given these data, Jacoby and Hayman (1987) suggested that printed word perception relies on episodic memory. Given these findings, it would be surprising if spoken word perception operated differently. In fact, relative to fonts, voices are more ecologically valuable and worthy of memory storage.

Human voices convey personal information, such as speakers' age, sex, and emotional state (Abercrombie, 1967). These aspects of speech are typically ignored in perceptual and linguistic theories, but they are clearly important. For example, pervasive changes in tone of voice are readily understood in conversation. Moreover, although early research (McGehee, 1937) indicated that long-term memory (henceforth LTM) for voices is poor, later researchers found reliable voice memory (Cartette & Barnebey, 1975; Hollien, Majewski, & Doherty, 1982; Papcio, Kreiman, & Davis, 1989). Indeed, Van Lancker, Kreiman, and Emmorey (1985; Van Lancker, Kreiman, & Wickens, 1985) reported that famous voices are easily recognized, even when played backward or when rate compressed. More recently, Remetz, Fellowes, and Rubin (1997) found that listeners can identify familiar voices, using only "sinewave sentences" as stimuli.

Memory for Words and Voices

As with printed words, researchers have previously assessed surface memory for spoken words. For example, Hintzman et al. (1972) played words to listeners in two voices. In a later recognition memory test, half of the words changed voices. Listeners discriminated between old and new voices well above chance (see also Cole, Coltheart, & Allard, 1974; Geiselman & Bellezza, 1976, 1977). Moreover, Schacter and Church (1992; Church & Schacter, 1994) recently found that implicit memory for spoken words retains very specific auditory details, including intonation contour and vocal pitch.

Martin, Mullennix, Pisoni, and Summers (1989) compared serial recall of word lists produced by 1 or 10 speakers. They found that LTM was reduced for 10-speaker lists and suggested that speaker variation induces normalization, usurping attention needed for rehearsal. However, Goldinger, Pisoni, and Logan (1991) later found that speaker variation interacts with presentation rate. When slow rates were used, recall from 10-speaker lists surpassed recall from 1-speaker lists (see also Lightfoot, 1989; Nygaard, Sommers, & Pisoni, 1992). Indeed, voice information appears to be an integral dimension of spoken words, as evidenced in a Garner (1974) speeded-classification task (Mullennix & Pisoni, 1990). Thus, attention to spoken words
logically entails attention to voices. Speaker variability may reduce recall at fast presentation rates by mere distraction (Aldridge, Garcia, & Mena, 1987). In a similar experiment, using 1- and 10-speaker lists, Goldinger (1990) examined self-paced serial recall. Volunteers controlled list presentation; they pressed buttons to play each word, pausing as long as they wished between words. Both the self-determined presentation rates and subsequent recall are shown in Figure 1. The recall data resembled the slow-rate data from Goldinger et al. (1991), and the listening times supported their account—speaker variation apparently motivates listeners to pause longer between words, allowing more rehearsal.

Of course, prior studies had established that voices are incidentally learned during word perception (Cole et al., 1974; Geiselman & Bellezza, 1976; Hintzman et al., 1972; Light et al., 1973). However, most used only two stimulus voices, usually a man’s and a woman’s. Thus, voice memory could reflect either analog episodes or abstract “gender tags” (Geiselman & Crawley, 1983). To address this, Palmeri, Goldinger, and Pisoni (1993) tested continuous recognition memory for words and voices. In this task, old and new words are continuously presented, minimizing rehearsal. Listeners try to classify each word as new on its first presentation and old on its repetition. The primary manipulation is the number of intervening words (lag) between first and second presentation of the words. Typically, recognition decreases as lag increases (Shepard & Teghtsoonian, 1961).

The Palmeri et al. (1993) study extended an earlier continuous-recognition study: Craik and Kirsner (1974) presented words to listeners in two voices (male and female). When repeated, half of the words switched voices. Same-voice (SV) repetitions were better recognized than different-voice (DV) repetitions across all lags, showing that voice details persist in LTM for 2–3 min. Unlike Craik and Kirsner, we used several levels of speaker variation. Participants heard 2, 6, 12, or 20 voices (half male and half female). This let us assess the automaticity of voice encoding: If listeners strategically encode voices, increasing from 2 to 20 speakers should impair this ability. Also, by including multiple speakers of both sexes, we could evaluate Geiselman and Crawley’s (1983) voice connotation hypothesis. By this view, male and female voices invoke different word connotations, so recognition should be sex dependent, not voice dependent. Finally, whereas Craik and Kirsner used lags up to 32 trials, we tested lags up to 64 trials.

The data were fairly decisive; First, the increase from 2 to 20 speakers had no effect, suggesting automatic voice encoding. Second, hit rates were higher for SV than for DV repetitions, regardless of sex. This suggested that word-plus-voice traces are formed in perception; only exact token repetition facilitates later recognition (i.e., the voice connotation hypothesis was not supported). Finally, the SV advantage was stable across lags, suggesting durable traces. Goldinger (1996) later extended this study in several respects: Episodic retention was assessed over longer delays by using both explicit and implicit memory measures (Musen & Treisman, 1990; Tulving, Schacter, & Stark, 1982). Also, the perceptual similarities among all stimulus voices were discovered by multidimensional scaling (MDS; Kruskal & Wish, 1978; Shepard, 1980). If episodic traces retain fine-grained perceptual details, then memory for old words in new voices should be affected by the similarity of the voices, even within genders.

In a recognition memory experiment, listeners heard 150 study words and 300 later test words. Participants heard 2, 6, or 10 voices in each session and waited 5 min, 1 day, or 1 week between sessions. Most important, half of the old words changed voices between study and test. As in continuous recognition, no effect of total variability was observed; accuracy was equivalent with 2, 6, or 10 voices. However, at delays of 5 min or 1 day, SV repetitions were recognized better than DV repetitions. The MDS data showed that performance to DV trials was affected by the perceptual distance between study and test voices, suggesting that study traces retain voice details with great precision. Voice effects diminished over time, however, and were absent after 1 week. In a similar implicit memory experiment, however, reliable voice effects were observed at all delays. Moreover, the MDS data showed that gradations of perceptual similarity affected performance for 1 full week. Together, the data suggest that detailed, lasting episodes are formed in spoken word perception.

![Figure 1. Self-paced serial recall data from Goldinger (1990). Top: self-determined presentation rates as a function of serial position. Bottom: subsequent recall.](https://example.com/figure1.png)
The Episodic Lexicon?

Given the preceding review, a natural question arises: If episodic traces of words persist in memory and affect later perception, might they constitute the mental lexicon? In many articles, Jacoby (1983a, 1983b; Jacoby & Brooks, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Jacoby & Witherspoon, 1982) has suggested nonanalytic word perception by comparison to stored episodes rather than to abstract nodes (see Feustel, Shiffrin, & Salasoo, 1983; Kirsner, Dunn, & Standen, 1987; Salasoo, Shiffrin, & Feustel, 1985). Although episodic theories of word perception have been frequently suggested, little formal modeling has occurred (except Salasoo et al., 1985).

Hintzman’s (1986, 1988) MINERVA 2

Several models cited earlier are hybrids, combining abstract and episodic representations. Indeed, such an approach may prove necessary to accommodate many linguistic processes (see the General Discussion). However, to assess the benefits of an episodic view, it is best to evaluate a “pure” model. If it fails, less extreme models are available. In the present research I tested Hintzman’s (1986, 1988) MINERVA 2. This model takes episodic storage to a logical extreme, assuming that all experiences create independent memory traces that store all perceptual and contextual details (cf. Underwood, 1969). Despite their separate storage and idiosyncratic attributes, aggregates of traces activated at retrieval create behavior. Thus, like Semon’s (1909/1923) theory, MINERVA 2 accounts for the specificity and generality of memory by using only exemplars. Indeed, simulations (Hintzman, 1986; Hintzman & Ludlam, 1980) reproduce behaviors typically considered hallmarks of abstract representations, such as long-lasting prototype effects in dot-pattern classification and memory (Posner & Keele, 1970).

Word perception in MINERVA 2 occurs as follows: For every known word, a potentially vast collection of partially redundant traces resides in memory. When a new word is presented, an analog probe is communicated (in parallel) to all traces, which are activated by the probe in proportion to their mutual similarity. An aggregate of all activated traces constitutes an echo sent to working memory (WM) from LTM. The echo may contain information not present in the probe, such as conceptual knowledge, thus associating the stimulus to past experience. Appendix A summarizes the formal model and details of the present simulations. Because the model’s operations are fairly intuitive, all text descriptions focus on the conceptual level.

Echoes have two important properties in MINERVA 2. First, echo intensity reflects the total activity in memory created by the probe. Echo intensity increases with greater similarity of the probe to existing traces, and with greater numbers of such traces. Thus, it estimates stimulus familiarity and can be used to simulate recognition memory judgments. Assuming that stronger echoes also support faster responses, inverse echo intensities were used to simulate response times (RTs) in the present research. Second, echo content is the “net response” of memory to the probe. Because all stored traces respond in parallel, each to its own degree, echo content reflects a unique combination of the probe and the activated traces. This is clarified by a relevant example: Assume that myriad, detailed traces of spoken words reside in LTM. If a common word is presented in a familiar voice, many traces will strongly respond. Thus, even if a perfect match to the probe exists in memory, all of the similar activated traces will force a “generic echo”—its central tendency will regress toward the mean of the activated set. However, if a rare word is presented in an unfamiliar voice, fewer traces will (weakly) respond. Thus, if a perfect match to the probe exists in memory, it will clearly contribute to echo content. Therefore, token repetition effects should be greater for unusual words or for words presented in unusual contexts (Graf & Ryan, 1990; Masson & Freedman, 1990).  

MINERVA 2 qualitatively replicates the recognition memory data from Goldinger (1996). In the model, “spoken words” are represented by vectors of simple elements, with values of −1, 0, or +1. The vectors were divided into segments denoting three major dimensions: Each word contained 100 name elements, 50 voice elements, and 50 context elements. When the model’s “lexicon” is created, every input creates a new trace. Some forgetting occurs over time, however, simulated by random elements reverting to zero (determined stochastically over forgetting cycles).

The simulations were fashioned after the six-voice condition. To mimic a person’s prior knowledge, I created an initial lexicon for the model: 144 words were generated and stored 20 times each. The name elements were identical for all 20 tokens of each word; voice and context elements were randomly generated. To approximate the experiment, I generated new tokens of all 144 words with identical context elements, and six configurations of voice elements denoted six “speakers.” The study phase was simulated by storing 72 words, once each (12 per voice). Intuitively, this allows the model to associate words in its lexicon with the specific context of the study phase, as would be necessary for a human participant. In a test phase, the model received all 144 words. Among the 72 old words, 36 had new voices (6 per voice). Between phases, the model completed 1, 3, or 10 forgetting cycles (for the study phases), representing three delay periods. The dependent variable was echo intensity, shown in Figure 2. As in the human data (top of Figure 2), the model’s hit rates were higher for SV trials, and the voice effect vanished over time.

Beyond this replication, the model provided a new prediction. In the test shown in Figure 2, all words had equal frequency (20 traces each). To better match the real experiment (Goldinger, 1996), I conducted another simulation with varying study word frequencies (i.e., the number of traces initially stored in the model’s lexicon). Instead of uniformly storing 20 traces, different words were represented by 2, 4, 8, 16, 32, or 64 traces (12 words per frequency value). As before, each word had

---

1 In general, for any model to predict repetition effects with common English words, contextual encoding must be assumed (Gillund & Shiffrin, 1984; Hintzman, 1988). Presumably, voice effects are observable in the laboratory because the study words are experienced in a unique setting for relatively unique purposes (see the General Discussion).

2 The use of vector representations has several advantages, including computational simplicity and theoretical transparency (Hintzman, 1986). If the model predicts data patterns without assuming complex representations, it likely reflects central processes rather than implementational details.
constant name elements across traces, but all traces had randomly generated voice and context elements. Once the variable-frequency lexicon was stored, the simulation was conducted with a constant “delay period” of three forgetting cycles.

The frequency manipulation produced an interesting new result: The SV advantage diminished as word frequencies increased. In terms of difference scores (SV minus DV trials, in echo-intensity units), the six frequency classes (2, 4, 8, 16, 32, and 64 traces) created mean SV advantages of .85, .58, .31, .25, .17, and .09, respectively. As noted, high-frequency (HF) words activate many traces, so the details of any particular trace (even a perfect match to the new token) are obscured in the echo. Thus,old HF words inspire “abstract” echoes, obscuring context and voice elements of the study trace. This model prediction motivated a post hoc correlation analysis on the Goldinger (1996) data, which confirmed stronger voice effects among lower frequency words ($r = -.35$, $p < .05$).

Episodes in Perception and Production

In the research reviewed earlier, lexical representations were examined by testing memory for spoken words. By contrast, in the present study I used a single-word shadowing (or auditory naming) task, in which participants hear and quickly repeat spoken words. The typical dependent measure in shadowing is the latency between stimulus and response onsets (Radeau, Morais, & Dewier, 1989; Slowiaczek & Hamburger, 1992). A seldom-used secondary measure is the speech output itself. The classic motor theory states that “speech is perceived by processes that are also involved in its production” (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, p. 452). Supporting research by Porter and Lubker (1980) showed that listeners could shadow syllables faster in a choice RT procedure than they could press a button in the same task (see also Porter & Castellanos, 1980). This suggests that shadowers may “drive” their articulators directly from speech input.

Acoustic measures are often examined in applied research, such as testing the effects of alcohol or noise on speech (Johnson, Pisoni, & Bernacki, 1990; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988) or the intelligibility of disordered speech (Geschwind, 1975). In basic research on lexical access, several researchers have examined spoken word durations: Wright (1979; also Geffen & Luszcz, 1983; Geffen, Stierman, & Tildesley, 1979) had volunteers read word lists aloud, finding longer durations of, and longer pauses between, low-frequency (LF) words (see Balota, Boland, & Shields, 1989). Whalen and Wenk (1993) reported that when people read homophones (e.g., time–thyme) aloud, LF spellings occasionally yield longer utterances (but only when blocked LF and HF lists were compared). These data suggest that, in certain conditions, cognitive aspects of lexical representation can affect speech acoustics.

Several years ago, I conducted an unpublished experiment in which volunteers shadowed words produced by 10 speakers. The hypothesis (borne largely of subjective experience) was that shadowers would “track” the stimulus voices. This vocal imitation was assessed by comparing acoustic parameters of shadowing speech to baseline speech (collected while participants read words aloud from a computer). As expected, shadowers tended to imitate the speakers, at least in terms of fundamental frequency and word duration. In a similar experiment, Oliver (1990) found that preschool children also track stimulus word durations in shadowing.

Testing MINERVA 2 by Spontaneous Imitation

By itself, imitation in shadowing reveals little about lexical representation. However, in MINERVA 2, new predictions may emerge. As noted, motor theory is based on a fundamental perception–production linkage, so the imitation prediction is emergent. On the other hand, MINERVA 2 cannot directly predict imitation, as it has no output mechanism. Given a probe stimulus, the model produces an echo—the researcher must decide how to translate this covert signal into overt behavior. However, imitation is both a natural and conservative prediction in MINERVA 2. Because echoes constitute the model’s only basis to respond, it is most economical to hypothesize that shadowers will generate a “readout” of the echo content. Indeed, by specifying both echo intensity and content, MINERVA 2 has a unique ability to predict both shadowing RTs and imitation.

Beyond allowing imitation to emerge as a plausible by-product, MINERVA 2 also makes principled predictions about the strength of imitation. Hintzman (1986) showed that echo content consists of blended information—new probes and stored episodes combine to form experience. Recall the hypothesized differences in echo content, depending on word frequency: HF

---

3 Marslen-Wilson (1985), however, showed that extremely fast shadowers conduct full-lexical, syntactic, and semantic analysis of speech. The results observed by Porter and his colleagues may be unique to meaningless syllabic input.
words excite many traces, so their idiosyncrasies are obscured ("generic" echoes). By contrast, echoes for LF words are strongly influenced by old traces resembling the probe. Because shadowing in MINERVA 2 is based on echoes, the model predicts that imitation will increase as word frequencies decrease. In this investigation, shadowing was examined in several ways. Of primary interest were comparisons between human data and MINERVA 2 simulations. As a grounding principle, it must be assumed that shadowing is based on perceptual—cognitive processes. That is, shadowing is not a shallow activity—words do not "travel directly" from the ears to the vocal tract in a reflex arc. This is clearly an assumption, but it finds support from prior investigations. For example, shadowing RTs are affected by word and neighborhood frequency (Luce, Pisoni, & Goldinger, 1990) and by phonemic priming (Slowiaczek & Hamburger, 1992). Also, when shadowing connected discourse, listeners are sensitive to word frequency, syntactic structure, and semantic context (Marslen-Wilson, 1985). If shadowing is a truly cognitive process, models like MINERVA 2 may predict performance. In the unpublished experiment summarized earlier, all words were presented twice in the shadowing condition. The model's prediction was tested by examining imitation to the second presentation of each word (the first shadowing trial creates the idiosyncratic memory trace necessary to influence later echo content). Post hoc analyses confirmed that imitation was stronger for lower frequency words ($r = -0.40, p < 0.05$), suggesting that shadowing speech is affected by episodic aspects of lexical representation.

**Experiments 1A and 1B: Shadowing English Words**

It is surely a coincidence that Hintzman (1986) chose the term echo for the key construct in his model. Nevertheless, from the perspective of testing MINERVA 2, a benefit of the shadowing paradigm is simultaneous assessment of echo intensity and content. Strong echoes (as for HF words) should yield fast responses. (Although Hintzman, 1986, did not model RTs, this is a natural assumption.) If the spoken response is considered a readout of the echo, its content may be estimated. Previous theories have related speech perception to production, usually positing connections by modular structures or abstract nodes (Cooper, 1979; MacKay, Wulf, Yin, & Abrams, 1993). Such models cannot make clear predictions regarding speech acoustics. Theories that propose an intimate perception—production linkage, such as motor theory (Liberman & Mattingly, 1985) or direct realism (Fowler, 1986, 1990b), may fare considerably better (see the General Discussion). Experiment 1A entailed manipulations of word frequency, number of token repetitions, and response timing. Also, the shadowing data were analyzed by "perceptual analysis" rather than by acoustic analysis. Each experimental manipulation was motivated by MINERVA 2; perceptual analysis was a pragmatic choice.

**Method**

For a detailed explanation of the method used in this experiment, see Appendix B.

**Word frequency.** A key diagnostic attribute in testing MINERVA 2 is word frequency. However, the words used by Goldinger (1996) came from the Modified Rhyme Test (House, Williams, Hecker, & Kryter, 1965) and did not ideally span frequency classes. For Experiment 1A, new words were selected with a better range and balance of frequencies—they were classified as high frequency (HF), medium high frequency (MHF), medium low frequency (MLF), and low frequency (LF). The words were recorded by multiple speakers, and experimental power was maximized by selecting speakers with a considerable "perceptual range" of voices. Fourteen volunteers recorded a short list of nonwords. Listeners rated the pairwise similarities of all voices, creating a matrix to analyze by MDS. With the scaling solution, 10 speakers who maximized perceptual variation were selected to record the full stimulus set.

**Repetitions.** Experiment 1A presented alternating blocks of listening trials and shadowing trials. In this manner, words were heard 0, 2, 6, or 12 times before shadowing. In theory, each repetition leaves an episodic trace, complete with voice and contextual details. Later presentations can then be tested for imitation. (It is also theoretically possible to observe imitation on the first presentation, especially for a LF or otherwise unique word.) If the stored traces are prominent in the echo used for shadowing, imitation should occur. This logic creates three predictions. First, as is typically observed, RTs should decrease as repetitions increase (Logan, 1990; Scarborough, Cortese, & Scarborough, 1977). In MINERVA 2, echo intensity will increase as more perfect matches to the stimulus token are compiled in memory. Second, imitation should increase as repetitions increase, as more traces resembling the stimulus token will contribute to echo content. Third, frequency effects should decrease with increasing repetitions, as occurs in printed word naming (Scarborough et al., 1977). Most models explain this interaction by short-term priming of canonical units, like logogens (Morton, 1969); HF words yield weak repetition effects because their thresholds are permanently near "floor." In MINERVA 2, with each repetition, echoes become increasingly characterized by context-specific traces created in the experiment. Thus, the model predicts a Frequency $\times$ Repetition interaction in both dependent measures—imitation and RT.

**Response timing.** One interpretive problem arises in this study; the imitation data are theoretically relevant only if they reflect a spontaneous response from memory to spoken words (i.e., if imitation reflects online perception). However, listeners may have a frivolous tendency to imitate voices, regardless of deeper lexical processes. The earlier results (such as the word frequency effect) cast doubt on such an atheoretic account, but the critical possibility of imitation as a general tendency demands consideration.

Experiment 1A included an immediate-shadowing condition, in which listeners shadowed words quickly after presentation. In this condition, participants may use echo content to drive articulation. Experiment 1A also included a delayed-shadowing condition (Balota & Chumbley, 1985), in which participants heard words but waited 3-4 s to speak. If people frivolously imitate voices while shadowing, they may persist in this behavior, despite waiting a few seconds. However, MINERVA 2 predicts that imitation will decrease over delays. The stimulus word should be recognized immediately. However, as the person holds it in WM, waiting to speak, continuous interactions occur between WM and LTM. This feedback loop will force a regression toward the mean of the stored category—each successive echo will "drift" toward the central tendency of all prior traces in LTM. Thus, idiosyncratic details of the original shadowing stimulus will be attenuated in the eventual echo used for output (see illustration in Hintzman, 1986, p. 416). Note that this is a progressive cycle: The first echo from LTM contains idiosyncrasies of the stimulus, but it is already somewhat abstract, as prior traces affect echo content. If the echo in WM is communicated to LTM again, the next echo will move closer to the central tendency of the stored category. After several seconds, the echo in WM—the hypothesized basis of a delayed-shadowing response—will be the lexical category prototype (perhaps the speaker's own voice). Thus, imitation should decline in delayed naming.

**Perceptual analysis.** The main dependent measure in Experiment 1A
was imitation of stimulus speakers by shadowing participants. However, “imitation” is quite difficult to define operationally. In the earlier experiment, acoustic parameters of the input and output utterances were compared, and imitation scores were derived. This approach had two major drawbacks. First, it is time consuming, severely limiting the data one can analyze. Second, the psychological validity of the imitation scores is unknown. Many acoustic properties can be cataloged and compared, but they may not reflect perceptual similarity between tokens—imitation is in the ear of the beholder.

If imitation scores miss the “perceptual Gestalt,” more valid measures may come from perceptual tests (Summers et al., 1988). Thus, each participants’ shadowing speech from Experiment 1A was used in Experiment 1B, an AXB classification task. On every trial, listeners heard two tokens of a word produced by a shadower: one from a baseline condition and one from the shadowing condition. These A and B stimuli surrounded the X stimulus—the original token that the shadower heard. AXB participants judged which stimulus, the first (A) or the third (B), sounded like a “better imitation” of the second (X). (Across groups, baseline tokens were counterbalanced across the first and third positions.) The percentage of listeners choosing the shadowed stimulus was used to estimate imitation in Experiment 1A.

In summary, Experiment 1A involved the collection of shadowing responses to words that varied in frequency, designated as LF, MLF, MHF, and HF words. Prior to shadowing, the words were heard (in listening blocks) 0, 2, 6, or 12 times. Additionally, words were either shadowed immediately on presentation or after a delay. All shadowing participants also recorded baseline tokens of all words by reading them aloud. After shadowing, each volunteer’s baseline and shadowing tokens were juxtaposed against the original stimulus tokens for AXB classification—listeners indicated which token (A or B) sounded like a better imitation of X. (Further methodological details are provided in Appendix B.) The expected results were (a) stronger imitation for lower frequency words, (b) stronger imitation with more repetitions, (c) an interaction of these factors, and (d) decreased imitation in delayed shadowing.

Results and Discussion

Experiment 1A. The “data” (i.e., the recorded tokens) from Experiment 1A were primarily used to generate stimulus materials for Experiment 1B. However, the shadowing RTs were also analyzed. When Figure 3 is examined, several key results are evident (statistical analyses for all data are summarized in Appendix C). The immediate-shadowing RTs (top of Figure 3) showed clear effects of frequency (faster RTs to higher frequency words) and repetition (faster RTs with increasing repetitions). The delayed-shadowing RTs (bottom of Figure 3) also showed a repetition effect, but no frequency effect. In general, the RTs suggested that the stimulus words were chosen and manipulated appropriately. Classic frequency and repetition effects emerged, with their usual interaction (Scarborough et al., 1977). Accordingly, these results provide a foundation to examine Experiment 1B.

Experiment 1B. Figure 4 shows the percentage of correct AXB judgments (collapsed across shadowing participants), as a function of word frequency, repetitions, and delay. In this study, “correct” AXB judgments were scored whenever a listener selected a shadowing token—rather than a baseline token—as the imitation. When Figure 4 is examined, several major effects are evident. When the tokens were produced in immediate shadowing, participants were far more likely to detect imitation, relative to tokens produced in delayed shadowing. Almost all cell means exceeded chance (50%) in immediate shadowing, but few exceeded chance in delayed shadowing. In addition to the delay effect, other predicted effects were observed: In both immediate and delayed shadowing, imitation increased when the tokens were lower frequency words, although the frequency effect was stronger in immediate shadowing. Also, in immediate shadowing, imitation increased with increasing repetitions.

The basic assumption needed to interpret these data concerns the nature of perception in the shadowing task and its bearing on speech acoustics. In MINERVA 2, echoes constitute the model’s only basis to respond. Hintzman (1986) showed that echo content consists of blended information—probes and stored episodes combine to form experience. If a response is made by using the first echo, its similarity to the probe should be considerable. This idea was supported in Experiment 1A; in immediate shadowing, certain trials (low frequency and high repetitions) invoked strong imitation. In contrast, if a response is generated slowly, the echo should cycle between WM and LTM, its content growing progressively less similar to the original probe. This prediction was also supported in Experiment 1A; in delayed shadowing, all imitation was reduced to near-chance levels.
Experiments 2A and 2B: Shadowing Nonwords in a Balanced Lexicon

Experiments 1A and 1B were encouraging; the data suggest that the acoustic content of shadowers' speech reflects underlying perceptual processes. Moreover, these processes are seemingly affected by detailed episodic traces. However, for several reasons, the results of Experiments 1A and 1B are equivocal. One challenge in this research is to ensure that vocal imitation in shadowing is a truly "lexical" response rather than a general tendency. Several precautions in Experiment 1A helped avoid this interpretive impasse. Words of several frequency classes were used and were repeated different numbers of times, and delayed shadowing was examined. Each factor modified the likelihood of imitation, which seems to rule out a simplistic "general tendency" account.

Unfortunately, although these precautions worked in Experiment 1A, none is sufficiently compelling. With respect to delayed shadowing, voice tracking may be a strategic process that makes immediate shadowing easier, but it does not help delayed shadowing. With respect to repetitions, hearing a token numerous times may create anticipation effects. For example, the early phonemes of a word may trigger a memory of its recent presentation. Participants may then imitate the speaker for any number of reasons. For these reasons, word frequency was the key to Experiment 1A. Relative to delay or repetition, the frequency manipulation was quite subtle. In theory, participants were oblivious to the differences, suggesting that frequency-sensitive imitation is a spontaneous effect. Unfortunately, other potential problems arose. To correct these, in Experiment 2A I examined nonword shadowing, using the same manipulations as before.

There were two main reasons to replicate Experiment 1A with nonwords. First, the use of nonwords with controlled frequencies should provide "cleaner" data to evaluate the simulation model. The Kučera and Francis (1967) frequency estimates predict data quite well, but they also introduce considerable noise. For example, some highly familiar words (e.g., violin and pizza) have very low-frequency estimates (Gernsbacher, 1984). By creating a "nonword lexicon" for participants, the shadowing and simulation data are more comparable than real words allow (see Feustel et al., 1983; Salasoo et al., 1985).

The second, more important reason to use nonwords in Experiment 2A was to remove a potential frequency-based confound. The words for Experiment 1A were originally recorded by cooperative volunteers who, presumably, tried to provide clear stimuli. Unfortunately, prior research shows that speakers tend to hyperarticulate LF words, at least with respect to duration (Wright, 1979). Thus, the original stimulus recordings for Experiment 1A may have contained systematic acoustic differences confounded with frequency. Following this logic to its dreary conclusion, if LF words were exaggerated in the stimuli, they may have induced greater imitation during shadowing. Also, imitation may be more easily detected in exaggerated words— if a bisyllabic LF word had a clear rise–fall intonation, it would be easy to judge whether its shadowed counterpart had the same intonation. If a bisyllabic HF word had a flat intonation, it would be difficult to judge if its shadowed counterpart matched. Two clear images are easier to compare than two noisy images.

The use of nonwords can ensure that stimulus confounds do not create frequency-based imitation differences. In terms of frequency, all nonwords should be roughly equivalent to recording volunteers, precluding systematic differences. Also, nonwords can be equally assigned to frequency conditions, eliminating all pronunciation differences across frequency classes. In Experiment 2A, the assignments of nonwords to frequency conditions were counterbalanced across shadowing participants. This was accomplished by presenting training and shadowing sessions on consecutive days. Using procedures from the listening blocks in Experiment 1A, I used the training sessions to create a nonword lexicon for shadowing participants. The only manipulated factor in training was exposure frequency: Nonwords were presented once each (LF), twice each (MLF), 7 times each (MHF), or 20 times each (HF). However, to avoid familiarizing listeners with the exact tokens used in shadowing, all training tokens were spoken by one novel speaker (whose voice was not used in test sessions). Shadowing sessions were completed on the second day, using the procedures of Experiment 1A (see Appendix B). As before, Experiment 2A was followed by an AXB classification test (Experiment 2B).
Method

For a detailed explanation of the method used in this experiment, see Appendix B.

Results and Discussion

Experiment 2A. The shadowing RTs closely resembled those from Experiment 1A (see top of Figure 5 and Appendix C). As before, immediate-shadowing RTs showed strong frequency and repetition effects (and their interaction). These effects were also evident, but attenuated, in delayed shadowing. As before, the RT data suggested that the key variables in Experiment 2A were manipulated over an acceptable range.

Experiment 2B. The mean ‘correct’ AXB classification rates for immediate- and delayed-shadowing tokens are shown at the top of both Figures 6 and 7, respectively. Imitation was virtually always detected in immediate shadowing, but it was rarely detected in delayed shadowing. As in Experiment 1B, robust frequency and repetition effects were observed in immediate shadowing. These effects were also observed, but attenuated, in delayed shadowing. However, unlike Experiment 1B, the frequency and repetition effects appeared additive in immediate shadowing rather than producing an interaction (see Appendix C for statistical analyses).

Simulation of Experiments 2A and 2B in MINERVA 2

As Hintzman (1986) noted, although MINERVA 2 is a quantitative model, it is best suited for qualitative analysis. If it predicts the major trends of the data, the model may constitute a reasonable account. To confirm that MINERVA 2 predicts the shadowing results, I conducted a simulation. To approximate a human participant, I initially stored a background lexicon of 1,000 “words” (random 200-element vectors), with randomly generated frequencies of 1–100 traces (only name elements were repeated across traces; voice and context elements were randomized). Next, 160 “nonwords” were generated. These were 200-element vectors, with 100 name elements (none matching background “words”), 50 voice elements, and 50 context elements. To mimic the training sessions of Experiment 2A, 40 HF nonwords were each stored 20 times, with constant name, voice, and context elements. Similarly, MRF, MLF, and LF nonwords were stored 7, 2, and 1 time(s), respectively. After training, the model completed three forgetting cycles, allowing random elements to revert to zero (see Appendix A).

Both dependent measures of Experiments 2A and 2B were simulated in tandem. Hintzman (1986, 1988) used echo intensities to model recognition memory and frequency judgments. In the present test, inverse echo intensities were assumed to provide reasonable RT estimates. Vocal imitation was estimated by echo content. In concrete terms, the model is given a 200-element probe vector with three basic elements: —1, 0, and 1. An echo may preserve the probe’s basic character, but it contains continuously valued elements between —1 and 1. To estimate imitation in the model, I converted these continuously valued elements back to discrete values by a program that rounded to whole values. (Values less than or equal to —.4 were converted to —1.

Figure 5. Immediate-shadowing response time (RT) data and MINERVA 2 simulation, Experiment 2A. HF = high frequency; MHF = medium high frequency; MLF = medium low frequency; LF = low frequency.

4 For reasons of expediency and validity, in the present study I used AXB classification (rather than acoustic analysis) to assess degrees of imitation. The AXB data confirmed that listeners detected imitation in the shadowers’ speech but did not reveal its perceptual basis. Although aspects of the speech signal making up imitation were not directly relevant to this research, it does pose an interesting question. Several acoustic factors seem likely candidates, including duration, amplitude, fundamental frequency (F0), and intonation contour. To examine which acoustic factors were compelling indicators of imitation, several tests were conducted, again using AXB classification. Fifty stimulus sets were selected that yielded high rates (92%) of “correct” AXB classification in Experiment 2B and were used to generate five new tests. In a control test, the stimuli were unchanged. In an equal duration test, all three nonwords per trial were modified by a signal processing package (CSL, by Kay Elemetrics) to have equal durations. Thus, duration cues could not be used to detect imitation. In similar fashion, three more AXB tests were generated in which mean amplitude, F0, and intonation contour were equated, respectively. (I am indebted to Joanne Miller and Keith Johnson for suggesting this method.) Groups of 10 listeners received each test. Predictably, the control test produced the best performance (87% correct), followed by the amplitude test (80%), F0 (78%), duration (63%), and intonation contour (59%) tests. The removal of any acoustic cue decreased the detectability of imitation, but only the duration and intonation tests reliably differed from control. From these data, it seems that temporal and melodic factors are particularly salient cues to imitation. However, pending a complete investigation (with acoustic factors tested in various combinations), this suggestion must be considered tentative.
and values greater than or equal to .4 were converted to 1. Intermediate values were converted to 0.) Imitation was then estimated by the proportion of position-specific voice elements with identical values. For the test session, another set of the same 160 nonwords was generated, with all of the name and context elements used in training. However, new configurations of voice elements denoted 10 new “speakers.” The simulation followed the experiment: 20 nonwords were presented once and their echoes were examined. Another 20 nonwords were presented twice; their echoes were examined after the second presentation. Echoes for 20 more nonwords were examined after their 6th presentation, and echoes for another 20 nonwords were examined after their 12th presentation. As in Experiment 2A, equal numbers of nonwords from each frequency class were included at each level of repetition.

The top of Figure 5 shows immediate shadowing RTs from Experiment 2A. The bottom of Figure 5 shows simulated RTs and clear qualitative agreement to the data. Figures 6 and 7 show simulated imitation data as proportions of “echoed voice elements” from LTM in response to probes. Figure 6 shows real and simulated AXB data from immediate shadowing; Figure 7 shows delayed shadowing. Delayed shadowing was simulated by feeding successive echoes back to the model 10 times after the first probe, allowing the resultant echo to drift toward the central tendency of the stored traces. (The selection of 10 cycles was fairly arbitrary, chosen in tandem with the forgetting parameter to provide noticeable forgetting, without complete erasure of stored information.) As both figures show, the model adequately predicted the basic trends of the imitation judgment data.6

Experiments 3A, 3B, and 3C: Shadowing Nonwords in a Skewed Lexicon

The use of nonword stimuli in Experiments 2A and 2B reinforced the prior results. In addition to alleviating possible stimulus confounds, Experiment 2A allowed more precise frequency manipulations than is possible with real words. In effect, the use of nonwords allows experimental creation of a participant’s “lexicon,” approximating the situation for MINERVA 2. Similar procedures are commonly applied to study perceptual categorization (e.g., Maddox & Ashby, 1993; Nosofsky, 1986; Posner & Keele, 1970). The use of nonwords as training and test stimuli confers another advantage—it is possible to shape the character of the stored categories. In Experiments 2A and 2B, items varied only in frequency; other aspects of the tokens (context of experience and voice characteristics) were held constant. In Experiment 3A, I again used nonwords introduced to participants in a training session. As before, the nonwords varied in training frequency and were presented for immediate or delayed shadowing after variable repetitions. However, in Experiment 2A, participants heard all nonwords in one training voice, ensuring fairly homogenous representations. Experiment 3A entailed more idiosyncratic training for each nonword. All 10 test voices were used in training but were not distributed within nonwords. Instead, the same voice was used for every repetition of any given nonword during training. In test sessions, voices were manipulated: Training voices were repeated in all listening blocks. However, during shadowing, half of the nonwords retained their training voices (SV), and half were presented in voices that were highly dissimilar to the training voice (DV), determined by the earlier MDS experiment. MINERVA 2 makes several interesting predictions for this procedure.

First, in immediate shadowing, participants should strongly imitate SV items, relative to DV items, and SV imitation should increase with repetitions. In SV trials, all stored tokens match the shadowing stimulus, making these predictions transparent. By contrast, DV items should show weaker imitation with in-

---

5 This estimation method was used for communicative clarity—it provides percentage scores, which are easily compared with the AXB classification data. However, given two vectors of equal length, an alternative (and perhaps more accurate) method is to compute dot products, which increase linearly with vector similarity. To test the validity of the present method, I also computed dot products (also called standard inner products). The results showed qualitative trends nearly identical to the present illustrations.

6 When the AXB data are compared to the simulations, note that chance is defined differently for each. Chance performance in AXB classification equals 50% correct. For the simulation, chance equals a random correlation of three-valued vector elements (−1, 0, +1) and is thus equal to 33% echoed voice elements.
creased repetitions, as memory amasses traces that will contradict the subsequent shadowing voice. Thus, the model predicts a Voice × Repetition interaction. Also, these effects should be sensitive to the nonword frequencies established in training. For SV trials, frequency effects should contradict the prior data—HF nonwords should now induce greater imitation than LF nonwords. In SV trials, the repetition and frequency manipulations are functionally identical; increases in either predicts greater imitation. By contrast, in DV trials, HF nonwords should be most resistant to imitation because many stored traces “work against” the shadowing stimulus. Thus, the model also predicts a Voice × Frequency interaction.

A second prediction involves delayed shadowing. In earlier experiments, imitation was expected to decrease in delayed shadowing. In Experiment 3A, this prediction was modified: In DV immediate-shadowing trials, echoes should partially reflect the probe stimuli, perhaps yielding some detectable imitation. However, in DV delayed-shadowing trials, responses may increasingly resemble the training stimuli, rather than the shadowing stimuli. As memory systems interact over the delay, each successive echo should drift toward the central tendency of the learned nonword category. In Experiment 3A, this central tendency was skewed toward the training voice. For the same reason, another prediction arose: In SV delayed-shadowing trials, there should be no decrease in imitation because all traces in WM and LTM support imitation. Thus, MINERVA 2 also predicts a Voice × Delay interaction.

As before, Experiment 3B was an AXB test juxtaposing baseline and shadowing tokens against shadowing stimulus tokens. However, to examine the unique predictions regarding training voices, I also conducted Experiment 3C. This was identical to Experiment 3B, but listeners heard training tokens (rather than shadowing stimulus tokens) as X stimuli. Thus, imitation of shadowing and training tokens was separately estimated.

Method
The methods for Experiments 3A, 3B, and 3C are summarized in Appendix B.

Results
Detailed results are presented in Appendix C. Thus, in the interest of brevity and clarity, the basic data patterns are reviewed in tandem with their associated simulations.

Simulation of Experiments 3A, 3B, and 3C in MINERVA 2
After the experiments, qualitative fits of MINERVA 2 to the data were examined. The simulations were conducted as previously described, with one exception: Half of the probes in shadowing sessions retained their training voice elements; half had new voice elements, taken from the set of 10 training voices. As before, RTs were estimated by inverse echo intensities, and imitation was estimated by proportions of echoed voice elements.

Experiment 3A. The top of Figure 8 shows the immediate-
shadowing RTs as a function of voice and repetitions (collapsed across nonword frequencies). Two key trends are shown—RTs decreased across repetitions (as before), and SV trials produced faster responses. The bottom of Figure 8 shows the simulated RTs, which showed the same major trends. Examining Experiment 3A further, Figure 9 shows real and simulated RTs as a function of voice and frequency, collapsed across repetitions. As shown, the model adequately predicts both the observed SV advantage and the frequency effect.

**Experiment 3B.** Figure 10 shows correct AXB classification rates for the immediate-shadowing tokens, shown as a function of voice and repetitions, collapsed across frequencies. Figure 11 shows the same data as a function of voice and frequency, collapsed across repetitions. Several main trends emerged in the data. First, imitation was stronger in SV trials. Second, imitation increased across repetitions, equivalently for SV and DV trials. Third, a predicted Voice × Frequency interaction emerged: Imitation slightly increased with frequency decreases in DV trials but showed the opposite trend in SV trials. As Figures 10 and 11 show, the model nicely predicts these qualitative data patterns.

The next simulations concerned the delayed-shadowing results. The top of Figure 12 shows AXB data for delayed-shadowing tokens as a function of voice and repetitions, collapsed across frequencies. Similarly, Figure 13 shows AXB data as a function of voice and frequency, collapsed across repetitions. In general, the data in Figure 12 resembled those in Figure 10, showing voice and repetition effects. However, these effects were both attenuated, relative to the immediate-shadowing condition. Similarly, the data in Figure 13 resembled the immediate-shadowing data in Figure 11, but with attenuated effects. As shown, MINERVA 2 predicted these effects and their diminishing magnitudes across delays.

**Experiment 3C.** Recall that Experiment 3C differed from the prior AXB tests by using training tokens—rather than shadowing stimulus tokens—as comparison standards. Accordingly, this change was applied to the Experiment 3C simulation: Echoes were compared with training stimuli, not test stimuli. Figures 14 and 15 show real and simulated AXB classification data for the immediate-shadowing tokens. As predicted, SV trials promoted robust imitation, in patterns similar to Experiment 3B. Figures 14 and 15 confirm that MINERVA 2 predicted the observed trends. The most interesting aspect of Experiment

---

7 In the simulations of Experiment 3C, chance performance was not defined as 33%, as before. Because a defined set of 10 voice vectors was available, their mean proportions of overlapping elements could be calculated; this value (41%) represents chance performance for the model to reproduce the training voice.
3C was the delayed-shadowing condition. Specifically, it was hypothesized that DV trials would reverse their prior pattern; after a delay, the shadowers' responses would come to resemble the training tokens. As shown in Figures 16 and 17, this prediction was supported; SV and DV trials produced nearly equivalent imitation. Moreover, the simulations shown in each figure verify the model's qualitative predictions.

General Discussion

The present findings, together with other data (Tenpenny, 1995), suggest an integral role of episodes in lexical representation (Jacoby & Brooks, 1984). Prior research has shown that detailed traces of spoken words are created during perception, are remembered for considerable periods, and can affect later perception—data most naturally accommodated by assuming that the lexicon contains such traces. The present study extends such prior research, showing episodic effects in single-word and nonword shadowing. Moreover, a strict episodic model (Hintzman, 1986) produced close qualitative fits to the data. Clearly, this does not mean the model is correct, but it provides some validation of the multiple-trace assumption.

The Speaker Normalization Hypothesis

Abstract representation is both an old and accepted idea in psycholinguistics. Indeed, Marslen-Wilson and colleagues (Gas-

---

Figure 11. Immediate-shadowing imitation data and MINERVA 2 simulation of Experiment 3B, shown as a function of voice and frequency, collapsed across repetitions. HF = high frequency; MHF = medium high frequency; MLF = medium low frequency; LF = low frequency.

Figure 12. Delayed-shadowing imitation data and MINERVA 2 simulation of Experiment 3B, shown as a function of voice and repetitions, collapsed across frequencies.
Kuhl, Meltzoff, and Stevens (1991) demonstrated a "cross-gender McGurk effect"—incongruous faces and voices fluently combine to yield the illusion (McGurk & MacDonald, 1976). Green et al. suggested that normalization occurs early in processing, allowing fusion of abstract representations, but they also noted that voice information remains.

Differences in the gender of the talker producing the auditory and visual signals had no impact on the integration of phonetic information. Thus, by the time the phonetic information was integrated from the auditory and visual modalities, it was sufficiently abstract as to be neutral with respect to the talker differences. Nonetheless, observers are very aware of an incompatibility in the cross-gender face–voice pairs. This suggests that the neutralization of talker differences for the purposes of phonetic categorization does not result in a loss of detailed information about the talker. (Green et al., 1991, p. 533)

Indeed, I contend that no published evidence shows that normalization reduces information. Several models posit perceptual compensation without information loss (Miller, 1989; Nearey, 1989; Syrdal & Gopal, 1986), showing that normalization and voice memory can peacefully coexist. However, is normalization theoretically necessary? Most theories treat it as a logical necessity because variable signals must be matched to summary representations. However, an episodic lexicon should support direct matching of words to traces, without normalization. Moreover, aside from null effects (e.g., Jackson & Morton, 1984), few data truly support normalization.

Consider vowel perception: Verbrugge and Rakerd (1986) presented "silent-center" syllables to listeners for identification. These /bVb/ syllables had their central 60% removed, leaving only the initial and final consonants with partial vocalic transitions. Listeners easily identified the missing vowels from these impoverished signals. In another condition, syllable pieces produced by men and women were spliced together, creating new silent-center stimuli. Although the speakers' vowel spaces differed widely, missing vowels were still easily identified. Verbrugge and Rakerd concluded that vowels are not identified by center frequencies, as most theories assume. Instead, speaker-independent articulatory information affords accurate perception (Fowler, 1986).

The Episodic Lexicon

Although many theories consider normalization a logical necessity, episodic models provide an alternative. As Jacoby and his colleagues have noted, many data suggest that episodes subserve perception. For example, Jacoby (1983b) suggested that word perception occurs nonanalytically, by comparison to prior episodes, rather than by decomposition into features. In the present research, an episodic model (MINERVA 2) was found...
AN EPISODIC LEXICON

Experiment 3C
Immediate Shadowing

Figure 15. Immediate-shadowing imitation data and MINERVA 2 simulation of Experiment 3C, shown as a function of voice and frequency, collapsed across repetitions. HF = high frequency; MHF = medium high frequency; MLF = medium low frequency; LF = low frequency.

to predict data from an ostensibly perceptual task. Thus, it seems parsimonious to suggest that episodes form the basic substrate of the lexicon.

Although MINERVA 2 was tested in this research, other models provide viable accounts of the data. For example, both the generalized context model (Medin & Schaffer, 1978; Nosofsky, 1986) and the SAM (search of associative memory) model (Gilund & Shiffrin, 1984) incorporate multiple-trace assumptions. MINERVA 2 was used here for pragmatic and theoretical reasons. On the pragmatic side, it is easily simulated, by virtue of simple representations and a small set of computations. On the theoretical side, MINERVA 2 has two benefits in the present application. First, it makes the extreme assumption of numerous, independent memory traces. Because the present goal was to assess the viability of an episodic lexicon, this unwavering assumption was desirable. Second, it makes simultaneous predictions regarding echo intensity and content, which naturally conform to the dependent measures in shadowing (RTs and speech acoustics).

Hybrid Models

MINERVA 2 is a purely episodic model that predicts prior results (Goldinger, 1996) and the present results. However, less extreme models may also work. Feustel et al. (1983; Salasoo et al., 1985) described a hybrid model in which both abstract lexical codes and episodic traces contribute to perception. By this view, words become codified by repetition—multiple episodes coalesce into units (similar to logogens). Episodes mediate token-specific repetition effects, but abstract codes provide the lexicon stability and permanence. In Klatt’s (1979) model of speech perception, phonetic variations are stored in memory, alongside lexical prototypes. Similarly, Tulving and Schacter (1990; Schacter, 1990) proposed a perceptual representation system (PRS) to identify objects, including words. PRS contains long-lasting traces of perceptual forms, with all details intact. Complementary central memory systems contain abstract information, such as category prototypes and conceptual associations.

In a particularly germane hybrid model, Kirsner et al. (1987) proposed a lexicon of abstract representations and episodic procedural records. In this model, word perception entails special processes that match stimuli to abstract lexical entries. Records of these processes are stored in memory, and surface details (such as voice) shape the record. On later word perception, past records are reapplied to the degree they resemble new inputs (although see Dean & Young, 1996). Regarding repetition effects, Kirsner et al. (1987) wrote the following:

The essence of our account is that word identification is achieved
by reference to a record. Similarity is the critical parameter. If the record collection includes an example that is similar to the current stimulus description, identification will be achieved easily and quickly. (p. 151)

The record-based model borrows logic from Kolers (1976; Kolers & Ostry, 1974), who suggested that fluent rereading of transformed text reflects memory for perceptual operations. Whereas Kolers studied strategic processes applied to a difficult perceptual task, Kirsner et al. (1987) assumed that procedural records arise for all perceptual processes, regardless of difficulty or salience. For example, recognizing a word in an unfamiliar voice will invoke normalization and matching procedures that are stored in a record. Later perception of a similar word will use the record, creating residual savings. With increased exposure to a certain voice (or handwriting, rotated text, foreign accent, etc.), the growing episode collection will support asymptotic (totally "normalized") performance. As a concrete example, Nygaard, Sommers, and Pisoni (1994) made listeners familiar with speakers' voices and found facilitated perception of new words produced by those speakers.

MINERVA 2 assumes that perceptual products (e.g., recognized words) are stored episodically. The record-based model assumes that perceptual processes are stored, alongside abstract representations. Clearly, these models are very difficult to discriminate— their central mechanisms and predictions may be formally identical. For example, it is commonly reported that voice (or font) effects in word perception are strongest when procedural cues are constant across study and test (Graf & Ryan, 1990; Masson & Freedman, 1990; Whittlesea, 1987; Whittlesea & Brooks, 1988; Whittlesea & Cantwell, 1987). On first consideration, such data appear to favor procedural models. Indeed, Ratcliff and McKoon (1996, 1997; Ratcliff, Allbritton, & McKoon, 1997) recently developed a process-based model of priming effects. In this model, perceptual processes are temporarily modified by stimulus processing, creating a bias to benefit later, similar stimuli. However, the same data are explicable by perceptual products (episodic traces) rather than by processes. Ratcliff and McKoon (1996) recognized this and postulated a potential role for episodes in the flow of information processing.

**Distributed Models**

Another alternative to pure episodic models are distributed models (e.g., Knapp & Anderson, 1984). In McClelland and Rumelhart's (1985) model, memory traces are created by activation patterns in a network. The trace for each stimulus is unique and can be retrieved by repeating its original pattern. The model develops abstract categories by superimposing traces, but its storage is more economical than MINERVA 2. McClelland and Rumelhart (1985) wrote the following:

> Our theme will be to show that distributed models provide a way to resolve the abstraction—representation of specifics dilemma. With a distributed model, the superposition of traces automatically results in abstraction though it can still preserve to some extent the idiosyncrasies of specific events and experiences. (p. 160)

The distributed model presents a reasonable compromise between episodic and abstract models. For example, it is easy to imagine how distributed networks derive central tendencies from exemplars. However, with all memory traces superimposed, it is unknown whether distributed models could display adequate sensitivity to perceptual details, as in the present data. Can repetition of an old word have a "special" effect after many similar words are combined in a common substrate? Presumably, if contextual encoding sufficiently delimits the traces activated during test (as in MINERVA 2), such results are possible.

**Motor Theory and Direct Realism**

Although this discussion has focused on models of lexical memory, the data are relevant to issues beyond episodic representations. The vocal imitation observed in shadowing strongly suggests an underlying perceiving—production link (Cooper, 1979; Porter, 1987) and is clearly reminiscent of the motor theory (Liberman et al., 1967; Liberman & Mattingly, 1985). In classic research conducted at Haskins Laboratories (New Haven, CT), it was discovered that listeners' phonetic percepts do not closely correspond to acoustic aspects of the speech signal. Instead, perception seems to correspond more directly to the articulatory gestures that create the signal. For example, the second-formant transition in the stop consonant /d/ varies dramatically across vowel environments, but its manifestations all sound like /d/. The motor theorists noted that perception fol-
lowered the articulatory action that creates a /d/—the tongue blade contacts the alveolar ridge. Given this stable action—perception correspondence, Liberman et al. (1967) suggested articulatory gestures as the objects of speech perception.

The original motor theory hypothesized that listeners analyze speech by reference to their own vocal tracts. The idea was that subphonemic features are specified by motions of semi-independent articulators. When this notion of feature specification was later found to be implausible (Kelso, Saltzman, & Tuller, 1986), the motor theory was revised (Liberman & Mattingly, 1985). The idea of “analysis by synthesis” was retained, but the goal was to retrieve a speaker’s “gestural control structures,” one level abstracted from physical movements. This process hypothesizes a few candidate gestures that may have created the speech signal, with corrections for coarticulation. Liberman and Mattingly (1985) wrote the following:

We would argue, then, that gestures do have characteristic invariant properties, as the motor theory requires, though these must be seen, not as peripheral movements, but as the more remote structures that control the movements. These structures correspond to the speaker’s intentions. (p. 23)

Although the mechanics of analysis by synthesis are not well specified, Liberman and Mattingly (1985, 1989) listed some necessary properties, which are easily summarized: Speech perception is a ”special” process, fundamentally different from general auditory perception. This is true with respect to decoding processes, neural underpinnings, and eventual products. To accommodate such a unique perceptual system, Liberman and Mattingly (1989) suggested that analysis by synthesis occurs in a module, independent of other perceptual or cognitive systems (Fodor, 1983, 1985). As has been argued elsewhere (Fowler & Rosenblum, 1990, 1991), this modularity assumption is fairly problematic. With respect to the present research, I have suggested that episodic memory traces are fundamentally involved in spoken word perception (cf. Jacoby & Brooks, 1984). However, a primary tenet of modularity is information encapsulation, which states that perception occurs without top-down influence. As such, it may be impossible to reconcile episodic perception with modularity.

A related theory that fares better is direct realism, described by Fowler (1986, 1990a, 1990b; Fowler & Rosenblum, 1990, 1991). As in motor theory, direct realism assumes the objects of speech perception are phonetically structured articulations (gestures). The term direct realism follows from Gibson’s (1966) view of visual event perception. A key aspect of Gibson’s theory is a distinction between events and their informational media. When people gaze on a chair, they perceive it via reflected light that is structured by its edges, contours, and colors. People do not perceive the light; it is merely an informational medium. Fowler’s suggestion for speech is very similar—articulatory events lend unique structure to acoustic waveforms, just as chairs lend structure to reflected light. Speech perception entails direct recovery of these articulatory gestures. Fowler (1990a) noted the following:

While it has taken speech researchers a long time to begin to understand coarticulation and suprasegmental layering, listeners have been sensitive to their structure all along. Listeners are remarkably attuned to talkers’ behavior in producing speech. (p. 113)

Although direct realism resembles the motor theory, there are important differences. Most notably, motor theory maintains that speech is subjected to computations that retrieve underlying gestures. In contrast, direct realism maintains that cognitive mediation is unnecessary—the signal is transparent with respect to its underlying gestures. As such, Fowler and Rosenblum (1991) suggested that modularity is unwarranted; general perceptual processes can recover the distal events in speech (see Porter, 1987, for a similar view).

According to Fowler (1986, 1990b), direct-realist speech perception is unmediated—it does not require inferences via mental representations, as in information-processing models. On first consideration, the assumption of unmediated perception is at variance with the present data. By definition, episodic perception is cognitively mediated. However, unlike motor theory, there is room for compromise in direct realism. Because it does not assume encapsulated processing, effects of perceptual learning are possible. Indeed, Sheffert and Fowler (1995) recently replicated the Palmeri et al. (1993) finding of voice memory in continuous recognition. They explained their data by combining direct realism with an episodic view of the lexicon.

Stored word forms may not be abstract representations stripped of information about the episodes in which they were perceived, but instead may be exemplars that contain speaker-specific information. An exemplar-based theory of the lexicon leads us to view normalization as a way of perceiving words that distinguishes invariant phonological information from invariant speaker information, but does not eliminate the latter information from memory for a word. When speakers produce words . . . different vocal tract actions structure the air distinctively [creating] the consonants and vowels of spoken words. In addition, however, the idiosyncratic morphology of the speaker’s vocal tract, the speaker’s affect, and other variables also structure acoustic speech signals distinctively. (Sheffert & Fowler, 1995, p. 682)

In essence, Sheffert and Fowler (1995) suggested that episodes created in word perception are gesturally based, which does not undermine the attractive properties of direct realism. Indeed, their logic is reminiscent of an insightful article in which Shepard (1984) attempted to reconcile Gibson’s direct realism with information-processing views of internal representation. Shepard noted that memory for perceptual invariants is a likely consequence of evolution, just as Gibson (1966) argued for sensitivity to invariants. Moreover, when signals are impoverished (or absent, as in dreaming), these internalized constraints of the physical world can support ”perception,” in various forms. Of particular relevance to the present article, Shepard (and Gibson, 1966) addressed internalized constraints that arise through individual learning. When stored representations are added to a theory of perception, researchers can apply a resonance metaphor (cf. Grossberg, 1980). Shepard suggested that ”as a result of biological evolution and individual learning, the organism is, at any given moment, tuned to resonate to incoming patterns” (1984, p. 433). Notably, the view of perception as a resonant state between signals and memories is precisely the view held in episodic memory models, including Semon’s (1909/1923) theory and Hintzman’s (1986) MINERVA 2.
Lexical Processes Beyond Perception?

Throughout this article, all references to “lexical processes” have implicitly been limited to perception of lexical forms. However, lexical processes outside the laboratory extend far beyond perception. Conversation requires syntactic parsing, ambiguity resolution, and so forth—processes that seem less amenable to episodic processing. This is a legitimate concern; simple models like MINERVA 2 cannot explain sentence or discourse processing. Moreover, people typically converse in a realm of ideas, without focusing on tangential information, such as voice details or environmental context. In short, perception seems abstract in natural language, relative to tasks such as single-word shadowing.

A related concern is the reliability of surface-specific effects in word perception. Both font- and voice-specific repetition effects have inconsistent histories in the literature (see Goldinger, 1996; Tenpenny, 1995). To observe robust effects, researchers typically need to contrive conditions that deviate from natural language experience. For example, voice and font effects are enhanced when attention is focused on surface attributes during study (Goldinger, 1996; Meehan & Pilotti, 1996) or when particularly salient attributes are used (Jacoby & Hayman, 1987; Kolvers, 1976). Surface-specific effects are also most evident when transfer-appropriate processing is applied in test sessions; episodic memory is strongly expressed when study operations are repeated at test (Blaxton, 1989; Graf & Ryan, 1990). This occurs with perceptual operations (such as translating rotated text) and with more abstract processes. For example, Whittlesea (1987; Whittlesea & Brooks, 1988; Whittlesea & Cantwell, 1987) has repeatedly shown that episodic effects in word or nonword processing are modulated by the purpose of experiences. When perceptual and contextual cues are repeated, they benefit processing. When perceptual cues are repeated in a new context (or new task), such effects are minimized. Taken together, the data suggest that episodic traces are not perceptual analogues, totally defined by stimulus properties. Rather, they seem to be “perceptual–cognitive” objects, jointly specified by perceptual forms and cognitive functions (Van Orden & Goldinger, 1994).

Beyond laboratory tasks, transfer-appropriate processing may help rationalize episodic models in several respects. For example, episodic models provide an intuitive account of token repetition effects, but they have generally weak intuitive appeal. Even when forgetting is assumed (Hintzman, 1986), it is difficult to imagine storing so many lexical episodes in memory. A related problem regards the ambiguous boundaries of linguistic events. In the laboratory, lexical episodes naturally conform to experimental trials. However, in real language, words are fairly subordinate entities. Because speech is typically used to converse, most episodes should emphasize elements of meaning, not perception. Ideas may be distributed over long or short utterances, which demands flexible episodic boundaries. This suggestion has empirical support: The attention hypothesis in Logan’s (1988) instance theory predicts that people will learn constellations of co-occurring features, provided they were attended. For example, attended word pairs are apparently stored as single episodes (Borovat & Logan, 1997; Logan & Etherton, 1994; Logan, Taylor, & Etherton, 1996). By extension, paying attention at the level of discourse will predict the creation of discourse-sized episodes. The episodic lexicon may not be a word collection; it may contain a rich linguistic history, reflecting words in various contexts, nuances, fonts, and voices.

This idea is reminiscent of Shepard’s (1984) reply to Gibson’s (1966) complaints about laboratory studies of vision. Gibson readily agreed that “laboratory vision” (e.g., tachistoscope studies) may rely on memory and perceptual inferences. However, he considered their likely contributions to “ecological vision” minimal, as viewers enjoy continuous illumination, eye movements, and so forth. Shepard (1984) later suggested that internal and external constraints can work in harmony, exercising a division of labor as the occasion requires. I suggest a similar role for linguistic episodes; in laboratory tests, isolated words are presented for idiosyncratic purposes. As a result, voice or font effects arise when the same unique contexts and stimuli are reinstated. However, other effects in word perception arise across virtually all procedures or participants. Examples of such robust effects are word frequency, semantic priming, and benefits of context.

If the natural units of episodic storage are stretches of real discourse, this data pattern is readily explained. Voice-specific repetition effects require access to unique memory traces. By contrast, word frequency and semantic priming effects should be supported by a groundswell of all stored traces. By experiencing a word in many contexts, a person will come to appreciate its high-frequency status, syntactic roles, and associative links to other words. A basic assumption in cognitive psychology is that sources of redundant information may trade-off in perception and memory (Neisser, 1967). By storing words in variable contexts, a person will amass myriad routes back to those words. Indeed, Hintzman (1986, p. 423) noted that by storing sentences as episodes, MINERVA 2 could explain lexical ambiguity resolution.

With respect to lexical representation, flexible episodic boundaries make a simple prediction: If words are usually stored as small pieces of larger sentences, any context-free retrieval will seem abstract, as Semon (1909/1923) predicted. Consider a common word, such as ride: Whether retrieved from the lexicon for production, or in response to an appearance on a computer screen, ride is a fairly generic character. The observer knows that ride can be a noun or a verb, that it rhymes with side, and so forth. However, in all likelihood, no particular voice-of- font-specific rides come to mind. Indeed, most words—even if they are represented epistemically—will be functionally abstract.

By contrast, a handful of words seem to be functionally episodic. Consider rosebud: Most people readily know that rosebud is a noun (and perhaps a spondee). However, they also know that rosebud was a sled and can probably imitate the famous utterance from Citizen Kane. Every culture has its share of popular catchphrases, but very few are composed of single words. Indeed, an informal survey at Arizona State University confirmed that examples of one-word, voice-specific “cultural earcons” are quite difficult to generate (in addition to rosebud, my volunteers provided stella and humbug). Notably, all of these examples are unique or LF words, which reflects their limited participation in discourse-sized episodes. This special set of words appears episodic, in both form and function.
Conclusion

Jacoby (1983a) noted that “there is a great deal of unexploited similarity between theories of episodic memory and theories of perception. . . . The difference is largely removed if it is assumed both types of task involve parallel access to a large population of memories for prior episodes” (pp. 35–36). Together with related findings, the present shadowing data suggest an episodic lexicon, with words perceived against a background of myriad, detailed episodes. Given episodes of sufficient complexity, and equivalent theoretical processes, researchers may account for behaviors beyond single-word laboratory tests.

References


Appendix A

MINERVA 2: The Formal Model and Simulations

This appendix summarizes the formal properties of MINERVA 2 and provides parameter values for the present simulations. This model description is an abbreviated version of the account provided by Hintzman (1986, pp. 413–414). As noted in the introduction, memory traces in MINERVA 2 are implemented as vectors, with units valued —1, 0, or +1. The model learns these traces by probabilistically storing each element of the vector, with likelihood of encoding given by parameter L. After learning, all nonzero elements may revert to zero, as determined by a forgetting parameter F. In the present simulations, these parameters were constant, with \( L = .90 \) and \( F = .15 \). (In the simulation of the Goldinger, 1996, data discussed in the introduction, these values were 1.00 and .25, respectively.) In “forgetting cycles,” each nonzero element is sampled and may change to zero, determined by a stochastic process in which probability \( F \) is used.

Once all traces are stored in LTM, model testing is accomplished by presenting a probe vector to WM. When this is done, each trace is activated to a degree commensurate with similarity to the probe. Assume that LTM contains \( m \) traces, each containing \( n \) vector elements, enumerated as \( g = 1 \cdots n \). Because position-specific similarity is the basis of activation, \( P(g) \) denotes probe element \( g \), and \( T(i,g) \) denotes the element at position \( g \) in trace \( i \). The similarity (S) of trace \( i \) to the probe is calculated as follows:

\[
S(i) = \frac{1}{N_x} \sum_{g=1}^{n} P(g)T(i,g),
\]

in which \( N_x \) is the number of nonzero elements in the trace. Similarity to the probe determines the degree of trace activation:

\[
A(i) = S(i)^3.
\]

As summarized in the introduction, echoes are composed by the collection of activated traces and have two primary characteristics. Echo intensity equals the summed activation levels of all traces:

\[
\text{Int} = \sum_{i=1}^{m} A(i).
\]

Finally, echo content is determined by summing the activation levels of all position-specific vector elements of all relevant traces:

\[
\text{Cont}(g) = \sum_{i=1}^{m} A(i)T(i,g).
\]

In the present research, all simulations were performed several times, to ensure that the random storage and forgetting functions did not create idiosyncratic results. Please note that although the model assumes parallel access to memory traces, all simulation processes are carried out in a serial manner.

Appendix B

Method: All Experiments

Participants

All three shadowing experiments (1A, 2A, and 3A) included different sets of 4 men and 4 women. All 24 participants were graduate students at Arizona State University and were native English speakers with normal (self-reported) hearing. In Experiment 1A, each participant received $20. In Experiments 2A and 3A, each participant received $40. The AXB classification experiments all included introductory psychology students. These students met the same inclusion criteria, and they received course credit for participation. Experiments 1B, 2B, 3B, and 3C included 80 participants each.

Stimulus Materials

Experiment 1A contained 160 English words that followed several basic constraints: Most important, 25% of the words fell into each of four frequency classes, defined as follows: High-frequency (HF) words were indexed >300 occurrences per million (Kucera & Francis, 1967), medium-high-frequency (MHF) words ranged from 150 to 250, medium-low-frequency (MLF) words ranged from 50 to 100, and low-frequency (LF) words were indexed <5. Half of the words in each frequency class were monosyllabic; half were bisyllabic. All frequency classes were balanced with respect to word-initial phonemes (equal proportions of stops, glides, etc.). All words and their frequencies are listed in Appendix D.

The words were recorded by 10 volunteers in a soundproof booth with an IBM computer, a Beyerdynamics microphone, and a Marantz DAT recorder. Words were shown on the computer; volunteers were asked to say each twice and to avoid lapsing into a monotone. The tapes were low-pass filtered at 4.8 kHz, digitized at 10 Khz (in a 16-bit analog-to-digital processor), and the subjectively clearer token of each word was stored in a digital file. Ten groups of 10 volunteers listened to the tokens; all were identified at or above 90%. The stimuli for Experiments 2A and 3A were 160 nonwords: half monosyllabic and half bisyllabic (see Appendix E). These were prepared in the manner described for the words.

Design and Procedure

Experiment 1A

Experiment 1A entailed four levels of word frequency, four levels of repetition, and two levels of delay—all manipulated within subject. To counterbalance all factors, I divided the words into 8 sets of 20 (5 words from each frequency bin), which were rotated across all conditions. Thus, across participants, all words were presented equally at each level of repetition and delay. Half of the participants performed immediate
shadowing first; half performed delayed shadowing first. In the baseline phase, all words were presented in random order. Participants were asked to speak each word quickly but clearly, pressing the space bar to continue. Instructions stressed speed and clarity equally, as in the later shadowing blocks. (It is imperative that volunteers experience comparable time pressure in the baseline and shadowing phases for the generation of a challenging AXB test. Faster naming responses are typically shorter and louder; Balota et al., 1989. Thus, AXB classification would be too easy if time pressure were only applied during shadowing.) Each participant wore Sennheiser HD-450 headphones with a built-in microphone; these were connected to the computer and DAT recorder, respectively. For each participant, baseline words were recorded in this initial block.

In the listening blocks, participants saw a matrix with a word in each cell. Depending on the block, 60, 40, or 20 words were shown. On each trial, a spoken word was presented at approximately 65 dB (sound pressure level); the participant had 5 s to click the word with the left mouse key. If the word was found in time, the next word played. If not, the word was highlighted in red for 250 ms, and the next word played. In blocks that repeated a word set several times, the response matrix was redrawn (with a new, random arrangement) after each iteration through the set. This “hear-and-find” procedure was used to maintain attention to the spoken words. (Correct identification rates were always greater than 80%. Participants reportedly always understood the words but could not always locate the box in time. Listening block data were not analyzed.)

In each trial of the shadowing blocks, participants saw a warning (***) for 500 ms, followed by presentation of a spoken word. Participants were instructed to repeat the word quickly and clearly, as in the baseline session. The headphone-mounted microphone relayed their speech to the DAT recorder; a standing microphone triggered a voice key, sending RTs to the computer. The delayed-shadowing blocks were identical, but each trial required the participant to wait for a tone before speaking. The tone occurred 3–4 s after the word, with any given delay determined randomly.

**Experiment 1B**

The recorded utterances from each participant in Experiment 1A were used to generate Experiment 1B (which actually consisted of eight sub-experiments—one per shadower—each administered to 10 AXB listeners). Each shadowing participant’s baseline and shadowing utterances were digitized and stored. Then, the stimulus token that the shadower heard was paired with these two utterances, as the X stimulus in the AXB design. Half of the trials presented the baseline token first; half presented it third. The participants judged which utterance, the first or third, was a “better imitation” of the second word.

The AXB participants made up groups of 5–8 students in a sound-attenuated room. All were seated in booths equipped with a computer, headphones, and mouse. Each trial began with a 500-ms warning (***), followed by two response boxes, labeled first and third. After 500 ms, three words were played, with a 750-ms silence between. The participant indicated whether A or B sounded more like X by clicking either box with the left mouse key. The experimental trials were preceded by 10 practice trials, generated with voices not used in the experiment.

**Experiment 2A**

Unlike Experiment 1A, Experiment 2A entailed training and test sessions, conducted on consecutive days. The training sessions were used to create a “nonword lexicon” for shadowing participants, using procedures similar to the listening blocks in Experiment 1A. Participants saw a matrix of 40 nonwords (which was rearranged after every 40 trials), listened to each nonword, and tried to click it within 5 s. The only factor manipulated in training was exposure frequency. Forty LF nonwords were presented once each, 40 MLF nonwords were presented twice each, 40 MHF nonwords were presented 7 times each, and 40 HF nonwords were presented 20 times each. This yielded 1,200 identification trials in the training session. However, to avoid familiarizing listeners with the exact tokens used in test sessions, I had all training tokens spoken by one novel speaker (whose voice was not used in later sessions). Across participants, all nonwords were equally assigned to each frequency class. Test sessions were completed on the second day, following the procedures of Experiment 1A.

**Experiment 2B**

All AXB procedures were identical to those of Experiment 1B.

**Experiment 3A**

Experiment 3A was mostly identical to Experiment 2A. However, in half of the shadowing trials, nonwords were presented in a voice that differed from all previous exposures. These DV trials always entailed changes from male to female voices, or vice-versa. The voices were chosen to maximize dissimilarity from training voices.

**Experiments 3B and 3C**

Experiment 3B was identical to Experiment 2B, presenting tokens recorded in Experiment 3A (baseline and shadowing), juxtaposed against shadowing stimulus tokens. In Experiment 3C, training tokens were used as X stimuli.

**Appendix C**

**Abbreviated Results: All Experiments**

**Shadowing RTs: Experiments 1A, 2A, and 3A**

The shadowing response times (RTs) were analyzed in analyses of variance (ANOVAs), always assuming a $p < .05$ significance criterion. Only the reliable main effects and interactions are listed here; other possible effects failed to surpass criterion.
AN EPISODIC LEXICON

Frequency: \( F(3, 21) = 71.7; \ MSE = 97.7 \)
Repetition: \( F(3, 21) = 229.7; \ MSE = 52.0 \)
Frequency × Repetition: \( F(9, 63) = 189.2; \ MSE = 52.0 \)
Delay: \( F(1, 7) = 9.3; \ MSE = 144.0 \)
Frequency × Delay: \( F(3, 21) = 33.7; \ MSE = 49.2 \)

As Figure 3 shows, these results reflect the predicted directions of effect: Shadowing RTs decreased when words were higher in frequency, or when they amassed repetitions. Frequency and repetition also produced their common interaction (Scarborough et al., 1977). The delay effect reflected generally faster responses in delayed shadowing (cf. Balota & Chumbley, 1985), and Frequency × Delay reflected the smaller frequency effects in delayed shadowing.

Experiment 2A

Across 8 shadowing participants, 24 errors were recorded. These trials were not analyzed or used in Experiment 2B. The immediate shadowing KIs are shown in Figure 5; the delayed shadowing RE are shown in Table C1.

The RTs were analyzed in a 4 X 4 X 2 ANOVA, in which frequency, repetition, and delay were examined. All RT data were taken together, and the effects listed below were reliable. The patterns (i.e., directions of effect) were identical to those just summarized in Experiment 1A.

<table>
<thead>
<tr>
<th>No. of repetitions</th>
<th>HF</th>
<th>MHF</th>
<th>MLF</th>
<th>LF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate shadowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 SV</td>
<td>649</td>
<td>655</td>
<td>679</td>
<td>710</td>
</tr>
<tr>
<td>2 DV</td>
<td>667</td>
<td>680</td>
<td>698</td>
<td>721</td>
</tr>
<tr>
<td>6 SV</td>
<td>646</td>
<td>653</td>
<td>659</td>
<td>668</td>
</tr>
<tr>
<td>12 DV</td>
<td>644</td>
<td>669</td>
<td>677</td>
<td>680</td>
</tr>
<tr>
<td>Delayed shadowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 SV</td>
<td>635</td>
<td>637</td>
<td>646</td>
<td>650</td>
</tr>
<tr>
<td>2 DV</td>
<td>650</td>
<td>647</td>
<td>661</td>
<td>669</td>
</tr>
</tbody>
</table>

Frequency: \( F(3, 21) = 27.1; \ MSE = 199.8 \)
Repetition: \( F(3, 21) = 59.2; \ MSE = 151.0 \)
Frequency × Repetition: \( F(9, 63) = 221.5 \)
Delay: \( F(1, 7) = 30.7; \ MSE = 239.2 \)
Frequency × Delay: \( F(3, 21) = 23.2; \ MSE = 191.6 \)

Experiment 3A

Across all shadowing participants, 31 recorded errors were excluded from the RT analyses and AXB experiments. The mean correct KIs in all conditions are shown in Table C2.

These RTs were analyzed in a 4X4X2X2 ANOVA, in which frequency, repetition, delay, and voice (same vs. different) were examined. The following effects were observed:

<table>
<thead>
<tr>
<th>Nonword frequency class</th>
<th>No. of repetitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate shadowing</td>
<td></td>
</tr>
<tr>
<td>0 SV</td>
<td>HF 649, MUF 655, MLF 679, LF 710</td>
</tr>
<tr>
<td>2 DV</td>
<td>HF 667, MUF 680, MLF 698, LF 721</td>
</tr>
<tr>
<td>6 SV</td>
<td>HF 646, MUF 653, MLF 659, LF 668</td>
</tr>
<tr>
<td>12 DV</td>
<td>HF 644, MUF 669, MLF 677, LF 680</td>
</tr>
<tr>
<td>Delayed shadowing</td>
<td></td>
</tr>
<tr>
<td>0 SV</td>
<td>HF 635, MUF 637, MLF 646, LF 650</td>
</tr>
<tr>
<td>2 DV</td>
<td>HF 650, MUF 647, MLF 661, LF 669</td>
</tr>
</tbody>
</table>

Frequency: \( F(3, 21) = 9.7; \ MSE = 212.0 \)
Repetition: \( F(3, 21) = 24.1; \ MSE = 211.8 \)
Delay: \( F(1, 7) = 111.0; \ MSE = 217.5 \)
Frequency × Delay: \( F(3, 21) = 3.0; \ MSE = 205.1, p < .06 \)
Frequency × Repetition × Delay: \( F(9, 63) = 18.40; \ MSE = 211.1 \)
Voice × Delay: \( F(1, 23) = 41.8; \ MSE = 210.1 \)

The effects of frequency, repetition, and delay (and their interactions) all reflected patterns similar to those in prior experiments. Although the main effect of voice was null, a Voice × Delay interaction was observed—a voice effect emerged in immediate shadowing, but not in delayed shadowing.

Imitation (AXB) Judgments:
Experiments 1B, 2B, 3B, and 3C

The mean percentage of "correct" AXB classifications (i.e., selections of shadowing tokens as imitations, rather than baseline tokens) was determined for all cells of each experimental design. Higher hit rates in AXB classification indicated more discernible imitation by the
shadowing participants. In each experiment, the hit rates were analyzed by ANOVAs and planned tests, and each cell mean was compared to a chance level of 50%.

Experiment 1B

The AXB classification data were shown in Figure 4 in the text. In immediate shadowing, most cell means surpassed chance (cutoff value = 64%); in delayed shadowing, few cell means exceeded chance (cutoff value = 63%). These data were analyzed in a 4 X 4 X 2 ANOVA, in which frequency, repetition, and delay were examined. The following effects were reliable:

- Frequency: \(F(3, 237) = 29.1; \text{MSE} = 8.2\)
- Repetition: \(F(3, 237) = 25.0; \text{MSE} = 9.2\)
- Frequency X Repetition: \(F(9, 711) = 14.0; \text{MSE} = 11.7\)
- Delay: \(F(1, 79) = 40.2; \text{MSE} = 8.6\)
- Frequency X Delay: \(F(3, 237) = 51.0; \text{MSE} = 12.8\)
- Repetition X Delay: \(F(3, 237) = 30.1; \text{MSE} = 13.3\)

As Figure 4 shows, listeners were more likely to detect imitations when the words were lower in frequency, or when they amassed repetitions. However, imitation was far stronger in immediate shadowing than in delayed shadowing. Indeed, all effects were attenuated in delayed shadowing.

Experiment 2B

The AXB classification data for immediate and delayed shadowing are shown at the top of Figures 6 and 7, respectively. All cell means exceeded chance in immediate shadowing (cutoff value = 63%), but few surpassed chance in delayed shadowing (cutoff value = 62%). As in Experiment 1B, robust frequency and repetition effects were observed in immediate shadowing. These effects were observed, but attenuated, in delayed shadowing. A 4 X 4 X 2 ANOVA verified the following effects:

- Frequency: \(F(3, 237) = 16.0; \text{MSE} = 5.7\)
- Repetition: \(F(3, 237) = 33.1; \text{MSE} = 5.1\)
- Delay: \(F(1, 79) = 85.8; \text{MSE} = 6.8\)
- Frequency X Delay: \(F(3, 237) = 2.6; \text{MSE} = 6.2, p < .065\)
- Repetition X Delay: \(F(3, 237) = 21.3; \text{MSE} = 7.0\)

Experiment 3B

To provide a clear account of the results, the AXB classification data from immediate and delayed shadowing were analyzed in separate 4 X

<table>
<thead>
<tr>
<th>Table C4</th>
<th>Percentage of Correct AXB Classifications in Delayed Shadowing, Experiment 3B</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of repetitions</td>
<td>Nonword frequency class</td>
</tr>
<tr>
<td></td>
<td>HF</td>
</tr>
<tr>
<td>0</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
<tr>
<td>2</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
<tr>
<td>6</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
<tr>
<td>12</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
</tbody>
</table>

Note. HF = high frequency; MHF = medium high frequency; MLF = medium low frequency; LF = low frequency; SV = same voice; DV = different voice.

<table>
<thead>
<tr>
<th>Table C5</th>
<th>Percentage of Correct AXB Classifications in Immediate Shadowing, Experiment 3C</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of repetitions</td>
<td>Nonword frequency class</td>
</tr>
<tr>
<td></td>
<td>HF</td>
</tr>
<tr>
<td>0</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
<tr>
<td>2</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
<tr>
<td>6</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
<tr>
<td>12</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td>DV</td>
</tr>
</tbody>
</table>

Note. HF = high frequency; MHF = medium high frequency; MLF = medium low frequency; LF = low frequency; SV = same voice; DV = different voice.
4 × 2 ANOVAs, in which frequency, repetitions, and delay (dropping the delay factor) were examined. In immediate shadowing, most SV means surpassed chance; few DV means exceeded chance (cutoff value = 64%). The percentage of correct AXB classifications in immediate shadowing are shown in Table C3.

The ANOVA conducted on these data revealed several effects: The frequency effect was null, but voice, $F(1, 79) = 80.20$, $MSE = 5.6$, and repetition, $F(1, 79) = 101.20$, $MSE = 4.90$, were robust. SV tokens generated stronger imitation, and all imitation increased across repetitions. A Voice × Frequency interaction, $F(1, 79) = 37.05$, $MSE = 6.82$, reflected the increased voice effect at higher frequencies.

In delayed shadowing, most SV (but few DV) means surpassed chance (cutoff value = 62%). The frequency effect was unreliable, but voice, $F(1, 79) = 49.00$, $MSE = 8.00$, and repetition, $F(1, 79) = 11.80$, $MSE = 9.10$, effects were observed. A Voice × Frequency interaction, $F(1, 79) = 5.10$, $MSE = 9.00$, reflected a larger voice effect at higher frequencies. The percentage of correct AXB classifications in delayed shadowing are shown in Table C4.

### Experiment 3C

The AXB data were analyzed as described for Experiment 3B. However, the general data pattern differed markedly from Experiment 3B. In immediate shadowing, 16 SV and 5 DV means reliably surpassed chance (cutoff value = 62%). The percentage of correct AXB classifications in immediate shadowing are shown in Table C5.

In immediate shadowing, a frequency effect was observed, $F(1, 79) = 73.40$, $MSE = 7.10$, but it was reversed, relative to prior experiments — higher frequency nonwords were more easily identified as imitations. This was true for both SV and DV words (null Frequency × Voice interaction), but a voice effect, $F(1, 79) = 39.10$, $MSE = 8.70$, reflected a persistent SV advantage. Although a repetition effect, $F(1, 79) = 18.10$, $MSE = 4.60$, was observed, repetition did not interact with voice.

#### Table C6

<table>
<thead>
<tr>
<th>No. of repetitions</th>
<th>HF</th>
<th>M HF</th>
<th>MLF</th>
<th>LF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>SV 65.1</td>
<td>DV 62.0</td>
<td>SV 61.8</td>
<td>DV 60.6</td>
</tr>
<tr>
<td>2</td>
<td>SV 61.8</td>
<td>DV 60.6</td>
<td>SV 61.6</td>
<td>DV 60.6</td>
</tr>
<tr>
<td>6</td>
<td>SV 70.2</td>
<td>DV 65.3</td>
<td>SV 66.7</td>
<td>DV 66.5</td>
</tr>
<tr>
<td>12</td>
<td>SV 69.8</td>
<td>DV 72.3</td>
<td>SV 70.5</td>
<td>DV 69.1</td>
</tr>
</tbody>
</table>

Note. HF = high frequency; M HF = medium high frequency; MLF = medium low frequency; LF = low frequency; SV = same voice; DV = different voice.

The percentage of correct AXB classifications for delayed-shadowing tokens are shown in Table C6.

In delayed shadowing, 10 SV and 6 DV means reliably surpassed chance (cutoff value = 63%). As in immediate shadowing, a "backward" frequency effect was observed, $F(1, 79) = 24.0$, $MSE = 8.20$, with higher frequency nonwords more easily identified as imitations. However, no voice effect (or interaction) was observed. Given a shadowing delay, all responses apparently sounded like training tokens. A repetition effect, $F(1, 79) = 20.90$, $MSE = 6.10$, was observed, but repetition did not interact with voice.

(Appendixes continue)
Appendix D

Stimulus Words (and Frequencies) Used in Experiment 1

<table>
<thead>
<tr>
<th>Bisyllabic</th>
<th>Frequency</th>
<th>Monosyllabic</th>
<th>Frequency</th>
<th>Bisyllabic</th>
<th>Frequency</th>
<th>Monosyllabic</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-frequency words (&gt;300)</td>
<td></td>
<td></td>
<td></td>
<td>Medium-low-frequency words (50–100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>442</td>
<td>school</td>
<td>492</td>
<td>symbol</td>
<td>54</td>
<td>rule</td>
<td>73</td>
</tr>
<tr>
<td>better</td>
<td>414</td>
<td>light</td>
<td>333</td>
<td>dozen</td>
<td>52</td>
<td>moon</td>
<td>60</td>
</tr>
<tr>
<td>system</td>
<td>416</td>
<td>church</td>
<td>348</td>
<td>handle</td>
<td>53</td>
<td>safe</td>
<td>58</td>
</tr>
<tr>
<td>second</td>
<td>373</td>
<td>group</td>
<td>390</td>
<td>cousin</td>
<td>51</td>
<td>bank</td>
<td>83</td>
</tr>
<tr>
<td>never</td>
<td>698</td>
<td>next</td>
<td>394</td>
<td>active</td>
<td>88</td>
<td>band</td>
<td>53</td>
</tr>
<tr>
<td>before</td>
<td>1,016</td>
<td>give</td>
<td>391</td>
<td>permit</td>
<td>77</td>
<td>crowd</td>
<td>53</td>
</tr>
<tr>
<td>social</td>
<td>380</td>
<td>white</td>
<td>365</td>
<td>career</td>
<td>67</td>
<td>phone</td>
<td>54</td>
</tr>
<tr>
<td>number</td>
<td>472</td>
<td>part</td>
<td>500</td>
<td>careful</td>
<td>62</td>
<td>chair</td>
<td>66</td>
</tr>
<tr>
<td>become</td>
<td>361</td>
<td>house</td>
<td>591</td>
<td>captain</td>
<td>85</td>
<td>tree</td>
<td>59</td>
</tr>
<tr>
<td>public</td>
<td>438</td>
<td>case</td>
<td>362</td>
<td>balance</td>
<td>90</td>
<td>bright</td>
<td>87</td>
</tr>
<tr>
<td>program</td>
<td>394</td>
<td>point</td>
<td>395</td>
<td>title</td>
<td>77</td>
<td>prove</td>
<td>53</td>
</tr>
<tr>
<td>country</td>
<td>324</td>
<td>side</td>
<td>380</td>
<td>forget</td>
<td>54</td>
<td>grass</td>
<td>53</td>
</tr>
<tr>
<td>matter</td>
<td>308</td>
<td>great</td>
<td>665</td>
<td>coffee</td>
<td>78</td>
<td>dust</td>
<td>70</td>
</tr>
<tr>
<td>between</td>
<td>730</td>
<td>work</td>
<td>760</td>
<td>novel</td>
<td>59</td>
<td>fresh</td>
<td>82</td>
</tr>
<tr>
<td>order</td>
<td>376</td>
<td>back</td>
<td>967</td>
<td>fashion</td>
<td>69</td>
<td>watch</td>
<td>81</td>
</tr>
<tr>
<td>power</td>
<td>342</td>
<td>state</td>
<td>808</td>
<td>favor</td>
<td>78</td>
<td>knife</td>
<td>76</td>
</tr>
<tr>
<td>city</td>
<td>393</td>
<td>last</td>
<td>676</td>
<td>garden</td>
<td>60</td>
<td>tone</td>
<td>78</td>
</tr>
<tr>
<td>later</td>
<td>397</td>
<td>door</td>
<td>312</td>
<td>listen</td>
<td>51</td>
<td>throat</td>
<td>51</td>
</tr>
<tr>
<td>people</td>
<td>847</td>
<td>place</td>
<td>569</td>
<td>master</td>
<td>72</td>
<td>speed</td>
<td>83</td>
</tr>
<tr>
<td>rather</td>
<td>373</td>
<td>young</td>
<td>385</td>
<td>vision</td>
<td>56</td>
<td>lake</td>
<td>54</td>
</tr>
<tr>
<td>Medium-high-frequency words (150–250)</td>
<td></td>
<td></td>
<td></td>
<td>Low-frequency words (&lt;5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>river</td>
<td>165</td>
<td>stage</td>
<td>174</td>
<td>bicep</td>
<td>1</td>
<td>germ</td>
<td>3</td>
</tr>
<tr>
<td>market</td>
<td>155</td>
<td>class</td>
<td>207</td>
<td>rustic</td>
<td>3</td>
<td>vest</td>
<td>4</td>
</tr>
<tr>
<td>police</td>
<td>155</td>
<td>sound</td>
<td>204</td>
<td>nectar</td>
<td>3</td>
<td>dire</td>
<td>1</td>
</tr>
<tr>
<td>figure</td>
<td>209</td>
<td>black</td>
<td>203</td>
<td>parcel</td>
<td>1</td>
<td>malt</td>
<td>1</td>
</tr>
<tr>
<td>beyond</td>
<td>175</td>
<td>floor</td>
<td>158</td>
<td>mingle</td>
<td>2</td>
<td>wilt</td>
<td>3</td>
</tr>
<tr>
<td>nature</td>
<td>191</td>
<td>book</td>
<td>193</td>
<td>staple</td>
<td>1</td>
<td>grub</td>
<td>2</td>
</tr>
<tr>
<td>father</td>
<td>183</td>
<td>cold</td>
<td>171</td>
<td>gusto</td>
<td>2</td>
<td>soot</td>
<td>1</td>
</tr>
<tr>
<td>spirit</td>
<td>182</td>
<td>town</td>
<td>212</td>
<td>forage</td>
<td>3</td>
<td>blur</td>
<td>3</td>
</tr>
<tr>
<td>music</td>
<td>216</td>
<td>ground</td>
<td>186</td>
<td>deport</td>
<td>1</td>
<td>crow</td>
<td>2</td>
</tr>
<tr>
<td>recent</td>
<td>179</td>
<td>north</td>
<td>206</td>
<td>pigeon</td>
<td>3</td>
<td>vine</td>
<td>4</td>
</tr>
<tr>
<td>table</td>
<td>198</td>
<td>girl</td>
<td>220</td>
<td>venom</td>
<td>2</td>
<td>mule</td>
<td>4</td>
</tr>
<tr>
<td>party</td>
<td>216</td>
<td>late</td>
<td>179</td>
<td>nugget</td>
<td>1</td>
<td>chunk</td>
<td>2</td>
</tr>
<tr>
<td>report</td>
<td>174</td>
<td>wall</td>
<td>160</td>
<td>garter</td>
<td>2</td>
<td>weed</td>
<td>1</td>
</tr>
<tr>
<td>picture</td>
<td>161</td>
<td>fire</td>
<td>187</td>
<td>portal</td>
<td>3</td>
<td>hoop</td>
<td>3</td>
</tr>
<tr>
<td>basis</td>
<td>184</td>
<td>bring</td>
<td>158</td>
<td>beacon</td>
<td>5</td>
<td>kelp</td>
<td>2</td>
</tr>
<tr>
<td>person</td>
<td>175</td>
<td>rest</td>
<td>163</td>
<td>patron</td>
<td>4</td>
<td>knack</td>
<td>4</td>
</tr>
<tr>
<td>value</td>
<td>200</td>
<td>lost</td>
<td>171</td>
<td>jelly</td>
<td>3</td>
<td>leash</td>
<td>3</td>
</tr>
<tr>
<td>common</td>
<td>223</td>
<td>care</td>
<td>162</td>
<td>cavern</td>
<td>1</td>
<td>fade</td>
<td>2</td>
</tr>
<tr>
<td>final</td>
<td>156</td>
<td>plan</td>
<td>205</td>
<td>hazel</td>
<td>2</td>
<td>stale</td>
<td>4</td>
</tr>
<tr>
<td>single</td>
<td>172</td>
<td>hard</td>
<td>202</td>
<td>wedlock</td>
<td>2</td>
<td>raft</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. Word frequencies are from Kučera and Francis (1967).
Appendix E

Nonwords Used in Experiments 2 and 3

<table>
<thead>
<tr>
<th>Bisyllables</th>
<th>Monosyllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>provate</td>
<td>welge</td>
</tr>
<tr>
<td>batoon</td>
<td>meach</td>
</tr>
<tr>
<td>vasult</td>
<td>cade</td>
</tr>
<tr>
<td>lactain</td>
<td>freem</td>
</tr>
<tr>
<td>daver</td>
<td>skave</td>
</tr>
<tr>
<td>meegon</td>
<td>nork</td>
</tr>
<tr>
<td>behick</td>
<td>bren</td>
</tr>
<tr>
<td>luding</td>
<td>serp</td>
</tr>
<tr>
<td>lexel</td>
<td>felp</td>
</tr>
<tr>
<td>redent</td>
<td>neep</td>
</tr>
<tr>
<td>erbow</td>
<td>rean</td>
</tr>
<tr>
<td>sagad</td>
<td>gink</td>
</tr>
<tr>
<td>elent</td>
<td>push</td>
</tr>
<tr>
<td>jandy</td>
<td>shoss</td>
</tr>
<tr>
<td>puxii</td>
<td>wurve</td>
</tr>
<tr>
<td>wanic</td>
<td>seck</td>
</tr>
<tr>
<td>ganet</td>
<td>tink</td>
</tr>
<tr>
<td>gisto</td>
<td>tupe</td>
</tr>
<tr>
<td>ensip</td>
<td>tunch</td>
</tr>
<tr>
<td></td>
<td>vint</td>
</tr>
<tr>
<td></td>
<td>lugre</td>
</tr>
<tr>
<td></td>
<td>reast</td>
</tr>
<tr>
<td></td>
<td>wug</td>
</tr>
<tr>
<td></td>
<td>zamp</td>
</tr>
<tr>
<td></td>
<td>zeeze</td>
</tr>
<tr>
<td></td>
<td>breeze</td>
</tr>
<tr>
<td></td>
<td>borse</td>
</tr>
<tr>
<td></td>
<td>bork</td>
</tr>
<tr>
<td></td>
<td>searl</td>
</tr>
<tr>
<td></td>
<td>mazz</td>
</tr>
<tr>
<td></td>
<td>zant</td>
</tr>
<tr>
<td></td>
<td>glesh</td>
</tr>
<tr>
<td></td>
<td>floak</td>
</tr>
<tr>
<td></td>
<td>loash</td>
</tr>
<tr>
<td></td>
<td>geel</td>
</tr>
<tr>
<td></td>
<td>pitch</td>
</tr>
<tr>
<td></td>
<td>chark</td>
</tr>
<tr>
<td></td>
<td>hine</td>
</tr>
<tr>
<td></td>
<td>glane</td>
</tr>
<tr>
<td></td>
<td>lisk</td>
</tr>
<tr>
<td></td>
<td>stamp</td>
</tr>
<tr>
<td></td>
<td>yamp</td>
</tr>
<tr>
<td></td>
<td>gurk</td>
</tr>
<tr>
<td></td>
<td>venr</td>
</tr>
<tr>
<td></td>
<td>gliss</td>
</tr>
<tr>
<td></td>
<td>prock</td>
</tr>
<tr>
<td></td>
<td>shak</td>
</tr>
<tr>
<td></td>
<td>dorve</td>
</tr>
<tr>
<td></td>
<td>horch</td>
</tr>
<tr>
<td></td>
<td>goip</td>
</tr>
<tr>
<td></td>
<td>shret</td>
</tr>
<tr>
<td></td>
<td>natch</td>
</tr>
<tr>
<td></td>
<td>clud</td>
</tr>
<tr>
<td></td>
<td>yole</td>
</tr>
<tr>
<td></td>
<td>croff</td>
</tr>
<tr>
<td></td>
<td>deese</td>
</tr>
<tr>
<td></td>
<td>plew</td>
</tr>
<tr>
<td></td>
<td>noast</td>
</tr>
<tr>
<td></td>
<td>murch</td>
</tr>
<tr>
<td></td>
<td>modge</td>
</tr>
<tr>
<td></td>
<td>fauze</td>
</tr>
<tr>
<td></td>
<td>trool</td>
</tr>
<tr>
<td></td>
<td>noil</td>
</tr>
<tr>
<td></td>
<td>rand</td>
</tr>
</tbody>
</table>

Received August 16, 1996
Revision received July 15, 1997
Accepted July 23, 1997

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the American Psychologist and APA Monitor. High School Teacher and International Affiliates receive subscriptions to the APA Monitor, and they may subscribe to the American Psychologist at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the Publication Manual of the American Psychological Association, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.