Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Phonetics 31 (2003) 63–79
Journal of Phonetics
www.elsevier.com/locate/phonetics

# Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English

Patrick W. Nye[a,*], Carol A. Fowler[a,b,c,d]

[a] *Haskins Laboratories, 270 Crown Street, New Haven, CT 06510, USA*
[b] *Department of Psychology, University of Connecticut, Storrs, CT 06269, USA*
[c] *Department of Psychology, Yale University, Box 208205, New Haven, CT 06520, USA*
[d] *Department of Linguistics, Yale University, Box 208205, New Haven, CT 06520, USA*

Received 3 December 2001; received in revised form 16 July 2002; accepted 17 September 2002

**Abstract**

Speakers imitate the speech they shadow. However, speech is not wholly imitative; speakers use their own speech habits or language knowledge in shadowing as well. We examined the interplay between the effects of input variables and of knowledge of the language on shadowing. We asked speakers to shadow utterances composed of phonetic sequences that varied in their order of approximation to English. Shadowing latency and errors reduced as order of approximation increased. This is consistent with the inference that knowledge of the language (e.g., speech habits, lexical or phonotactic knowledge) guides shadowing. To assess the interplay between this knowledge and the effect of imitation of the input on shadowing, we asked whether imitative fidelity varied with order of approximation. We used an AXB test in which X was a shadowed utterance, A (or B) was a shadowed response to X and B (or A) was a read version of the same utterance produced by the speaker of A (or B). Listeners were asked which of A or B was a better imitation of X. Generally, they chose the shadowed utterance; however, they did so significantly more frequently when the utterance was a 1st than a 12th order of approximation to English.


## 1. Introduction

Studies designed to probe the psychological mechanisms engaged in speech and language processing have frequently employed the task of verbal shadowing and have also used statistically manipulated stimulus materials. Papers on verbal shadowing, published in the 1960s by Chistovich and colleagues (Chistovich, 1960; Chistovich, Fant, de Serpa-Leitao, & Tjernlund,

*Corresponding author. Tel.: +1-203-865-6163; fax: +1-203-865-8963.
*E-mail address:* nye@haskins.yale.edu (P.W. Nye).

0095-4470/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
PII: S0095-4470(02)00072-4

1966a; Chistovich, Fant, & de Serpa-Leitao, 1966b; Chistovich & Kozhevnikov, 1969) did much to draw the attention of other investigators (Rosenberg & Lambert, 1974; Marslen-Wilson, 1975, 1985; Kent, 1973, 1979; Porter & Lubker, 1980; Goldinger, 1998) to the insights that can be gained from studies of the shadowing of either synthetic or natural utterances. Similarly, with respect to the choice of stimulus materials, concepts in communication theory (Shannon & Weaver, 1949) and/or statistics influenced many psychologists (Miller & Selfridge, 1950; Taylor & Moray, 1960; Lawson, 1961; Davis, Moray, & Triesman, 1961; Salzinger, Portnoy, & Feldman, 1962; Triesman, 1965a; Pitt & McQueen, 1998; Vitevitch & Luce, 1999) in the development of stimuli for studies of human attention (Moray & Taylor, 1958; Underwood, 1974), lexical access and verbal memory.

This paper explores the issue of how familiarity with the *phonotactic* patterns of English influences shadowing behavior. Our method of exploration draws on the concept of *probabilistic phonotactics*. Defined by Trask (1996), the term describes the statistical frequencies with which syllable segments and sequences of syllable segments appear in syllables and words. Thus, we examine whether people who are denied the opportunity to draw on their knowledge of the probabilistic phonotactics of a spoken passage, produce shadowed speech that is more strongly influenced by the phonetic details of the passage than would otherwise be the case. Much of the evidence for this hypothesized difference in behavior comes from Goldinger (1998), who recently showed that people engaged in a shadowing task have a tendency to imitate what they hear. Moreover, the lower the frequency of the utterances being shadowed the greater is the tendency to imitate. Extrapolating from this observation, we predict that, the more language knowledge a shadower can use, the less imitative his utterances are likely to be. In essence, the more shadowers depend on memory for an utterance, a lexical representation for example, to shadow spoken input, the less they may imitate the input itself.

We make a distinction between *imitation* and *shadowing*.[1] In a typical shadowing task, participants are asked to repeat the utterances they hear as quickly as possible. This instruction will usually result in articulations whose phonetic and prosodic structure is influenced by both a shadower's regional dialect and his or her personal vocal mannerisms. By imitation, on the other hand, we refer to the features of a shadower's utterances that offer evidence of a tendency, whether conscious or not, to suppress personal speech habits and more precisely reproduce phonetic and/or non-phonetic aspects of the target utterance. In Experiment 1 we elicit shadowed speech under conditions that allow differential use of knowledge about the statistics of spoken English. We then look for evidence of the use of that knowledge in error rates and latency measurements. In Experiment 2, we look for an increase in the tendency of shadowers to imitate the target as the ability to use language knowledge is curtailed.

### *1.1. Previous studies*

#### *1.1.1. Sensitivity to linguistic knowledge in memory during language processing*

Our research is designed to bring together two findings in the literature. One is the finding that, by a variety of measures, listener/speakers are sensitive to the redundancies of their language, and

[1] We acknowledge that the word shadowing can be understood to encompass a gamut of articulatory behavior from precise mimicry to utterances that have the very crudest resemblance to a target. However, we wish to treat as evidence of imitative behavior all those features of a utterance that make it resemble the target and at the same time differ from the way in which a speaker would normally pronounce that utterance in the absence of external acoustic influences such as when reading aloud from a phonetic text.

they use the redundancies to guide phonetic speech processing in speech perception as well as in production. Their ability to use the redundancies, particularly in on-line speech tasks suggests rapid access to learned knowledge of language. The second finding is that speakers imitate one another; however, their doing so is mediated by their access to their knowledge of language.

Over the decades, language users' access to, and use of, stored knowledge of language has been demonstrated in a variety of ways and by examining a variety of kinds of language knowledge. We review just a few illustrative studies.

Some early research investigated language users' sensitivity to redundancies in prose by varying the order of approximation to English of word sequences. For example, Salzinger, Portnoy, and Feldman (1962) constructed 50-word printed passages that represented zero and 1st through 7th orders of statistical approximation to English word order in addition to plain text.[2] From each passage, they deleted words at regular intervals (e.g., every fifth word) and replaced it by underlined blank spaces (cf. Taylor, 1953, 1956). Subjects guessed the identities of the missing words. Salzinger et al. found that the subjects correctly guessed a greater proportion of the missing words from the higher than from the lower order passages.

These early studies (see also Triesman, 1965a, b) demonstrate that language users are sensitive to redundancies in prose. However, they do not show that language users can access the information quickly enough for it to have an impact on "on-line" language processing. To examine earliness of access to language knowledge, Marslen-Wilson (1973, 1975, 1985) developed a shadowing task in which participants heard word sequences or isolated words and repeated them as quickly as they could. Grammatical regularities, the semantic coherence of an utterance, and real words as contrasted with nonsense words constitute redundancies that, if shadowers have rapid access to them, can facilitate the shadowing task. Marslen-Wilson found that both "close" shadowers (with latencies to shadow normal prose that averaged approximately 250–300 ms) and "distant" shadowers had shorter latencies to shadow normal prose than prose that was either semantically uninterpretable, or semantically and syntactically deviant. Close shadowers showed effects of these variables despite the fact that, when they shadowed isolated words and nonwords, they showed weak or absent effects on latencies and errors of variations in lexical status, word length and word frequency. These latter results suggest that close shadowers begin shadowing a word before they know what it is. Even at these short latencies, in connected speech, they take advantage of syntactic and semantic redundancy in normal prose, suggesting that access to this higher order language information occurs quite rapidly.

Recent research using shadowing has suggested that listeners are continuously sensitive not only to such higher order linguistic variables as grammatical and semantic coherence, and the lexical status of a sequence of consonants and vowels, but also to variations in the frequency or probability with which phoneme sequences occur in words of the language. Vitevitch and Luce (1998, 1999) collected shadowing (naming) latency data from participants who heard isolated words and nonwords. The nonwords consisted either of phone sequences with high positional

[2] A 1st order approximation to English word order contains a random sequence of words that appear with the same frequency as they are found in ordinary English prose. Higher orders (2nd, 3rd, 4th and so on), are assembled from random sequences of word-pairs, triples and four-word sequences that appear in English. In each passage the 2-, 3- or 4-word sequences appear with the same frequencies that they appear in intelligible English.

frequency (e.g., /sɛp/) where /s/ occurs frequently in initial position, /ɛ/ in medial position and /p/ in final position; or sequences with low positional frequencies (e.g., ðiʃ/). They found shorter latencies to high than to low positional frequency nonwords.

*1.1.2. Imitation of speech*

Infants have been found to imitate speech (vowels) as young as 12 weeks of age (Kuhl & Meltzoff, 1996). This imitative tendency, which appears in facial gesturing as well (e.g., Meltzoff & Moore, 1997), might be interpreted as a means by which young humans learn to become competent participants in their cultural community. They imitate already-competent members.

Whether or not that is so, the imitative tendency persists into adulthood in speech (Goldinger, 1998; Giles, Coupland, & Coupland, 1991; Sancier & Fowler, 1997), in facial gesturing (McHugo, Lanzetta, Sullivan, Masters, & Englis, 1985), and perhaps in other domains as well (e.g., Shockley, Santana, & Fowler, in press). Sancier and Fowler (1997) reported, for example, that a bilingual speaker of Brazilian Portuguese and English had shorter voice onset times (VOTs) in both her Portuguese and her English voiceless stops (more Portugese-like VOTs) when she had just returned to the US after a several month stay in Brazil—where she heard only Portuguese—than when she had been in the US for several months—where she heard only English. The finding was striking not only in confirming a tendency to imitate speech, but also its very small (but statistically highly reliable) magnitude of effect. This suggested that the speaker's utterances reflected two influences: first a tendency to imitate the ambient speech, and second a tendency to speak in a way consistent with her history of speaking and hearing speech. Following Goldinger (1998), we propose to look at the interaction of these variables.

Goldinger (1998) elicited isolated word utterances from speakers under two conditions. In one condition, they read isolated words presented visually on a computer screen under instructions to read the words quickly but clearly. Next they heard words spoken by another talker and shadowed them under instructions to do so quickly but clearly. Goldinger asked whether the latter productions were imitations of the shadowed words. To assess that, he used an AXB discrimination procedure. X was always a word that talkers had shadowed. A (B) was the shadowing response to that word by one of the talker subjects; B (A) was the same word type as A(B) and the same talker, but it was produced as a naming response to print on the computer screen. Listeners were asked which of A or B sounded more like X. They were able to do the task and chose the shadowing response rather than the read word as the better imitation of X a greater than chance proportion of the time. However, the departure from chance was greater for low than high frequency words. This suggests that shadowers accessed their memories for the words they shadowed, and that the more influential the memories the weaker the imitative fidelity of the shadowing response. Goldinger invoked a theory of lexical memories as token memories (in which each time a word is experienced a distinct memory trace for the word is stored) to explain the word frequency effect.

Our aim is to extend these findings by asking whether lower-level knowledge of the language, similar in level to Vitevich and Luce's variations in the positional frequencies of different phone sequences in English, likewise is accessed when words are perceived. Specifically, we ask whether imitative fidelity of shadowing responses, as indexed by Goldinger's AXB task, is greater for lower- than higher-frequency sequences.

## 2. Experiment 1

In Experiment 1, we elicit shadowed speech under conditions that allow differential use of knowledge about the statistics of spoken English. We then look for evidence of the effect of that knowledge on error rates and latencies of shadowing productions. In Experiment 2, we look for an increase in the tendency of shadowers to imitate the target as the ability to use language knowledge is curtailed.

### *2.1. Method*

#### *2.1.1. Stimulus materials*

Construction of the spoken target stimuli was undertaken in four steps.

Step 1 involved the assembly of a 50,000-word database (a total of 289,140 symbols including spaces, commas, periods and stress marks). Components of the database were 2342 alphabetically transcribed sentences from the DARPA TIMIT CD-ROM (1990) and the transcripts of several broadcast discussions of news events obtained from the Internet. Only one copy of each of the repeated sentences was selected at random from the ARPA disk. This database ensured that all phones occur in English.

In step 2, our database was translated into a broad phonetic format by a speech synthesizer (DecTalk$^{TM}$). Upon receiving the alphabetic input, this synthesizer generated (in addition to speech) the corresponding phonetic transcription of the database using *Arpabet* phonetic symbols. This string of phonetic symbols, which included periods and commas, was subsequently captured by a computer and became the phonetic database stored in a hard disk file.

Then, in step 3, the method described by Hultzén, Allen, and Miron (1964) was employed to generate, from the database, texts representing different orders of approximation to normal English phonetic sequences. For the purpose of text generation, spaces, commas and periods were defined to be phonetic symbols. The first-order approximation contained phonetic symbols that randomly appear with the same frequency that they occur in the database. The second order approximation was generated by the procedure of (a) picking a two-symbol sequence at random from a frequency-weighted tabulation of all symbol-pairs occurring in the database, (b) sending that pair of symbols to the output (c) entering the database at a random point and finding the first instance of the two-symbol sequence, (d) appending the symbol that follows to the output and to the two-symbol sequence itself, and (e) discarding the first symbol of the sequence to form a new two-symbol sequence that now forms the target of the next search that begins at step (c). The generation of third and higher orders followed the same procedure with sequences of three or more symbols obtained from their respective frequency-weighted tables. Using this procedure, two sets of nine 900 symbol-long phonetic texts were constructed that represented the orders 1, 2, 3, 4, 5, 6, 8, 10 and 12. Each of these texts was printed with IPA Doulos phonetic symbols substituted for Arpabet symbols. Samples of these texts are shown in Table 1.

Finally, step 4 employed two professional phoneticians to make speech recordings of two sets of phonetic texts each covering 9 orders of approximation to English. One set of recordings, the practice set, was used to train a group of volunteer shadowers and a second set, the test set, was used to test them. The phoneticians practiced their readings of the texts by articulating short strings of 18–25 IPA symbols plus spaces. Before recording each string of symbols the

Table 1
Samples from statistically constructed texts

*First order*

æn hætu əmɛnzniəs kɪz bi hulɪft ðə lɪksɛn hi n ə k ðeniæ ðə lɛi fajft ə tb oɪmɛkʌnd prən vʊʃəl grɛi pot ðəliə ɪn soraj wə pɚɛkom æz dz nʌstrɛinəl t wi ən kə ɪpʊ ʌtm sɛistmzumi t bi æt ɛsɪn ɑt wəm.

*Second order*

ʃi stɔkmɛi rɪtəl fortæ nd ðə lirzæ nd əv fok ə kwɛs, bəlɪv lɔŋ. θʃornɪtm ɑj wətɪdɛiʃənz korfəl, hæz krɛt trɛindʒrɪstɚ əv pʌblədəmz tapt mɛnsɪz doɚ hɚ sækstɛits sɪsəpoəθ.

*Third order*

ɵart nɛkdəts, majkroənz. kɪmbəlɪʃmənsuəli dont bɪn hɚ ə fɚətɛrɪkən əv ɪŋkrisənt prajz ɪn ðə ɛɚrzæ nd pɛdɪd tʊ ɛni lʊkt əv nat tʊ hɪmsɛlf baj formənt bi ʃud flɪk əv sʌlvd əs, ænd ə ...

*Fourth order*

ðer on kʌlɚ əkstrəfisɪks ænd wɛn ɑj æm ɑj majt haw ɪt wiθ lajf ɑjv gat æz ðə ʊnajtɪd nɛiʃən fəivɚ əv lɛibəl tʊ hɚ ɪn mɪliən ə təmpɚəitɪŋ ɪ tɔf ə nʌmbɚz ə lɔsutsa r ənformәs.

*Fifth order*

ænd orgənizɛiʃənz, bæskɪtbɔl, mɛni jirz, plʌs sɛlɪbrɛit jɚ fɛiθ gruts. ɪd ɛniwʌnz ɪmədʒɪnɛiʃən. ɑjm goɪŋ tʊ ðis gudəlajn, so kampɚəbəl tʊ mɛik ðer wɛi θru ðə gɪlto rɪn ðə nansən.

*Sixth order*

wəgənh æd kɔt fajr ɔlso givən. ʊv gat no bɪznɪs ʌp hɪrw əz mor tʊ du. ɑj hæv wɪθ ðer kəmpækt dɪsks, ænd wɪθ ðə lɛdʒəndh æz ə lat əv dʒuz æz wɛl pɛid æz læmps, ɑj hæv hɚ ə lʊd, wɪl ...

*Eighth order*

ɪldɚz, æz sʌmwʌn wɛl, ɪt wəz ðer ðæt ʃi wʊd hæv bɪn ɔlmost kəntɪnʊd tʊ kʌm wɛn radʒɚz tʊθ fɛl awt. ɵlænd pɛrənθʊd orgənɪzɛiʃənz hæv hædwɪθ fæmi ænd frɛndz.

*Tenth order*

ɮajk hɪz glasi blæk hɛr. ø jɛs, hid tɔkt əbawt pæʃən ænd drajv, ðæt ʊ hæv sɛt jɚ sajts ɑn ətʃivɪŋ ə goəl ænd jɚ wɪlɪŋ tʊ sort əv gɛt æt ðer tru fɪlɪŋgz. ɮɛts hir frəm kɛiθ wʊdz.

*Twelfth order*

ŋat maj gɛim. hiz ə pɪnto ænd hi fotəgræfs wʌndɚfəlli. hawz jɚ lʌk hʌni. ʃtɛik mi, ʃi sɛdm ɚnd lɛt mi æd, ʊtopiənɪzəm, ɔlso. wi əv ðə lɪbrəl lɛd wɚld gat ɔl sɛt fɚ pɪs ænd rɪhəbəlɪtəiʃən.

phoneticians listened to a recording of the DecTalk device producing speech from the *Arpabet* version of the same symbol-string at a rate of 90 wpm.[3] In addition to providing a guide to intonation, the primary role of the DecTalk output was to indicate a target rate of production and insure that each segment would be produced at the same pace within each order and across all orders. The phoneticians were under strict instructions to reproduce each of the phonetic elements specified by the IPA symbols in their texts, and to avoid being influenced by any deviations in the phonetics of the synthesized speech. A considerable number of practice repetitions and re-recordings were required to achieve fluent accurate results on every one of the short strings, particularly at the lowest orders of approximation. The task became progressively less difficult for the phoneticians as order increased. With the aid of a waveform editor (SoundEdit$^{TM}$16), all the recorded strings from a given order of approximation were assembled into one continuous recording. And, using the editor's two-channel mode, the total length of the recording of a given

[3] The slowest rate of delivery available from the DecTalk was estimated as 110–120 wpm. The rate used in this study was reduced to 90 wpm by digitizing the signal and using the menu command Effect/Tempo available in the waveform editor SoundEdit$^{TM}$16.

Table 2
$H$ values for each level of analysis

| Level | Uncertainty | | Redundancy | |
|---|---|---|---|---|
| | $H$ | $H_{rel}$% | $R$ | $R_{rel}$% |
| 0 | 5.58/5.43 | — | — | — |
| 1 | 4.75/4.64 | 85.2/85 | 0.82/0.79 | 14.6/15 |
| 2 | 3.73/3.47 | 66.9/64 | 1.85/1.96 | 33.1/36 |
| 3 | 2.80/2.42 | 50.2/45 | 2.78/3.01 | 49.8/55 |
| 4 | 2.12/1.65 | 38.0/30 | 3.46/3.78 | 62.0/70 |

Data from Hultzén et al. (1964) appear after the / symbol.

order was checked against the synthesizer's production of that same order. When rate discrepancies were found, they were corrected by means of the editor's "Tempo" command that effected the necessary adjustment with no perceptible change in pitch. Upon completion, the two sets of 9 recordings ranged from 90 to 100 s in length. To assess the constraints reflected in the frequencies of the phoneme sequences at different levels of analysis, we followed the procedure of Hultzén et al. (1964) for calculating the $H$ (uncertainty) statistic, a figure of merit for our database.[4] In Table 2, the results of an analysis performed on the present 50,000-word database have been tabulated side-by-side with the figures given by Hultzén et al. The latter results were obtained from a database of only 20,032 phonemes and hence $H$ declines more rapidly at higher levels.

The table also shows the relative uncertainty ($H_{rel} = 100 \times H_x/H_0$; levels $x = 1, 2, 3, 4$), the redundancy ($R = H_0 - H_x$) and the relative redundancy ($R_{rel} = 1 - H_{rel}$). Attempts to extend the analysis beyond the 4th order failed due to computer memory limitations.

### 2.1.2. *Participants*

Twelve adults aged between 30 and 50 years with self-assessed normal hearing volunteered to undergo the practice schedule and perform the test shadowing task. All were native English speakers. The subjects were linguists, psychologists, programmers and secretaries. One subject CF, an author of this paper, was certainly aware of both the literature and the purpose of the experiment. However, it is difficult to envisage how this might have influenced her performance given the demands of the task. With the exception of subject CF, the proposed method of analysis was not discussed with any of the participants prior to the task.

[4] The accuracy with which texts can be made to reflect different orders of approximation to English depends on the size and content of the database used as a source. The uncertainty $H = -\sum p_i \log_2 p_i$ is a measure of the properties of the database. In this expression, $p_i$ is the probability of occurrence of each particular symbol or group of symbols ($i$) among all the individual symbols or groups of symbols of the same length that are present in the database. The magnitude of $H$ indicates the variability in the probabilities of finding each of the available groupings of phonetic symbols. Thus, $H_0$ (see Table 1: Level 0) constitutes a baseline in which all the symbols are assumed to appear with equal probability; $H_1$ is obtained from the actual symbol frequencies in the database; $H_2$ is obtained from all the available combinations of two symbols; $H_3$ from all combinations of three symbols and so on.

### 2.1.3. Procedure

Participants were instructed, in the case of all orders, to shadow the target speaker's utterances with a consistent delay that was as short as they could possibly achieve on a sustainable basis. In particular, the shadowers were told not to adopt the strategy of periodically attending to the target for short periods and then delivering their responses in rapid bursts of speech. Each person was assigned a row in a $9 \times 9$ latin square the rows of which specified the order in which the participants shadowed the different orders of approximation to English. Participants 10, 11 and 12 were randomly assigned to one of the already occupied rows. Each participant spent 40 min repeatedly shadowing, in the order specified by the assigned row, 9 practice recordings representing one of each of the 9 orders of approximation. Following this preparation, and a brief break, the volunteers then shadowed each of the 9 test recordings, once only, in the order given by their assigned row in the Latin square. The voices of both the target speaker and the shadower were digitized and stored on a hard disk for analysis.

Analyses of response latencies and errors were conducted with the assistance of a two-channel waveform display program specially prepared for that purpose using MATLAB. This program applied a time shift to the shadower's response that permitted an observer to simultaneously view on a computer monitor both the target waveform and the shadower's delayed response. The time shift advanced the shadower's waveform display by a fixed amount that approximated the average response latency. However, notwithstanding the on-screen appearance of short (and occasionally negative) measurement intervals, all of the recorded latency measurements included the fixed time shift. Each target recording was divided into syllable-like units. The total number of syllables varied by plus or minus ten from order to order and yielded an average of about 220 per order. Syllables or syllable strings from the target and shadowed waveform displays were selected with a cursor and played in sequence for aural verification and comparison. On the basis of such comparisons, a syllable was selected as a possible imitation when, in the opinion of the analyst, the speaker had produced an utterance that, while not indistinguishable from the target, was a linguistically acceptable token of the target in the subject's dialect. Thus, the onsets of all the acceptably articulated syllables were identified and the corresponding response latencies for these syllables were calculated. All the omitted and incorrectly articulated syllables were counted as errors.

## 2.2. Results

The analysis of the shadowers' recordings showed that most of the volunteers fell into the category defined by Marslen-Wilson (1985) as *distant* shadowers with an average shadowing latency of 500 ms or more. However, three of the shadowers performed inconsistently. On some orders they performed as *distant* shadowers and on others as *close* shadowers with high error rates and mean latencies ranging from 250 to 300 ms. In general, the articulations of shadowers who fell into Marslen-Wilson's *distant* category were more precise than those in the *close* category of shadowers. The latter frequently adopted a strained, slurred and unnatural manner of speech that made it very difficult for the analyst to decide whether their utterances were acceptable or not. Consequently, the measurements obtained from the three inconsistent shadowers were abandoned and only the measurements obtained from the remaining nine distant shadowers were subsequently analyzed.

Table 3
Average latencies for correct syllables

| | Orders of approximation to English | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Subj | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 |
| AF | 856 | 809 | 791 | 815 | 764 | 820 | 679 | 742 | 774 |
| CF | 773 | 744 | 606 | 634 | 594 | 581 | 595 | 556 | 547 |
| DW | 765 | 838 | 882 | 840 | 844 | 792 | 728 | 643 | 556 |
| EM | 651 | 624 | 643 | 639 | 466 | 622 | 534 | 471 | 531 |
| JW | 723 | 761 | 703 | 809 | 793 | 706 | 745 | 733 | 695 |
| KP | 834 | 922 | 839 | 866 | 783 | 752 | 679 | 636 | 670 |
| LK | 457 | 453 | 395 | 467 | 386 | 366 | 343 | 466 | 324 |
| MD | 789 | 774 | 875 | 892 | 851 | 762 | 767 | 771 | 647 |
| SG | 944 | 776 | 839 | 871 | 1006 | 825 | 806 | 775 | 733 |
| Mean | 755 | 745 | 730 | 759 | 721 | 692 | 653 | 644 | 609 |

Because a reliance on the auditory and/or linguistic acuity of a single analyst ran the risk of bias, the opinion of a second analyst was sought. The second analyst examined three of the nine analyzable recordings, and his results were compared with those of the principal analyst. This comparison showed that differences between the numbers of syllables accepted by the two analysts was significant ($t(26) = 4.103$, $p = 0.0004$). However, the most extreme mean difference in any given order of approximation did not exceed 28 ms. The coefficient of correlation between the average latencies measured by the two analysts at each order of approximation was 0.997. Averaged over all orders, the mean difference was 7.11 ms with a standard deviation of 12.63. Furthermore, a $t$-test also showed that the differences in mean latencies obtained by the two analysts were significant at the 2% level ($t(26) = 2.46$; $p = 0.021$). Table 3 below contains the results obtained by the principal analyst. The mean latencies are plotted in Fig. 1.

An analysis of variance (with Order treated as the independent variable and subjects' latencies identified as dependent measures) was performed. It shows that the null hypothesis (shadowing latency is independent of the order of approximation to English) is firmly rejected ($F(8, 64) = 7.87$, $p < 0.0001$). Also, the observed reduction in Latency as Order increases can be approximated by linear regression (see Fig. 1).

Mean Latency $= 780.45 - 14.071 \times$ Order, $R^2 = 0.92$, $F(1, 7) = 82.35$ and $p < 0.0001$. The significance of the differences between all possible pairs of Orders was also examined by Fisher's Protected Least Significant Difference (PLSD) test. That test reveals that among 36 comparisons between Orders, 18 differences in mean latencies are significant at the 5 percent level. No significant differences were found between numerically adjacent Orders. The closest pair of Orders for which a significant difference ($p < 0.0166$) was found was the pair composed of Orders 4 and 6. This was followed by significant differences at the 5 percent level between Order-pairs 5 and 8 and 4 and 8.

The number and percentages of shadowing errors are shown in Table 4. Included as *errors* are all the omitted as well as mispronounced syllables. The second row from the bottom of Table 4 shows the average number of errors made per participant and the bottom row shows those same

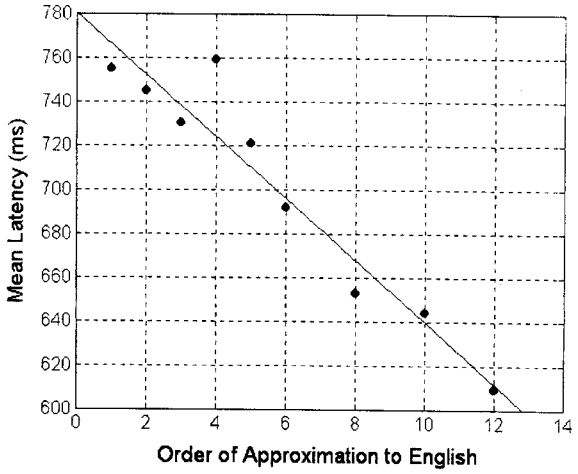


Fig. 1. Mean shadowing latency (ms) of 9 participants for 900-phonetic symbol length passages vs. their order of approximation to English. The regression line is linear. Mean latency = 780.45 − 14.071 × Order.

Table 4
Errors per subject for all orders

| | Orders of approximation to English | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Subj | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 12 |
| AF | 92 | 41 | 67 | 59 | 22 | 33 | 47 | 44 | 15 |
| CF | 97 | 43 | 46 | 30 | 71 | 55 | 62 | 40 | 52 |
| DW | 169 | 82 | 48 | 67 | 52 | 27 | 34 | 30 | 78 |
| EM | 100 | 53 | 55 | 31 | 196 | 11 | 20 | 20 | 19 |
| JW | 115 | 50 | 67 | 48 | 16 | 31 | 32 | 15 | 11 |
| KP | 115 | 52 | 73 | 47 | 13 | 6 | 5 | 8 | 18 |
| LK | 106 | 81 | 100 | 71 | 74 | 52 | 57 | 38 | 54 |
| MD | 139 | 71 | 75 | 72 | 48 | 36 | 29 | 25 | 29 |
| SG | 111 | 38 | 54 | 33 | 15 | 11 | 9 | 3 | 11 |
| Mean | 116 | 56 | 65 | 51 | 56 | 29 | 33 | 25 | 32 |
| Percent | 46 | 24 | 28 | 21 | 25 | 13 | 14 | 11 | 14 |

errors expressed as a percentage of the total number of possible errors (i.e., the number of syllables available in each text). In Fig. 2 the data from the bottom row of Table 4 are plotted.

The ANOVA provides evidence of a substantial error effect as a function of the Order of approximation ($F(8,64) = 9.31$, $p<0.0001$). However, in this instance, a regression analysis shows the relationship between Orders and Errors to be better described by a logarithmic rather than by a linear function. The RMS residual for a linear regression is larger than that for a logarithmic (6.92 vs. 4.59). The latter gives the function:

$$\text{Mean Errors} = 40.218 - 12.314 \times \log_e(\text{Order}),\ R^2 = 0.841,\ (F(1,7) = 36,924,\ p = 0.0005).$$

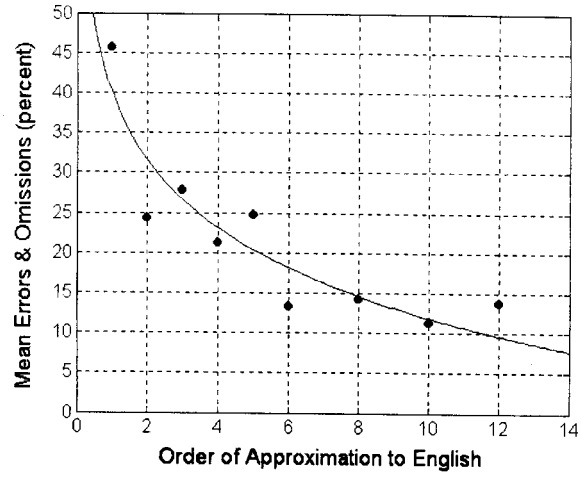


Fig. 2. Percentage of the mean number of errors and omissions made by 9 participants when shadowing 9 passages representing different orders of approximation to English. The regression line is logarithmic. Mean errors = $40.218 - 12.314 \times \log_e$ (Order).

Moray and Taylor (1958) found that a similar logarithmic expression provided the best fit to error data obtained from a shadowing study performed under somewhat different conditions.

The data in Fig. 2 can be interpreted in another way, however. Specifically, a visual inspection will reveal the fact that the error data cluster into three error regions, with order 1 serving as the first cluster, orders 2–5 serving as the second and orders 6 and higher the third. From this perspective, we might ask whether these clusters are indicative of a tendency for the size of familiar shadowed sequences to grow in a categorical rather than a continuous manner. That is, are familiar sequences largely the length of single phones in order 1, syllables in orders 2–5, and words in higher orders? The reader can judge for him- or herself whether the sequences in Table 1 have this categorical character. Although resolving the issue is not critical for our project, it may be relevant to questions concerning whether listeners are sensitive to the probabilistic or statistical properties of speech (e.g., Vitevitch & Luce, 1998, 1999) or to more abstract and qualitative properties (e.g., Frisch & Zawaydeh, 2001). Our data do not speak clearly to this issue because the latencies do not have the categorical character that the errors show. In any case, both latencies and error rates show a decline with order of approximation to English and hence with the familiarity of the sequences being shadowed. We next ask whether familiarity affects the fidelity of imitations that occur when speech is shadowed.

## 3. Experiment 2

While errors increased substantially at low orders of approximation there is an aspect of the shadowing performances at low orders that might be expected to improve, namely, imitative fidelity. Major support for this assertion comes from Goldinger (1998) who observed that imitative behavior during shadowing increases as the frequency of the word being shadowed

declines. Experiment 2 makes the assumption that listeners who know what a speaker is about to say are inclined to be less attentive to acoustic information than listeners who do not. Therefore, it is a plausible hypothesis that shadowers will imitate the lower order approximations to English with greater fidelity than the higher order (more familiar) approximations.

### 3.1. Method

#### 3.1.1. Stimulus materials

To test this hypothesis we adopted the method previously used by Goldinger (1998). We selected two recordings from those made by the shadowers employed in Experiment 1. The criteria were that the chosen shadowers should have both low error scores (particularly for order 1) and phonetic reading skills. Thus, recordings of both the 1st order and 12th order performances of shadowers AF and CF were selected. From those utterances scored as correct in shadower AF's two recordings, 12 multi-syllabic utterances were picked at random from each recording. When shadower CF scored correctly on the same utterance as AF (which was the case for 11 of the 12 utterances selected from AF's recording of order 1), those utterances were also selected from CF's recording. And, to bring CF's total for order 1 to 12, a new utterance was selected. Then, to provide a basis for judging the two shadowers' imitative fidelity, the same utterances were extracted from the phonetician's recordings that had been the target of their shadowing performances. Also, to provide controls, the same two sets of utterances were excised from recordings by the two shadowers of the 1st- and 12th-order phonetic texts (cf. Goldinger, 1998); these recordings were made as the shadowers read the texts. The synthesizer recordings were presented, as they had been to elicit the phonetician's utterances, to guide only the speaking rate and intonation of these readings. The readers were also firmly instructed to read the phonetic script and not to imitate the synthetic speech. The procedure was, therefore, identical to that used to obtain the target recordings. Finally, all the excised multi-syllabic utterances were assembled in AXB format to form four listening tests. That is, each trial of the test presented three occurrences of the same phonetic sequence. The middle one, X, was always a production by the phonetician. A (or B) was a shadower's shadowed response to X. B (or A) was the same speaker's read production.

Each of the four test sequences (representing the 1st and 12th order texts by speakers 1 and 2) consisted of 96 presentations. Thus, each utterance appeared eight times in random order, four times in the order *read/target/shadowed* and four times in the reverse order.

#### 3.1.2. Participants

Thirty five undergraduates at University of Connecticut participated in the experiment for course credit. Of that number, 19 listened to orders 1 and 12 as produced by shadower 1, and 16 listened to the same orders produced by shadower 2.

#### 3.1.3. Procedure

Participants were instructed to listen to each AXB presentation and to decide whether token A or token B bore the closest resemblance to the target X. The listeners were told that they should weigh such features as stress and intonation in addition to phonetic quality before making a decision. The listeners were permitted to hear each presentation as many times as they wished. A computer reproduced the stimuli from 16-bit digitized samples delivered at a rate of 22,050

Table 5
Mean imitations identified vs. mean of random selection

| Speaker | Order | Mean | Std. dev. | DF | $t$ | $p$ |
|---|---|---|---|---|---|---|
| AF | 1 | 54.56 | 10.39 | 15 | 2.53 | 0.0233 |
| AF | 12 | 45.56 | 7.22 | 15 | −1.35 | 0.1969 |
| CF | 1 | 55.84 | 7.49 | 18 | 4.56 | 0.0002 |
| CF | 12 | 50.84 | 6.39 | 18 | 1.94 | 0.0683 |

per second and the listeners registered their decisions by applying a mouse-click on one of two buttons (button A or button B) that they saw on the computer monitor.

### 3.2. Results

The total number of shadowed utterances identified by each listener as imitations (from the sets of 8 repetitions × 12 utterances) was calculated for the two 1st- and two 12th-order presentations. Then, the data were submitted to a single-sided paired $t$-test in which the number of correctly chosen imitations was compared to the chance value of 48. The null hypothesis under examination was that the mean of the distribution of tokens identified as imitations is less than or equal to the mean to be expected from a purely random selection. The numbers of shadowed utterances from 1st order and 12th order passages identified as imitations are given in Table 5.

For the 1st order utterances, the $t$-test shows that listeners are able to identify imitated utterances with considerably higher reliability than would be expected by chance. For the 12th order, utterance reliability drops. The test shows that, between the 1st and 12th orders, the probability of achieving the observed recognition rates by chance rises substantially.

An ANOVA with factors Order (1, 12) and Shadower (AF, CF) confirmed a significant effect for Order ($F(1,33) = 14.91$, $p = 0.0005$), with performance on Order 1 more accurate than on Order 12 as predicted and for Shadowers ($F(1,33) = 2.72$, $p = 0.1083$) with Shadower CF eliciting slightly more accurate judgments than Shadower AF. The interaction was not significant.

## 4. Discussion and conclusions

Our two experiments provide information about shadowing and the impact on it of familiarity with the phonetic sequences being shadowed. Experiment 1, in which we collected shadowing productions, showed that shadowing latency declines as the familiarity of sequences increases. We can ask why shadowing latencies decline as the order of approximation to English increases. The results provide one indication that speaker/listener's shadowing behavior is guided, not only by the material being shadowed, but by something about their familiarity with the language. If the effects are on their articulation alone, the source of the effect might be articulatory practice. Talkers may produce sequences that they have frequently produced more fluently than unfamiliar sequences. Alternatively, or in addition, the effect on latency may stem from their knowledge of the language—for example, their knowledge of words of the language, of permissible or frequent phone sequences or of both kinds of knowledge. This might allow them to anticipate what is

coming and so to know what to say sooner. The results of the shadowing test of Experiment 1 are certainly consistent with this latter account.

With respect to the listening task, the objective of our experiment has been to test a related hypothesis that shadowers who hear phonetic sequences that diverge from normal English phonotactics tend to mimic such utterances more precisely than utterances that structurally conform to English. This prediction derives from two observations. First, when individuals shadow speech, they imitate it; secondly imitative fidelity increases as word frequency decreases (Goldinger, 1998). The results of our AXB listening tests have shown that when utterances extracted from shadowed targets are compared with utterances obtained from reading the target strings aloud (paced by the synthesized string), the 1st order sequences are more frequently judged to be closer to their targets, than the 12th order sequences. This may happen because, with familiar sequences, information in memory, such as lexical knowledge and knowledge of permissible phone sequences, has a larger influence in guiding the shadowers' speech than with unfamiliar speech. This, of course, converges with the second of our two accounts for the latency effect in Experiment 1.

In turn, the effect of lexical knowledge may occur for a variety of reasons having to do with the nature of lexical memory and the ways in which listeners access it. Marslen-Wilson (e.g., 1978) has suggested that listeners generally identify a word as early as possible in the course of its production. Some words can be identified before they have been entirely produced by a speaker at a point that Marslen-Wilson calls the recognition point. In that case, listeners do not attend closely to the final part of the word as spoken; rather they use their lexical knowledge to complete it. One index of this is that, particularly in contexts where words are predictable, shadowers are less sensitive to mispronunciations that occur late in the word than to mispronunciations that occur early (Marslen-Wilson, 1978). This characterization of lexical access might account for our findings, because 12th order approximations include mainly real words of the language. First order sequences include very few. Accordingly, there are more opportunities for shadowers to cease attending closely to the 12th than to the 1st order sequences.

The foregoing account incorporates the conventional idea that lexical memory is a memory of abstract word types. However, Goldinger (1998) argued that his findings that low frequency words are imitated more faithfully than high frequency words provide evidence that lexical memory is an exemplar memory system. Our findings can be accommodated in this theoretical framework as well.

In an exemplar memory, words are stored, not as abstract types, but as individual tokens that listeners have experienced. As such, they retain information in them in addition to information about the linguistic segments that compose the word. For example, they include information about the voice of the speaker who produced the word. In the version of the Minerva exemplar memory model (Hintzman, 1986) that Goldinger used to characterize this kind of lexical memory, when a listener hears a word, exemplars of words in memory are activated to the extent that they share features with the input word. These activated memory traces coalesce with each other and with the input word and form a representation of the word in which features that are consistent across the activated traces are amplified whereas those that are inconsistent cancel each other out. This is called an "echo." Goldinger suggests that the echo guides the shadowers' shadowing responses. When shadowers hear a low frequency word, few exemplars are activated, because there are few in memory. Accordingly, the shadowing response is guided by an echo on which the input word has had a large impact, and so the shadowed response is imitative of the input word.

When they hear a high frequency word, many traces coalesce with the input word to form an echo that, therefore, preserves few features specific to the input, and imitative fidelity is reduced. The same account can be applied to our findings. Phone sequences that constitute a 1st order approximation to English activate very few traces in lexical memory, if any, and so imitative fidelity of the input word is high. Phone sequences that constitute a 12th order approximation do activate words and phrases in memory, and so imitative fidelity is lower.

The two accounts agree that the results reflect the mediating role of lexical memory in the imitation of speech, and we take this to be the main conclusion of our research. Interestingly, shadowers imitate the speech they hear, and elsewhere we have speculated on why this may be (e.g., Fowler, Brown, Sabadini, & Weihing, submitted). However, that is not all that they do. Their shadowing shows evidence of guidance not only by the speech being shadowed immediately but also by the shadowers' past history of experience hearing the shadowed sequence. Shadowers bring their knowledge of their language to bear on their perceptually guided speech actions.

Our findings do not distinguish the specific account in terms of recognition point from that in terms of exemplar lexical memory. However, they do appear to be distinguishable in principle. The recognition point account, but not the exemplar memory account, predicts that imitative fidelity will be reduced after the recognition points of words. This can be tested, in principle, by testing the imitative fidelity of word fragments that come before or after a word's recognition point. A more practical test, however, would be to test the imitative fidelity of whole words that do or do not have recognition points before they end.

The observation that lexical memory guides speech production may also explain an earlier empirical observation that we have made. Sancier and Fowler (1997) found that the voice onset times of the voiceless stops of a bilingual speaker changed depending on her recent language experience. After two months in Brazil (where she heard and spoke her native Portuguese with its unaspirated voiceless stops), her voiceless VOTs were shorter than after a similar interval in the United States. Possibly she, like the shadowers of this study, imitated the speech she heard. The major point of relevance here, however, is how small the effects were. They averaged about 6 ms. Accordingly, the major impact on her speech derived from her lifetime of experience with the language (or her many years of experience in the case of her second language). A small impact occurred from the ambient language. The impact of the ambient language occurred, Sancier and Fowler concluded, because speakers imitate one another. The impact of lexical memory was to mitigate imitative fidelity, a finding consistent with Goldinger's (1998) that low frequency words are imitated with greater fidelity than high frequency words. A second, somewhat less relevant, point is that the effect occurred in parallel in both languages. That is, her English VOTs as well as her Portuguese VOTs were affected when she heard Brazilian Portuguese; her Portuguese and English VOTs were affected when she heard English. This may be of interest in addressing issues about how lexical memory is accessed in bilingual speech perception; it suggests that the lexicons of both languages may be accessed by input from either language.

## Acknowledgements

The authors thank Julie Brown for assistance in collecting imitation judgements and Jeffrey Weihing for help in preparing the database, analyzing its properties and making latency

measurements. We are also grateful to Arthur Abramson and Douglas Honorof who generously contributed the time required to carefully rehearse and speak the target texts. Gary Chant of the City University of New York Graduate Center was extremely helpful in arranging the loan of a DecTalk™ speech synthesizer. The research was supported by NIH grant DC-03782.

## References

Chistovich, L. A. (1960). Classification of rapidly repeated speech sounds. *Akusticheskii Zhurnal, 6*, 392–398 (English translation in *Soviet Physics Acoustics, New York*, 1961, *6*, 393–398).

Chistovich, L. A., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966a). *Mimicking of synthetic vowels.* Quarterly progress status report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, vol. 2, pp. 1–18.

Chistovich, L. A., Fant, G., & de Serpa-Leitao, A. (1966b). *Mimicking and perception of synthetic vowels, part II.* Quarterly progress status report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, vol. 3, pp. 1–3.

Chistovich, L. A., & Kozhevnikov, V. A. (1969). Some aspects of the psychological study of speech. In L. D. Proctor (Ed.), *Biocybernetics of the central nervous system* (pp. 305–321). Boston: Little Brown.

DARPA TIMIT. (1990). Acoustic-phonetic continuous speech corpus. NIST Speech Disc 1-1.1, National Technical Information Service PB91-505065.

Davis, R., Moray, N., & Triesman, A. (1961). Imitative responses and the rate of gain of information. *Quarterly Journal of Experimental Psychology, 13*, 78–89.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (submitted). Do listeners to speech perceive gestures? Evidence from choice and simple response time tasks. *Journal of Memory and Language.*

Frisch, S., & Zawaydeh, B. (2001). The psychological reality of the OCP-place in Arabic. *Language, 77*, 91–106.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: developments in applied sociolinguistics. Studies in emotion and social interaction* (pp. 1–68). New York: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Hultzén, L. S., Allen, J. H. D., & Miron, M. S. (1964). *Tables of transitional frequencies of English phonemes.* Urbana, IL: University of Illinois Press.

Kent, R. D. (1973). The imitation of synthetic vowels and some implications for speech memory. *Phonetica, 28*, 1–25.

Kent, R. D. (1979). Imitation of synthesized English and nonEnglish vowels by children and adults. *Journal of Psycholinguistic Research, 8*, 43–60.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*(4), 2425–2438.

Lawson, E. A. (1961). A note on the influence of different orders of approximation to the English language upon eye–voice span. *Quarterly Journal of Experimental Psychology, 13*, 53–55.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*, 522–523.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science, 189*, 226–227.

Marslen-Wilson, W. (1978). Processing interactions and lexical access during word reception in continuous speech. *Cognitive Psychology, 10*, 29–63.

Marslen-Wilson, W. (1985). Speech shadowing and speech comprehension. *Speech Communication, 4*, 55–73.

McHugo, G. J., Lanzetta, J., Sullivan, D., Masters, R., & Englis, B. (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality & Social Psychology, 49*(6), 1513–1529.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: a theoretical model. *Early Development & Parenting, 6*, 179–192.

Miller, G. A., & Selfridge, J. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176–185.

Moray, N., & Taylor, A. (1958). The effect of redundancy in shadowing one of two dichotic messages. *Language and Speech, 1*, 102–108.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

Porter, R. J., & Lubker, J. F. (1980). Rapid reproduction of vowel–vowel sequences: evidence for a fast and direct acoustic–motoric linkage in speech. *Journal of Speech and Hearing Research, 23*, 593–602.

Rosenberg, S., & Lambert, W. E. (1974). Contextual constraints and the perception of speech. *Journal of Experimental Psychology, 102*, 178–180.

Salzinger, K., Portnoy, S., & Feidman, R. S. (1962). The effect of order of approximation to the statistical structure of English on the emission of verbal responses. *Journal of Experimental Psychology, 64*, 52–57.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421–436.

Shockley, K., Santana, M.-V., & Fowler, C. A. (in press). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly, 30*, 415–433.

Taylor, W. L. (1956). Recent developments in the use of cloze procedure. *Journalism Quarterly, 33*, 42–48.

Taylor, A. M., & Moray, N. (1960). Statistical approximations to English and French. *Language and Speech, 3*, 7–10.

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. London: Routledge.

Triesman, A. M. (1965a). Verbal responses and contextual constraints in language. *Journal of Verbal Learning and Verbal Behavior, 4*, 118–128.

Triesman, A. M. (1965b). The effects of redundancy and familiarity on translating and repeating back a foreign and native language. *British Journal of Psychology, 56*, 369–379.

Underwood, G. (1974). Moray vs. the rest: the effects of extended shadowing practice. *Quarterly Journal of Experimental Psychology, 26*, 368–372.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408.