Research Report

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
22(9) 1113–1119

DOI: 10.1177/0956797611417723
http://pss.sagepub.com
SAGE

Finding Levels of Abstraction in Speech Production: Evidence From Sound-Production Impairment

Adam Buchwald[1] **and Michele Miozzo**[2,3]
[1]Department of Communicative Sciences and Disorders, New York University; [2]Department of Experimental Psychology, University of Cambridge; and [3]Department of Cognitive Science, Johns Hopkins University

Abstract

A widely held view in linguistics and psycholinguistics is that there are distinct levels of processing for context-independent and context-specific representations of sound structure. Recently, this view has been disputed, in part because of the absence of clear evidence that there are abstract mental representations of discrete sound-structure units. Here, we present novel evidence that separate context-independent and context-specific representations of sound structure are supported by distinct brain mechanisms that can be selectively impaired in individuals with acquired brain deficits. Acoustic data from /s/-deletion errors of 2 aphasic speakers suggest both a phonological level of processing at which sound representations (e.g., /p/) do not specify context-specific detail (e.g., aspirated or unaspirated) and a distinct level at which context-specific information is represented. These data help constrain accounts of sound-structure processing in word production and crucially support the claim that context-independent linguistic information affects language production.

Keywords

phonemes, allophones, speech production, aphasia, phonology

Received 7/23/10; Revision accepted 4/18/11

Whether there are abstract mental representations of experience in addition to representations that encode more detailed aspects of experience is a key question in many areas of cognition. This question is particularly important in those areas in which sensory and motor systems interface, of which the domain of spoken language is a primary example. Theories in linguistics (Browman & Goldstein, 1986; Chomsky & Halle, 1968; Prince & Smolensky, 1993/2004) and psycholinguistics (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999) have traditionally distinguished between two types of mental representations of speech sounds: representations encoding context-specific articulatory and acoustic details of the measurable speech signal and more abstract representations encoding context-independent information about speech sounds. In these theories, abstract representations specify context-independent elements (e.g., phonemes, segments, or gestures) that underlie sound structure and are stored in long-term memory, and these elements are mapped to corresponding context-specific representations (e.g., allophones or coordinated gestures) that more directly encode the continuous speech signal (e.g., position in a syllable and temporal and spatial coordination among adjacent elements). Both the nature of this mapping (e.g., rules vs. violable constraints) and the content of these representations have been debated extensively.

The debate, however, has recently extended to the fundamental distinction between context-specific and context-independent representations: This distinction has been disputed in accounts that posit only context-specific representations (Port, 2010; also see Wickelgren, 1969). According to these accounts, lexical representations are richly detailed exemplars of the acoustic, articulatory, and perceptual experiences of words, and there exists no level of representation encoding more abstract elements (but see Pierrehumbert, 2006). Whether context-independent representations are necessary has thus become a pressing question in the field of language production. We addressed this question by investigating selective speech impairments resulting from acquired brain lesions.

Much of the psycholinguistic evidence cited in support of context-independent representations in speech production has come from the speech errors of neurologically intact and damaged individuals (Dell, 1984; Fromkin, 1971, 1973; Shattuck-Hufnagel, 1987; Stemberger, 1990). For example, Fromkin (1973) reported that when a sound is erroneously substituted

Corresponding Author:
Adam Buchwald, Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Suite 900, New York, NY 10012
E-mail: buchwald@nyu.edu

or exchanged for another sound, it takes on the context-specific properties corresponding to the new context; this result has been explained by claiming that the error occurs at a level of context-independent units that are mapped to context-specific representations only after the error occurs. However, recent data resulting from the widespread use of acoustic and articulatory measures show that many speech errors are best described as combinations of motor units corresponding to the target and error sounds, rather than as complete errors (e.g., Hardcastle & Edwards, 1992; Mowrey & MacKay, 1990; Pouplier & Hardcastle, 2005). These data revealing partial errors are consistent with theories eschewing abstract representations because the errors reflect differences that cannot be described with reference to larger abstract units (e.g., phonemes); such data therefore underscore the relevance of analyses based on instrumental measures rather than transcription.

Studies of acquired language deficits subsequent to brain damage (e.g., aphasia) offer a unique opportunity to investigate whether there are indeed separate levels for processing context-independent and context-specific representations, or whether such a distinction is unwarranted. Evidence that individuals with an acquired impairment make errors reflecting only one or the other level of processing would support theories that include such a distinction. Accounts that do not assign a level of processing for context-independent representations predict that such individuals will make only errors with context-specific properties.

To test these alternative accounts, we investigated consonant clusters, which are a common source of speech errors for individuals with acquired phonological deficits (Romani & Galluzzi, 2005). In the research we report here, we focused on the speech errors made by D. L. E. and H. F. L., 2 American English speakers with aphasia subsequent to stroke. D. L. E. and H. F. L. systematically delete /s/ from onset clusters consisting of /s/ plus a stop, as in *sp*ill, *st*ill, and *sk*ill. The corresponding words without /s/ (*pill*, *till*, and *kill*) begin with the English voiceless stops /p/, /t/, and /k/, and these stops are pronounced with aspiration—that is, their release is accompanied by a burst of air (traditionally indicated as [p^h], [t^h], and [k^h]; Fig. 1a). By contrast, *spill*, *still*, and *skill* contain the

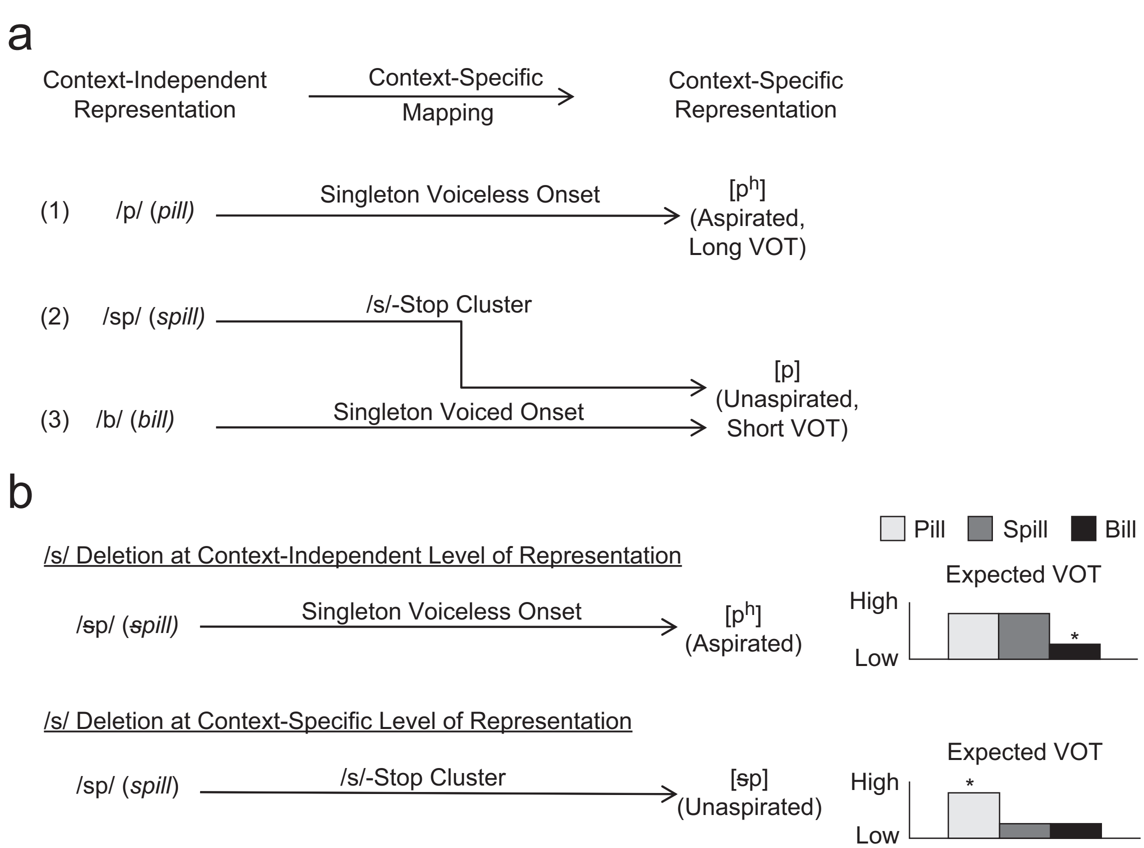

Fig. 1. Predictions of the traditional account of sound-structure processing, according to which there exist separate levels for abstract, context-independent representations and detailed, context-specific representations. In English, context-specific mapping of context-independent representations results in an aspirated stop with long voice-onset time (VOT) in the case of a word-initial voiceless stop and results in an unaspirated stop with a short VOT in the case of a voiceless stop following word-initial /s/ or a word-initial voiced stop (a). The bar graphs (b) illustrate how the level at which the deletion of a word-initial /s/ (a common speech error in individuals with acquired deficits) occurs would be expected to affect the VOT of an immediately following voiceless stop. If the deletion occurs at a level with context-independent representations, the stop would be expected to be aspirated and to have a VOT significantly longer than that of a word-initial voiced stop. In contrast, if the deletion occurs at a level that does include context-specific detail, the stop would be expected to be unaspirated and to have a VOT significantly shorter than that of a word-initial voiceless stop (predicted significant differences are indicated by asterisks).

unaspirated variant of these consonants, which—all else being equal—are phonetically realized as English voiced stops in word-onset clusters (Klatt, 1975). That is, the stop in *spill* is phonetically equivalent to the /b/ in *bill*. Browman and Goldstein (1986) explained this variation in aspiration of voiceless stop consonants as due to an influence of context on the timing of the glottal gesture associated with voicing and aspiration.

Given a system with context-independent representations at one level and context-specific representations at another, deletion of word-initial /s/ could result in two distinct patterns: If the deletion occurs at a level where context-specific details are not specified (i.e., if the representation of the word is context independent), then the resulting voiceless stop following the /s/ should have the timing associated with aspiration (i.e., the timing used in the case of a word-initial voiceless stop), so that *spill* would be pronounced like *pill*. Alternatively, if the deletion occurs at a level encoding context-specific representations (i.e., if the /s/ is deleted from a representation that includes contextual details), then the timing should be based on the presence of a cluster, and the stop should be unaspirated, as in *bill* (Fig. 1b). However, if speakers represent sounds only with respect to their context-specific (surface) form, all word-initial voiceless stops following a deleted /s/ should be unaspirated, as in *bill*.

We tested these contrasting predictions by using a repetition task to generate /s/-deletion errors and then using acoustic analyses on those tokens to determine whether critical consonants were produced with aspiration. In addition, we used two minimal-pair discrimination tasks to establish that the /s/-deletion errors stemmed from production deficits rather than perceptual deficits.

Participants

D. L. E. (born in 1934) is a left-handed male college graduate and retired CEO. In 2001, he experienced a left middle cerebral artery (MCA) infarct. MRI scans revealed a large lesion extending over posterior-frontal and temporal-parietal cortex, insula, and thalamus. H. F. L. (born in 1950) is a right-handed male who earned a medical doctorate and worked as a radiologist. In 2007, he experienced a left MCA infarct, which left him with right-sided hemiplegia. Each individual was diagnosed with nonfluent aphasia.

The speech of both participants was dysfluent and contained hesitations and pauses, and H. F. L.'s speech was limited to single-word utterances and a few formulas (e.g., "thank you"). D. L. E. and H. F. L. were moderately impaired in tasks routinely administered to assess oral-motor deficits, and each was diagnosed with mild-to-moderate apraxia of speech. Both participants were more impaired in naming a picture than in repeating a picture name spoken to them (D. L. E.: 67% vs. 87% correct, respectively; H. F. L.: 10% vs. 80% correct, respectively), a pattern suggesting impaired lexical access in addition to a deficit affecting postlexical, phonological processing. D. L. E.'s speech errors were most often phonological errors, that is, errors in which the target and spoken word overlapped in form by at least 50% (e.g., *cannon* → *gannon*). Such errors were the only speech errors produced by H. F. L.

Repetition Task

We used a repetition task to elicit spoken responses from D. L. E. and H. F. L. Words were presented in isolation by the experimenter and were immediately repeated by the participant, and these stimuli and output were recorded. We chose this task because both participants could perform it easily, and it permitted us to directly examine their phonological impairment (rather than their lexical-access impairment). Although our main aim was to investigate /s/-onset words, most of the words presented for repetition were fillers introduced to vary the onsets of the target words.

Two factors conspire in making the /s/ of word onsets consisting of /s/ plus a stop (e.g., *spill*) particularly vulnerable to deletion among individuals with phonological deficits. One factor relates to the common observation that consonant clusters are vulnerable to error in individuals with aphasia. Both of our participants produced consonants less accurately when they occurred in clusters (e.g., *blink*) than when they occurred as singletons (e.g., *link*). For D. L. E., the average accuracy in repeating cluster consonants was 45.1% (533/1,182), and the average accuracy in repeating singletons was 81.4% (1,060/1,302), $\chi^2(1, N = 2,384) = 353.7$, $p < .0001$, whereas for H. F. L., the average accuracy in repeating cluster consonants was 52.9% (717/1,356), and the average accuracy in repeating singletons was 81.5% (594/729), $\chi^2(1, N = 2,085) = 165.0$, $p < .0001$. The second factor relates to sonority (a correlate of amplitude). A preference for sonority profiles that rise from the onset to the nucleus has been observed both in cross-linguistic patterns and in language-internal frequency (Clements, 1990). In general, onset consonant clusters in which the second consonant is more sonorous than the first (e.g., /sl/) are preferred over those in which the second consonant is less sonorous than the first (e.g., /sp/, /st/, /sk/). Sonority profiles affect error rates in some aphasic patients (Romani & Calabrese, 1998), a finding replicated with our participants. Both participants made fewer errors repeating clusters with rising sonority profiles (i.e., /sl/) than repeating clusters consisting of /s/ plus a stop—D. L. E.: $\chi^2(1, N = 525) = 25.77$, $p < .0001$; H. F. L.: $\chi^2(1, N = 860) = 42.60$, $p < .0001$ (Fig. 2). Deletion of /s/ was the most common error for both participants in the words included in this analysis.

Minimal-Pair Discrimination

Participants may have deleted onset /s/ in the repetition task because of failures to perceive this sound. To rule out this possibility, we tested our participants' perception of /s/ using minimal-pair discrimination tasks requiring same/different decisions. One task was from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay,

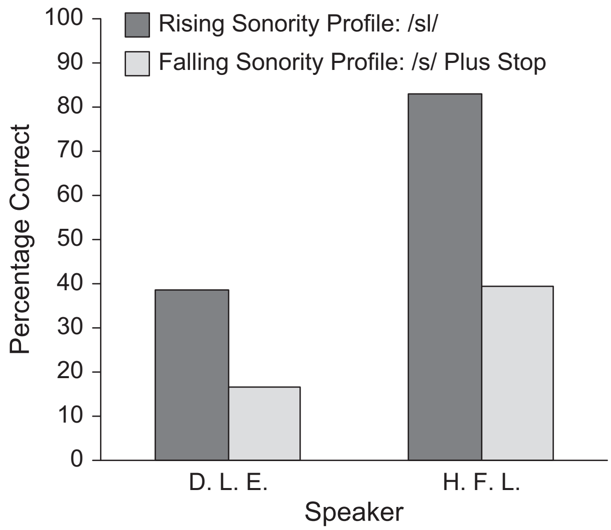

Fig. 2. The 2 participants' accuracy in repeating onset clusters with rising sonority (/sl/) and with falling sonority (/s/ plus a stop).

Lesser, & Coltheart, 1992), which involves minimal-pair discrimination of nonwords, not specifically targeting /s/; the other task was developed specifically for this study and required discriminating between words beginning with /s/ plus a stop and their stop-initial counterparts (either voiced, such as *spill* vs. *bill*, or voiceless, such as *spill* vs. *pill*). The results (see Table 1) indicate that both participants' perception of spoken words was intact.

Acoustic Analyses of Voice-Onset Times

To examine the nature of the /s/-deletion errors participants made during the repetition task, we measured the VOT of the word-onset stop consonants. VOT is the duration from the release of the stop constriction to the onset of voicing (Lisker & Abramson, 1964). For unimpaired English speakers, word-initial voiceless stops (/p/, /t/, /k/) are produced with aspiration and long VOTs, whereas word-initial voiced stops (/b/, /d/, /g/) are produced without aspiration and with short VOTs (Fig. 1). In words beginning with /s/ plus a stop, the stop is produced without aspiration, and the VOT is similar to that of a word-initial voiced stop (i.e., the VOT of /p/ in *spill* is similar to the VOT of /b/ in *bill*). Thus, if only surface forms of words (i.e., with context specified) are encoded in the lexicon, the /p/ in *spill* would be pronounced as an unaspirated stop with a relatively short VOT after /s/ deletion. However, as shown in Figure 1, under the traditional view, which includes separate levels for context-independent and context-specific representations, two separate patterns could be observed: A speaker who deletes the /s/ of /sp/ from a representation at the context-independent level would produce the aspirated stop [p^h] (long VOT), whereas a speaker who deletes /s/ from a representation at the level encoding context-specific representations would produce the unaspirated stop (short VOT).

We tested these predictions with comparisons within minimal triads (e.g., *spill*, *pill*, *bill*) composed of monosyllabic words that were presented in the repetition task. Each triad consisted of one word beginning with an /s/ plus a stop, one word beginning with a voiceless singleton onset stop, and one word beginning with a voiced singleton onset stop. The words in a triad were not always perfect matches (e.g., *stoop*, *toot*, *dupe*), but they were matched for both the stop's place of articulation and the identity of the following vowel, because a consonant's place of articulation and the height of the following vowel influence VOT in English stops (Klatt, 1975). Words in a triad were never presented in succession. For each token in which /s/ deletion occurred, we compared the VOT of that stop (e.g., of /p/ in *~~s~~pill*) with the VOTs of the stops in the other (correctly repeated) triad members (e.g., the /p/ in *pill* and /b/ in *bill*). Note that /s/ deletions could lead to real words (*~~s~~pill* → *pill*) or to nonwords (*~~s~~toop* → *toop*).

Two independent raters identified /s/ deletions on the basis of acoustic and spectrographic evidence. Only tokens that both raters identified as /s/ deletions were considered for VOT analyses (D. L. E.: $N = 195$; H. F. L.: $N = 58$). We performed one-factor repeated measures analyses of variance (ANOVAs)

Table 1. D. L. E.'s and H. F. L.'s Accuracy on Tests Assessing Minimal-Pair Discrimination

	D. L. E.		H. F. L.	
Task and stimuli	*Same* trials	*Different* trials	*Same* trials	*Different* trials
PALPA Subtest 2	35/36 (z = −0.69)	36/36 (z = 0.45)	36/36 (z = 0.58)	36/36 (z = 0.45)
Minimal-pair discrimination task				
Pairs involving voiceless initial stops (e.g., *spill-pill*)	59/60 (98%)	60/60 (100%)	80/81 (99%)	35/36 (98%)
Pairs involving voiced initial stops (e.g., *spill-bill*)	58/60 (97%)	49/60[a] (82%)	80/81 (99%)	32/32 (100%)

Note: Each cell of this table indicates the number of word pairs correctly identified as being the same or different out of the total number of word pairs presented. For the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA), the table presents z scores based on norms from unimpaired participants (Kay, Lesser, & Coltheart, 1992).
[a]The confusions D. L. E. occasionally made when listening to words like *spill* and *bill* cannot explain his repetition errors, as he did not make production errors of the kind in which *spill* is pronounced like *bill.* Thus, if he had a mild auditory perception deficit, this deficit would not have led to the pattern discussed in the Acoustic Analyses section.

Table 2. Results of Voice-Onset Time Analyses Comparing Stops in Utterances With /s/-Deletion Errors With Stops in Utterances Correctly Pronounced With Singleton Stop Onsets

Participant and place of articulation	Analysis of variance	Comparison with correct voiceless stops		Comparison with correct voiced stops	
		t test	*d*	*t* test	*d*
D. L. E.					
Labial	$F(2, 140) = 11.48, p < .001$	$t(70) = -1.53, p = .13$	—	$t(70) = 2.81, p < .007$	0.70
Coronal	$F(2, 140) = 11.45, p < .001$	$t(70) = -0.51, p = .61$	—	$t(70) = 3.51, p < .002$	0.61
Velar	$F(2, 34) = 10.70, p < .001$	$t(17) = -1.70, p = .11$	—	$t(17) = 2.89, p < .01$	0.73
H. F. L.					
Labial	$F(2, 78) = 323.83, p < .002$	$t(39) = -21.39, p < .001$	3.79	$t(39) = -0.51, p = .61$	—
Coronal	$F(2, 32) = 82.52, p < .001$	$t(16) = 8.75, p < .001$	3.44	$t(16) = 1.26, p = .22$	—

with VOT as the dependent variable and onset type (error onset vs. correct voiceless onset vs. correct voiced onset) as the independent variable. Separate ANOVAs were conducted for each participant and each place of articulation (labial: /sp/, /p/, /b/; coronal: /st/, /t/, /d/; velar: /sk/, /k/, /g/). H. F. L. could not regularly produce singleton velars or /sk/ clusters, so they were not included in his tests. Each ANOVA yielded a significant result (see Table 2) that we further explored with paired *t* tests comparing VOT of error onsets with (a) VOT of correct voiceless onsets and (b) VOT of correct voiced onsets. The VOT data showed different patterns for the 2 participants: D. L. E.'s stops in words with /s/ deletions were similar to his aspirated correct voiceless stops (and significantly longer than his unaspirated correct voiced stops), whereas H. F. L.'s stops in words with /s/ deletions were similar to his unaspirated correct voiced stops (and significantly shorter than his aspirated correct voiceless stops; see Fig. 3). Post hoc analyses revealed similar patterns irrespective of whether deletions led to real words (*~~s~~pill → pill*) or nonwords (*~~s~~toop → toop*; see Table 3).

We also compared the VOT of the stops in accurately produced /s/-onset clusters with the VOT of the stops in /s/ deletions and with the VOT of singleton voiced onset stops (e.g., *spill* vs. *~~s~~pill* vs. *bill*). This was possible only for H. F. L., who, unlike D. L. E., produced a sufficient number of correct /s/-stop onsets to permit this comparison. The VOT of the stops in correctly produced clusters (labial: $M = 12.34$ ms, $SD = 5.70$; coronal: $M = 27.99$ ms, $SD = 5.40$) did not differ significantly from the VOT in clusters with deleted /s/ (see Fig. 3) or the VOT of correctly repeated singleton voiced stops (see Fig. 3; all *t*s > 1). These data confirm that H. F. L.'s stops were not produced differently when the /s/ was deleted (e.g., *~~s~~pill*) than when it was correctly produced (e.g., *spill*) or when the stop was a correctly repeated voiced singleton (e.g., *bill*).

Discussion

The stops that surfaced in D. L. E.'s deletion errors were produced in a way similar to that of his word-initial voiceless aspirated stops ([p^h], [t^h], [k^h]), whereas the stops that surfaced in H. F. L.'s deletion errors were produced without aspiration (like /b/, /d/, and /g/). Given that a stop immediately following a word-initial /s/ typically surfaces without aspiration, D. L. E.'s deletion errors occur at a point in the production system where representations of sound structure contain context-independent elements, whereas H. F. L.'s errors occur at a point where representations encode context-specific detail. For each individual, the same patterns were obtained for each place of articulation. This result is critical for two reasons. First, the consistency of these patterns renders it unlikely that the errors resulted from difficulty with a particular articulatory sequence.

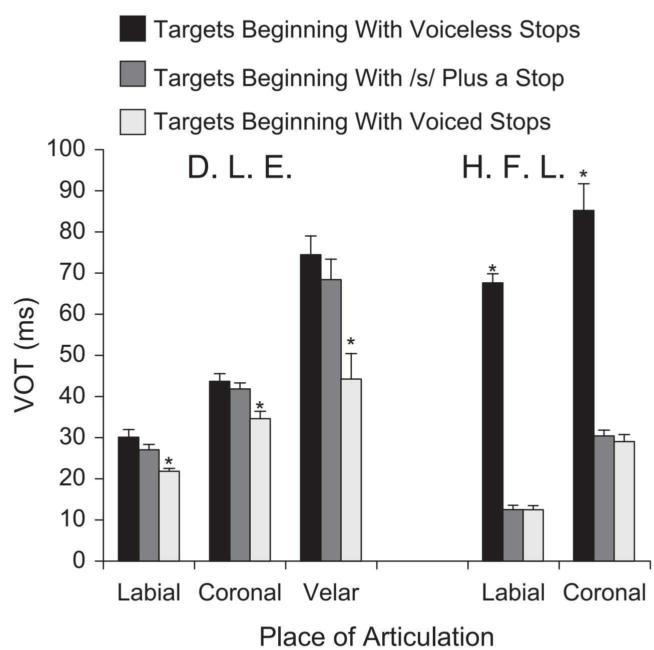

Fig. 3. Voice-onset time (VOT) in the repetition task for D. L. E. and H. F. L. as a function of the onset consonant's place of articulation and target word type. For target words beginning with /s/ plus a stop, only tokens in which the /s/ was deleted are included. The other two target types were words beginning with singleton voiceless and voiced stops. Error bars represent standard errors of the mean. Asterisks denote VOTs significantly different from the VOT for tokens in which the /s/ was deleted, within a given place of articulation ($p = .05$). H. F. L. could not regularly produce singleton velars or /sk/ clusters, so they were not included in his tests. See Table 2 for results of statistical tests.

Table 3. Results of Voice-Onset Time Analyses Comparing Stops in Utterances With /s/-Deletion Errors With Stops in Utterances With Correctly Pronounced Singleton Stop Onsets, Restricting the Analyses to Target Words in Which /s/ Deletion Creates Nonwords

Participant and triad type	Voiceless stops	Voiced stops
D. L. E.		
/s/ deletion creates a nonword with voiceless onset	$t(45) = -0.85, p = .41$	$t(45) = 2.03, p = .05$
/s/ deletion creates a nonword with voiced onset	$t(38) = -1.08, p = .29$	$t(38) = 3.01, p = .006$
H. F. L.		
/s/ deletion creates a nonword with voiceless onset	$t(13) = -9.42, p < .001$	$t(13) = 0.67, p = .55$
/s/ deletion creates a nonword with voiced onset	$t(14) = -13.31, p < .001$	$t(14) = 1.47, p = .17$

Note: Separate analyses were conducted for triads in which an /s/ deletion resulted in a nonword starting with a voiced stop (e.g., *star* became *tar* or **dar*; *spine* became *pine* or **bine*) and triads in which an /s/ deletion resulted in a nonword starting with a voiceless stop (e.g., *stoop* became **toop* or *dupe*; *spite* became **pite* or *bite*). Asterisks in the examples indicate nonwords resulting from /s/ deletions.

Second, this finding indicates that these mappings among levels of sound-structure processing apply at a level removed from the speech signal (e.g., they affect all stops regardless of place of articulation).

The data reported here are consistent with an account of spoken production containing at least two processing levels that can be selectively impaired by brain damage: one processing stage with context-independent representations of sounds and another with context-specific representations. Issues requiring further research include determining the exact nature of the information in these representations and whether these levels are sequentially related. More broadly, our results suggest that the cognitive system implicated in spoken production is not limited to computing representations that reflect the embodiment of speech (context-specific articulation and acoustics); this system also processes more abstract elements that are context independent.

Much of the literature suggesting that people have only detailed representations of embodied experience of sound structure (Port, 2010) has focused on findings regarding perception, such as those indicating that memory for spoken words includes articulatory details (Palmeri, Goldinger, & Pisoni, 1993) and that speech perception is facilitated by familiarity with details of a speaker's voice (Nygaard, Sommers, & Pisoni, 1994). The data presented here indicate that both context-independent and context-specific representations of sound structure are computed in speech production. To the degree that speech sound processing relies on (at least some) common cognitive mechanisms for perception and production (e.g., Liberman & Mattingly, 1985), both types of representations may be active in perception as well as production (as argued by Lahiri & Marslen-Wilson, 1991).

Acknowledgments

We thank Cristina Sanchez, Cettina Chiarelli, and Colleen Herlihy for preparing materials, testing participants, and scoring the data, and Kristina Kmieciak, Katherine Canonico, and Jacqueline Irwin for helping with data analysis. We also thank Brenda Rapp and Sarah Hawkins for fruitful discussion of the results, and Matt Goldrick and two anonymous reviewers for feedback on this report. Finally, we thank D. L. E. and H. F. L. for their willing participation in the study.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by National Institutes of Health Grant DC006242.

References

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *6*, 219–252.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper and Row.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In M. Beckman & J. Kingston (Eds.), *Papers in laboratory phonology 1* (pp. 283–333). Cambridge, England: Cambridge University Press.

Dell, G. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 222–233.

Dell, G. (1986). A spreading activation theory of retrieval in sentence processing. *Psychological Review*, *93*, 283–321.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52.

Fromkin, V. (1973). *Speech errors as linguistic evidence*. The Hague, The Netherlands: Mouton.

Hardcastle, W. J., & Edwards, S. (1992). EPG-based description of apraxic speech errors. In R. D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 287–328). Amsterdam, The Netherlands: John Benjamins.

Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia*. Hove, England: Erlbaum.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*, 686–706.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the mental lexicon. *Cognition*, *38*, 245–294.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, *22*, 1–75.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, *88*, 1299–1312.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, *34*, 516–530.

Port, R. F. (2010). Rich memory and distributed phonology. *Language Sciences*, *32*, 43–55.

Pouplier, M., & Hardcastle, W. J. (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica*, *62*, 227–243.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. (Original work published 1993)

Romani, C., & Calabrese, A. (1998). Syllabic constraints on the phonological errors of an aphasic patient. *Brain & Language*, *64*, 83–121.

Romani, C., & Galluzzi, C. (2005). Effects of syllabic complexity in predicting accuracy of repetition and direction of errors in patients with articulatory and phonological difficulties. *Cognitive Neuropsychology*, *22*, 817–850.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17–52). Hillsdale, NJ: Erlbaum.

Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition*, *35*, 123–157.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*, 1–15.