# The Role of Subphonemic Units in Sub lexical Speech Errors: Insights from Resource Models of Working Memory

## Jenah Black* and Nazbanou Nozari

Department of Physiology, Carnegie Mellon University, Pittsburgh, PA, USA

*Corresponding author: Jenah Black, Department of Physiology, Carnegie Mellon University, Pittsburgh, PA, USA; E-mail: jenahb@andrew.cmu.edu

## Abstract

**Background:** Sublexical errors (e.g., "tent"→"dent") may result from phoneme substitution (/t/→/d/) or from subphonemic processes, such as a change to a phonetic feature (voiceless→voiced). We employed resource based theories of working memory to disentangle these sources by analyzing the distance between errors and their targets in the articulatory phonetic space. If segmental encoding re lects the active assembly of phonetic features, precision should vary with word length. If, on the other hand, segmental encoding re lects the selection of whole phonemes as units, degree of precision should not vary with word length, as a phoneme is either correct or incorrect.

**Methods and findings:** 1,798 phonological errors were collected from four individuals with aphasia. Group level analyses showed that increased word length led to greater distance in phonetic space between the target and response phonemes (β=0.171, t=14.60, p<.001). This pattern was also clear in individuals data. A second set of analyses showed that this effect was not driven by articulatory simpli ication.

**Conclusions:** These data provide strong support for the role of subphonemic units in the generation of sublexical errors in aphasia as well as during segmental encoding in word production.

**Keywords:** Sublexical errors; Precision; Lemma; Electromyography; Institutional Review Board (IRB); Western Pennsylvania Patient Registry (WPPR)

## Introduction

In neurotypical adult speakers, speech is often seamless and error free. But when errors do happen, their systematic patterns greatly inform our understanding of the hierarchical nature of the language production system. While the origin of lexical errors (e.g., saying 'dog' when you mean to say 'cat') is well-understood, the origin of sublexical errors (e.g., saying 'mat' for 'cat') is less clear. An outstanding question is whether such errors arise primarily from non-decompositional phonemes or instead from bundles of phonetic features. Answering this question allows us to better conceptualize the structure of the language production system, as well as the in luences that act upon that system at different stages of processing. This paper utilizes a new approach to analyzing sublexical errors from individuals with aphasia to shed light on this issue.

## The architecture of the language production system

The language production system can be generally understood as a set of hierarchical levels, moving from abstract representations (e.g., lexical items) to more grounded representations (e.g., motor codes). There are at least two distinct stages of processing: in the irst stage, semantic features of an object or concept activate the corresponding lexical item, or lemma, along with the other lexical items that share some of those features. The second stage begins by selecting a lemma and proceeds by mapping it on to representations which translate it into sound. This stage itself includes multiple layers of representation, including the phonological level, the motor planning level, and the motor execution level [1]. At the phonemic level, phonemes activate their respective phonetic features and phonological rules are applied, generating allophonic variations that are sensitive to context. Finally, phonetic representations are translated into a plan of motor movements, which are executed to produce speech.

The multi level architecture of the language production system is further evidenced by the different properties of speech errors. For example, the fact that lexical errors (e.g., 'dog' for 'cat') respect grammatical category while segmental errors (e.g., 'mat' for 'cat') do not suggests that sublexical errors arise at a later stage, a ter the application of syntactic constraints [2]. Disentangling sublexical layers of representation has been more challenging. An important debate is whether phonemes exist as holistic representations or whether what is linguistically considered to be a phoneme is merely a combination of a number of phonetic features. In comprehension, the evidence from selective adaptation studies supports the latter view. Selective adaptation is a perceptual phenomenon in which repeated exposure to a categorical

stimulus makes it less likely to be perceived in the future when an ambiguous token is heard [3]. Multiple studies have demonstrated that selective adaptation is not position independent, as would be expected if position invariant phonemes were critical to perception [4]. Most recently, Samuel exposed listeners to a series of words that contained the sounds /b/ or /d/ in word final or word initial position. When listeners adapted to the exposure words that were /b/or/d/ final, they did not demonstrate adaptation when presented with syllables that were /b/or/d/ initial. From this, he concluded that phonemes may not be the right unit of processing in perception. However, production is different from comprehension in a number of ways, including the need for selecting an accurate representation in order to issue and execute a precise motor command. It is thus possible that phonemes, as non-decompositional units, are represented more strongly in the production system.

There is some support for the psychological reality of phonemes in production. For example, Oppenheim and Dell found that errors made in inner speech display a lexical bias (errors result in words more often than non-words; Baars, et al.; Nozari and Dell) but not a phonetic similarity effect (the tendency for phonemes that share features to interact; Caramazza, et al., MacKay, Shattuck-Hufnagel and Klatt). From this pattern of results, the authors concluded that representations in inner speech are impoverished compared to spoken language and do not interface with phonetic features, suggesting a level of processing with access to phonological representations but not the feature information that drives the phonetic similarity bias [5]. Additionally, Roelofs found participants were able to name a set of pictures more quickly when the initial sounds of the items were identical (that is, the same phoneme), compared to when the initial sounds were dissimilar (e.g., /d/ versus /f/). There was, however, no facilitation when the phonemes merely shared many features: /b/ and/p/. He concluded that the facilitation occurred at the phonological level, supported by phoneme representations that are insensitive to subphonemic information. Finally, the primacy of phonological representations for production is demonstrated in accommodation, or the fact that the allophonic environment often changes to accommodate an error. For example, if a consonant is erroneously voiced, the preceding vowel will lengthen, such that the utterance remains well-formed by the rules of English (e.g., 'track cows' for target 'cow tracks', where the slipped /s/ accommodates the final /w/ and becomes voiced; Fromkin. This pattern suggests that these errors are generated before phonemes are translated into phonetic features.

The presumed prevalence of accommodated errors naturally led to the claim that the majority of segmental errors must occur at the phonological level, rather than the phonetic or motor levels [6]. However, more recent researchers would contest the idea that sublexical errors are rare, claiming self-reported errors (which was a common method of collecting data in early studies) did not show sublexical errors because such errors are significantly harder to perceive, remember, and transcribe than phonological errors. Indeed, evidence using acoustic recordings and measurements of articulatory movements (in lieu of transcription) confirmed that gradient, sublexical errors are commonly produced and that those errors show traces of the intended target, suggesting competition. For example, the Voice Onset Time (VOT) of a voiced consonant was longer if it was erroneously produced in the place of a voiceless consonant. Pouplier, et al. reported complementary results when examining accommodation in simplified consonant clusters. Not only was accommodation inconsistently applied (28% of errors failed to demonstrate accommodation), but the resulting acoustics also displayed a gradient pattern. Gormley reported similar findings with accommodation of vowel length in the context of incorrect voicing of a following consonant (40.7%of errors were unaccommodated, though 51% of errors were not able to be classified). Finally, Mowrey and MacKay used electromyography recordings of muscle movement to examine errors, specifically errors made on (s) in tongue twisters like "she sells seashells." Of the 48 errors observed, 43 displayed a continuous change across features (primarily the intrusion of the labial feature from /ʃ/) while only 5 displayed a categorical phoneme deletion. Thus, gradient errors were not only common but in fact greatly outnumbered the categorical errors. The authors further claim that these errors were not categorical feature changes, but rather should be characterized as subfeatural.

To summarize, the evidence is mixed and the degree to which phonemes, as holistic non-decompositional representations, drive segmental encoding in language production is still an open question. However, one thing is clear: Speech errors are a powerful tool to shed light on this question. The challenge is that neurotypical adult speakers do not produce many phonological errors unless induced to do so through artificial constraints, such as high phonemic similarity in tongue twisters, which may itself bias the nature of such errors. A complementary source of phonological errors is individuals with aphasia. In the next section, we briefly review what is known about speech errors in this population and propose a novel method for examining the nature of their sublexical errors.

## Analyzing aphasic errors

Reflecting the general organization of the language production system discussed in the previous section, aphasic errors fall under two general categories, lexical and sublexical. The general similarity in the pattern of errors between individuals with aphasia and neurotypical speakers has been the main argument for the "continuity hypothesis": The idea that aphasic errors reflect the same core processes as non-aphasic errors, only in much greater quantities [7]. This, in turn, implies that aphasic errors can be a great asset for studying the architecture of the language production system.

In aphasia research, sublexical errors have often been viewed as a miss election of phonemes. For example, in the most influential computational model of aphasic word production, the interactive two step model, phonological errors arise from the noisy mapping of lexical items on to phonemes. This idea has gained traction with the demonstration that the strength of such lexical to phonological mapping determines both the accuracy of

auditory word repetition and the number of phonological errors in naming in people with aphasia [8]. These models do not contain subphonemic representations such as phonetic features. Therefore, the representations of phonemes /b/ and /p/ are no more similar than the representations of phonemes /b/ and /s/, despite the much greater overlap in the phonetic representation of the former pair. This would not be a problem, if sublexical errors are indeed largely independent of phonetic features.

However, there is evidence that some sublexical errors produced by individuals with aphasia do interface with phonetic features. For example, Buchwald and Miozzo (2011, 2012) described two speakers diagnosed with non-fluent aphasia and apraxia of speech (participant D.L.E. and participant H.F.L.). Both individuals tended to simplify word-initial consonant clusters (e.g., /sp/ to /p/). In Buchwald and Miozzo, D.L.E. showed a robust pattern of accommodation such that a voiceless stop became aspirated after the initial fricative was deleted, abiding by the phonotactic rules of English. However, H.F.L. did not demonstrate accommodation, and the new onset remained unaspirated. In Buchwald and Miozzo, the authors examined the simplification of consonant clusters made up of /s/ nasal onsets and nasal singletons (e.g., 'small' and 'mall'). Prototypically, nasal singletons are produced with a longer duration compared to nasals in a cluster. D.L.E produced simplified clusters with a longer duration, similar to how the nasal would be produced in a nasal singleton. H.F.L. produced the shorted duration expected from a consonant cluster. Overall, D.L.E. produced a pattern of errors that would be expected if the errors were generated at the phonological level, while H.F.L. produced a pattern of errors that would be expected if the errors were generated during motor planning. The results across both papers support a double origin of sublexical errors in aphasia, one occurring at the level of phonemes and one at the level of context specific features.

Errors of the kind examined in Buchwald and Miozzo, although extremely useful in probing the origins of segmental errors, constitute a very small proportion of aphasic errors. It thus remains an open question: Do the majority of segmental errors reflect a problem at the phonological or the phonetic level? To answer this question, we propose a new approach that critically hinges on the deficits of working memory. A subset of individuals with aphasia show sublexical errors that reflect a specific impairment of phonological working memory. Phonological working memory is important in language production because selected linguistic representations (e.g., phonemes) must be stored in the correct sequence in order to be produced correctly. Individuals with phonological or graphemic working memory deficit typically show a length effect (more sublexical errors on longer words) and a decreased phonological working memory capacity measured independently [9]. Most studies of phonological working memory in aphasia tacitly assume phonemes as the to be remembered units. This is aligned with older views of working memory as a fixed capacity system with a number of "slots" to hold the units of operation. If the number of units exceeded the number of slots, errors are likely. In the current context, an individual with aphasia may have a decreased working memory capacity of 3 units. The slot based theory would predict that this person would be fine producing 3-phoneme words, but as soon as the number of phonemes exceeds 3, the additional phonemes will no longer be represented in working memory and are thus pulled randomly. In contrast to slot based models, resource based models do not propose a concrete cap on the number of items represented in memory. Rather, items in memory pull from a shared set of resources; as the set size increases, each item receives less of the resource. This mechanism introduces a new notion, *i.e.,* the precision with which an item is remembered. In the same person described above, additional phonemes would still be represented in working memory albeit in a less precise manner.

How can we define "precision" for working memory representations in language production? An analogy from visual processing may help here: Instead of asking participants to recall a color patch as being either "blue" or "green", we could present them with a color spectrum between blue and green and ask them to pick the exact color. The closer their pick to the original color patch, the more "precise" their choice. Using this technique, Wilken and Ma showed that as the number of color patches increased, participants precision in recalling the exact colors decreased. The same logic can be applied to sounds. Hepner and Nozari played participants a set of sounds that varied continuously between two phoneme endpoints (e.g.,/ ga/-/ka/ or /ra-la/). At the end of the trial, the participant marked along a slider where a particular sound was located between the two endpoints. Analogously to vision, the deviation between the true location of the sound and the remembered location continuously increased as the set size increased, reflecting a reduction in precision. This increase was evident even when the set size changed from a single sound to two sounds, supporting the idea that a shared resource underlies this system rather than a definite number of slots. Thus, phonological working memory is subject to the assumptions of a resource based model of working memory: As the number of remembered items increases, there should be a corresponding decrease in the precision of the remembered representations.

## The current study

The current study uses the notion of precision in resource models of working memory to examine the nature of sublexical errors in aphasia. Depending on whether sublexical errors represent non-decomposable units (phonemes) or a bundle of phonetic features, different patterns can be expected in individuals with a phonological working memory deficit. Specifically, the concept of precision does not apply to a discrete unit such as a phoneme. A /b/ should be equally likely to be substituted with a /p/ as with a /s/, as long as phonotactic constraints are not violated [10]. On the other hand, precision is well-defined and easily measured for units composed of phonetic features: a /b/ is closer to a /p/ than a /s/, because it shares more phonetic features with the former than the latter. By this logic, an increase in the working memory load, defined here as the word length, should systematically increase the deviation between the target and the error if the unit is a bundle of features, but no such systematicity should be observed if the unit is a non-decompositional phoneme.

To measure the deviation between the target and error, we used distance in the phonetic space, measured as the ALINE distance [11]. Briefly, the ALINE algorithm takes two strings of phonemes a target (e.g., "toothbrush", /tuθbɹʌʃ/) and a response (e.g., /puftʌs/) and finds an alignment that minimizes the total cost of the edit operations (*i.e.,* the sequence of deletions, insertions, and substitutions) needed to get from the target to the response. The ALINE distance between each pair of aligned phonemes is then calculated, which is the sum of the differences in their features, weighted by the perceptual salience of those features.

Figure 1 shows the alignment selected for /tuθbɹʌʃ/→/puftʌs/ and the ALINE distances between the aligned phonemes. As can be seen, the distance score is lower when the target and error are closer in the phonetic space (/θ/and/f/) and larger when they are farther apart (e.g., /b/and/t/).



**Figure 1:** An example of an alignment of a target and a response selected by the ALINE algorithm, with the ALINE distance between each pair of aligned phonemes shown below. In this example, the response/puftʌs/ was produced for the target /tuθbɹʌʃ/ ("toothbrush").

**The prediction is then straightforward:** If sublexical errors are truly phonemic with minimal influence of phonetic features, there should be no systematic relationship between length and ALINE distances. If, on the other hand, such errors are phonetic in nature, or represent units that are readily decomposed into phonetic features, we would expect an increase in ALINE distances as word length increases. To test these predictions, we selected four individuals with phonological working memory deficit and tested them on a large picture naming task [12]. We then computed ALINE distances for 1000+ sublexical errors and investigated the relationship between word length and such distances.

Observing a correlation between length and ALINE distance would point to the sensitivity of sublexical errors to distributed phonetic features. But what is the nature of such features? One possibility is that phonetic features are put together during segmental encoding before motor planning. Alternatively, phonetic features may be put together during the motor planning itself. To distinguish between the two, we take advantage of the phenomenon of articulatory simplification, often observed in deficits of motor planning, such as speech apraxia. Articulatory simplification refers to changing certain sounds to others in a manner that reduces articulatory complexity and effort, e.g., increasing the degree of constriction. Galluzzi, et al. demonstrated that among a group of people with aphasia, those who were additionally apraxia showed significantly more simplifications than those who primarily showed phonological deficits (aphasia without apraxia). This is in line with the broader literature showing that individuals with apraxia generally make errors on marked phonemes, resulting in a subsequently unmarked production [13]. Thus, it is possible that continuous changes in precision may be the result of simplification errors. If that is this case, errors should display directional changes, and such changes should increase with length. Otherwise, the results can be taken to support a decompositional view of phonemes, in which sublexical errors arise from the miss election of phonetic features during phonological assembly but before motor planning.

## Materials and Methods

### Participants

Four native English speaking individuals with post stroke chronic aphasia (1 female, 3 males; ages 32–64 years) participated in the study. Three were recruited from the Snyder center for aphasia life enhancement, while one (P2) was recruited from the Western Pennsylvania Patient Registry (WPPR). The basic attributes of the participants are described in Table 1. All participants were consented under protocols approved by the Institutional Review Board (IRB) for Johns Hopkins university or Carnegie Mellon university and received monetary compensation for their participation.

**Table 1:** Participants demographic information.

| Participant | Age | Gender | Premorbid handedness | Education | Years handedness |
|---|---|---|---|---|---|
| P1 | 59 | Male | Right | High school | 7 |
| P2 | 47 | Male | Right | Unknown | 16 |
| P3 | 32 | Female | Right | Associates | 3 |
| P4 | 64 | Male | Right | High school | 2 |

### Background tests and inclusion criteria

**Participants were selected to have the following inclusion criteria:** (a) Good semantic lexical comprehension so they would be able to follow task instructions and have a strong knowledge of labels for concepts, (b) Impaired phonological processing, so they would produce the data suitable for the planned analyses, and (c) Evidence of a phonological working memory problem, so they would be sensitive to the manipulation of working memory load.

Semantic lexical abilities were measured with two word to picture matching tasks, one easy (175 pictures from the Philadelphia naming test presented as target with unrelated foils) and one more difficult (30 pictures presented as targets along with semantically and phonologically related foils), previously normed and used [14]. Accuracy was above 98% for each participant in the easy task (Except P2 for whom this score

was unavailable) and above 77% in the difficult task, indicating good semantic lexical abilities (Table 2).

Phonological processing was measured by the number of lexical or nonlexical errors that were phonologically related to the target in picture naming using the PNT and in auditory word repetition using the repetition version of the PNT, the Philadelphia Repetition Test (PRT). Participants overall production abilities ranged from 39-71% accurate on the PNT and slightly higher, 47-89% on the PRT. Importantly, in each of the four participants, at least 25% and 60% of the commission errors in naming and repetition, respectively, were phonologically related to the target, marking some level of impairment in phonological processing (Table 2).

**Table 2:** Results of the background tests used as inclusion criteria. See text for the description of tests.

| | Semantic lexical comprehension | | PNT | | PRT | | Phonological working memory |
|---|---|---|---|---|---|---|---|
| Participant | Accuracy on the easy task | Accuracy on the difficult task | Overall accuracy | Phonological errors (% out of commission) | Overall accuracy | Phonological error count (% out of commission errors) | Rhyme probe scores |
| P1 | 100% | 100% | 41% | 82 (81%) | 47% | 87 (95%) | 2.58 |
| P2 | - | 77% | 39% | 29 (31%) | 50% | 85 (98%) | 1.58 |
| P3 | 100% | 93% | 71% | 11 (24%) | 89% | 13 (77%) | 0.5 |
| P4 | 98% | 90% | 49% | 35 (47%) | 65% | 39 (64%) | 1.41 |

Finally, phonological working memory was assessed using a modified version of the rhyme probe task [15]. In this task, participants heard a list of words followed by a test probe. They indicated whether or not the test probe rhymed with any of the words in the proceeding list. The score is a rough proxy for how many items can be kept in phonological working memory. Participants' performance on the rhyme probe task ranged between 0.5 and 2.58, indicating that, on average, the participants could not hold more than 3 items in phonological working memory. Table 2 presents the full results of the background tests.

## Material and Procedures

A large scale picture naming task was used [16]. The word set consisted of 444 items, comprised of color photographs obtained from online repositories. The entire set was administered twice, in two different pseudorandom orders, for a total of 888 trials. To reduce the semantic blocking effect, items from the same semantic category were at least 12 items apart.

Participants were given 20 seconds to name each picture and no feedback was given, except if the participant misidentified the target word (e.g., the participant responds "hand" to a

picture of a finger, and the experimenter prompts: "What part?"). Testing was broken into as many sessions as was needed to complete the task. Sessions were recorded and saved for offline transcription.

### Segmental error coding and ALINE distance calculation

All targets, along with the first lexical response on each trial, were transcribed into IPA from the recorded audio. If the participant committed a segmental error on an identifiable semantic error, the target was considered to be the semantic error. For example, if in response to the picture of an orange, the participant responded/æbl/, the target was considered to be the semantically related "apple". Additionally, any attempt on which the target word could not be determined was removed (9.9%). The data was transcribed and coded by two independent coders (k=.88). All transcriptions and codings were checked and discrepancies were reconciled. A third coder double checked the final codes for consistency. All coders were native speakers of American English.

To measure the deviation (and thus precision) of the representations stored in phonological working memory, we

used the ALINE distance between the aligned consonants. The phonemes in each pair of target and response transcriptions were aligned using PyALINE, an implementation of the ALINE algorithm in Python. The ALINE algorithm consists of two sub operations: One to calculate the distance between the target and the response phonemes and one that uses the distance measure to generate the optimal alignment. The constants used in these calculations can be found in the difference score is determined by the sum of the differences between the feature vectors of the target and response phonemes. These vectors consist of multivalued features (for place, manner, height, and back) and binary features (for nasal, retroflex, syllabic, long, lateral, aspirated, round, and voice.) For instance, the place feature is coded as 1.0 if the phoneme is bilabial and if glottal, with intermediate values assigned to phonemes with places of articulation in the middle of the mouth. Each feature difference is weighted by the salience of that feature. For example, the place feature is more salient than the aspiration feature and is thus considered to a greater degree when differences are summed. Different sets of features are used for vowels (syllabic, nasal, retroflex, high, back, round, and long) and consonants (syllabic, manner, voice, nasal, retroflex, lateral, aspirated, and place.) If one of the phonemes is a vowel and the other a consonant, the consonant features are used. The equation for calculating phoneme distance is as follows:

$$distance_{target,response} = \sum_{f \in features} |(target_f - response_f)| * salience_f$$

The distance metric is then used to calculate the optimal alignment of the target and response phoneme sequences. Traversing incrementally across both sequences, all possible phoneme substitutions, expansions, or skips are considered by subtracting the distance measure between the phonemes from a base value for a potential substitution, expansion, or skip, resulting in a matrix of possible alignments in which higher values indicate better alignments. The optimal alignment is then selected by recursively searching the matrix, following the path of minimal alignment 'cost'. The precise algorithms used to determine the alignment can be found in the OSF repository and are detailed in Kondrak. The final ALINE distance metric is computed based on the optimally aligned phonemes of the target and response, using the above distance equation.

The resulting distance metric can be thought of as a measure phonetic precision, where a larger distance corresponds to a less precise representation of the target. If no error was made (*i.e.,* the target and response phonemes are the same), the distance is 0. Figure 1 illustrates an example of an aligned target and response word with the corresponding distances between aligned phoneme pairs.

**Articulatory simplification coding:** Articulatory simplification could be observed at multiple levels. For example, at the syllabic level, a consonant cluster may be simplified into a single consonant. In the current analysis, phonological simplifications (in which a phoneme becomes less marked) will be examined. Galluzzi, et al. identified four dimensions along which phonemes become less marked: voicing (consonants tend to become voiceless), manner (consonants tend to become more constricted; e.g., fricative changing to a stop), place (consonants tend to move further forward in the mouth e.g., an alveolar changing to a bilabial), and trills (tend to become /l/). Voicing and trills are less relevant dimensions of simplification in English (participants in Galluzzi, et al. spoke Italian, which, unlike English, requires native VOT and trills) and so only manner and place will be assessed in the current study.

For place, if the place of articulation for the response phoneme was further forward than the target phoneme, the error was coded as 1. If the response was further back the error was coded as -1. If there was no change in place the error was coded as 0. A similar coding scheme was used for manner where the error was coded as 1 if the response was more constricted,-1 if the response was less constricted, and 0 if there was no change in manner.

## Statistical analysis

Unless stated otherwise, analyses were conducted using linear mixed effects models with the imer test package in R [17]. In the main analyses, ALINE distance and word length were the dependent and independent variable, respectively. The log of the ALINE distance for each phoneme in a word was used so that the distance distribution approximates normality. Word length was determined as the number of phonemes in a word and was centered and scaled. To ensure that the effect of length of ALINE distance was not altered by the phoneme's position in the word, the analyses were repeated with a control variable, the relative position of a phoneme to the word boundaries, henceforth referred to simply as "position". The motivation for including this variable comes from a large literature on positional effects in tasks that require serial recall from memory. In such tasks, items at the beginning and the end of a list are typically remembered better than items in the middle positions, creating a U shaped distribution [18]. A similar effect is observed in orthographic tasks, where accuracy in usually highest for the first and last letters than the middle letters in a word, mimicking the same U shaped distribution and showing the clear application of memory principles to segmental encoding. Position was calculated as the distance to the nearest boundary (*i.e.,* number of phonemes to the start or end of the word, whichever is closest) divided by the total length of the word. For example, in /tɹʌmpət/, the /p/ phoneme is two positions away from the end of the word (versus four positions from the start) and the word length is seven, so it's relative to a boundary are 29. Onset and coda /t/ each have a distance of 0 to the nearest edge, and so on and so forth, capturing the U shaped distribution referred to above. Finally, the random effect structure was tailored to the model: In the aggregate analyses, both the random intercept of items and subjects were included. In the individual analysis, only the random intercept of items was included.

The effect of length on place and manner simplification was examined using the same linear mixed effects models, except with the dummy coded place and manner changes as the dependent variable. The 'clmm' function in 'ordinal' package was used in order to perform linear mixed effects models on ordinal data. The models were fit to the manner and place

simplifications separately, and p values were corrected for multiple comparisons using the Bonferroni method.

# Results

A total of 1798 phonemes with errors were obtained for

analysis. Table 3 displays the distribution of words and errors that were used to calculate ALINE distances.

**Table 3:** The number of words and errors elicited from each participant used to calculate ALINE distances.

| Participant | Words | Words with an error | Phonemes with an error |
|---|---|---|---|
| P1 | 766 | 485 | 875 |
| P2 | 1157 | 354 | 516 |
| P3 | 794 | 61 | 76 |
| P4 | 881 | 221 | 331 |

## Effect of word length on ALINE distance

First, we examined the effect of world length on ALINE distance to determine if increasing the number of phonemes in working memory resulted in greater distance between errors and targets (Table 4). Figure 2 illustrates the average ALINE distance calculated over all errors for the four participants, which grows monotonically as a function of word length.
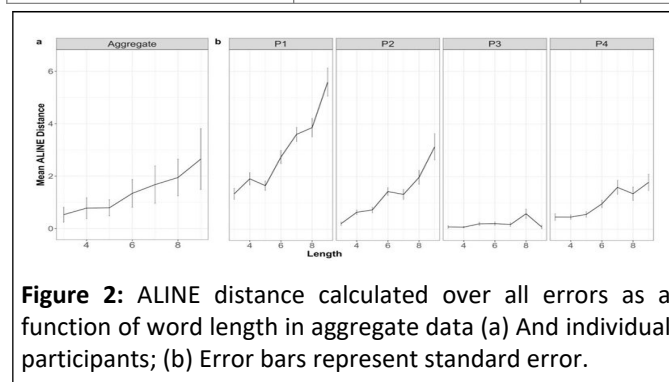
The first pass model with only word length as the independent variable showed a significant increase in ALINE distance as a function of increasing word length ($\beta$=0.171, t=14.60, p<.001). The second pass model added the control variable position and its interaction with length. The effect of length remained significant in this model ($\beta$=0.17, t=14.80, p<.001). Additionally, there was a significant effect of position

($\beta$=-0.02, t=-3.19, p <.001), and an interaction between length and position ($\beta$=0.03, t=4.05, p<.001). This interaction shows that the impact of length on ALINE score increases the farther the phoneme is from the edges.

To confirm this pattern was not an artifact of aggregating the data, the same analysis was carried out on each individual. In the first pass model, the effect of word length on ALINE distance was significant in three and marginal in the fourth participant, showing a common trend in all participants. The inclusion of the position variable and its interaction with length in the second pass model did not change this pattern.

**Table 4:** The effect of length on ALINE distance for individual participants over all errors.

| Participant | Coefficient | SE | t | p value |
|---|---|---|---|---|
| P1 | 0.21 | 0.02 | 10.99 | <.001 |
| P2 | 0.19 | 0.02 | 9.03 | <.001 |
| P3 | 0.05 | 0.03 | 1.97 | 0.05 |
| P4 | 0.16 | 0.02 | 6.69 | <.001 |



**Figure 2:** ALINE distance calculated over all errors as a function of word length in aggregate data (a) And individual participants; (b) Error bars represent standard error.

To summarize, group-level analyses showed that increased word length led to higher ALINE scores. This pattern was also obvious in individuals data, supporting the idea that increased working memory load led to a greater distance between the error and the target in the phonetic space. This finding, in turn, suggests that segmental errors are sensitive to units smaller than phonemes. The changes in the sub-phonemic properties of errors in longer words may have two sources: (1) The decreased precision of representations in working memory when the load increases, or (2) Articulatory simplification when a longer word

must be executed. If the first source, we would not expect a shift from certain articulatory features to other features in a systematic manner. However, if the errors result from articulatory difficulty, we would expect a shift in a certain direction, *i.e.,* in a way that makes articulation simpler. The next set of analyses tests the second possibility.

## Errors as a result of articulatory simplification

Figure 3 shows the mean of simplifications for the place and manner of articulation for each participant. A mean greater than zero suggests that the participant simplified the articulation more often than the reverse. One sample t-tests were used to determine significance. P1 and P2, but not P3 or P4, showed statistically significant fronting of their responses (t(633)=6.08, p<.001; t(350)=6.08, p<.001, respectively), which remained significant after correcting for comparisons across the four participants (α=.0125). None of the four participants showed a significant simplification effect on the manner of articulation after correcting for multiple comparisons.

Since P1 and P2 both showed a significant place simplification, separate linear mixed effect models with length as a fixed effect were applied to assess whether the length effect observed above can be explained by their tendency for simplification. After correcting for two comparisons (α=.025), P1 did show a significant effect of length on place simplification (β=0.24, t=2.71, p=.007), but P2 did not (β=-0.04, t=-0.23, p=.82). Collectively, these results suggest that, while articulatory simplification may explain the length effect in some individuals with aphasia, it is not the only, or the main, source of the observed length effect.
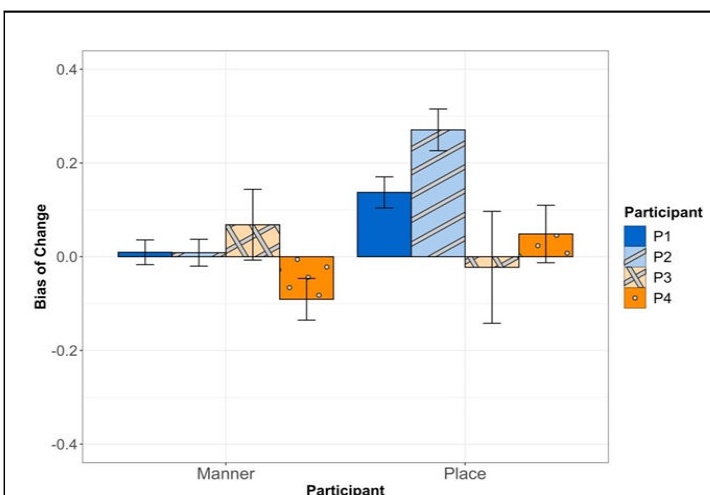


**Figure 3:** The bias of change toward simpler articulations for consonant errors, split by manner simplifications (the constriction in the mouth becoming less open) and place simplifications (the constriction occurring further forward in the mouth.) Positive means indicate a trend toward simplification, negative means indicate a trend away from simplification. Some participants tend to simplify the articulations necessary to make a sound, resulting in a change from a marked articulation.

## Discussion

We utilized sublexical errors in conjunction with modern theories of phonological working memory to address a classic question in the language production literature: To what degree do phonetic features, as decompositional representations, drive segmental encoding? We found that increased working memory load resulted in a systematic decrease in precision of segmental encoding indexed by an increase in the distance between the phonetic features of the target and the error, suggesting that phonetic features are critically involved in segmental encoding.

## Implications for models of speech production

These results dovetail nicely with previous findings implicating phonetic features in the generation of sublexical errors. For example, phonemes that share many phonetic features are more likely to interact during sublexical errors than more distance phonemes (*i.e.,* the phonetic similarity effect; Shattuck-Hufnagel and Klatt). This suggests that phonetic features are involved in encoding at the processing point at which the error occurred [19]. Further, the prevalence of gradient errors supports the notion that decompositional phonetic features are involved in the generation of sublexical errors. As such, these results are at odds with models that posit a purely categorical view of sublexical errors [20]. In such a model, phonetic features are only incorporated later in the speech production process, when speech is executed. However, errors that result from the miss election of non-decompositional phonemes would not be reflected as systematic change in phonetic space. Rather, a categorical change in phoneme would result in sporadic changes in phonetic space. Thus, the current results refute a strict categorical model.

One interpretation of the current findings is that phonetic features alone drive segmental encoding. However, this view is inconsistent with the evidence reviewed earlier in this paper the provides support for the psychological reality of phonemes. For example, Oppenheim and Dell demonstrated that representations in inner speech did not interface with phonetic features, as sublexical errors in inner speech did not abide by the phonetic similarity bias. Additionally, the occurrence of accommodated errors (though less prevalent than previously believed) does provide evidence for errors that originate before the application of context specific phonetic rules. It is thus very unlikely that these findings negate the existence of phonemes as computationally relevant units.

The second interpretation of the current results is that errors may stem from different layers of the sublexical system, with phonemes and phonetic features as two independent sources. For example, Frisch and Wright examined speech errors elicited by a tongue twister task, where it was likely that subjects would make errors on /s/ and /z/. Degree of voicing was examined using acoustic analysis. While many of the errors could be explained as gradient variation driven by competition of phonological representations, the rate of categorical shifts between voiceless to voiced was too high to be explained away as extreme instances of continuous change. Further, as discussed earlier, Buchwald and Miozzo presented two

individuals with aphasia whose speech errors arose from different levels of representation, either the phonological level or the motor planning level. These results point to a system in which errors can arise from multiple sources both as categorical changes at the phonological level and as gradient changes at lower levels. The results of the current study can be accommodated within this view: Although all four individuals in this study demonstrated the length effect compatible with the involvement of phonetic features, it is possible that there are other individuals who do not demonstrate the effect. It is also true that within the studied individuals, the effect was not observed in every sublexical error. One limitation here is that transcription may distort what was actually said. Double transcription alleviates this problem to some degree, but the technique remains noisier than acoustic measurements of speech. Another possible limitation is that targets are not always clearly identifiable. Care was taken to identify the most plausible targets for the analysis, and when this was not possible, to exclude the trial. Although this practice does not entirely exclude the possibility of target misidentification, it substantially reduces its probability. It is thus possible that miss election has truly happened at different levels of processing, even within the same individual.

The third and most parsimonious account of the current data is a system with cascading and interactivity between phonemes and phonetic features. In such models, while the phonological level remains the primary source of sublexical errors, it is continuously influenced by lower level representations *via* feedback. Figure 4 shows a schematic of such a system. When a given phoneme, e.g., /k/ is activated *via* the lemma "cat", it activates its articulatory phonetic features, e.g., velar and stop, *via* cascading activation. Those features in turn activate other relevant phonemes. In this case, both features also converge on phoneme/g/, while only the stop feature activates /d/. Thus /g/ which is closer in phonetic space to /k/ due to more shared features gains greater activation and is more likely to be erroneously selected for production than the more distant phoneme/d/. This dynamic naturally explains the phonetic similarity effect, explained earlier. It further explains the current findings: When working memory load is limited, each segment receives more resources and can be maintained with greater activation compared to when many phonemes must be simultaneously maintained in working memory. This means that under low working memory load, the selected segment is likely to be either correct, or if not, systematically related to the target through the activation of shared features, *i.e.,* closer in the articulatory phonetic space. Under high working memory load, on the other hand, each segment receives less resources and the weak activation of phonemes dissipates too quickly for the process of cascading and feedback to effectively activate a near phoneme. Therefore, if a phoneme is lost, it is more likely to be substituted by a random phoneme, one that has perseverated from before, or one that is better supported by context (e.g., intervocalic consonants may be more likely to be erroneously voiced, taking up the voiced feature of the surrounding vowels) without this process being systematically related to the distance between the target and error in the articulatory phonetic space.
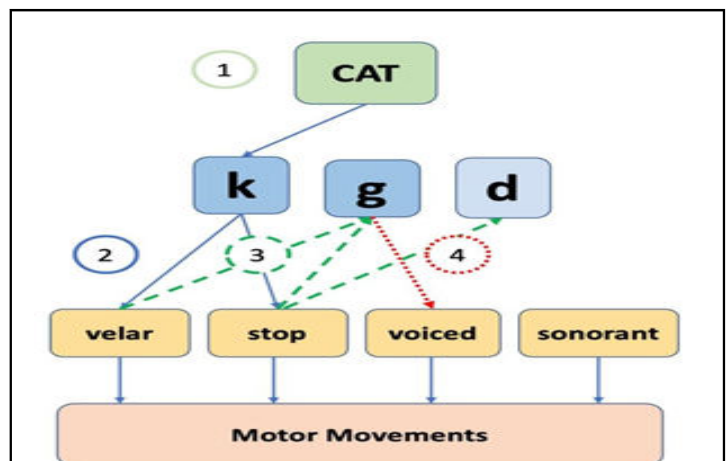


**Figure 4:** The mechanism for producing the onset /g/ for target /k/ with greater probability than the onset/d/ in a system with cascading and interactivity. (1) Activation of the 'Cat' lexical item, (2) Feed forward activation of phonetic features *via* the phoneme /k/, (3) Feedback activation of the phonetically related phoneme /g/ as well as the competitor phoneme /d/. While the /d/ phoneme is partially activated by the 'stop' feature, the /g/ phoneme is activated more strongly because it shares more phonemes with /k/. (4) Feed forward activation of the phonetic features of /g/, including the erroneous 'voiced' feature, leading to the production of /g/ instead of /k/.

## Phonetic features as representations

The first set of analysis showed that phonetic features are involved in the generation of sublexical errors. We argued that these results are best accommodated within a model in which phonemes and phonetic features interact. But what is the nature of phonetic feature representations? One possibility is that phonetic features are motor features and as such are directly involved in motor planning. Another possibility is that these are pre motor representations. In order to distinguish between these accounts, we looked to the apraxia literature. Individuals with apraxia have a known articulatory motor planning deficit. Critically, they tend to simplify their articulations due to difficulties with articulatory planning. This property can be used as a blueprint for what to expect if phonetic features are fundamentally motor representations: if changes in the phonetic features reflect something about motor planning, they should be in the direction of motor simplification.

Errors elicited in the picture naming task were analyzed to see if any of the subjects tended to simplify their productions. Only two subjects tended to simplify their productions, and only one simplified their articulations more often as the length of the word increased. This pattern supports the basic assumption that certain representations involved in motor planning do change systematically as a function of length, but also demonstrates that motoric simplification cannot explain the pattern of errors seen across all four subjects in picture naming task. In at least in three of the four participants, phonetic features drove the effect independently of motor planning units. This finding motivates a separation between phonetic features and motor planning units

as levels of representations in the language production system.

## A useful tool for future research

The novelty of this study is in utilizing a new tool to address an old question. The ALINE distance and its sensitivity to working memory load (*i.e.,* length) provides a powerful tool for large scale analysis of sublexical errors in the phonetic space. While we presented an example of such an application in aphasia, the tool can be used in other populations as well. The interactive model that we endorsed in the current study makes testable predictions regarding when a length effect on articulatory phonetic distance should or should not be obtained. For instance, situations that significantly reduce the time necessary for feedback (e.g., by imposing short response deadlines; e.g., Dell) should eliminate the length effect on phonetic distance. Such a prediction leaves open an avenue for the extension of the current method to speech errors elicited from neurotypical speakers. The tool can be similarly used to analyze other aspects of the system, such as the influence of semantic similarity on sublexical errors in the articulatory phonetic space.

## Conclusion

This study aimed to assess the origin of sublexical errors produced by individuals with aphasia in light of modern theories of working memory and using a tool from computational linguistics. We found that sublexical errors are sensitive to units smaller than phonemes, suggesting that subphonemic representations play a critical role in the emergence of sublexical speech errors, and more generally, the process of segmental encoding. We further showed that this unit is pre motor. Collectively these findings support a system with cascading and interactivity between phonemes and pre motor phonetic representations.

## Acknowledgments

## References

1. Ades AE (1974) How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception and Psychophysics 16:61–66

2. Atkinson RC, Shiffrin RM (1971) The control of short term memory. Sci Am 225:82–91

3. Baars BJ, Motley MT, MacKay DG (1975) Output editing for lexical status in artificially elicited slips of the tongue. J Verbal Learn Verbal Behav 14:382–391

4. Buchwald A, Miozzo M (2011) Finding levels of abstraction in speech production: Evidence from sound production impairment. Psychol Sci 22:1113–1119

5. Buchwald A, Miozzo M (2012) Phonological and motor errors in individuals with acquired sound production impairment. J Speech Lang Hear Res 55: 1573-1586

6. Buchwald A, Rapp B (2006) Consonants and vowels in orthographic representations. Cogn Neuropsychol 23:308–337

7. Buchwald A, Rapp B (2009) Distinctions between orthographic long-term memory and working memory. Cogn Neuropsychol 26:724–751

8. Caramazza A, Miceli G, Villa G (1986) The role of the (output) phonological buffer in reading, writing, and repetition. Cogn Neuropsychol 3:37–76

9. Caramazza A, Miceli G, Villa G, Romani C (1987) The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. Cogn 26:59–85

10. Cera ML, Ortiz KZ (2010) Phonological analysis of substitution errors of patients with apraxia of speech. Dement Neuropsychol 4:58–62

11. Corley M, Brocklehurst PH, Moat HS (2011) Error biases in inner and overt speech: Evidence from tongue twisters. J Exp Psychol Learn Mem Cogn 37:162–175

12. Cowan N (2001) The magical number 4 in short term memory: A reconsideration of mental storage capacity. Behav Brain Sci 24:87–114

13. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA (1997) Lexical access in aphasic and nonaphasic speakers. Psychol Rev 104:801–838

14. Eimas PD, Corbit JD (1973) Selective adaptation of linguistic feature detectors. Cogn Psychol 4:99–109

15. Freedman ML, Martin RC (2001) Dissociable components of short-term memory and their relation to long-term learning. Cogn Neuropsychol 18:193–226

16. Frisch SA, Wright R (2002) The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. J Phon 30:139–162

17. Fromkin VA (1971) The non-anomalous nature of anomalous utterances. Language 47:27–52

18. Galluzzi C, Bureca I, Guariglia C, Romani C (2015) Phonological simplifications, apraxia of speech and the interaction between phonological and phonetic processing. Neuropsychologia 71:64–83

19. Gathercole SE, Willis CS, Baddeley AD, Emslie H (1994) The children's test of nonword repetition: A test of phonological working memory. Memory 2:103–127

20. Goldrick M, Blumstein SE (2006) Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. Lang Cogn Process 21:649–683