Acoustic Characteristics of Speech Error Repairs

Chelsea Sanker
Stanford University

This study examines the acoustic characteristics of error repairs within natural speech in English. Speech error repairs provide a line of evidence for the effects of competitors on lexical and phonological planning and subsequent phonetic aspects of speech. Some work suggests that the existence of salient lexical competitors results in words being phonetically more distinct from those competitors, at least in certain tasks. However, the patterns within speech error repairs seem to suggest prosodic emphasis rather than contrastive hyperarticulation. Repairs have higher F0 and higher intensity than fluent productions of the same words, with mixed evidence for duration effects. There was no evidence for the acoustic characteristics of speech error repairs increasing the acoustic differentiation between a word and its competitors. I argue that speech error repairs receive a distinctive type of emphasis, with relatively limited duration effects in order to differentiate the repair from the preceding disfluent slowdown.

Keywords: speech errors, repairs, acoustic correlates, prosodic prominence

1 Introduction

How do planning and lexical competition impact phonetic characteristics of words? Some work suggests that the existence of salient minimal pairs results in "contrastive hyperarticulation", in which words are produced to be more phonetically distinct from competitors than they would be based on the typical realization of their component sounds. One proposal for why these patterns exist is that inhibiting lexical competitors inhibits the phonetic characteristics associated with the phonemes present in those competitors.

1.1 *Contrastive hyperarticulation* There is a body of work which finds that the existence of lexical competitors impacts the phonetic realization of words. One of the main lines of evidence comes from comparing words which have minimal pair competitors to words which do not have such competitors. Many studies find that acoustic correlates of phonological contrasts are larger in words with a minimal pair for that contrast, e.g. the voice onset time (VOT) for the initial stop is longer in *pie* (cf. *buy*) than in *pipe* (e.g. Baese-Berk & Goldrick 2009, Wedel et al 2018). Studies also often find that making the minimal pair salient within the task can increase differences among words which have minimal pair competitors (e.g. Schertz 2013, Seyfarth et al 2016).

There are a few different phonetic characteristics that have been examined in studies of contrastive hyperarticulation. The most frequently investigated phonetic measurement is VOT in word-initial long-lag VOT stops, which is generally found to be longer in words for which a short-lag competitor exists (Baese-Berk & Goldrick 2009, Goldrick et al 2013, Wedel et al 2018) or if the competitor has been made salient, among words or nonce words for which such competitors exist (Kirov & Wilson 2012, Schertz 2013, Buz et al 2016, Stern & Shaw 2023). Kirov and Wilson (2012) demonstrated that the effect of the competitor is specific to the phonological contrast; VOT in aspirated onsets is longer when the word is presented with an onset competitor rather than a competitor that differs in the vowel or the coda. However, Ohala (1994) did

* I would like to thank the audience at WECOL 2023 for their helpful questions and comments.

Proceedings of *WECOL 2023*

not find a significant effect of making the competitor salient.

Some of these studies also look at word-initial unaspirated stops in English. Some English speakers produce prevoicing for this category of stops, but many do not. Wedel et al (2018) found shorter positive VOT for prevoiced/short-lag VOT word-initial stops for which long-lag minimal pair competitors exist; they note that prevoiced stops were rare in their corpus. Schertz (2013) found that a salient aspirated competitor increased negative VOT in prevoiced/short-lag stops, caused in part by prevoicing being more common in this condition. Goldrick et al (2013) separately analyzed prevoiced and short-lag VOT realizations; they did not find a significant effect of minimal pair competitors for either realization of this category. Seyfarth et al (2016) looked at voicing in coda fricatives; voiced fricatives were voiced through more of their duration when a voiceless minimal pair competitor was salient.

Vowel duration as conditioned by coda voicing exhibits some effects of minimal pairs, though studies vary in whether minimal pairs seem to cause shortening before voiceless codas or lengthening before voiced codas. De Jong (2004) found that vowels in both coda voicing environments were longer under corrective focus, but that the size of the voicing-conditioned duration difference is larger than in non-focused contexts. Goldrick et al (2013) found that vowel duration was longer among words with minimal pair competitors for words with voiced coda stops, while there was no significant effect for words with voiceless coda stops. Seyfarth et al (2016) found that vowel duration was shorter before /s/ when a salient minimal pair competitor with /z/ was present, while there was no significant effect for words with /z/. Vowel duration differences associated with vowel quality have also been observed to increase when minimal pair competitors are made salient, e.g. lengthening of tense vowels with lax competitors (Schertz 2013), shortening of /ɛ/ and lengthening of /æ/ under corrective focus (de Jong 2004).

An effect of minimal pair competitors has also been observed for vowel formants. Vowels have sometimes been found to be more distinct – i.e. have greater Euclidean distance – in words that differ between those vowels, e.g. /ɛ/ versus /æ/ in *dead* (cf. *dad*) as compared to /ɛ/ in *chess* (Clopper & Tamati 2014, Wedel et al 2018). However, not all studies looking for such effects have found them, e.g. with corrective focus within a sentence (de Jong 2004) or repetitions correcting a word that was mis-identified by a listener (Ohala 1994).

1.2 *Mechanism of lexical competition* Some of these studies vary in their methodology. While there is no clear pattern in results differing based on task design, the factors influencing speakers' productions might differ depending on the task. Some studies used a natural speech corpus (e.g. Wedel et al 2018), looking at whether words have minimal pairs or not. Some studies elicited words in a reading task (e.g. Baese-Berk & Goldrick 2009 [Experiment 1], Goldrick et al 2013), similarly looking at whether words have minimal pairs or not. In some studies, target words were said for a human listener or speech recognition system, which then identified the word (e.g. Kirov & Wilson 2012, Baese-Berk & Goldrick 2009 [Experiment 1], Seyfarth et al 2016); these studies compare whether the minimal pair competitor was shown as one of the selection options or not, influencing the salience of that competitor. In some studies, the listener or speech recognition system asked for a repetition or asked for a clarification that indicated a misheard word, e.g. "did you say *beat*?" when the word was 'bit' (e.g. Ohala 1994, Schertz 2013, Buz et al 2016). De Jong (2004) had speakers produce the target word in a corrective focus context (e.g. "He said BAT, not BAD.").

It is also possible that there are deliberate listener-driven effects which differ from speaker-internal effects. While lexical competition might produce some automatic consequences in the process of lexical retrieval and subsequent production, deliberate attempts to produce a clear utterance for a listener are likely to involve different mechanisms. Such effects are most likely to be present in studies which make minimal pair competitors salient by having a listener indicate mishearing the utterance as that competitor.

It is also possible that the differences observed across different words are caused by other phonotactic or lexical characteristics that correlate with the existence of minimal pairs, particularly in a natural speech corpus where there might be additional effects due to the context. In studies that examine the same word in different contexts, such factors would be eliminated.

One of the potential characteristics correlated with the existence of minimal pairs is neighborhood density, i.e. the number of lexical neighbors that a word has. This form of lexical competition has been shown to have a range of effects, such as slower lexical access for words with higher neighborhood density (e.g. Vitevitch & Luce 1999, Luce & Large 2001). Slower access for words with more neighbors might predict less phonetic reduction, including the longer VOT in words like *pie*. The existing work on how the number of lexical competitors of a word impacts phonetic characteristics of speech production is generally consistent with that prediction. Most studies have found more vowel dispersion in words with more lexical neighbors (Munson & Solomon 2004, Scarborough 2010), though Gahl et al (2012) found less vowel dispersion in words with more lexical neighbors. Studies have also found longer VOT in long-lag stops for words with more lexical neighbors (Fox et al 2015, Fricke et al 2016). When separately controlling for neighborhood density and presence of a minimal pair, Fox et al (2015) found that only neighborhood density was a significant predictor of VOT, though on the other hand, Wedel et al (2018) included both as factors in their model and only found a significant effect of the existence of a minimal pair competitor. A more general effect of neighborhood density might suggest that the observed effects are not due to hyperarticulation of specific phonetic characteristics based on which cues are crucial for distinguishing between lexical competitors.

1.3 *Disfluencies and repairs* The studies described above look at lexical and phonological competition as it affects productions that are still within the intended phonological category. However, sometimes phonological competition results in errors. In an experiment using tongue twisters, Goldrick and Blumstein (2006) demonstrated that the target category influences acoustic characteristics of errors: VOT is longer than the typical short-lag VOT in a short-lag stop that was intended to be long-lag and shorter in a long-lag stop that was intended to be short-lag. Natural speech errors follow the same pattern – erroneous productions of short-lag VOT categories have longer VOT than fluent productions where the short-lag category was intended, and erroneous productions of long-lag VOT categories have shorter VOT than fluent productions (Alderete et al 2021). These patterns are explained as the target sound still being activated to some degree, even though the erroneous sound achieved higher activation, so both categories contribute to the motor plan for VOT. Gradient errors have also been observed in other contrasts, e.g. fricative voicing (Frisch & Wright 2002).

Based on the effects of competitors in speech errors, we might expect non-errorful speech to also be more similar to competitors. Baese-Berk and Goldrick (2009) argue that errors and non-errors differ substantially in degree of activation. In non-errorful speech, the target word has much higher activation than competitors, so they can rapidly be suppressed, whereas in errorful speech, the target word and the erroneous word both have high activation, and the target word is not fully suppressed.

Repairs of speech errors can be an informative category for examining the effect of competitors when the speaker produces the correct word but previously had strongly activated a competitor. This line of evidence may help in evaluating the mechanisms underlying some of the observed competitor effects. Cutler (1983) found that repairs of speech errors are prosodically marked, mentioning intensity, F0, and duration as characteristics, but without quantifying the patterns of each characteristic. Subsequent studies have reported higher pitch and higher intensity in repairs than in the erroneous elements preceding them (Nooteboom 2010, Carter and Plug 2014). Some explanations for why these acoustic characteristics exist focus on the role of the listener; the speaker emphasizes the repair because it contrasts semantically with the error, making the correction salient (Levelt & Cutler 1983). Nooteboom (2010) proposes that the prosodic marking is used to draw attention away from the error by emphasizing the correct word.

The prosodic marking of repairs might depend on the type of repair. Some work proposes that lexical errors are more likely to be prosodically marked, whereas phonological errors are often not marked at all (Cutler 1983, Shattuck-Hufnagel & Cutler 1999). Some work on contrastive hyperarticulation is consistent with lexical and phonological levels of planning producing different effects of competition, e.g. Stern and Shaw (2023) found a smaller effect of lexical competitors on VOT in nonce words than has been found previously in real words. However, some work on speech error repairs has found that lexical and phonological repairs are marked in the same way (e.g. Carter and Plug 2014). Some of the variation across studies might be due to ambiguity of categorizations, e.g. Shattuck-Hufnagel and Cutler (1999) found that phoneme-level errors resulting in real words exhibit patterns like phoneme-level errors that do not result in real words. There might also be further differences within these broader categories. Some work has found that prosodic marking is more common when the repair was a replacement with a more suitable word rather than replacing a word that was semantically entirely inaccurate (Levelt & Cutler 1983), which has been explained as higher informativity increasing the likelihood of prosodic prominence (Plug 2011).

1.4 *Emphasis* Do the effects of lexical competition necessarily reflect contrastive hyperarticulation, or could they be explained by a more general process of hyperarticulation or emphasis, not specific to the phonetic characteristics that distinguish a word from its minimal pair competitor? Much of the existing work which has found effects of minimal pair competitors was measuring phonetic characteristics associated with slower speech: increased VOT in long-lag stops (Baese-Berk & Goldrick 2009, Kirov & Wilson 2012, Schertz 2013, Goldrick et al 2013, Buz et al 2016, Wedel et al 2018, Stern & Shaw 2023), longer prevoicing in voiced stops (Schertz 2013), longer vowel duration before voiced codas (de Jong 2004, Goldrick et al 2013), longer vowel duration in tense vowels (Schertz 2013), and longer vowel duration in low vowels (de Jong 2004).

There are some studies which looked at characteristics that might be impacted by slower speech rate or "careful speech" styles despite not being direct measurements of overall duration. Some studies have found significant effects of minimal pair competitors for these characteristics, e.g. more dispersed vowel spaces (Clopper & Tamati 2014, Wedel et al 2018) and voicing through a longer percentage of voiced fricatives (Seyfarth et al 2016). However, some studies have looked at vowel formants and not found an effect (Ohala 1994, de Jong 2004).

The clearest evidence for contrastive hyperarticulation would be in characteristics that are shortened to increase contrastiveness when a minimal pair competitor exists or is made more salient. A few studies have found significant effects in this direction: Shorter positive VOT for short-lag stops (Wedel et al 2018), shorter vowel duration before voiceless codas (Seyfarth et al 2016), and shorter vowel duration in /ɛ/ when contrasted with /æ/ (de Jong 2004). However, there are also studies which looked for these effects and failed to find them: No VOT shortening in short-VOT stops (Ohala 1994, Goldrick et al 2013), and no shortening of lax vowels (Schertz 2013).

The mixed results for contrastive hyperarticulation in sounds which are not characterized by increased length might suggest that the effects are primarily driven by overall increased prominence rather than exaggeration of phonologically contrastive characteristics, though this might interact with task design and the particular contrast. Consistent with an explanation based on increased prominence, Oh and Byrd (2019) showed that corrective focus results in lengthening for both long-VOT and short-VOT categories in Korean. Some studies have found that lengthening occurs to some degree in both categories with shorter duration and longer duration and is just larger in the longer category, e.g. duration of vowels before voiced versus voiceless codas (de Jong 2004) and duration of tense versus lax vowels (Schertz 2013); this might reflect hyperarticulative lengthening being proportional to the duration of the segment rather than absolute.

If some of the effects of minimal pair competitors are due to emphasis, other phonetic characteristics might be expected, regardless of whether they are cues to a phonological contrast. Notably, other lexical characteristics like frequency and predictability in context have phonetic correlates, e.g. duration, F0, intensity (e.g. Gahl et al 2012, Seyfarth 2014, Tang & Shaw 2021). Some of these effects may be due to the greater likelihood of low frequency words receiving emphasis, as more predictable words are less likely to be stressed (Pan & Hirschberg 2000). Many acoustic characteristics associated with focus stress are also characteristics associated with lower lexical frequency, e.g. longer duration, higher F0, and higher intensity (Breen et al 2010, Katz & Selkirk 2011).

Even if words are not focused, some might be more prone to the broad hyperarticulation patterns associated with "clear speech". When speakers are trying to speak clearly and carefully, they produce words with longer duration, higher F0, and higher intensity, among other characteristics (Maniwa et al 2009, Granlund et al 2012). Notably, this style can be elicited with methods that are the same as many of the contrastive hyperarticulation studies, i.e. giving a speaker feedback indicating that their listener or speech recognition system had not recognized the word that was said (Maniwa et al 2009). The characteristics of clear speech do not seem to be sensitive to language-specific patterns of contrasts, e.g. vowels are not more dispersed in clear speech in a language with more vowels in its phonological inventory (Bradlow 2002). However, some studies have found differences across languages in the realization of clear speech, e.g. VOT shortening for the English short-lag /b/ but not for the similar short-lag /p/ in Finnish – which does not have a contrast with a corresponding long-lag stop (Granlund et al 2012).

However, even if some results reflect emphasis of misheard words, some of the contrastive hyperarticulation results do not seem to have this explanation. Schertz (2013) addressed emphasis as a possibility, but did not find that asking the speaker to repeat the word had a significant effect on vowel duration or intensity.

1.5 *This work* This study examines the acoustic characteristics of error repairs within natural speech in English. Analyses of repairs have often compared the repair and the preceding error (e.g. Cutler 1983, Nooteboom 2010, Carter and Plug 2014), which can make it difficult to evaluate which aspects of the results are due to characteristics of the repair and which are due to characteristics of the error. In the current study, comparisons are made between the repair and instances of the same or phonologically matched words in fluent contexts. In addition to avoiding potential confounds due to phonetic characteristics of speech errors, this method makes it possible to include repair items regardless of how much of the erroneous word was present.

2 Methods

2.1 *Data* Data for this study comes from the Fisher corpus of American English telephone conversations (Cieri et al. 2005): 334 segments in 173 speech error repairs, all paired with instances of the same speaker producing the same or phonologically matched words in fluent contexts.

Items were included for analysis regardless of whether the erroneous word was identifiable or not, but were only included if there was enough of the erroneous word that at least one segment was clearly identifiable.

Repairs were identified as instances where the speaker stopped mid-word or immediately after a word and replaced that individual word with a different word. Utterances were excluded if the disfluency was associated with restructuring the sentence rather than replacing the word, or if the original word and the replacement were synonymous.

The paired fluent items were matched exactly wherever possible, i.e. to another instance of the same speaker saying the same word. When there wasn't an instance of the speaker saying the same word in a fluent context, a close match was chosen based on matching in word length, stress, and neighboring segments for the segments being measured. If there was no close match, the item was omitted.

2.2 *Analysis* Results come from mixed-effects regression models calculated with the lme4 package in R (Bates et al. 2015); p-values were calculated with the lmerTest package (Kuznetsova et al. 2015).

Words were manually segmented and analyzed in Praat. Several dependent variables were analyzed: VOT, onset consonant duration, vowel duration, F0, and vowel intensity. To reduce potential effects of ambiguous boundaries, the vowel portions included neighboring sonorants if either the repair or its paired item did not exhibit a clear boundary between the vowel and that sonorant.

Repairs were also labelled for whether the error was lexical (e.g. *public* for *private*) or phonological errors (e.g. *noca(l)* for *local*); however, many errors could not be categorized, either because there was not enough of the erroneous word to allow an identification or because the error was ambiguous.

3 Results

3.1 *VOT* Table 1 presents the summary of a mixed effects regression model for VOT. The fixed effects were Utterance Type (Fluent, Repair), Voicing Category (Voiced, Voiceless), and the interaction between Type and Voicing. There were random intercepts for speaker and for segment.

Table 1 Regression model for VOT. *Reference Levels: Type = Fluent, Voicing = Voiced.*

	Estimate	SE	t-value	p-value
(Intercept)	22.9336	6.7863	3.379	0.00913
Type Repair	-0.2832	5.7206	-0.050	0.96060
Voicing Voiceless	39.3082	8.1606	4.817	0.00500
Type Repair : Voicing Voiceless	-1.2582	6.2909	-0.200	0.84184

There was no significant difference in VOT between fluent speech and repairs, either for voiced stops or voiceless stops. The expected VOT difference between voiced and voiceless stops is observed as a main effect; voiceless stops have longer VOT than voiced stops.

An additional factor of interest is whether the type of error impacts repairs. Table 2 presents the summary of a mixed effects regression model for VOT, restricted to items for which error type could be categorized. This model adds fixed effects of Error Type (Lexical, Phonological) and its interaction with Utterance Type. There were random intercepts for speaker and for segment.

Table 2 Regression model for VOT, restricted to items with identifiable error type. *Reference Levels: Type = Fluent, Voicing = Voiced, ErrorType = Lexical.*

	Estimate	SE	t-value	p-value
(Intercept)	20.049	10.084	1.988	0.0815
Type Repair	1.315	8.086	0.163	0.8713
Voicing Voiceless	40.993	11.959	3.428	0.0221
ErrorType Phonological	11.152	5.912	1.886	0.0621
Type Repair : Voicing Voiceless	-5.580	8.219	-0.679	0.4995
Type Repair : ErrorType Phonological	-1.553	6.321	-0.246	0.8066

There was no evidence for different types of errors resulting in different VOT values in corresponding repairs. Note that the marginal main effect for Error Type is for fluent productions; this might suggest that there are characteristics of words which make them more or less prone to lexical or phonological errors and which also impact VOT.

3.2 *Consonant duration* Table 3 presents the summary of a mixed effects regression model for consonant duration in word-initial consonants other than stops. The fixed effect was Utterance Type (Fluent, Repair). There were random intercepts for speaker and for segment.

Table 3 Regression model for duration of word-initial consonants that were not stops. *Reference Levels: Type = Fluent.*

	Estimate	SE	t-value	p-value
(Intercept)	80.866	6.079	13.302	< 0.0001
Type Repair	16.653	5.848	2.848	0.00657

Word-initial consonants were significantly longer in repairs than in fluent speech. This effect is driven both by an overall shifted distribution as well as a longer tail end of particularly long consonants, as illustrated in Figure 1.

Figure 1. Density plot for consonant durations in each type of item.

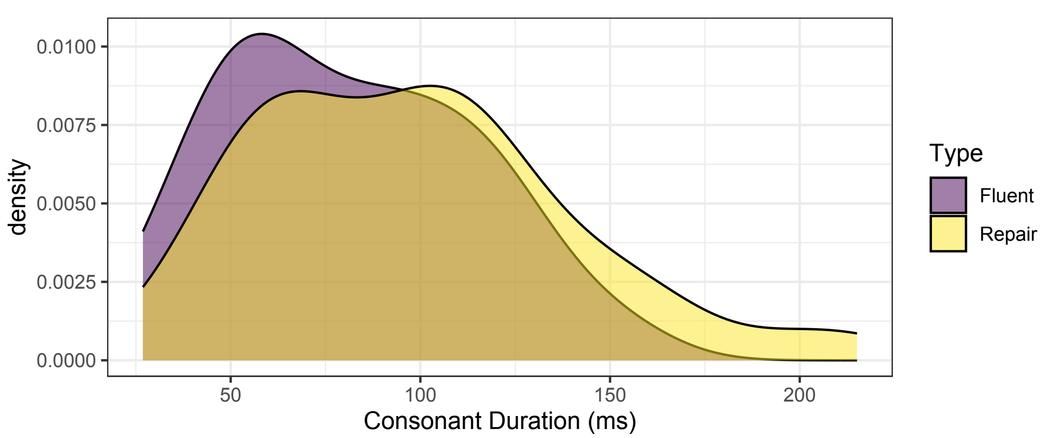

Table 4 presents the summary of a mixed effects regression model for duration in word-initial consonants other than stops, restricted to items for which error type could be categorized. This model adds fixed effects of Error Type (Lexical, Phonological) and its interaction with Utterance Type. There were random intercepts for speaker and for segment.

Table 4 Regression model for consonant duration, restricted to items with identifiable error type. *Reference Levels: Type = Fluent, ErrorType = Lexical.*

	Estimate	SE	t-value	p-value
(Intercept)	80.283	8.203	9.788	< 0.0001
Type Repair	13.075	6.794	1.925	0.062
ErrorType Phonological	8.469	11.955	0.708	0.482
Type Repair : ErrorType Phonological	2.157	10.331	0.209	0.836

The interaction between Utterance Type and Error Type was not significant; there was no evidence for different types of errors resulting in different word-initial consonant durations in the corresponding repairs. However, it is important to note that there were relatively few items that could be included in this model; only 37 pairs of items in which the error could be categorized began with consonants that were not stops.

3.3 *Vowel duration* Table 5 presents the summary of a mixed effects regression model for vowel duration. The fixed effect was Utterance Type (Fluent, Repair). There were random intercepts for speaker and for segment.

Table 5 Regression model for vowel duration. *Reference Levels: Type = Fluent.*

	Estimate	SE	t-value	p-value
(Intercept)	140.797	14.454	9.741	< 0.0001
Type Repair	4.588	4.337	1.058	0.291

There was no significant difference in vowel duration between fluent speech and repairs. However, it may be informative to also consider whether vowels were word-initial or not, because the lengthening observed in consonants might be constrained to the initial segment of a repair.

Table 6 presents the summary of a mixed effects regression model for vowel duration, with added fixed effects of Vowel Position (Word-Initial, Non-Initial) and the interaction between Vowel Position and Utterance Type. There were random intercepts for speaker and for segment.

Table 6 Regression model for vowel duration, including vowel position as a factor. *Reference Levels: Type = Fluent, VowelPosition = Non-Initial*

	Estimate	SE	t-value	p-value
(Intercept)	142.013	14.482	9.806	< 0.0001
Type Repair	2.348	4.491	0.523	0.6016
VowelPosition Initial	-16.024	16.377	-0.978	0.3285
Type Repair : VowelPosition Initial	29.589	16.324	1.813	0.0715

There was a trend towards an interaction between vowel position and utterance type, with longer word-initial vowels in repairs than in fluent speech, whereas vowels in other positions exhibited little effect. However, it is important to note that there were very few word-initial vowels in this dataset; there were only 14 pairs of items. Thus, the trend must be interpreted cautiously.

Given that vowel duration is a major cue to coda voicing in English, we might predict an interaction between utterance type and coda voicing (cf. de Jong 2004, Goldrick et al 2013, Seyfarth et al 2016). Table 7 presents the summary of a mixed effects regression model for vowel duration, adding fixed effects of Coda Voicing (Voiced, Voiceless) and its interaction with Utterance Type to the model from Table 5. There were random intercepts for speaker and for segment.

Table 7 Regression model for vowel duration, including coda voicing as a factor. *Reference Levels: Type = Fluent, CodaVoicing = Voiced.*

	Estimate	SE	t-value	p-value
(Intercept)	156.587	11.180	14.006	< 0.0001
Type Repair	10.939	9.332	1.172	0.2452
CodaVoicing Voiceless	-29.736	15.273	-1.947	0.0541
Type Repair : CodaVoicing Voiceless	-2.807	14.222	-0.197	0.8441

There was no evidence for different effects of repairs on vowels before voiced and voiceless codas. There was only a marginal main effect for coda voicing, with shorter vowels before voiceless codas. The main effect of coda voicing is expected based on previous work demonstrating voicing-conditioned vowel duration in English.

Table 8 presents the summary of a mixed effects regression model for vowel duration, restricted to items for which error type could be categorized. This model adds fixed effects of Error Type (Lexical, Phonological) and its interaction with Utterance Type to the model from Table 5. There were random intercepts for speaker and for segment.

Table 8 Regression model for vowel duration, restricted to items with identifiable error type. *Reference Levels: Type = Fluent, ErrorType = Lexical.*

	Estimate	SE	t-value	p-value
(Intercept)	135.938	15.866	8.568	< 0.0001
Type Repair	6.221	6.811	0.913	0.363
ErrorType Phonological	5.485	12.694	0.432	0.666
Type Repair : ErrorType Phonological	-4.409	10.303	-0.428	0.669

There was no evidence for different types of errors resulting in different vowel durations in corresponding repairs. Neither lexical errors nor phonological errors exhibited evidence of an effect on vowel duration in repairs.

3.4 *F0* Table 9 presents the summary of a mixed effects regression model for F0 mean within vowels. The fixed effect was Utterance Type (Fluent, Repair). There were random intercepts for speaker and for segment.

Table 9 Regression model for F0 mean within vowels. *Reference Levels: Type = Fluent.*

	Estimate	SE	t-value	p-value
(Intercept)	181.409	4.133	43.893	< 0.0001
Type Repair	8.793	3.261	2.697	0.00761

F0 mean was significantly higher in repairs than in fluent speech. The overall shape of the distribution is similar for both types, but the tails on each end differ in their width, as illustrated in Figure 2.

Figure 2. Density plot for F0 means in each type of item.

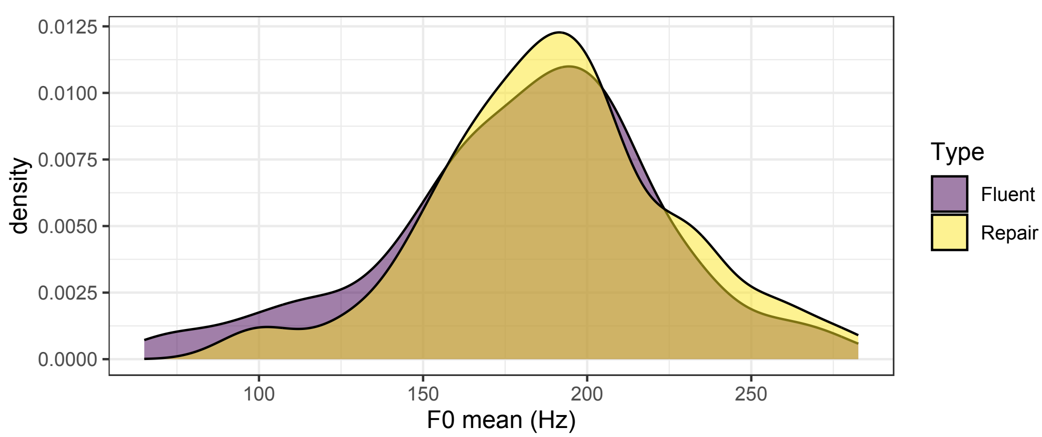

Table 10 presents the summary of a mixed effects regression model for F0 mean, restricted to items for which error type could be categorized. This model adds fixed effects of Error Type (Lexical, Phonological) and its interaction with Utterance Type. There were random intercepts for speaker and for segment.

Table 10 Regression model for F0 mean, restricted to items with identifiable error type. *Reference Levels: Type = Fluent, ErrorType = Lexical.*

	Estimate	SE	t-value	p-value
(Intercept)	178.609	5.296	33.724	< 0.0001
Type Repair	14.406	5.085	2.833	0.0053
ErrorType Phonological	4.022	7.010	0.574	0.5667
Type Repair : ErrorType Phonological	-12.090	7.691	-1.572	0.1183

There was a trend towards higher F0 in repairs of lexical errors than in repairs of phonological errors, as seen in the interaction between Utterance Type and Error Type. However, the difference did not reach significance.

3.4 *Intensity* Table 11 presents the summary of a mixed effects regression model for mean intensity within vowels. The fixed effect was Utterance Type (Fluent, Repair). There were random intercepts for speaker and for segment.

Table 11 Regression model for mean intensity within vowels. *Reference Levels: Type = Fluent.*

	Estimate	SE	t-value	p-value
(Intercept)	63.9360	0.8805	72.613	< 0.0001
Type Repair	1.4564	0.4635	3.142	0.00193

Intensity was significantly higher in repairs than in fluent speech. While both distributions had similar upper ends, they differed in their peaks and the lower side of those peaks, as illustrated in Figure 3.

Figure 3. Density plot of intensity in each type of item.

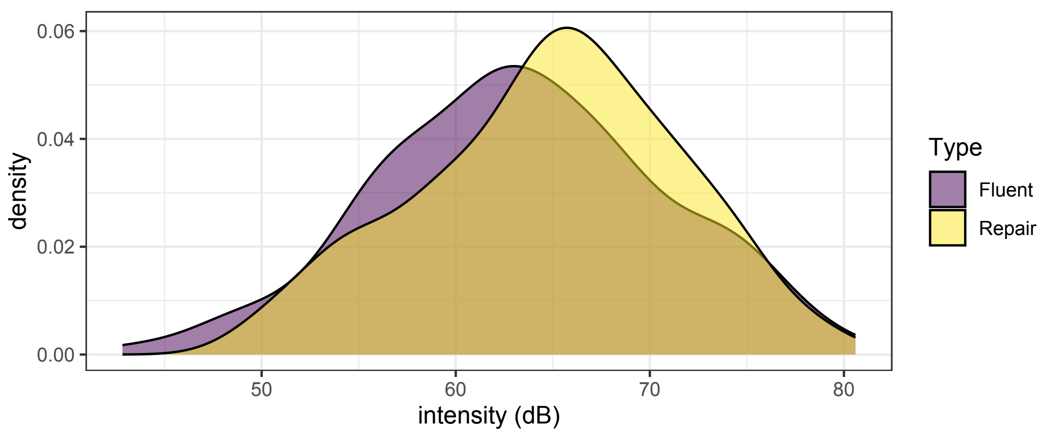

Table 12 presents the summary of a mixed effects regression model for intensity, restricted to items for which error type could be categorized. This model adds fixed effects of Error Type (Lexical, Phonological) and its interaction with Utterance Type. There were random intercepts for speaker and for segment.

Table 12 Regression model for intensity, restricted to items with identifiable error type. *Reference Levels: Type = Fluent, ErrorType = Lexical.*

	Estimate	SE	t-value	p-value
(Intercept)	63.3251	1.1471	55.206	< 0.0001
Type Repair	1.5831	0.7493	2.113	0.0364
ErrorType Phonological	1.4514	1.2677	1.145	0.2538
Type Repair : ErrorType Phonological	-0.7880	1.1334	-0.695	0.4881

There was no evidence for different types of errors resulting in different intensity in corresponding repairs. However, it is worth mentioning that the difference in intensity between fluent speech and repairs was only significant for repairs of lexical errors.

4 Discussion

The results align with the previously reported correlates of speech error repairs: Higher pitch and higher intensity (Nooteboom 2010, Carter and Plug 2014). The current study confirms that those patterns are indeed due to characteristics of the repairs themselves rather than being driven by characteristics of the erroneous elements preceding them, which they have been compared to in previous studies.

These results are consistent with corrections eliciting focus prominence of some sort (cf. Cutler 1983, Shattuck-Hufnagel & Cutler 1999), in order to draw attention to the repair that is replacing the previous erroneous elements. However, in contrast to previous descriptions of prosodic focus (e.g. Breen et al 2010, Katz & Selkirk 2011), there wasn't strong evidence of repairs overall having longer vowel duration than fluent productions of the same words. The lack of vowel duration effects might be due to competing pressures of emphasis versus signaling resolution of the preceding disfluency; speech errors often have longer duration than fluent speech (Schriberg 2001) and faster speech rate is associated with higher confidence (Scherer et al 1973), so longer duration would be a poor cue to focus in the context of a disfluency.

Although there was no overall effect of repairs on vowel duration, there was an effect on consonant duration. Most of the consonants in the analysis were word initial, so greater length in repairs could reflect lexical competition slowing down lexical access (cf. Vitevitch & Luce 1999, Luce & Large 2001). Lengthening seems to be very local, only impacting the first segment in the word. This lengthening effect could indicate that speakers frequently start producing the repair word before lexical access is entirely complete, so the beginning of the word may reflect the same slowdowns that characterize speech errors (Shriberg 2001). Based on the lack of lengthening in VOT, this lengthening is restricted just to the first gesture or subpart of the initial segment; the constriction of these word-initial stops might be lengthened, but the duration of the constriction is not measurable for many word-initial stops. Few words in the data begin with vowels, so this lengthening is not apparent overall in vowels, but there is some evidence that the few word-initial vowels do exhibit vowel lengthening in repairs. Another possible explanation of the lengthening in initial segments of a repair could be domain-initial strengthening (cf. Keating et al 2004); despite the usual syntactic influences on prosodic structure, a repair might begin a new phonological phrase or higher level prosodic unit.

There was no evidence for contrastive hyperarticulation. The acoustic characteristics associated with speech error repairs do not seem to differ based on the phonological contrasts which distinguish the repair words from competitors, so they are not consistent with inhibition of lexical competitors causing productions which are less like those competitors. Most of the observed effects (consonant duration, vowel intensity, vowel F0) are not increasing distinctiveness between the correction and the error, and are not even primary cues to phonological contrasts of English. However, there was not enough data of particular featural errors to clearly evaluate whether contrastive hyperarticulation might exist specifically in repairs of single-segment replacement errors.

There was no evidence within this data for a difference between repairs of lexical and phonological errors. The lack of effect might be due to the relatively small dataset; the number of tokens is even smaller when restricted to items for which the categorization of the error as lexical or phonological was clear. Some items did not have a clear categorization because the error was a simple substitution but also resulted in a semantically plausible word (e.g. *million* for *billion*); others could not be categorized because not enough of the erroneous word was present to enable identification. If repairs really do exhibit the same acoustic characteristics regardless of the type of error, this might suggest that the acoustic characteristics reflect speakers emphasizing the repair in order to draw listeners attention to the correct word, rather than acoustic characteristics being the result of suppressing lexical or phonological competitors.

5 Conclusions

The results provide evidence that repairs receive emphasis, marked by increased F0 and increased intensity. These prosodic effects do not seem to be driven by speakers increasing the phonetic differences between the target word and the error.

Duration might not be a characteristic of repair-marking emphasis, due to the slowdowns in preceding errors. The observed lengthening of initial consonants in repairs might reflect repairs starting a new prosodic domain, or might indicate that the slowdown during an error continues into the repair.

References

Alderete, John, Melissa Baese-Berk, Keith Leung, & Matthew Goldrick. 2021. Cascading activation in phonological planning and articulation: Evidence from spontaneous speech errors. *Cognition* 210. Article 104577.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1-48.

Baese-Berk, Melissa, & Matthew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24(4). 527-554.

Bradlow, Ann. 2002. Confluent talker- and listener-oriented forces in clear speech production. In Carlos Gussenhoven & Natasha Warner (Eds.), *Laboratory Phonology* 7, 241-273. Berlin: Mouton de Gruyter.

Breen, Mara, Evelina Fedorenko, Michael Wagner, & Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25. 1044-1098.

Buz, Esteban, Michael Tanenhaus, & T. Florian Jaeger. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language* 89. 68-86.

Carter, Paul & Leendert Plug. 2014. Prosodic marking, pitch and intensity in spontaneous phonological self-repair in Dutch. In Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia, & Doris Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production*, 69-72.

Cieri, Christopher, David Graff, Owen Kimball, Dave Miller, & Kevin Walker. 2005. Fisher English Training Part 2, Speech LDC2005S13. Web Download. Philadelphia: Linguistic Data Consortium.

Clopper, Cynthia & Terrin Tamati. 2014. Effects of local lexical competition and regional dialect on dialect production (L). *Journal of the Acoustical Society of America* (Letters to the Editor) 136(1). 1-4.

Cutler, Anne. 1983. Speakers' conceptions of the function of prosody. In Anne Cutler & D. Robert Ladd (Eds.), *Prosody: Models and measurements*, 79-91. Berlin: Springer.

Fox, Neal, Megan Reilly, & Sheila Blumstein. 2015. Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language* 83. 97-117.

Fricke, Melinda, Melissa Baese-Berk, & Matthew Goldrick. 2016. Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience* 31(5). 639-645.

Frisch, Stefan & Richard Wright. 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30, 139-162.

Gahl, Susanne, Yao Yao, & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66. 789-806.

Goldrick, Matthew, & Sheila Blumstein. 2006. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 21. 649-683.

Goldrick, Matthew, Charlotte Vaughn, & Amanda Murphy. 2013. The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America* 134(2). EL172-EL177.

Granlund, Sonia, Valerie Hazan, & Rachel Baker. 2012. An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals. *Journal of Phonetics* 40, 509-520.

De Jong, Kenneth. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32. 493-516.

Katz, Jonah, & Elisabeth Selkirk. 2011. Contrastive focus vs discourse-new: Evidence from phonetic prominence in English. *Language* 87(4). 771-816.

Keating, Patricia, Taehong Cho, Cécile Fougeron, & Chai-Shune Hsu. (2004). Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden, & Rosalind Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 145-163.

Kirov, Christo & Colin Wilson. 2012. The specificity of online variation in speech production. *Proceedings of the Annual Meeting of the Cognitive Science Society* 34. 587-592.

Kuznetsova, Alexandra, Per Bruun Brockhoff, & Rune Haubo Bojesen Christensen. 2015. *lmerTest: Tests in linear*

mixed effects models. https://CRAN.R-project.ord/package=lmerTest. R package version 2.0-29.
Levelt, Willem & Anne Cutler. 1983. Prosodic marking in speech repair. *Journal of Semantics* 2(2), 205-217.
Luce, Paul & Nathan Large. 2001. Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes* 16(5-6). 565-581.
Maniwa, Kazumi, Allard Jongman, & Travis Wade. 2009. Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*. 125(6). 3962-3973.
Munson, Benjamin, & Nancy Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47. 1048-1058.
Nooteboom, Sieb. 2010. Monitoring for speech errors has different functions in inner and overt speech. In Martin Everaert, Tom Lentz, Hannah de Mulder, Øystein Nilsen, & Arjen Zondervan (Eds.), *The Linguistics Enterprise: From Knowledge of Language to Knowledge in Linguistics*, 213-233. Amsterdam: John Benjamins.
Oh, Miran, & Dani Byrd. 2019. Syllable-internal corrective focus in Korean. *Journal of Phonetics* 77. Article 100933.
Ohala, John. 1994. Acoustic study of clear speech: A test of the contrastive hypothesis. In *Proceedings of the International Symposium on Prosody, 18 September 1994, Yokohama, Japan*, 75-89.
Pan, Shimei & Julia Hirschberg. 2000. Modeling local context for pitch accent prediction. In *Proceedings of the 38th annual meeting of the Association for Computational Linguistics*, 233-240.
Peirce, Jonathan. 2007. PsychoPy–Psychophysics software in Python. *Journal of Neuroscience Methods* 162(1-2). 8-13.
Plug, Leendert. 2011. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *Journal of Phonetics* 39, 289-297.
Scarborough, Rebecca. 2010. Lexical and contextual predictability: Confluent effects on the production of vowels. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio, & Nathalie Vallée (Eds.), *Laboratory Phonology* 10, 557-586. Berlin: Mouton de Gruyter.
Scherer, Klaus, Harvey London, & Jared Wolf. 1973. The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality* 7(1). 31-44.
Schertz, Jessamyn. 2013. Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics* 41(3-4). 249-263.
Seyfarth, Scott, Esteban Buz, & T. Florian Jaeger. 2016. Dynamic hyperarticulation of coda voicing. *Journal of the Acoustical Society of America* 139. EL31-EL37.
Shattuck-Hufnagel, Stefanie & Anne Cutler. 1999. The prosody of speech error corrections revisited. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, & Ashlee C. Bailey (Eds.), *The 14th International Congress of Phonetic Sciences*, 1483-1486. Berkeley: University of California.
Shriberg, Elizabeth. 2001. To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31(1), 153-169.
Stern, Michael, & Jason Shaw. 2023. Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language* 132. Article 104443.
Tang, Kevin & Jason A. Shaw. 2021. Prosody leaks into the memories of words. *Cognition* 210. Article 104601.
Vitevitch, Michael, & Paul Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40(3). 374-408.
Wedel, Andrew, Noah Nelson & Rebecca Sharp. 2018. The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language* 100. 61-88.