Articulatory Perspectives on Errors*

Marianne Pouplier
University of Edinburgh and Haskins Laboratories

In speech error research discrepant results have been obtained depending on the research tool employed. While traditionally errors have been identified perceptually, acoustic and articulatory measurements have come to different conclusions about the characteristics of errors. This has led to renewed concerns about perceptual biases in transcription records on the one hand and the comparability of laboratory elicited and naturally occurring errors on the other hand. The present paper presents an overview of the results of articulatory error studies and discusses their relation to transcription based error work. We also discuss briefly how future research avenues may further our understanding of errors.

1. Introduction

That speech error arguments are mostly a matter of emphasis has been made explicit many years ago by Anne Cutler (1981); she classified the types of arguments which have been used to arrive at generalizations over error data into "Some Errors," "More Errors," and "No Errors" arguments. The first type is chiefly a characterization of which kinds of errors occur at all, and it allows us to draw conclusions about the organization of the lexicon and speech production. For instance, hypotheses about the psychological reality of segments are founded on the observation that individual consonants or vowels can participate in errors as separate units. "More Errors" arguments, on the other hand, consider the quantitative patterning of errors, and have been a prominent type of argument for segmental, as opposed to featural representations (Shattuck-Hufnagel and Klatt 1979). More, and indeed most speech errors it seems, are serial misorderings of segments, as for example when the intended *take my bike* is erroneously pronounced as *bake my bike* (Fromkin 1973). A classic example of a "No Errors" argument is Rulon Wells's (1951) First Law which famously states that "A slip of the tongue is practically always a phonetically possible noise." (cf. also Boomer and Laver 1968; Fromkin 1971). Although the argument, as Cutler points out, has never been one of "entirely absent," but more one of "almost completely absent," that is, a matter of emphasis. Yet how strongly this empha-

* Work supported by the 6th European Framework Programme and NIH grant HD-01994 to Haskins Labs. Thank you to Carson Schütze and Vic Ferreira for organizing an interesting and stimulating workshop.

MIT Working Papers in Linguistics 53, 115–132
The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop
Carson T. Schütze and Victor S. Ferreira (Eds.)

sis on certain regularities of speech errors has emerged from decades of transcription based error research can be for example appreciated from the fact that for instance Dell et al. (1993) calibrated the output of their parallel-distributed-processing model against "standards from the speech error literature" and assumed 99% of errors to be phonotactically well-formed.

In recent years, the discussion of which aspects of error data should receive "some," "more" and "no" emphasis has gained new momentum: acoustic and notably articulatory investigations of sublexical errors have rendered results different from those obtained from transcription records of errors. Concerns about inherent limitations of transcription as a research tool in this context have been confirmed in that transcription records may, at least to a certain degree, be qualitatively and quantitatively inaccurate. However, while instrumental studies have highlighted significant methodological problems of transcription, they suffer from their own limitations, albeit different ones from transcription records. Also the comparability of laboratory elicited and naturally occurring errors needs to be reconsidered and has emerged as a key issue. Another limiting factor in gauging the theoretical significance of instrumental error work is that only relatively few studies have examined errors articulatorily or acoustically at this point, leaving it unclear which aspects of transcription records in detail may be inaccurate.

The purpose of the present paper is to provide an overview of the state of the art of error research specifically from an articulatory perspective. We will consider what properties of errors articulatory and also acoustic studies have revealed and how these relate to transcription studies. We will conclude with suggestions for future research avenues which may help resolve some of the most critical issues in the current debate.

2. Phonological and phonetic errors

A major, "classic" type of speech production model (the key models being Dell 1986; Dell et al. 1993; Levelt 1989; Levelt et al. 1999) advocates that word form retrieval in speech production proceeds through independent phonological and phonetic processing modules. Decades of error research seemed to suggest fairly unequivocally that errors happen at a symbolic segmental (or featural) planning level (e.g., Levelt et al. 1999; Meyer 1992; Shattuck-Hufnagel 1979). While phonological features, bound to a segmental timing slot, specify place and manner of articulation, they do so in a purely symbolic fashion: there are no actual spatio-temporal parameters associated with phonological representations, there is no temporal dimension beyond the linear sequencing of segments. While the translation of these symbolic phonological representations into articulator movements during phonetic implementation is generally not part of these types of models, it is precisely speech error data that seem to force the conclusion that at some stage, word form retrieval and encoding operate on the basis of purely symbolic units. In sublexical errors, articulatory properties seem to be computed only after the error has occurred, since the coarticulatory properties of a shifted segment are appropriate to the *new* position. For instance in a mispronunciation of *slumber party* as *lumber sparty* (Fromkin 1973), the [p] is appropriately unaspirated in the errorful pronunciation. Phonetic implementation does not "know," as it were, whether the input sequence of phonological segments contains errors in serial order or not; the segments are implemented appropriately to their actual position in the segmental string. According to transcription studies, nothing in errors suggests that the units erroneously shifted in their serial order have spatio-temporal specifications. Any potentially present deviation from a canonical articulatory pattern is thus not directly caused by the error process itself, but can only be indirectly linked to the error, for example through monitoring effects. Any errors occurring during phonetic implementation have been hypothesized to occur independently of the phonological encoding stage (Dell et al. 1993; Levelt et al. 1999), yet an explicit account of how these errors may come about is generally not part of these models.

Several instrumental studies have, however, provided an increasing body of evidence that casts doubt on the hypothesis that phonological and phonetic errors can be distinguished on the basis of normal vs. anomalous execution, that is, it may not be the case that the majority of errors are wholesale segmental shifts which are then executed normally. These studies could show that errorful articulations can frequently cover a range of movement amplitudes which are not observed in non-errorful productions of the same utterance (Boucher 1994; Goldstein et al. in press; Mowrey and MacKay 1990; Pouplier in press; Frisch 2007; cf. also Frisch and Wright 2002 and Goldrick and Blumstein 2006 for similar results in acoustic error studies). The main finding that has emerged from articulatory studies has been that speech errors are often not a segment or feature (gradiently) substituting for another one, but instead, in many cases, the intended and intruding unit are produced at the same time (cf. also Laver 1979 on vowel errors). For example, we found in an articulography (EMMA)[1] study of errors that in phrases like *top cop*, a [k]-like tongue dorsum gesture appeared in the prevocalic position of *top*, without the tongue tip raising gesture for [t] disappearing—both were articulated simultaneously in the same prevocalic position (similar errors were also observed for postvocalic positions). These gestural intrusion errors may or may not lead to phonotactically illegal constellations, depending on the gestural composition of the particular consonants involved. Crucially, Goldrick and Blumstein (2006), in their acoustic study of voicing errors, present evidence that the observation of gradience in acoustic/articulatory measures during speech errors is indeed due to the partial activation of multiple phonological targets, and is not due to inherently increased variability in alternating (error-triggering) environments as opposed to non-alternating (control) environments.

We have accounted for the coproduction errors we identified in our EMMA experiments within a gestural speech production framework (Browman and Goldstein 1992). Errors in gestural coordination may lead to rhythmic synchronization (i.e., simultaneous production) of gestures at 1:1 frequency and in-phase. Errors in this view arise from the interplay of different lexically stable coordination modes and the extra-linguistically stable 1:1, in-phase coordination mode (cf. Turvey 1990). In error triggering environments, a stable dynamic attractor is weakened to an extent that a different dynamically stable attractor

[1] For a brief introduction to articulography (EMMA), cf. Stone 1997.

comes to dominate over the intended utterance. This attractor may be linguistically or extra-linguistically stable (for a detailed discussion cf. Goldstein et al. in press; Pouplier in press). These intrinsically stable coordination modes have previously been argued to play a role in speech and grammar (e.g., Browman and Goldstein 2000; Kelso et al. 1986; Stetson 1951). In this view, phonotactically legal and illegal errors are inherently the same phenomenon in terms of errorful coupling relations that gestures enter into. This approach obviates the necessity of separate phonological and phonetic representations with their respective distinct error types.

An alternative account has been put forward by Goldrick and Blumstein (2006). They interpret their finding of gradient voicing errors in voiced-voiceless stop interactions within a cascading activation model. This model is a strictly feedforward model, yet activation of competing candidates cascades down throughout the speech production system, including phonetic encoding. Goldrick and Blumstein point out that their results can be accounted for in any model that allows partial activation of competing phonological candidates (independently of assumptions about the nature of phonological representations in terms of gestures, segments, features, etc.), yet they reject the assumption of strictly discrete phonological and phonetic processing stages in which only a single phonological output representation serves as input to the phonetic processing component.

3. "Some" errors or "most" errors?

There has always been a general acknowledgement that some errors do not fit the dominant pattern of normal execution, but here the emphasis that different types of arguments place on the data has played a pivotal role. Errors that are not phonologically well-formed have been assumed to be by and large edited out before surfacing, while audibly anomalous execution in errors has been taken to be a hallmark property of errors which originate during phonetic implementation (cf. e.g., Levelt et al. 1999; Meyer 1992). Generally both types of errors have been observed (that is, perceived) to occur only rarely—Laver (1979) is one of the few early studies that reported systematically occurring subphonemic errors and explicitly discussed how they might originate. Also Butterworth and Whittaker (1980) found errors violating phonotactics, yet these studies have had no deeper impact on the architecture of the major speech production models. The possibility that errors may happen at the interface between phonological and phonetic encoding or at the implementation level has always been acknowledged, yet, with the few exceptions already mentioned, no actual account of these errors has been proposed within the traditional speech error literature.

At the heart of the current speech error debate lies thus the question of what to make of these newly revealed properties of errors. How do they relate to the generalizations over error patterns which have emerged from transcribed data and what clues do these errors give us as to their origin? In the main speech production models, different levels of word form encoding make use of different representations—so far, the premise has been that normal execution is the key piece of evidence which can tell us where in the speech production system a given error originated: Symbolic segmental shifts point to phonological planning errors while anomalous execution is a characteristic of phonetic errors, with the latter occurring far less frequently than the former. Several further arguments which have traditionally been used to draw the line between phonological and phonetic errors shall now be considered in the light of the results from instrumental error studies.

3.1 Frequency of occurrence

The "most errors" argument has been the decisive one in the segment versus feature debate: most errors can be analyzed either at the level of the segment or the feature (e.g., ***p**its and **b**eeses* for ***b**its and **p**ieces*, which could be a [voice] feature or a segment-level exchange; Shattuck-Hufnagel and Klatt 1979). Yet due to the rare occurrence of unambiguous feature errors it has been generally accepted that ambiguous errors should be treated as segmental rather than featural substitutions (Shattuck-Hufnagel and Klatt 1979; but see Dell et al. 1993; Guest 2001; Roberts 1975). Particularly with respect to the possibility that gestures in the sense of Articulatory Phonology (Browman and Goldstein 1992) may be the primitive units of speech production which participate in errors, we encounter a "no errors" argument: Nothing in speech errors (as analyzed through transcription) seemed to suggest that the units that come to be serially misordered have any articulatory, spatio-temporal specifications. To the contrary: errors appeared to provide evidence for the *necessity* of symbolic segmental representations (Fowler 1995; Levelt et al. 1999). The predominance of segmental (as opposed to smaller unit) errors has further cast doubt on underlying gestural representations. Yet it seems that the limitations of perceptual data evaluation have at least partly conditioned these results: the majority of errors we have observed in our EMMA study were intrusion errors (leading to a coproduction of the intended and the intruding gesture), not substitutions (cf. Figure 1). We could also show in the context of interactions between nasal and non-nasal consonants that errors can systematically involve units below the level of the segment. For example, in the phrase *kim kid* the velum gesture of the labial nasal [m] may intrude during the coronal stop [d] independently of the lip gesture and vice versa (Goldstein et al. in press).[2] These results show for one that the assumption that the vast majority of errors are wholesale segmental substitutions is, at least for laboratory studies and the elicitation paradigms employed, not correct (a point discussed further below). While most instrumental studies have used rapid overt repetition to elicit errors, also with a non-repetition based elicitation technique (SLIP, Motley and Baars 1976a), a significant amount of ill-formed errors could be identified on the basis of tongue kinematics (Pouplier in press).

Figure 1 gives the error type distribution according to substitution errors or coproduction errors for the data reported in Goldstein et al. in press,[3] for a

[2] In terms of the size of the interacting units, these findings are also compatible with the assumption of feature errors. That most errors should be analyzed as feature errors has been argued, among others, by Roberts (1975), Guest (2001) and Dell et al. (1993).

[3] We also observed errors which were neither intrusion nor substitution errors, but reduction errors without accompanying intrusion. These were comparatively rare and are not reported here but see Goldstein et al. in press for a detailed discussion.

repetition-based elicitation paradigm on the left and for the data reported in Pouplier (in press) for the SLIP elicitation paradigm on the right. Both experiments used articulography (EMMA) to investigate errorful [t]–[k] interactions.

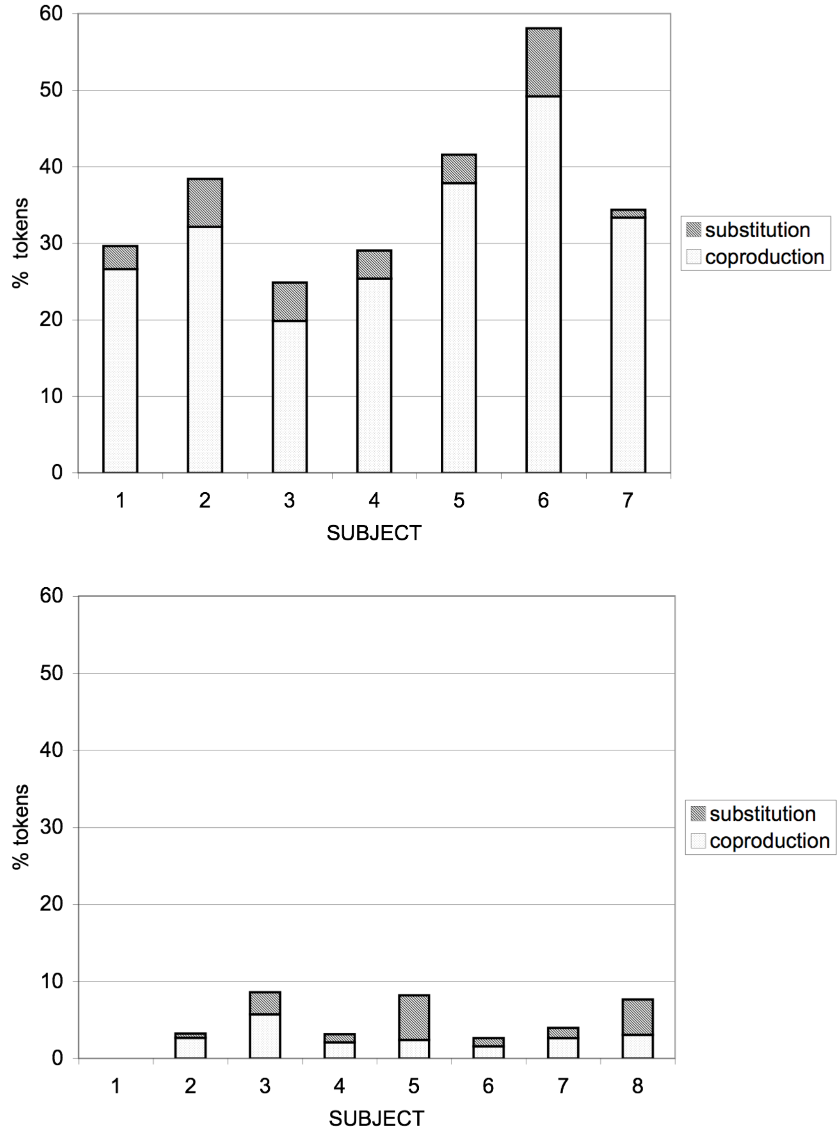

Figure 1
Percent of tokens exhibiting substitution or coproduction errors
(The lighter portions of the bars represent coproduction errors, and the darker portions substitution errors. Results from a repetition-based elicitation technique are above, those from the SLIP technique below.)

We can observe in Figure 1 that not only does the error rate drop dramatically between the two elicitation paradigms, but also the proportion of coproduction errors relative to substitution errors changes. (Note, however, that also substitution errors could be of gradient magnitude in that a partial intrusion could be accompanied by a partial reduction.) However, in both experimental paradigms a statistically significant intrusion bias emerged. The SLIP technique has traditionally been hypothesized to elicit phonological planning errors since it relies on priming rather than overt repetition to elicit errors. Since errors elicited with this paradigm are frequently not categorical segmental substitutions, this supports the assumption that also phonological planning errors are not necessarily executed normally. Alternatively, it may have to be reconsidered which level of representation the SLIP technique taps into, a point we will return to below.

3.2 Contextuality

Speech errors can been classified as being *contextual* or *noncontextual*, the former denoting an error in which the intruding segment (gesture/feature) comes from neighboring material (e.g., *cup of coffee* → *cuff of coffee*; Fromkin 1971), while for the latter, no source for the error can be identified in the immediate neighborhood of the surface utterance (e.g., *a three day thing* → *a three way thing*; Fromkin 1973). Non-contextual errors are of interest here because they presumably arise from (semantically and phonologically) competing plans, and inner speech is often assumed to be phonological in nature (e.g., Dell 1986; Dell and Repka 1982; Wheeldon and Levelt 1995). Therefore the classic speech production models predict non-contextual errors to display characteristics of single segment shifts (that is, normal execution). If several units are coproduced in noncontextual errors, this is consistent with the assumption that normal execution is not a hallmark property of phonological planning errors.

In normal speakers, non-contextual errors are relatively rare: according to Schwartz et al. (1994), contextual errors account for 70% or more of normal speakers' errors (cf. also Talo 1980), yet in disordered speakers, it is non-contextual errors which occur at a rate of about 70–80%. It has long been known that speech errors in normal and disordered speakers share certain key characteristics, notably that some errors can be analyzed as symbolic segment shifts (e.g., Blumstein 1973; Wilshire 2002). Wood (1997) investigated errors in patients with Broca's aphasia (two with and two without apraxia of speech) using electropalatography (EPG) and found many errors to be coproductions of several segments (one intended, one or more intruding). For instance, the target phrase *a key* was heard to be erroneously pronounced as [ə t^hi], but the EPG contact pattern revealed that both a coronal and a velar closure were produced at the same time, while only the coronal gesture was perceived. In the context of apraxia of speech these types of errors had earlier been termed *misdirected articulatory gestures* (MAGs) by Hardcastle and Edwards (1992). Wood further could show that errors are not confined to double articulations; she also observed patterns of simultaneous coronal–velar–labial articulations (cf. also Wood and Hardcastle 1999). For instance, during the intended phrase *the book*, a coronal–velar contact pattern was present in prevocalic position, while the intended labial closure was still in place in that same prevocalic position (for a more detailed comparison between these types of errors in disordered and normal speakers, cf. Pouplier and Hardcastle 2005). In the Wood (1997) study, for 25% of the errors classified as clear phonemic substitutions by trained transcribers (speech therapists), additional linguo-palatal contact patterns could be identified in the EPG data (cf. also Pouplier and Goldstein 2005 on the perception of ill-formed speech errors).

Just like the instrumental error studies for normal speakers, these data for disordered speakers suggest that also errors arising at the phonological planning level can lead to phonotactically illegal outcomes in that interacting consonants in inner speech can come to be simultaneously produced, similar to experiments in which errors are triggered through overt repetition (cf. also the SLIP results mentioned earlier).

3.3 Lexicality effects

Transcription records of errors have revealed that errors are more likely if the outcome of an error is a word as opposed to a nonword (Dell and Reich 1981; Motley and Baars 1976b). The lexicality effect has sometimes been taken to support feedback (e.g., Dell and Reich 1981), but also monitoring accounts have been proposed to account for this phenomenon (Hartsuiker et al. 2005; Levelt 1989; Levelt et al. 1999). Frisch and Wright (2002) found in their acoustic study that both partial and categorical (substitution) errors are more likely if the result of the error was a word. Goldrick and Blumstein (2006), in their acoustic investigation of voicing errors, discovered a tendency for the lexical status of the error outcome to influence the extent to which traces of the originally intended segment were found in the errorful utterance. Gradient errors (i.e., VOT values intermediate between those for canonically voiced or canonically voiceless tokens) occurred predominantly in cases where the outcome of the error was a nonword; for word outcomes, categorical errors predominated (although only a small number of word-outcome errors could be included in this particular analysis). The authors interpret this as a suppression effect of the intended target caused by the lexical representations for word-outcome errors. The phonological representation of the word-outcome error receives supporting activation from its lexical representation (since the outcome of the error is a word). In the case of a nonword outcome, the result of the error has no lexical representation, thus the phonological representation of the error outcome will be comparatively weak. This will in turn lead to a situation where two competing phonological representations (the intended one and the errorful one) will have more or less equal strength and will hence both be traceable in the resulting articulation. While at this point the data on the lexicality effect are sparse, the data currently available suggest that the lexicality effect is not confined to normally executed substitution errors, but also emerges in gradient errors.

3.4 The role of time

It is well-known that tongue twisters become increasingly difficult with each repetition, that is, errors build up over time. Due to their high effectiveness in triggering errors, many laboratory elicitation methods draw on the continuous repetition of phrases with alternating consonants (or vowels), similar to tongue twisters (e.g., Wilshire 1999; Croot et al. 2006). This factor is of particular interest in the current context since this build-up of errors over time may reflect physiological constraints on motor execution. This build-up may then possibly prove revealing as to the origin of the errors we observe: it might be expected that errors at the very beginning of trials could be of a different type from those occurring after many repetitions. In particular, wholesale substitutions could be predominant at the beginning of trials, while there may be a prevalence of phonetically ill-formed errors towards the end of trials. In Goldstein et al. (in press) we analyzed errors as to this effect and showed a build-up of error-rate as a function of repetition number in a continuous repetition task. Looking at substitution errors only (i.e., excluding intrusion errors), the overall picture remained the same: Also for substitutions we observed error rate to increase as a function of repetition number (cf. Figure 2).

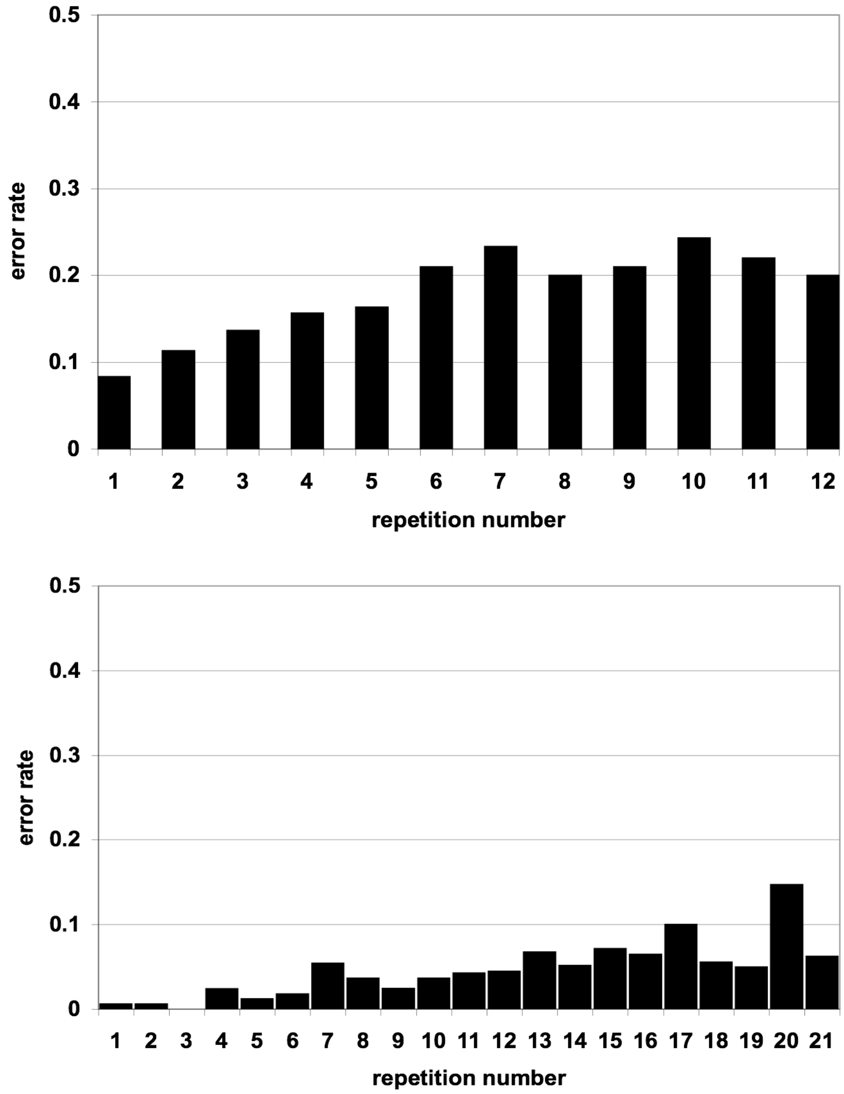

Figure 2
Error rate as a function of repetition number for both well-formed and ill-formed errors (above), and for substitution errors only (below)

These data are consistent with the hypothesis that both well-formed and ill-formed errors have a common origin: conditions that cause error rate to increase (such as repetition number or increased speaking rate) equally affect all types of errors, ill-formed or not. We have proposed that this build-up of errors is due to the strength of coupling relations evolving over time. An interpretation of errors in terms of errorful gestural coordination patterns allows us to relate this effect to the fact that coupling relations take time to come into effect, independently of whether the particular coupling relation is phonotactically legal.

3.5 Timing of coproduced gestures

While the main speech production models generally incorporate parallel activation of multiple competing candidates, the selection stage at the end of the phonological encoding process chooses a single candidate and the output is a linear sequence of segments. This view becomes problematic if some errors indeed result in a simultaneous production of two consonants, as we have claimed. An intrusion bias in errors has previously been identified in transcription studies (e.g., Butterworth and Whittaker 1980; Stemberger and Treiman 1986), yet these studies do not identify errors as *simultaneous* productions of the intended and the intruding segment but as errors creating a *sequential* consonant cluster. It should thus be considered whether these coproductions of source and target consonants can, at least in some cases, alternatively be interpreted as rapid self-corrections, that is, the consonants are in fact encoded and produced sequentially, but implemented along a continuum of overlap values on the surface. Conceivably, in the phrase *top cop*, an observed coproduction of /t/ and /k/ during an intended *top* could result from an erroneously planned *cop* being rapidly corrected to the intended *top*, yet too late to prevent errorful output. It is not immediately apparent why such a scenario would lead to a continuum of overlap values in which in some cases the gestures are produced simultaneously while in others either the errorful or the intended gesture slightly precedes the other one.

In the context of disordered speech, Hardcastle and Edwards (1992) informally observed variation in the timing patterns of errorfully coproduced gestures (identified on the basis of EPG) and suggest these may have diagnostic significance: an intruding gesture rapidly followed by the target gesture was taken to be indicative of error correction, that is, intact sensory feedback monitoring. A target gesture being rapidly followed by an intruding gesture or a simultaneous production of intruding and target gestures were interpreted as uncorrected errors, and taken to be indicative of a disruption of sensory feedback monitoring. What the thresholds should be that divide a continuum of timing values into different categories and different error mechanisms remains unclear in their paper. This interpretation can also not be directly extended to normal speakers, since there is no reason to assume disrupted sensory feedback (although e.g. high processing demands could conceivably perturb the processing of sensory feedback). Whether a continuum of timing values should indeed be interpreted as reflecting different error mechanisms thus remains to be investigated (cf. also McMillan et al. 2005 on feedback and editing in errors and Hartsuiker and Kolk 2001 for a discussion of time intervals for cut-off and repair times in speech production).

3.6 Error elicitation

Now we have come to a pivotal question in our consideration of the state-of the-art of error research. Due to the low occurrence of errors in natural speech (.1–.2%, Garnham et al. 1981), laboratory studies of errors employ by necessity error elicitation techniques. Many prominent (transcription or instrumental) studies of phonological errors have used rapid word repetition tasks, often similar to tongue twisters, to obtain a meaningful number of errors per subject, assuming that errors forced in the laboratory are comparable to those occurring spontaneously in natural settings. Stemberger (1992) cautioned against the use of tongue twisters in laboratory elicitation, yet Shattuck-Hufnagel (1983) obtained similar results from corpus data and tongue twister-like error elicitation: for both, errors could be described as segmental substitutions (cf. also Wilshire 1999). Renewed concerns that the occurrence of ill-formed errors is inflated by laboratory elicitation techniques are not unwarranted, yet several arguments have to be kept in mind. Laboratory elicitation techniques have been a popular research tool in speech error research since the 1970s, and the elicitation techniques employed by instrumental studies are hardly new (cf., among others, Croot et al. 2006; Dell et al. 2000; Goldrick 2004; Guest 2001; Shattuck-Hufnagel 1983; Wilshire 1999 for transcription studies using repetition based elicitation methods for the investigation of cognitive word form encoding). To the extent that there are caveats about what types of errors can be elicited in the laboratory, these apply to all laboratory studies alike, independently of whether errors are identified by means of transcription or acoustic/articulatory measurements. Note that Goldrick and Blumstein (2006) distinguished errorful from nonerrorful utterances on the basis of auditory evaluation first before performing their acoustic analyses on the two groups. They obtained comparable results to the other instrumental studies, in which errors were identified on the basis of acoustic or articulatory measures only, without taking perceptual criteria into account (Frisch and Wright 2002; Goldstein et al. in press; Mowrey and MacKay 1990; Pouplier in press). All instrumental studies have found that errors often are not wholesale substitutions but show properties of both the intended and the intruding consonant along a continuum of values.

Importantly, not all instrumental records of errors were obtained by employing overt continuous repetition to elicit errors: As mentioned above, coproduction errors could also be identified during utterances elicited with the SLIP paradigm—an elicitation method which is quite far removed from free speech occurring in natural settings, but nonetheless has traditionally been hypothesized to elicit phonological planning errors. A paper by Boucher (1994) is of special interest in this context: the errors he reports are the only articulatory (X-ray) observations of errors which are not from an error elicitation study, they are errors which occurred spontaneously during the recordings for Rochette's (1973) study on French consonants. Not surprisingly given the low frequency of spontaneously occurring errors, Boucher's body of evidence comprises two errors by a single speaker. His data again fit into the general pattern of results that has emerged from instrumental studies: Both the intended and the errorful articulation are present at the same time; it is not the case that one substitutes for the other—albeit only one of the articulations may be perceived. Also Wood's (1997) work on disordered speakers comes up again in this context, since she collected data during a sentence production task in which subjects had to spontaneously produce sentences on the basis of a picture or target word. For both simple picture naming as well as sentence production tasks, Wood found a significant number of MAGs.

The theoretical significance of the results of instrumental error studies overall hinges on the question of which kind of errors are elicited in the laboratory. However, irrespective of the outcome of that question transcription error studies need to be re-evaluated to some extent on the basis of the results instrumental studies have revealed so far: the evidence strongly suggests that popular laboratory elicitation methods may not accomplish what we have thought they do, yet at the same time we know that transcription studies of naturally occurring errors underrepresent at least certain types of errors. An estimate of how big (or small) a discrepancy there may be between error patterns observed under current laboratory elicitation methods and naturally occurring errors thus remains a critical topic for future research.

We have begun to address this central issue by developing novel methods for laboratory error elicitation which are designed to require successively more complex speech planning mechanisms (Pouplier et al. 2005). Tasks start with highly rhythmic repetitions of short tongue twister-like phrases (e.g., *top cop*), and gradually expand the complexity of planning required to include the semantic, syntactic, prosodic and phonological/phonetic decisions of spontaneous communicative speech. We test the hypothesis that the proportion of sublexical errors involving replacement of whole segments will increase, and the proportion of phonotactically illegal errors will decrease, as the planning process comes to resemble the demands of producing normal communicative utterances. The goal is to understand whether speech errors which occur at different stages in the speech production process will have different properties. If planning operates over purely symbolic segmental units, there should be at least some errors that arise in planning due to the replacement of one symbolic unit with another. Our data so far from an EMMA experiment with a single subject showed different error rates for elicitation conditions which differed in whether a prosodic phrase boundary was present or absent. Our results so far suggest that the amount of prosodic planning required does influence error rates as well as the ratio of different error types to each other. While we observed more substitution errors compared to coproduction errors in the increased prosodic planning conditions, in all conditions coproduction errors outnumbered substitution errors.

4. Conclusions

Instrumental studies of speech errors elicited in the laboratory have shown that the outcome of an error is often not executed normally. The most significant contribution of instrumental studies of slips of the tongue has been to show that many errors involve the production of an “extra,” erroneous speech unit simultaneously with the intended one and thus both the intended and intruding unit can be present (to varying degrees) in the speech output. These errors may go perceptually undetected, or, if they are perceived, the nature of the articulatory event and thus the error may not always be identifiable. That errors show properties of both the intended and the intruding segment is most problematic for speech production models which posit a strict dichotomy between phonological and phonetic encoding, but is predicted by models which do not uphold this sharp distinction between phonological and phonetic planning, such as the gestural model or a cascading activation model.

Instrumental investigations have shown that ill-formed errors are of significant frequency (in the statistical as well as theoretical sense), and that it is thus precisely these errors which can afford novel insights into the architecture of the speech production process. However, the amount of data available at the moment from instrumental studies is sparse, especially so in comparison with decades of highly productive transcription based work. For a better understanding of the origins of both ill-formed and well-formed errors, we need to increase our understanding of which kinds of errors are triggered in the laboratory by current elicitation methods. Also articulatory, acoustic and perceptual aspects of errors will ultimately have to be integrated in order to gain a more complete picture of the nature of anomalous utterances.

References

Blumstein, Sheila (1973). *A phonological investigation of aphasic speech*. The Hague: Mouton.

Boomer, Donald S. and John D. M. Laver (1968). Slips of the tongue. *British Journal of Disorders of Communication* 3, 2–12.

Boucher, Victor J. (1994). Alphabet-related biases in psycholinguistic enquiries: Considerations for direct theories of speech production and perception. *Journal of Phonetics* 22, 1–18.

Browman, Catherine and Louis Goldstein (1992). Articulatory phonology: An overview. *Phonetica* 49, 155–180.

Browman, Catherine and Louis Goldstein (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5, 25–34.

Butterworth, Brian and Steve Whittaker (1980). Peggy Babcock's relatives. In *Tutorials in motor behavior,* vol. 1, ed. G. E. Stelmach and J. Requin, 647–656. Amsterdam: North-Holland.

Croot, Karen, Claudia Au and Amy Harper (2006). Prosodic structure in tongue twister errors. Paper presented at Laboratory Phonology X, Paris.

Cutler, Anne (1981). The reliability of speech error data. *Linguistics* 19, 561–582.

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

Dell, Gary, Cornell Juliano and Anita Govindjee (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17, 149–195.

Dell, Gary, Kristopher Reed, David Adams and Antje Meyer (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26, 1355–1367.

Dell, Gary and Peter Reich (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20, 611–629.

Dell, Gary and Renee J. Repka (1982). Errors in inner speech. In *Experimental slips and human error: Exploring the architecture of volition,* ed. B. J. Baars, 237–262. New York: Plenum.

Fowler, Carol (1995). Speech production. In *Handbook of perception and cognition: Speech, language and communication,* vol. 11, ed. J. Miller and P. Eimas, 29–61. New York: Academic Press.

Frisch, Stefan A. (2007). Walking the tightrope between cognition and articulation: The state of the art in the phonetics of speech errors. This volume.

Frisch, Stefan A. and Richard Wright (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30, 139–162.

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

Fromkin, Victoria A. (ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Garnham, Alan, Richard C. Shillock, Gordon D. A. Brown, Andrew I. D. Mill and Anne Cutler (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics* 19, 805–817.

Goldrick, Matthew (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51, 586–603.

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

Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot Saltzman and Dani Byrd (in press). Dynamic action units slip in speech production errors. *Cognition*.

Guest, Daniel James (2001). Phonetic features in language production: An experimental examination of phonetic feature errors. Ph.D. dissertation, UIUC.

Hardcastle, William and S. Edwards (1992). EPG-based descriptions of aphasic speech errors. In *Intelligibility in speech disorders: Theory, measurement, and management,* ed. R. D. Kent, 287–328. Philadelphia: John Benjamins.

Hartsuiker, Robert, Martin Corley and Heike Martensen (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language* 52, 58–70.

Hartsuiker, Robert and Herman H.J. Kolk (2001). Monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42, 113–157.

Kelso, J. A. Scott, Elliot Saltzman and Betty Tuller (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 14, 29–59.

Laver, John (1979). Slips of the tongue as neuromuscular evidence for a model of speech production. In *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler,* ed. H. W. Dechert and M. Raupach, 21–26. The Hague: Mouton.

Levelt, Willem (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, Willem, Ardi Roelofs and Antje Meyer (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–75.

McMillan, Corey, Martin Corley, Robin Lickley and Robert Hartsuiker (2005). Relative contributions of feedback and editing in language production: Behavioural and articulatory evidence. *Journal of the Acoustical Society of America* 117, 2618 [abstract].

Meyer, Antje (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition* 42, 181–211.

Motley, Michael T. and Bernard J. Baars (1976a). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly* 24, 28–34.

Motley, Michael T. and Bernard J. Baars (1976b). Semantic bias effects on the outcomes of verbal slips. *Cognition* 4, 177–187.

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

Pouplier, Marianne (in press). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*.
Pouplier, Marianne and Louis Goldstein (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics* 33, 47–75.
Pouplier, Marianne and William Hardcastle (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica* 62, 227–243.
Pouplier, Marianne, Mark Tiede, Stefanie Shattuck-Hufnagel, Man Gao and Louis Goldstein (2005). The role of intonational phrasing in the elicitation of speech errors. *Journal of the Acoustical Society of America* 117, 2622 [Abstract].
Roberts, E.W. (1975). Speech errors as evidence for the reality of phonological units. *Lingua* 35, 263–296.
Rochette, Claude E. (1973). *Groupe de consonnes en français: étude de l'enchaînement articulatoire à l'aide de la radiocenématographie et de l'oscillographie*. Québec: Presses de l'Université Laval.
Schwartz, Myrna F., Eleanor M. Saffran, Diane E. Bloch and Gary S. Dell (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language* 47, 52–88.
Shattuck-Hufnagel, Stefanie (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper and E. C. T. Walker, 295–342. Hillsdale, NJ: Lawrence Erlbaum.
Shattuck-Hufnagel, Stefanie (1983). Sublexical units and suprasegmental structure in speech production planning. In *The production of speech,* ed. P. F. MacNeilage, 109–136. New York: Springer.
Shattuck-Hufnagel, Stefanie and Dennis Klatt (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18, 41–55.
Stemberger, Joseph (1992). The reliability and replicability of naturalistic speech error data. A comparison with experimentally induced errors. In *Experimental slips and human error: Exploring the architecture of volition,* ed. B. J. Baars, 195–215. New York: Plenum Press.
Stemberger, Joseph and Rebecca Treiman (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language* 25, 163–180.
Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in action*. Amsterdam: N.-Holland.
Stone, Maureen (1997). Laboratory techniques for investigating speech articulation. In *The handbook of phonetic sciences,* ed. William J. Hardcastle and John Laver, 11–32. Oxford: Blackwell.
Talo, Ewa Söderpalm (1980). Slips of the tongue in normal and pathological speech. In *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand,* ed. Victoria A. Fromkin, 81–86. New York: Academic Press.
Turvey, Michael T. (1990). Coordination. *American Psychologist* 45, 938–953.
Wells, Rulon. (1951). Predicting slips of the tongue. *Yale Scientific Magazine* 26(3), 9–30.
Wheeldon, Linda and Willem Levelt (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34, 311–334.
Wilshire, Carolyn E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech* 42, 57–82.
Wilshire, Carolyn E. (2002). Where do aphasic phonological errors come from? Evidence from phoneme movement errors in picture naming. *Aphasiology* 16, 169–197.
Wood, Sarah (1997). Electropalatographic study of speech sound errors in adults with acquired aphasia. Ph.D. dissertation, Queen Margaret University College, Edinburgh.

Wood, Sarah and William Hardcastle (1999). EPG study of lingual errors in adults with acquired aphasia: Implications for models of speech production. *Proceedings of the XIVth International Congress of Phonetic Sciences* (ICPhS), San Francisco, 1341–1344.

Linguistics and English Language
University of Edinburgh
Adam Ferguson Building
40 George Square
Edinburgh EH8 9LL
United Kingdom

pouplier@ling.ed.ac.uk