Self-Monitoring and Self-Repair in Spontaneous Speech

Stefanie Pillai
English Language Department, Faculty of Languages and Linguistics
University of Malaya, Kuala Lumpur, Malaysia
e-mail: stefanie@um.edu.my

***Abstract**: This study explores what repairs in the spontaneous production of speech reveal about the psycholinguistic processes of self-monitoring and self-repair. Three intervals were examined: error-to-cut off; cut off-to-repair; error-to-repair. The intervals indicate support theories of internal speech monitoring, and also indicate that the planning of speech-repairs can take place pre-articulatorily as well*

***Key words**: error-detection, Perceptual Loop Theory, self-monitoring, speech production, self-repairs.*

Self-repairs are self-initiated corrections of one's own speech within the same speaking turn (Postma, 2000; Schegloff, Jefferson, & Sacks, 1977). They are a normal phenomenon in spontaneous speech, and are produced in response to a linguistic problem, such as the inability to retrieve lexical items, and the incorrect use of pronunciation, lexis or syntax. These problems can be overtly detected, but it has also been posited that they can be detected in inner or pre-articulatory speech by some form of speech monitoring mechanism inherent in the speech production process (Laver 1969 & 1980; Levelt, 1983 & 1989; van Wijk & Kempen, 1987). Repairs can be produced with the related problem being partly produced, where speakers cut their speech off in the midst of a word. Alternatively, speakers may produce hesitation in their speech, such as filled pauses (e.g. *ah*, *ahm*, *er*), silent pauses and prolonged segments (Pillai, 2004). A simplified diagram of the process of self-monitoring is shown in Figure 1.

Problem detected → (Speech interrupted) → (Hesitation) → (Self-Repair produced)

Note: The items in parenthesis are optional

Figure 1. The Process of Error-Detection, Hesitation and Self-Repair in Speech Production

It has been shown that self-repairs can be regarded as a manifestation of a "quality control" (Hieke, 1981, p. 148), a mechanism present in the process of speech production to correct pre-articulatory or post-articulatory errors. In order for self-repairs to take place, there must be an awareness that an error is about to be, or has been produced by the speaker. Hence, the concept of self-repair is consistent with the idea that self-monitoring occurs in the process of speech production. There have been attempts to explain the relationship between monitoring and self-repair in speech, particularly to account for the following phenomena:

(i) how and when errors are detected and corrected
(ii) how soon after error detection speech is interrupted

One of the main theoretical models explaining how speech is monitored and repaired is the Perceptual Loop Theory (Levelt, 1983; 1989). Based on a corpus of repairs made in the spontaneous speech of adult speakers of Dutch, Levelt (1983, 1989) formulated a theory to account for both monitoring and repairing in speech. The theory is based on the premise that speakers monitor their own speech just as they monitor the speech of others (Levelt, 1983; Levelt, Roelofs, & Meyer, 1999). Levelt divided self-repairs into three major phases:

(i) monitoring and interrupting speech whenever trouble is detected
(ii) hesitating and pausing (characterised by the use of silent or filled pauses)
(iii) repairing disfluent speech.

In Levelt's speech production model, the generation of an idea or message of an intended utterance occurs at the Conceptualizer. At this stage of conceptualization, the message can be monitored, for example for appropriateness. The speaker might need to decide, for instance, on the right choice of word to express a particular idea based on his knowledge of the social rules governing language use. If the process of monitoring at this stage finds the message to be inappropriate for some reason, a new message can be generated. A preverbal message that goes through this stage goes in as input into the Formulator, which turns this concept into a linguistic structure. This is done through the process of lemma selection, where a lemma is retrieved from the mental lexicon (where information about the lemma's meaning, syntactic, morphological and phonological features are available). Thus, the lemma can be grammatically encoded, producing a surface structure for the message. This process involves, accessing information about grammatical form and the features associated with it such as person, number, tense and aspect.

The surface structure is then phonologically encoded, where syllabification of the structure is thought to take place (Levelt, 2001; Levelt, Roelofs, & Meyer, 1999). At this stage, the individual sound segments that make up the intended word as well as its syllabic structure are thought to be put together. The phonological word is then phonetically encoded, which results in a phonetic plan or "gestural" (Levelt, Roelofs, & Meyer, 1999, p. 33) or "articulatory score" (Levelt, 2001). This plan specifies how the word is to be articulated by the speech organs. This phonetic plan is referred to as "internal speech". Since the process of language production occurs at an extremely fast pace, there is thought to be an "Articulatory Buffer" (Levelt, 1989, p. 12) between the Formulator and the Articulator, where the phonetic plan is temporarily stored. While the Articulator is executing a phonetic plan, the next message to be articulated will be stored in this buffer, while waiting to be retrieved and executed (Levelt, 1989, p. 13).

Internal speech then goes into the Articulator, and once the articulators go into motion, audible speech is produced. According to Levelt's (1989) model, audible self-produced overt speech goes through the Speech-Comprehension System, where it is processed in the same way we process other people's speech that we hear. This system also processes internal speech or pre-articulatory speech, and it is suggested that this is where error detection for both internal and overt speech occurs. Levelt, Roelofs and Meyer (1999, p. 33) illustrate this point with the following examples:

i) Entrance to yellow …er, to grey
ii) We can go straight to the ye-… to the orange dot

In the first example, it is suggested that overt speech is being monitored since the speaker can hear the error, that is, *yellow*. In contrast, in the second example, the speaker interrupts after the first syllable of the word *yellow*, suggesting that the error was already detected prior to its utterance. This detection is possible presumably because internal speech is being monitored. Partially intercepted errors (Levelt, 1989, p. 474), such as in the second example above, which are auditorily perceived as fragments (Blackmer & Mitton, 1991; Nakatani & Hirschberg, 1994; Shriberg, 1994), and the interruption of speech by various forms of hesitation in speech, are said to be indications of pre-articulatory editing. Other evidence of pre-articulatory editing comes from studies, where subjects reported that they detected errors in their inner speech (Dell & Repka, 1992). Postma and Noordanus (1996) also found that apart from normal speech production, subjects reported errors in silent, mouthed, and noise-masked speech production, implying that speakers need not hear their errors to correct them.

To understand the process of self-monitoring and self-repair, the time intervals involving errors, the point of interruption and self-repairs have been previously investigated. The intervals, as shown in Figure 2 are as follows:

(i) error-to-cut off
(ii) cut off-to-repair
(iii) error-to-repair

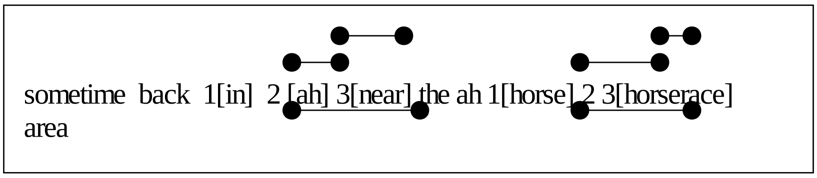


Figure 2. Time Intervals in Self-Repairs (Adapted from Blackmer & Mitton, 1991)

The interval between 1 (the error) and 2 (perceived cut off or interruption point) is the error-to-cut off time. The interval between 2 and 3 (the onset of repair) is the cut off-to repair time. The interval between 1 and 3 is the error-to-repair time. There can be an editing phase between 2 and 3 containing a silent or filled pause.

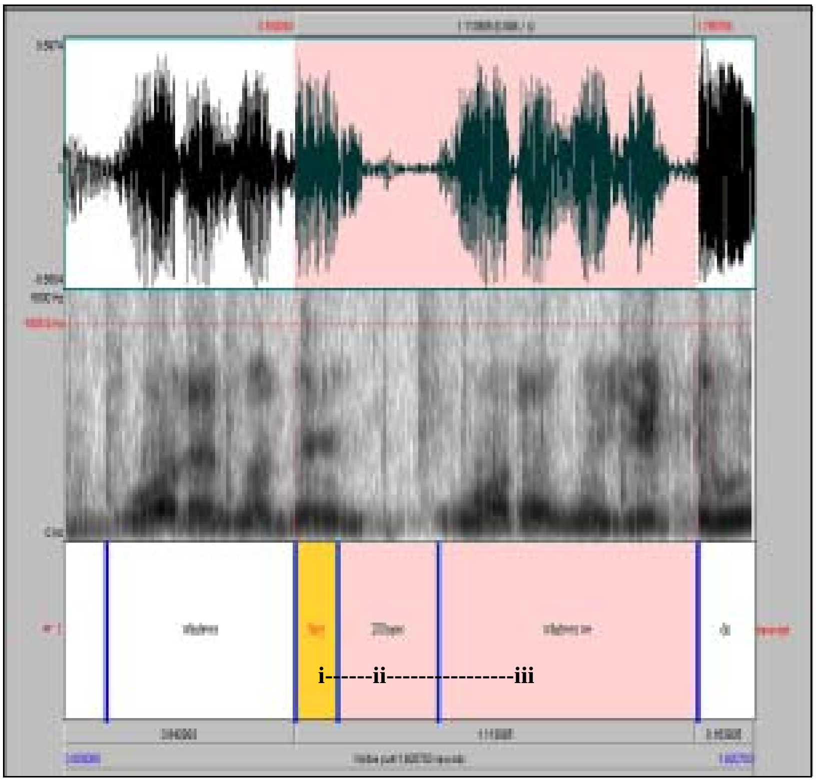
While the Perceptual Loop Theory accounts for short error-to-cut off intervals (as low as less than 100msec) because of the presence of inner speech monitoring, it does not adequately account for short cut off-to-repair and error-to-repair intervals. Further, the Perceptual Loop Theory assumes that repair-planning begins only upon interruption of speech which would not explain short error-to-repair intervals.

In relation to these inadequacies in the theoretical understanding of the processes of self-monitoring and self-repair in speech production, this study examines self-repairs in naturally occurring speech, as opposed to experimentally induced speech. It is felt that a study on naturally occurring self-repairs can contribute empirical data to support or discount current theories of self-monitoring in speech. More specifically, this study examines what the three time intervals reveal about the processes of self-monitoring and self-repair in the spontaneous production of speech.

METHODOLOGY

The data consist of recordings of sixty-seven callers to a radio show. The conversation between the presenters and the callers were transcribed orthographically. However, only the utterances of the callers were analysed for the purposes of this study since their speech was more likely to be unscripted. Most of the transcription conventions used in this study were adapted from Jefferson's transcription system as outlined in Atkinson and Heritage (1984).

To understand the mechanisms involved in self-monitoring and self-repair, the three intervals presented in Figure 2 were measured based on spectrograms and waveforms as well as auditory examination of the related utterances using *PRAAT* (Boersmal & Weenink, 2005). As shown in Figure 3, measurements were taken from the onset to the offset of an error for error-to-cut off intervals (see i-ii in Figure 3). The repair was deemed to begin from the offset of an editing phase. In the absence of an editing phase, it was taken to begin at the offset of an error. Thus, cut off-to-repair intervals were measured from the onset of an editing phase or offset of the error (or prolonged segment, if there was one) to the onset of a repair (see ii-iii in Figure 3). Error-to-repair intervals were measured from the onset of an error to the onset of a repair (see i-iii in Figure 3). Utterances which only had within-utterance hesitations but no overt errors or repairs were considered as *possible-repairs*, since there was no direct evidence to suggest that a repair had been made (Hartsuiker & Kolk, 2001; Levelt, 1983). For such repairs only the cut off-to-repair interval was measured since there was no visible error in the utterance. For instances of repeats, the items preceding the cut off point was considered as the *error*, and the one following this point or the editing phase was considered as the *repair*.



(i-ii = error-to-cut off; ii-iii = cut off-to-repair; i-iii = error-to-repair)

Figure 3. Measurements for Intervals

FINDINGS AND DISCUSSION

There were a total of 264 of the utterances that were interrupted by some form of hesitation. No repair was made after hesitations in 138 instances, which are referred to as *possible-repairs* in this study. The other 126 instances that were interrupted by some form of hesitation were regarded as instances of self-repair. An additional of 113 repairs did not have any form of hesitation. Thus, in total there were 239 self-repairs and 138 possible-repairs found in the data.

Error-to-Cut off Intervals

In this study, the error-to cut off intervals for all 239 self-repairs had a mean of 347msec, a median of 314msec, a mode of 175msec and a standard deviation of 231msec. The shortest error-to-cut off interval was 15msec, while the longest was 1785msec. Most of the error-to-cut off intervals (about 97%) were 800msec and below, and long intervals above 1 sec rarely occurred. Although all self-repairs had intervals below 100msec, only approximately 7% of error-to-cut off intervals in self-repairs were below 100msec, and about 14% were below 150msec.

It was more likely for speech to be cut off within 800 msec after an error had been produced. In approximately half (51%) of self-repairs, speech was interrupted at about 400msec and below after the production of the error. The implication of this is that speech is not stopped immediately upon detection of a problem or production of an error. This is because there needs to be a time-frame for the stop signal to be sent to the articulators upon error-detection, keeping in mind that inner speech recognition is thought to take about 150 to 200msec (Levelt, 1989). Given that it is estimated to take about 180 to 200msec to stop articulation (Ladefoged, Silverstein, & Papcurn, 1973), speakers seemed to have a tendency to go on speaking a little while longer before they interrupted themselves. This means that in most cases, the *Main Interruption Rule* (Nooteboom, 1980), that speech is interrupted *immediately* upon detection of an error, was not applied. The term *immediately*, however should be taken within the context of the estimated speed at which speech can be instructed to stop (Hartsuiker & Kolk, 2001, p. 118). Perhaps this is also because, as suggested by Seyfeddinipur and Kita (2001), pre-articulatory repair-planning has commenced and is going on while speakers continue their utterance.

For short error-to-cut off intervals, error detection must have taken place pre-articulatorily, as explained by Levelt's Perceptual Loop Theory. Empirically, evidence for pre-articulatory monitoring also stems from the fact that a high percentage of self-repairs with error-to-cut off times of less than 150ms were fragment repairs (67%). More than a quarter of these fragmented repairs had error-to-cut off times of less than 150ms. This means that, in these cases, the error was interrupted mid-segment within 150msec, suggesting that the decision to stop speech must have been made earlier. This in turn indicates that error-detection must have occurred prior to speech being interrupted.

Monitoring of inner speech is not a new finding as it has been reported in other studies. For instance, experimental studies by Dell and Repka (1992), Levelt, Roelofs and Meyer (1999), and Postma and Noordanus (1996) have reported that speakers are able to monitor their internal speech, perhaps even as early as at the abstract phonological level.

Cut off-to-Repair Intervals

The intervals for the possible-repairs and self-repairs had a mean of 257msec, median of 90msec, mode of 0msec and a standard deviation of 364msec. Individually, self-repairs had a mean of 131msec, median and mode of 0msec, and a standard deviation of 261msec. In comparison, possible-repairs had a mean of 472msec, median of 391msec, mode of 0msec and a standard deviation of 415msec. Most of the intervals for possible-repairs were 600msec and below compared to most of them being 200msec and below for self-repairs. In fact, a *Mann-Whitney U Test* showed that there was a significant difference between the cut off-to-repair intervals for self-repairs and possible-repairs, U ($N_1 = 239$, $N_2 = 138$) 6637, $p < .001$.

More than 50% of self-repairs had 0msec cut off-to-repair intervals compared to only about 15% of the possible-repairs. Similarly, all self-repairs had a much higher percentage of intervals that were below 100msec (71%) and 250msec (79%), compared to possible-repairs, where the intervals, in general, tended to be longer.

In possible-repairs, it is assumed that the error has been detected early enough for it to be prevented from being articulated. According to Levelt's model (1989, p. 473), given that it takes about 300-350msec from the delivery of the phonetic plan to the start of articulation and about 150 to 200msec for inner speech recognition, the internal loop has between 0-100msec to instruct the articulator to stop, if an error is detected in inner speech. This results in an interruption in the flow of speech through the use of hesitation devices such as silent pauses, filled pauses or lengthening of sounds. Thus, possible-repairs presuppose both pre-articulatory error-detection and repair-planning.

Indeed such short intervals in repairs question the notion that post-articulatory error-detection and repair-planning only commences upon the cessation of speech. Instead, these processes must have been ongoing even during the production of speech prior to the cut off point. (Blackmer &

Mitton, 1991). If we once again consider the times estimated for error detection (100msec) and the stop signal to be sent to the articulator (180msec), and take away the total of this from the total duration of the intervals in this study, it would leave approximately 170msec and below for most of the intervals for re-planning.

Nakatani and Hirschberg (1994) suggest that perhaps this interval between the production of an error to the production of the repair need not correspond to re-planning time, implying that repair-planning may have begun earlier. Hartsuiker and Kolk's (2001) experimental studies suggest that the process of interruption and repair are simultaneously triggered when a problem is detected, which means that a repair can be ready upon or soon after interruption.

Error-to-Repair Intervals

The interval for all the self-repairs had a mean duration of 506msec, a median of 373msec, a mode of 265msec and a standard deviation of 537msec. Similar to error-to-cut off and cut off-to repair intervals, long error-to-repair intervals were not common, with most of the intervals falling below 1 second. Error-to-repair intervals of less than 200msec accounted for 19% of all self-repairs, and 20% (22) of self-repairs if repeats were excluded. The latter figure is almost twice the percentage cited by Blackmer and Mitton (1991), who found that 10% of their overt repairs had error-to-repair intervals below 200msec. The mean duration of self-repairs without the repetitions of 595msec is shorter than the mean of 838msec for overt repairs found in Blackmer and Mitton (1991).

Short error-to-repair times again indicate that the planning of the repair is going on while speech is being produced (Blackmer and Mitton, 1991). Blackmer and Mitton (1991) suggest that speech stored in the Articulatory Buffer means that there is time for inner speech to be monitored, which allows the decision to interrupt speech to be made and subsequently the repair is ready for articulation by the time or soon after the error is articulated. Hartsuiker and Kolk's (2001) monitoring model, where the process of interruption and repair-planning are triggered off simultaneously upon error-detection, could also be used to explain fast error-to-repair intervals.

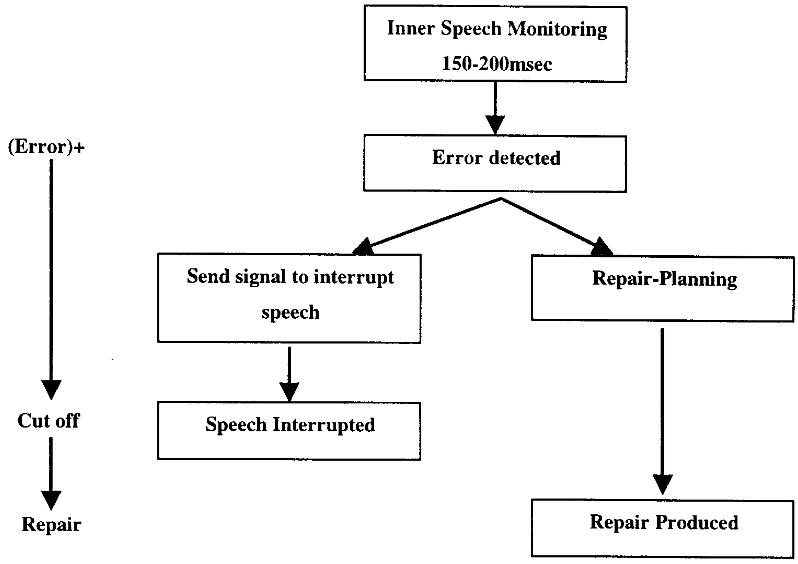
CONCLUSION

Short error-to-cut off intervals show that the decision to interrupt speech seems to have been made soon after the detection of an error, consistent with the Main Interruption Rule. This means that even if errors have been detected, they may still be overtly produced if the articulators had already begun to produce the part of the message containing the error before it received the signal to interrupt speech. Thus, overt errors, particularly those that are cut off as fragments, may have been detected pre-articulatorily rather than post-articulatorily.

The fact that almost half or more of all self-repairs had 0msec error-to-repair intervals means that once speech was interrupted, speakers were able to produce repairs almost immediately, which strongly suggests that the repair-planning must have commenced prior to, and not upon interruption. Thus, perhaps as suggested by Hartsuiker and Kolk (2001), the processes of error-detection, speech interruption and repair-planning are not serial, carry forward type of processes but parallel ones, where in the process of monitoring internal speech, error detection may trigger both the instructions to stop speech and to commence repair-planning. It was also not uncommon to find error-to-repair intervals of below 200msec, further indicating the possibility of pre-articulatory error-detection and of re-planning commencing before interruption. This possible process of error-detection, speech-interruption and repair-planning is summarised in Figure 4.

Future Directions

The time intervals found in this study indicate the presence of both pre-articulatory error-detection and repair-planning commencing before the point of interruption. While the former can be explained by production the Perceptual Loop theory, more cross-linguistic and experimental research is needed to further study the mechanisms involved in speech-monitoring and repair-planning. A combination of data comprising points of brain activation and outward production of disfluencies may be able to provide a better picture of the entire process of speech-monitoring, including error-detection, speech-interruption and repair-planning.



+ the error will be audible if arrival of the speech interruption signal precedes its articulation

Figure 4. The Process of Error-Detection, Speech-Interruption and Repair-Planning

REFERENCES

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39,* 173-194.

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer.* Retrieved December 30, 2005 from http://www.praat.org/

Atkinson, J. M., & Heritage, J. (1984). Jefferson's transcript notation. In A. Jaworski & N. Coupland (Eds.) (1999), *The discourse reader.* (pp.158-166). New York: Routledge.

Dell, G. S., & Repka, R.J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition.* (pp. 237-262). New York: Plenum Press.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42,* 113-157.

Hieke, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech, 24,* 147-160.

Kormos, J. (2000). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition, 22,* 145-167.

Ladefoged, P., Silverstein, R., & Papcun, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America, 54,* 1105-1108.

Laver, J. D. M. (1969). The detection and correction of slips of the tongue. In V. A. Fromkin (Ed.) (1973), *Speech errors as linguistic evidence*. (pp. 132-143). The Hague: Mouton.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 287-306). New York & London: Academic Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 14-104.

Levelt, W. J. M. (1989). *Speaking: From Intention to articulation*. Cambridge, Massachusetts: MIT Press.

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

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair in spontaneous speech. *Journal of the Acoustical Society of America 95*(3), 1603-1616.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* (pp. 87-96). New York & London: Academic Press.

Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research, 30,* 163-184.

Pillai, S. (2004). *Patterns of disfluencies and the process of self-monitoring in spontaneous speech.* Unpublished doctoral dissertation. University of Malaya**,** Malaysia.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*(2), 97-132.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39*(4), 375-392.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361-382.

Seyfiddinipur, M., & Kita, S. (2001). *Proceedings of the Disfluency in Spontaneous Speech ISCA Tutorial and Workshop*. Gesture as an indication of early error detection in self-monitoring of speech. Edinburgh: University of Edinburgh.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Doctoral thesis. Unpublished doctoral dissertation. University of California, Berkeley.

van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology, 19*, 403-440.