Journal of Memory and Language 69 (2013) 417–428

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Parallels between self-monitoring for speech errors and identification of the misspoken segments

Sieb G. Nooteboom *, Hugo Quené

Utrecht University, Utrecht Institute of Linguistics OTS, Trans 10, 3512 JK Utrecht, The Netherlands

A R T I C L E   I N F O

Article history:
Received 17 December 2012
revision received 19 April 2013
Available online 22 May 2013

Keywords:
Speech errors
Self-monitoring
Speech perception
Identification
Reaction times

A B S T R A C T

This paper investigates self-monitoring for speech errors by means of consonant identification in speech fragments excised from speech errors and their correct controls, as obtained in earlier experiments eliciting spoonerisms. Upon elicitation, segmental speech errors had been either not detected, or early detected or late detected and repaired by the speakers. Results show that misidentifications are rare but more frequent for speech errors than for control fragments. Early detected errors have fewer misidentifications than late detected errors. Reaction times for correct identifications betray effects of varying perceptual ambiguity. Early detected errors result in reaction times that are even faster than those of correct controls, while late detected errors have the longest reaction times. We speculate that in early detected errors speech is initiated before conflict with the correct target arises, and that in both early and late detected errors conflict between competing segments has led to detection.

Introduction

Speakers occasionally produce erroneous speech sounds. Does the speech sound resulting from an error constitute a categorically different speech sound, or is it a blend of competing speech segments? Speakers may also correct their speech errors. Does self-monitoring involve inspecting production processes during speech preparation or does it involve inspecting the end products of production processes by employing speech perception? In order to answer these questions, this paper describes an attempt to investigate aspects of speakers' self-monitoring for speech errors in an indirect way. To that end we had listeners identify speech fragments containing segmental speech errors and fragments containing corresponding correctly spoken control fragments. The errors had either been not detected (*bood geer*), or early detected and repaired (*boo...good beer*) or late detected and repaired (*bood geer...good beer*) by the original speakers. Listeners' error rates and reaction times in identifying these segments, taken from various types of speech errors, may provide answers to the questions above.

For a survey of models of self-monitoring the reader is referred to Postma (2000) and Nozari, Dell, and Schwartz (2011). In the course of this paper we will mainly focus on the differences between perception-based monitoring as exemplified in the perceptual loop theory of self-monitoring by Levelt (1989) and Levelt, Roelofs, and Meyer (1999) on the one hand and conflict-based monitoring as proposed by Nozari et al. (2011) on the other. We assume that from the perceptual loop theory one may infer that self-monitoring employs perceptual properties of speech sounds in error detection. We also assume that self-monitoring mainly employs perceptual comparison between error form and correct target form, as suggested by Nooteboom (2005a, 2005b) and Nooteboom and Quené (2008). It should be noted that the latter position has been criticized by McMillan and Corley (2010). They have

* Corresponding author. Address: Cor Ruyslaan 20, 3584 gd Utrecht, The Netherlands.

E-mail address: S.G.Nooteboom@uu.nl (S.G. Nooteboom).

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.04.006

difficulty to see how such a comparison between error and correct target could fit in the cascading framework they propose, in which competing segments can be simultaneously activated and can activate simultaneous conflicting articulatory gestures. In that framework segmental errors are not all-or-none. This would complicate comparison between intended targets and error forms. McMillan and Corley also wonder why, if the correct target is available for comparison with the target form, an error form was generated in the first place. However, these objections notwithstanding, if self-monitoring would employ comparison between error form and target form, then one would predict that the probability of error detection increases with perceptual distance between error form and target form. This is different for conflict-based monitoring as proposed by Nozari et al. (2011).

Nozari et al. (2011) reject perception-based monitoring because of the often reported double dissociation between speech error detection and speech perception in aphasic patients (e.g. Butterworth & Howard, 1987; Liss, 1998; Marshall, Rapaport, & Garcia-Bunuel, 1985; Marshall, Robson, Pring, & Chiat, 1998). Their conflict-based monitor for speech errors is computationally implemented in the two-stage word production model described by Dell (1986), it is production-based, and monitors for conflict of activation between simultaneously activated units during speech preparation. This proposed self-monitoring system would fit in well with the cascading framework proposed by McMillan and Corley (2010). These authors suggest that the conflict among multiple simultaneously active segments, competing for the same slot in inner speech, may cascade down to articulation. This would then result in articulatory blending. The theory by Nozari et al. predicts that (in normal speakers) the probability of error detection increases with increasing amount of conflict between simultaneously activated units. Combining the conflict-based theory of Nozari et al. (2011) and the cascading of activation proposed by McMillan and Corley (2010), we see that according to the conflict-based theory of monitoring the probability of error detection increases with amount of articulatory blending. This is interesting because whereas it is difficult to measure the amount of conflict of activation (but see Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, Botvinick, & Cohen, 2004), we can in principle assess the amount of articulatory blending by means of a perception experiment, as we will report below.

Most speech errors are errors against single speech segments; roughly half of these segmental speech errors are detected and repaired by the speakers (cf. Nooteboom, 1980, 2005a; Nooteboom & Quené, 2008). Until not very long ago people studying segmental speech errors seemed to work from the assumption that most segmental speech errors are categorical in nature, that is that they consisted of the substitution, deletion or addition of complete segments. This despite the fact that in the last few decades it has been shown repeatedly that blends of simultaneously pronounced speech sounds are not infrequent. For example, Mowrey and MacKay (1990) demonstrated that segmental errors of speech elicited in the laboratory, every now and then contain electromyographic evidence of simultaneous competing segments. In an acoustic study of elicited confusions between /s/ and /z/, Frisch and Wright (2002) found that both categorical and gradient errors occurred, although categorical errors were more frequent than one would expect if they were just extreme examples of gradient voicing errors. Goldrick and Blumstein (2006), focusing on voice onset time in voiced and voiceless consonants, also found that in elicited segmental speech errors acoustic traces of the competing segments can be found. From these studies it seems apparent that many speech errors are gradient and not categorical, although the results still do not exclude the possibility that the majority of speech errors is categorical. More recently it has been demonstrated that (at least in a particular experimental setting) most segmental speech errors are not categorical errors but rather blends of competing articulatory gestures (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Pouplier, 2007; McMillan & Corley, 2010). It is at this stage not clear what causes the discrepancy between the earlier studies and the more recent studies, but all these studies agree that gradient segmental speech errors are frequent. The fact that, despite the relative frequency of articulatory blending, in the past canonical speech errors were assumed to be categorical instead of gradient may be attributed to perceptual illusions during transcription (Pouplier & Goldstein, 2005): Our perception is categorical also when the perceived produced speech segments are not.

If most speech errors indeed are articulatory blends of competing segments, this is likely to have consequences both for detecting such speech errors in self-monitoring and for the perceptual properties of these errors. In this paper we present an experiment exploring parallels between self-monitoring for segmental speech errors and perceptual identification of the misspoken segments, and testing some hypotheses stemming from the supposed blended origin of segmental speech errors. It should be noted that the prevalence among segmental speech errors of articulatory blends supports a cascading model of speech preparation in which segments may compete for the same slot in the speech plan (McMillan & Corley, 2010). This, in turn, makes the proposal by Nozari et al. (2011) of conflict-based monitoring for speech errors seem realistic.

The basic idea underlying this paper is the following. If Goldstein, Pouplier, Chen, Saltzman, and Byrd (2007) and McMillan and Corley (2010) are right, then speech segments resulting from segmental errors of speech often must carry the acoustic consequences of the articulatory blending of speech sounds. These acoustic consequences of articulatory blending must in turn have perceptual consequences, even if very often these consequences are not reflected in auditory transcription (cf. Pouplier & Goldstein, 2005; McMillan, 2008). If we excise speech fragments containing the erroneous segments from elicited speech errors and offer these speech fragments, together with speech fragments excised from correct controls (no speech errors), in a simple speech segment identification experiment, then the perceptual consequences of the assumed articulatory blending may become apparent in two dependent measures, viz. frequency of misidentifications and reaction