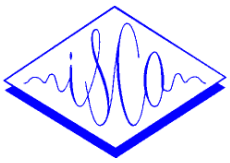

ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

Application of EXPLAN theory to spontaneous speech control

Peter Howell and James Au-Yeung

Department of Psychology
University College London
p.howell@ucl.ac.uk james@psychol.ucl.ac.uk

Abstract

Problems for theories that explain speech errors by a monitoring process are discussed. EXPLAN theory is based on a proposal about planning and execution time, not on how errors arise. This theory is outlined and support from characteristics of fluency failure and altered feedback studies given.

1. Introduction

Spontaneous speech output differs from read output of a prepared text insofar as fluency failure is more likely. This difference may arise because a plan is supplied when a text has to be read whereas the plan has to be generated on the fly in spontaneous productions. The material that has to be read is usually syntactically correct and such a "grammatically-correct" model is taken as the benchmark against which to assess spontaneous productions that are likely to be ungrammatical (for instance Levelt's well-formedness rule for speech repairs) [1]. Most studies on read text use English, but English orthography allows little scope for prescribing the prosodic plan other than a crude indication of location of pauses, stress etc. in the punctuation. This may be why the prosodic structure differs between a text produced spontaneously and a transcript of this text read by the same speaker [2]. Two conclusions can be drawn: (1) Clarification is needed about how fluency failure arises, that looks at the dynamics of fluency failure with respect to ongoing production rather than by reference to exceptions relative to a grammatical model. (2) The read/spontaneous difference suggest that planning has an important role in shaping the characteristics of speech output. Our view is that olanning primarily restricts the availability of future speech segments. The way speakers use segments prior to a point of fluency failure reflects how they are tackling the oncoming fluency failure.

We define "fluency failure" as occurring when ".. speech control falters even though the speaker does not produce an overt error" [3]. This differs from previous definitions. Bernstein Ratner suggests that fluency failure in stuttering arises "from a disturbance in construction of the prearticulatory plan" (p.97) where the disturbance in some of the models she discusses, involves speech error [4]. Of course, the difference in definition of fluency failure depends primarily on how error is defined. We use error to refer only to incorrect phoneme selection and anticipation (not perseveration) of a preceding word or phrase. So phoneme blends (e.g. " *tab"* for *taxi* + *cab*) and transpositions (e.g. *"slow and sneet"* for *"snow and sleet"*) in Spoonerisms would be errors. The prolonged, broken and repeated parts of words, pauses of all types and perseverated whole words in *"I got on, on the seven … fffifty … three train t.to Mac…clesfield."* are all fluency failures but not speech errors according to our definitions. Another important point, implicit in what has already been said, is that planning limitations on future segments (mentioned under point 2 above) are not necessarily, or even usually, errors in planning future segments. More often than not, planning limitations are evident when the plan is available late rather than whether the correct one is or is not available.

The contrast between our viewpoint and the generally accepted viewpoint engenders a different perspective on the phenomena of spontaneous speech control. In particular, the difference in definitions of error and fluency failure highlight whether an explanation of the surface characteristics of spontaneous speech should be sought in the limitations during planning or error breakdown relative to a normative model. Our perspective is not without precursors but it is a perspective that has received less attention than error-approaches [2, 5]. We begin defending our minority outlook by considering a theory that illustrates the alternative position highlighting its main attraction, namely how it is able to unify several strands of evidence, and then go on to discuss some inherent problems of such a perspective.

2. A standard approach to explaining these issues

Levelt proposed a model in which a message goes through several stages before it is output, and the pattern of errors at output is regarded as providing evidence for these stages [6]. A message starts in the conceptualizer that is responsible for generating a message and monitoring to ensure that it is delivered appropriately. Monitoring devices like that in the conceptualizer take the output of a process and compare it with what the process was intended to produce (the initial input to the conceptualizer). If there is a discrepancy (an error) an adjustment is made to the output to reduce the error. The message next goes through the formulation stage. At the output of the formulation stage, the message is represented as a phonemic string. In Levelt's model, as in many others [7, 8], two sub-stages to the formulation stage are identified where the message is represented in lemma and phonological forms respectively (however, see [9] for data questioning two sub-stages). Articulation is the final step in translating the abstract phonemic representations to overt speech. The overt speech is sent via an external loop through the speech perception system to the monitor in the conceptualizer. In Levelt's model, then, one monitor suffices for checking events between intention and output action.

Repairs, as in *"Turn left at the, no, turn right at the crossroads"*, are interpreted as support for the internal monitoring process. According to a repair interpretation, the speaker gives the wrong direction, realizes this and exchanges the *reparandum* (*"left"*) with the *alteration* (*"right"*). The

substitution is not the only thing that happens: The speaker overshoots the reparandum (goes on with the words *"at the"*) and *retraces* to *"turn"* when the message recommences (the word before *"left"* that is not in error). The repair contains an interruption (here the word *"no"* though this could be a pause or a short phrase). (Most of these parsed components are optional.) Different forms of error on the reparandum support monitoring at syntactic, lexical and phonemic levels and speech can be interrupted, corrected and restarted at any point between conceptualization and articulation. Levelt also considers that errors at levels before speech output can be detected and repaired before they have been spoken overtly. These are called *covert* repairs and an example is *"Turn right at the, at the crossroads"*. Such repairs are problematic to interpret as there is no outward sign indicating the error (i.e. why the speaker hesitated after *"turn"*), nor even whether an error occurred at all. Other constructions, described as repairs, are different (D) repairs in which the topic is changed (like non sequitors) and repairs in which the speech is not pitched at an appropriate level for the addressee (appropriateness, A, repairs).

A different source of evidence, altered auditory feedback AAF, is usually cited in support of the external loop that features in models like Levelt's. Feedback is used as a term to refer to the route by which output is returned to a monitor so the message can be compared with that originally intended. Levelt's model allows the same monitor to process internal and external feedback by routing external feedback to the monitor via the speech comprehension system. The effects of AAF procedures on speech control can be illustrated with delayed auditory feedback (DAF). When this form of feedback is applied a speaker hears the voice a short time after it is spoken. DAF slows speech mainly by elongating vowels, and the speech produced has a monotone pitch and high amplitude (with the last two effects again focused on vowels) [10, 11]. The effects of DAF on fluent speakers can readily be explained by interruption to the external loop by the DAF. In this type of explanation, speakers continue to use the altered sound for voice control even though the sound is delayed. The problems observed earlier arise because the monitor acts on the misleading feedback that leads the speaker to think an error has been made. In particular, DAF leads speakers to adjust output timing even though no adjustment was needed.

The main differences between our viewpoint and internal and external monitoring explanations of these aspects of speech control are: (1) We consider that the focus on errors is limiting: (2) The patterns of speech interpreted as a result of monitoring mislead theorists and constrain them into ways of explaining events in a monitoring framework when other, more general, interpretations are possible: (3) There are problems using AAF evidence in support of the external loop.

3. Detailed critique of internal and external monitoring accounts of spontaneous speech control

Errors are infrequent events in language with some estimates as low as 0.1% [12]. Fluency failures, on the other hand, are common. For instance, Howell, Au-Yeung and Sackin estimate that fluent speakers produce around 2.57% fluency failures on function words and 0.97% on content words [13].

A monitor takes corrective action when a difference is detected between intended and actual versions (the actual version assumed to be in error). To establish whether a difference has occurred or not, like needs to be compared with like. Consequently, the monitor must have a representation of intended forms produced as actual output at any of the stages up to and including output of a message. Why should an errorless version of the intended form at lower levels be available in the conceptualizer for comparison when these lower levels generate an erroneous output? If the conceptualizer can generate a correct version for monitoring, why can this not be done (or this representation used) for output of lower stages?

Empirically, the repairs described earlier, viewed from the perspective of our definition, actually provide little support for an internal error-monitoring process. Evidence that an error has occurred is only obtained when speech output reveals this. However, this selfsame evidence for an error would suggest either that no monitoring takes place or that the monitor has not worked on this occasion as the speech goes right through to output. The only type of repair where there may be evidence for the operation of an internal monitor is where errors are intercepted and the result is a covert repair. However, in these repairs another interpretation is that no error occurred in the first place (see the later discussion of stalling fluency failures for an alternative explanation of some surface-form features considered to represent components of repairs). In the case of D repairs, the message is abandoned rather than repaired, and A repairs are a matter of style rather than anything else. Abandonment and restart of a message (as in D repairs) may be a general process (applying to all repairs) reflecting anticipated planning difficulties rather than arising from internal monitoring for errors and on-line repair to remove any detected discrepancy.

There are also some specific problems for the proposal that the external loop is monitored using AAF evidence. Borden pointed out that auditory processing takes time: A segment has to finish, then its auditory output has to be processed to establish that the sound was produced correctly before the next one can be initiated [14]. Marslen-Wilson and Tyler estimate that recognition of running speech is about 200 ms and a processing delay of this order would lead to slower speech rates than speakers achieve [15]. Second, the information provided over the external loop would have to be a veridical record of what was produced; otherwise establishing if and what error has occurred by a monitor with the intention of correcting it, would not be possible [16]. However, the representation of articulatory output provided over this loop is not veridical with respect to the intended message. The auditory representation the speaker receives while speaking is affected by internal (mainly bone-conducted) sound and external noise sources (for instance,

variation caused by each unique speaking environment) during transmission.

4. The EXPLAN model of fluency failure

We turn, now, to a brief description of our model before describing some ways it has been evaluated. Cognitive planning (PLAN) and articulatory execution (EX) of speech are independent processes in the EXPLAN model. From the outset, the contrast with auditory monitoring accounts is apparent as PLAN and EX are interdependent in monitors (EX leads to auditory output that, if error is detected, would restart or tune PLAN processes in an auditory monitor). PLAN and EX operate as a chaining process in fluent speech (when one word finishes EX, the next PLAN is picked up) and the process is intrinsically timed [17].

The situation in which speech is fluent and one form of fluency failure when PLAN cannot keep up with EX is shown in Figure 1. Time is along the abscissa and time for planning and execution are indicated by the length of the bar. So, in the fluent speech example, two segments that are quick to plan are followed by one that takes longer to plan. Execution lags production by one segment (starting after the end of the first planned segment) and execution time determines when the plan of the next segment needs to be ready. As long as there is sufficient time for the execution of one segment for the next one to be planned (or has been in the preceding sequence), speech will proceed fluently. Though not essential to the theory, we have used Selkirk's phonological words to formulate some tests on English [18]. Selkirk defines a phonological word as consisting of a content word and an optional number of function words preceding and following it. Function words are simpler than content words in English and if the simplicity is reflected in planning time, the fluent example in Figure 1 would represent a function-function-content word sequence (e.g. "in a trice"). As planning of "trice" can take longer than the time to execute "a" allows, "a" can finish execution before the plan for "trice" is ready. The speaker needs more time for planning. In fluent speakers, this can arise by pausing when the plan runs out or by repeating words that immediately precede the difficult word. Au-Yeung, Howell and Pilgrim have shown that function word repetition always occurs on those preceding content words that would buy planning time [19].

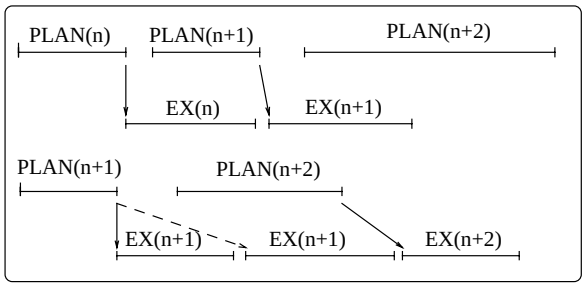

Figure 1: *The EXPLAN model for fluent control and word repetition*.

The property of English content words that may lead to them posing difficulty could be in the phonological structure at onset. Fluency failure occurs most often at word onset and "tip-of-the tongue" studies show that these parts are generated first. We have developed a characterization for difficulty of word onsets that incorporates whether the onset has a consonant string and whether consonants in this string are difficult to produce (indexed by the average age at which children acquire them) [20].

Another factor that can lead to variation in fluency is rate. At the outset, we noted that a text supplies a plan. According to EXPLAN, this would reduce the chance of speech execution getting ahead of planning. In contrast, if speech is negotiated at too fast a rate in local sections like "in a trice", the chance of planning getting ahead of execution is increased. Howell, Au-Yeung and Pilgrim have shown that a rapid rate in local stretches increases the likelihood of stuttering (an acute form of fluency failure) [21].

As EXPLAN does not allow interactions between execution and planning, it predicts that feedback-monitoring processes do not operate in fluency failures. This prediction appears to stand at odds with the disruption caused by alteration to auditory feedback that, as we saw, is readily explained as due to interference with monitoring processes. EXPLAN proposes that fluent speech control operates without external timing though there is an external timekeeper that regulates speech rate under specified circumstances. One of these is when DAF is presented. DAF is input direct to the timekeeper rather than transmitted from execution back to planning processes. This timekeeper is governed by number and timing relationship between inputs. Under DAF there is one extra asynchronous input that affects the timekeeper's ability to follow the response beat [22]. The lengthened period caused by the delay, adjusts an oscillator coupled to the EXPLAN process that also lowers the centre frequency of responses generated in the EXPLAN process [23]. This slows the speech of fluent speakers under DAF. For full details see [16].

The way this addresseses the two problems that were discussed in the introduction concerning an auditory monitoring account are considered with respect to whether they are problems for EXPLAN. The processes described do not require complete analysis of auditory output for speech content, they propose that any serial input to the timekeeper can affect its operation. As speech analysis is not performed on these inputs, the slowing problem that occurs in an auditory monitor does not arise. Also, the speech does not have to be veridical of the plan, only to represent rhythmic input to the timekeeper.

This work was supported by the Wellcome Trust.

5. References

[1] Levelt, W. J. M., "Monitoring and self-repair in speech", *Cognition, Vol. 14, 1983, pp 41-104*.

[2] Howell, P. and Kadi-Hanifi, K., "Comparison of prosodic properties between read and spontaneous speech", *Speech Communication, Vol. l0, 1991, pp 163-169*.

[3] Howell, P. and Au-Yeung, J., "The EXPLAN theory of fluency control and the diagnosis of stuttering", in *Current Issues in Linguistic Theory: Clinical Linguistics: Language Pathology, Speech Therapy, and Linguistic Theory*, John Benjamins, Amsterdam, in press.

[4] Bernstein Ratner, N., "Stuttering: A psycholinguistic perspective", in Curlee, R. F. and Siegel, G. M. (eds) *Nature and Treatment of Stuttering*, Allyn & Bacon, Boston, 1997.

[5] MacWhinney, B. and Osser, H., "Verbal planning function in children's speech", *Child Development, Vol. 48, 1977, pp 978-985*.

[6] Levelt, W. J. M., *Speaking: From Intention to Articulation*, MIT Press, Cambridge, MA, 1989.

[7] Dell, G. S., "A spreading-activation theory of retrieval in sentence production", *Psychological Review, Vol. 93, 1986, pp 283-321*.

[8] Dell, G. S. and O'Seaghdha, P. G., "Stages of lexical access in language production", *Cognition, Vol. 42, 1992, pp 287-314*.

[9] Caramazza, A. and Miozzo, M., "The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon", *Cognition, Vol. 64, 1997, pp 309-343.*

[10] Black, J., "The effect of side-tone delay upon vocal rate and intensity", *Journal of Speech and Hearing Disorders, Vol. 16, 1951, pp 50-56*.

[11] Lee, B. S., "Effects of delayed speech feedback", *Journal of the Acoustical Society of America, Vol. 22, 1950, pp 824-826*.

[12] Shallice, T. and Butterworth, B., "Short-term memory impairment and spontaneous speech", *Neuropsychologia, Vol. 15, 1977, pp 729-735*.

[13] Howell, P., Au-Yeung, J., and Sackin, S., "Exchange of stuttering from function words to content words with age", *Journal of Speech, Language and Hearing Research, Vol. 42, 1999, pp 345-354*.

[14] Borden, G. J., "An interpretation of research on feedback interruption in speech", *Brain & Language, Vol. 7, 1979, pp 307-319.*

[15] Marslen-Wilson, W. D. and Tyler, L. K., "Central processes in speech understanding", *Philosophical Transaction of the Royal Society of London Series B, Vol. 259, 1981, pp 297-313.*

[16] Howell, P., "The EXPLAN theory of fluency control applied to the treatment of stuttering by altered feedback and operant procedures", in *Current Issues in Linguistic Theory: Clinical Linguistics: Language Pathology, Speech Therapy, and Linguistic Theory*, John Benjamins, Amsterdam, in press.

[17] Fowler, C. A., "Coarticulation and theories of extrinsic timing", *Journal of Phonetics*, Vol. 8, 1980, pp 113-133.

[18] Selkirk, E., *Phonology and syntax: The relation between sound and structure*, MIT Press, Cambridge, MA, 1984.

[19] Au-Yeung, J., Howell, P., and Pilgrim, L., "Phonological words and stuttering on function words", *Journal of Speech, Language, and Hearing Research, Vol. 41, 1998, pp 1019-1030*.

[20] Howell, P., Au-Yeung, J., and Sackin, S., "Internal structure of content words leading to lifespan differences in phonological difficulty in stuttering", *Journal of Fluency Disorders, Vol.* 25, 2000, pp 1-20.

[21] Howell, P., Au-Yeung, J., and Pilgrim, L., "Utterance rate and linguistic properties as determinants of speech dysfluency in children who stutter", *Journal of the Acoustical Society of America, Vol. 105, 1999, pp 481-490*.

[22] Howell, P., Powell, D. J., and Khan, I., "Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks", *Journal of Experimental Psychology: Human Perception and Performance, Vol. 9, 1983, pp 772-784*.

[23] Large, E. W. and Jones, M. R., "The dynamics of attending: How people track time-varying events", *Psychological Review, Vol. 106, 1999, pp 119-159*.