ACOUSTIC PROPERTIES OF DISFLUENT REPETITIONS

E. E. Shriberg
Institute for Perception Research (IPO), Eindhoven, The Netherlands
and SRI International, Menlo Park, CA, USA

ABSTRACT

Acoustic properties of disfluent repetitions are examined to investigate two proposed functions of repeating. Repeating may serve as a filler while hesitating; alternatively, repeating may function to bridge a gap when speech resumes after a break. Classification of tokens based on pause patterns reveals that: (1) most cases fit the bridging function; and (2) duration and F0 properties support the hypothesis of two distinct types, as well as the proposed associated functions.

INTRODUCTION

Speakers often repeat words in spontaneous speech, resulting in lexical disfluencies such as that shown in Figure 1.

in the the Senate
R1 R2 Continuation

Figure 1. Example of a disfluent repetition and terminology.

Little is known, however, about *why* speakers utter the repeated instance (R2).

Heike [1] suggested two alternative functions of repeating, which could be distinguished based on the presence of an unfilled pause following R2. Possible surface patterns for the hypothesized functions are summarized in Figure 2.

Prospective: R1 (...) R2 ...Continuation
Retrospective: R1 ... R2 Continuation

Figure 2. Surface patterns for proposed functions. "..."=pause; "( )"=optional.

He termed cases in which R2 was followed by a pause ***prospective repeats***, suggesting such repeats serve a stalling function, to hold the floor during hesitation. Cases in which R2 was not followed by a pause (but preceded by a pause) were termed ***retrospective repeats,*** and were proposed to function as bridging devices to connect a continuation with preceding material after a break in fluency.

Although these proposed functions are reasonable theoretical possibilities, in practice there is little if any empirical evidence to support the claim that there are two types of repeats. For purposes of this paper, we will assume that there *are* two different functions of repeating, and that the functions can be distinguished on the basis of whether or not R2 is followed by an unfilled pause before the continuation. Leaving the issue of function aside, the terms "prospective" and "retrospective" will be used simply to refer to the classification of the repetition based on its surface pause characteristics.

Given these assumptions as a starting point, this paper seeks to answer two questions about the different types of repeats. First, are both types actually found in spontaneous speech data, and at what relative rates? Second, if we look more closely at acoustic properties of repetitions, can we find characteristics other than unfilled pauses that pattern differently for the two types?

METHOD

Single-word disfluent repetitions were extracted from the speech of six speakers (three male, three female) in the SWITCHBOARD corpus of telephone conversations [2]. In this corpus, speakers conversed with an unfamiliar partner on a chosen topic. Despite the somewhat contrived task, conversations were rated as highly natural-sounding by transcribers. Selection of repetitions was limited to cases with no other disfluency either between or directly following the repetition.

Hand-labeling of 242 such cases was conducted using the GIPOS speech analysis package developed at IPO. For each example, five time values were recorded: the onset and offset of R1, the onset and offset of R2, and the onset of the continuation. For examples with adequate F0 tracks for both R1 and R2, and in which both R1 and R2 showed a roughly linear F0 trajectory, the first good F0 and last good F0 of R1 and of R2 were also recorded.

RESULTS AND DISCUSSION

Frequency of types

Figure 3 shows the frequency of types (as classified based on the presence of a pause following R2). As shown, both

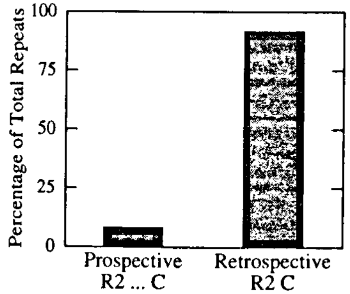

Figure 3. Relative frequency of types. "R2"=repeated instance, "C"=continuation, "..."= unfilled pause.

patterns occur in the speech data examined. However, the clear majority of cases were classified as retrospective, i.e. cases in which R2 is hypothesized to serve a bridging rather than stalling function.

Acoustic properties of types

To address the second question posed in the introduction, duration and F0 properties of R1 and R2 were examined in an aim to provide evidence other than simple pauses to support Heike's claims.

Duration. In Figure 4, the duration of R1 is plotted against the duration of R2 for all tokens (over all words and speakers). Different symbols denote the two different repeat types. The equivalence line (y=x) is indicated for reference.

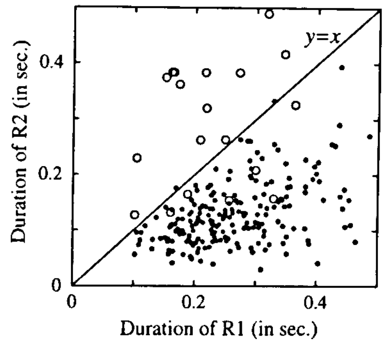

Figure 4. Duration of R1 versus R2 for all tokens; ○ = prospective repeat, ● = retrospective repeat.

As shown, there is a clear difference between the two types. Data for prospective repeats occur both above and below the equivalence line. The data for retrospective repeats, however, are nearly all below the equivalence line, indicating that R1 is systematically longer than R2.

In order to interpret these data, however, we must know whether R1 is lengthened, or R2 shortened (O'Shaughnessy suggested that both effects occur [3]). To address this issue, a small study controlled for speaker and word was conducted. Durations for R1, R2, and unrepeated (fluent) instances of the word "the" were compared for the two speakers with the largest amount of data. The fluent examples were chosen from those conversations which also contained the repeated instances. For speaker 1, 19 repeated tokens and 40 unrepeated tokens were obtained; for speaker 2, 12 repeated and 33 unrepeated tokens were obtained. Results are shown in Figure 5.

Despite the small sample sizes, there is a significant difference between R1 and the other two conditions, as can be inferred from the error bars. Also, importantly, R2 does not appear to be shortened since it appears actually slightly longer than unrepeated instances.

Thus, for retrospective repeats, there is lengthening at R1 and no lengthening at R2; this is consistent with the bridging

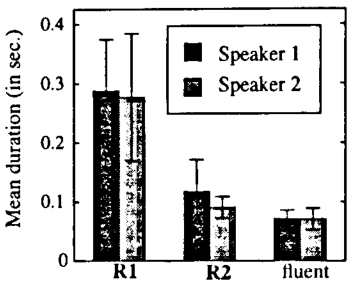

Figure 5. Mean duration of R1, R2, and unrepeated instances of "the" for two speakers.

function proposed by Heike for such cases, since R2 resembles a fluent resumption.

Given these results, inspection of absolute durations in Figure 4 also suggests that R2 is lengthened for most prospective repeats, consistent with the proposal that in these cases R2 functions as a hesitation device.

Fundamental frequency. A difference between prospective and retrospective repeats was also found in F0 properties of R1 and R2. When the four measured F0 points described in the method section were plotted at equal intervals (i.e. not taking into account the duration of the words), results showed roughly parallel trends at different F0 ranges, although this requires modifying the linear scale (an appropriate scaling model is described in [4]). Thus, a representative picture of the relationships between these four points can be conveyed by plotting the mean values for each speaker. Such values are plotted for retrospective repeats in Figure 6. Values are shown separately for males and females in order to display results on appropriate scales.

As shown, both words (in nearly all cases words were unaccented) fall in F0. A consistent tendency across speakers is that the onset of R2 is reset to a value about equal to that of the onset of R1. It is also notable that R2 falls in F0 like R1,

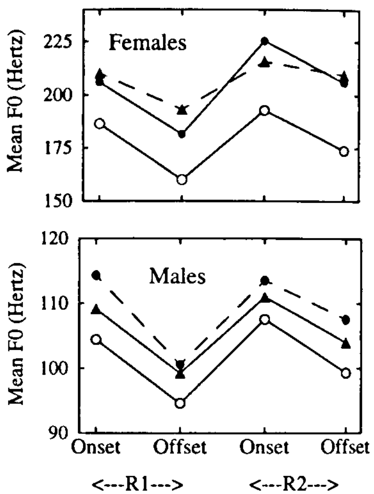

Figure 6. Mean onset and offset F0 of R1 and R2 in retrospective repeats, by speaker.

but not to as low a value; this is probably due to the much shorter duration of R2.

It was not possible to obtain data for all speakers for the set of prospective repeats due to low sample size. However, in Figure 7, results are shown for one female speaker. The data for the retro-

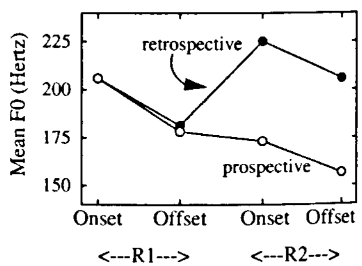

Figure 7. Mean onset and offset F0 of R1 and R2 in retrospective and prospective repeats for a single speaker.

spective repeats from the same speaker (from Figure 6) have been replotted in Figure 7 for comparison.

As shown, the two types show different patterns. Unlike retrospective repeats, prospective repeats show an R2 that con-

tinues to fall in F0 from R1. Similar results were obtained for other speakers. The continuous fall in F0 for the prospective repeats is consistent with an observation for filled pauses, which are also proposed to serve a stalling function: it was noted informally in past work that sequential filled pauses (e.g. "uh uh uh") showed successively lower starting F0 values.

Although some prospective repeats showed a reset, retrospective repeats rarely lacked the reset. The reset for the retrospective repeats is consistent with findings by Levelt and Cutler [5], who observed that repairs tend to be uttered at the same F0 values as the material they replace.

SUMMARY AND FUTURE WORK

This research found that when repetitions were classified based on pause characteristics, the majority of cases showed no pause between R2 and the continuation. Such cases are consistent with a bridging, rather than a stalling function.

Furthermore, analyses revealed that the classification of tokens based on pauses correlated with durational and F0 properties of R1 and R2. This result adds weight to the the proposal that there are two different types of repeats. In addition, the duration and F0 properties pattern in ways that are consistent with Heike's proposed functions.

An important next step in this line of research is to investigate the question of function directly, for example by conducting controlled elicitation or perceptual experiments. Goals for such future work include gaining knowledge about factors (e.g., syntactic, prosodic, task-related, speaker-related) that influence the production of repeats, as well as gaining an understanding of how repetitions function both for speakers and listeners.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Marc Swerts for valuable comments and discussion. This work was supported by an NSF-NATO postdoctoral fellowship. It is also part of ongoing research funded by the Advanced Research Projects Agency under NSF Grant IRI-9314961, and by NSF Grant IRI-8905249. The views and conclusions contained in this document are those of the author and should not be interpreted as reflecting the official policies of the funding agencies.

REFERENCES

[1] Heike, A.E. (1981). A content-processing view of hesitation phenomena. *Language and Speech, 24* (2), pp. 147-160.

[2] Godfrey, J.J., Holliman, E.C. & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proc. ICASSP*, pp. 517-520.

[3] O'Shaughnessy (1993). Analysis and automatic recognition of false starts in spontaneous speech. *Proc. ICASSP*, pp. 521-524.

[4] Shriberg, E.E. (1994). Preliminaries to a theory of speech disfluencies. Unpublished Ph.D. thesis, University of California at Berkeley.

[5] Levelt, W.J.M. & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics, 2* (2), pp. 205-217.