The Phonetics of Repetition in Other-Initiated Repair Sequences

Traci S. Curl

University of York, Department of Language and Linguistic Science
tsc3@york.ac.uk

ABSTRACT

This research describes the different phonetic forms of self-repetitions used in other-initiated repairs (i.e., repetitions following *what?; huh?* by another speaker). Phonetic analysis of the repairs contradicts published accounts of the phonetic correlates of clear speech, and descriptions of the phonetics of repetitions. The phonetic realization of the repetition is shown instead to be systematically related to differences in the turn-taking properties and sequential relevance of the original utterance. This study emphasizes the importance of combining attention to sequential analysis with analysis of fine phonetic detail to fully understand the orderliness in everyday conversation.

1 INTRODUCTION

A growing body of research investigating the interrelationship of phonetic structure and sequence organization is showing how systematically produced clusters of phonetic events are manipulated and responded to in both shaping and interpreting talk in natural conversation (e.g., [1] *inter alia*, [2]). This study continues this methodological approach, integrating Conversation Analysis (CA) and impressionistic and instrumental phonetics. The concern is to show how particular activities in a conversation are managed and oriented to *by the participants themselves*.

This study focusses on the phonetic and interactional structure of the activity of other-initiated repairs. In these sequences, Speaker A's talk is treated as somehow problematic – a trouble source (TS). The evidence for this is Speaker B's production of *what?; huh?*, providing for Speaker A to somehow repair his/her problematic talk in the following turn. In the cases examined here, Speaker A's repair of the TS turn is done by lexical repetition. The production of repetitions in other-initiated repair sequences provides a way to examine the claims of published research on the phonetic realization of both repetition and clarification, using naturally-occurring talk-in-interaction as the data.

Previous research makes opposing predictions about how the phonetics of repetition repairs (REPs) would be realized. This study shows that neither set of predictions is supported by the data, and that rather, the phonetic parameters of loudness, pitch range, duration, and articulatory settings are systematically manipulated according to the sequential fittedness of the TS turn.

2 REPETITION AND CLARIFICATION

Two well-known bodies of research exist which delineate the phonetic features commonly found in repetitions and in clear/highly intelligible speech (see [3], [4] and references therein). REPs potentially incorporate both phenomena, however: the utterances are produced as lexical repetitions of the prior turn, but in response to prompting from another participant. That is, one participant has indicated difficulty in understanding a prior utterance, giving the speaker a chance to clarify that utterance. Since the same words are used, we might surmise that the REP will be uttered 'more clearly' than the TS utterance.

Research into the phonetics of repetition has found that repetition correlates with shorter word durations and phonological reduction processes. The phonetic correlates of 'clear speech' have also been investigated experimentally; this work finds that highly intelligible speech (produced both with and without instructions to speak 'more clearly') has an expanded vowel space, an expanded pitch range, and a slower tempo than other, or conversational, speech.

H & H theory [5] encapsulates these two approaches. According to this theory, speakers must choose between the default state of producing utterances with least effort (hypo-articulation), and the principle of informativeness – making their utterances sufficiently discriminable. Since repetitions reproduce information already given in the conversation, they are redundant, and therefore repetition repairs could be hypo-articulated. However, repetition repairs are produced as responses to a display by the listener that he or she requires 'help' in understanding the prior utterance. This argues for their hyper-articulation.

This creates a paradox for predicting the phonetic realization of repetitions used in clarification (if that is in fact the activity being done in other-initiated repair sequences). It also highlights a problem found in many experimental and corpus-based studies of spontaneous speech – the unexplicated use of labels such as repetition and clarification, with no principled method for investigating the participants' use and understanding of the talk being analyzed.

3 DATA

The data come from the Callhome corpus of American English. This corpus of telephone calls between family members and friends was collected in the late 1990s as part of an

ISBN 1-876346-48-5 © 2003 UAB

unrelated research project on speech recognition, and was made available through the Linguistic Data Consortium.

Participants were recruited in the United States via the internet, advertising, and word-of-mouth. The calls were recorded in stereo, with separate channels for each speaker. Because each speaker was recorded in a separate channel, overlapping talk does not present a problem for either impressionistic or acoustic analysis. Over 50 calls, totalling 25 hours of talk, were used to compile the collection of 69 fragments discussed in this report. Each REP sequence and surrounding context was copied from the CD-ROMs into separate WAV files using the Praat software program, which was also used in the instrumental analysis described below.

4 PHONETIC ANALYSIS

This research combines the methodological stance of Conversation Analysis (CA) with more traditional methods of phonetic analysis. One goal of CA is to discover what categories are demonstrably relevant to the participants themselves; therefore, rather than beginning with a priori assumptions of what parameters to concentrate on, the collected data fragments were examined with as little prejudice as possible to discover which, if any, phonetic parameters differed between the speakers' productions of the trouble source turns and the REPs. Four parameters were found to vary systematically between the two utterances: loudness, pitch range, duration, and articulation.

4.1 LOUDNESS

Impressionistic observations of the loudness of the REPs compared to the TS turns were checked against intensity measurements taken from Praat. The measurement of intensity in real-life speech settings is fraught with technical problems; in addition, speakers sometimes laugh or cough mid-utterance, or background noise interferes with the signal.When intensity measurements were possible to obtain, however, they corresponded with the impressionistic observation of the relative loudness of the two turns.

4.2 PITCH

Using the pitch extraction function of Praat, F_0 traces of the TS turns and REPs were prepared, and hand-corrected where necessary. Each speaker's baseline F_0 was computed based on inspection of a 5-minute stretch of talk taken from the middle of each conversation; this information was used to convert the F_0 traces to semitones, a more accurate representation of the perception of pitch than a Hertz scale. Although nearly all the TS turns and REPs had similar intonation contours, the variation in pitch range covered by each turn was noted and recorded.

4.3 DURATION

Some REPs were found to have more syllables than the TS turns (and vice versa). Since the REPs were all repetitions of the same words, the unit 'word' (rather than syllables) was used for segmentation and measurement of both utterances.

The duration of each word in the TS and REP turns was measured by hand from the waveforms and spectrograms, and plotted together on a logarithmic scale to normalize different magnitudes of change. Using these graphs, the relative difference in duration of each word in a REP could be compared to that word's duration in the TS turn.

4.4 ARTICULATION

Detailed phonetic records of both the TS and REP turns were prepared, using the techniques of analytic parametric listening described in [6, 30ff]. Attention was paid to the (small) differences between the two productions of 'the same' articulatory complexes, e.g. noting any differences in the consonantal and vocalic articulations of the TS turn and the REP. This method of relativistic listening relates the characteristics of sounds at one place in structure (i.e. the REP) to those at another place in structure (the TS turn). Additionally, the details of the sound qualities of longer stretches of the material were noted.

4.5 SUMMARY

The phonetic differences between the TS turns and REPs grouped into two clusters: some REPs were louder, had expanded pitch ranges, longer durations on repeated words, and long domain articulatory resettings relative to the TS turns. Other REPs were quieter, had compressed pitch ranges, shorter durations, and similar articulatory and phonatory settings to the TS turns.

5 INTERACTIONAL ANALYSIS

The methodology of CA offers a principled way of describing how participants manage the temporally unfolding interactional structure of conversation. CA researchers have shown repeatedly that the sequential organization of conversation is not arrived at accidentally; that there is 'order at all points' [7]. The principled examination of massive amounts of talk-in-interaction has shown speakers' orientations to the normative pressures exerted by certain kinds of structures found recurrently within conversation. A full explication of the methods of CA is not possible here; for good introductions see [8], [9].

Each turn in a conversational sequence displays a speaker's understanding of what activity is taking place in the sequence up to that point. With this understanding in mind, certain differences can be noted within the data fragments. Some of the TS turns seemed to be appropriately designed and placed to match the structure set up by the prior turn and/or sequence; that is, they were 'fitted'. Fitted turns continued an activity already in progress, or began a new sequence after a collaborative closing of a prior. An example of such a fitted TS turn is given below.

(1) TWINS 5245CHAm

1 A: an um (.) what else
2 (0.3)
3 A: um:: the t[wins are getting] °big°
4 B: [oh I had this]
5 (0.7)
6 B: what?
7 (.)
8 A: the twins are getting bi:g

The turn treated as a TS occurs in line 3. "An um (.) what else" in line 1 displays Speaker A's understanding that whatever she and Speaker B were discussing is possibly complete, and signals her search for a 'next topic'. Speaker B passes on the opportunity to begin a new sequence herself – note the 0.3 second silence at line 2. In line 3, Speaker A begins the projected new sequence by proferring a topic: "um:: the twins are getting big", an utterance fitted to the place in which it occurs.

Other TS turns seemed disjunct where they occurred. The term 'disjunct' covers problematic incursion into the turn space of another speaker (that is, overlapping talk), and/or an observable lack of relevance on the part of the TS turn (i.e., 'inappropriateness' of the turn in relation to the prior). Fragment 2 provides an example of a disjunct TS turn.

(2) BATHROOM 4431CHAm

1 A: well god damn it
2 (1.8)
3 A: I'm sick of gettin' trounced on
4 (0.7)
5 B: you inna bathroom?
6 (0.4)
7 A: huh
8 B: you inna bathroom?

Prior to the talk shown in this fragment, Speaker A has been telling her brother, Speaker B, about an argument she had with their parents. She completes her telling with a summary of her feelings in lines 1 and 3. In spite of the lengthy pauses between these utterance, Speaker B makes no move to affiliate with her, nor even to display any awareness that the story has come to an end. When he does speak, it is merely to enquire about Speaker A's whereabouts. This turn does not provide a sequentially relevant response to the prior turn. Instead, it begins a new, unrelated sequence,with no display of a link to the prior nor any marker of its unrelated status. Since each turn in a sequence is assumed to coherently continue the sequence in progress, special devices such as *by the way* need to be employed to mark out disjunct sequences. No such markers are employed here.

One result of the interactional analysis is to show that what looks superficially like the same conversational structure may be sequentially different just below the surface. Although all the repetitions occurred in other-initiated repair sequences, the trouble source turns which were repeated turn out to be involved in (at least) two different sorts of activities: the fitted TS turns which perform the relevant next actions, and the disjunct trouble source turns which do not. This difference in the turn engendering the repair initiation is reflected in the phonetic structure of the REP itself.

6 COMBINING PHONETIC AND INTERACTIONAL ANALYSES

Differences in the parameters of loudness, pitch range, duration, and articulatory settings are distributed systematically according to the place in the structure of the conversation where the TS turn occurs. REPs of fitted turns were louder, had expanded pitch ranges, had longer durations on repeated words, and had long-domain articulatory resettings. REPs of disjunct turns, on the other hand, were quieter, had compressed pitch ranges, had shorter durations on repeated words, and had similar articulatory and phonatory settings.

The methodological stance adopted here is a qualitative, case-by-case analysis for which the quantified findings which follow are not intended as a substitute. They should be taken instead as an abridged version of a more complete, and more subtle, analysis.

6.1 LOUDNESS

For the fitted TS turns, for which intensity could be reliably measured, 16 REPs are louder than their corresponding TS turns; one is quieter, and two are similar. Assuming a normal distribution of louder and quieter repetitions, this distribution is highly unlikely to be due to chance (sign test, $p<0.001$). Among the group of disjunct TS turns and REPs, the opposite distribution was found. Twenty-nine REPs were *quieter*, and 13 were louder, again a significant difference in distribution (sign test, $p=0.0195$). The admittedly crude sign test is offered here simply to reinforce the difference in the distribution of louder and quieter repetitions between the REPs of fitted and disjunct TS turns.

6.2 PITCH RANGE

The pitch range of the REPs of fitted TS turns are all expanded between 1 and 3.8 semitones, while 85% of the REPs of disjunct TS turns have pitch ranges which are compressed (some as much as 7.5 semitones). An increase or decrease in pitch range is often treated as paralinguistic; that is, a 'higher high' is said to possibly convey greater emphasis, but not to affect the identity or function of the contour. In this data, however, the manipulation of pitch range corresponds to differences in the sequential placement of the TS. In other words, the interactional function performed by these REPs *does* correspond to the identity of the contour, as expressed not by a categorical (and therefore classed as linguistic) manipulation of high and low tones, but by the gradient (and therefore classed as paralinguistic) manipulation of pitch

ISBN 1-876346-48-5 © 2003 UAB

range. This highlights the difficulty, and possible futility, of maintaining a distinction between the paralinguistic and linguistic, especially when such distinctions are drawn before examining the data and seeing what parameters are relevant for and oriented to by the participants themselves.

6.3 DURATION AND ARTICULATION

The durational measurements and articulatory differences are not easily quantified, given that they rely on relative differences between each TS and its corresponding REP. Therefore, representative examples of each will be presented.

Figure 1 shows the duration of a fitted TS and REP plotted on a logarithmic scale. The y-axis is duration; thus plot points that are higher along the y-axis represent longer durations. The empty circles represent words in the TS, the filled circles words in the REP.

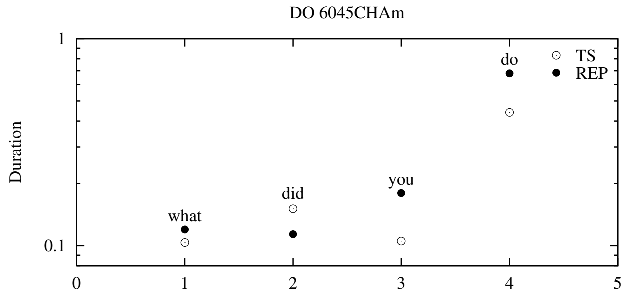

Figure 1: Word duration of TS and Rep: DO 6045CHAm

This figure shows that three of the four words in the REP have longer durations than they did in the TS, exemplifying the normal pattern for fitted TS sequences. As is shown in this graph and others (including those of disjunct TS turns and REPs in which the duration of words in the REP is less than of the TS), it is overwhelmingly the 'content' words (e.g., nouns and main verbs) which exhibit durational differences; note that here the auxiliary verb "did" does not pattern with the rest of the turn.

Finally, it was noted that generally, nothing in the REPs suggested a 'more careful' rendering of articulatory targets such as we would expect in clarification; nor was there consistent evidence of reduction processes at work as suggested by prior work on repetitions. Rather, REPs of fitted TS turns showed long-domain changes to articulatory settings, while REPs of disjunct TS turns were produced to be maximally similar to the original utterance, down to similarities between the vowel spectra, energy and duration of fricatives, and the method of closure release.

An example of articulatory re-setting is provided by the fitted TS and REP turns discussed above, "what did you do?" Here, the major difference in articulation between the TS and REP is liprounding. This rounding is evident throughout the REP, but had a restricted domain in the TS. In both utterances, "what" is articulated with a labiovelar approximant; however, in the TS liprounding is maintained only until the onset of the first vowel, with no further liprounding evident until midway through the final vowel. In the REP, however, liprounding is maintained throughout.

7 CONCLUSION

This study shows that repetitions employed as repairs are produced with neither the characteristic phonetics of 'clear speech' nor that of other, previously studied repetitions. It argues against considering repetition as a homogeneous phenomenon, and shows that the manner in which a certain cluster of phonetic parameters is manipulated in the repetition is dependent on the sequential fittedness of the TS turn.

REFERENCES

[1] E. Couper-Kuhlen and M. Selting, Eds., *Prosody in Conversation: Interactional Studies*, Cambridge University Press, Cambridge, 1996.

[2] R. Ogden, "Turn transition, creak and glottal stop in Finnish talk-in-interaction," *Journal of the International Phonetic Association*, vol. 31, no. 1, pp. 139–152, 2001.

[3] L. W. Shields and D. A. Balota, "Repetition and associative context effects in speech production," *Language and Speech*, vol. 34, no. 1, pp. 47–55, 1991.

[4] A. R. Bradlow, M. Torretta, and D. B. Pisoni, "Intelligibility of normal speech I: global and fine-grained acoustic-phonetic talker characteristics," *Speech Communication*, vol. 20, pp. 255–272, 1996.

[5] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech production and speech modelling*, W.J. Hardcastle and A. Marchal, Eds., pp. 403–439. Kluwer, Dordrecht, 1990.

[6] J. Kelly and J. Local, *Doing Phonology*, Manchester University Press, Manchester, 1989.

[7] H. Sacks, *Lectures on Conversation*, vol. I, Blackwell, Oxford, 1992.

[8] S. C. Levinson, *Pragmatics*, Cambridge University Press, Cambridge, 1983.

[9] J. Heritage, "Current developments in conversation analysis," in *Conversation: An Interdisciplinary Perspective*, D. Roger and P. Bull, Eds., pp. 21–47. Multilingual Matters, Clevedon, 1989.