

Low-Pitch Regions as Dialog Signals? Evidence from Dialog-Act and Lexical Correlates in Natural Conversation

Nigel Ward

University of Tokyo

Abstract

In earlier work, we identified a 110 millisecond region of low pitch as a prosodic feature which seems to bear the dialog function of encouraging back-channel feedback from the listener. In this paper, we¹ examine the ways in which this prosodic feature co-occurs with semantic, pragmatic, and lexical events. Both subjective analysis and statistical analysis suggest that low-pitch regions are associated with the completion or near-completion of the transmission of some unit of information, the occurrence of a disfluency, and the occurrence of back-channel feedback. We take this as evidence that low-pitch regions are real prosodic features.

1 Introduction

In earlier work, we identified ‘low-pitch regions’ as prosodic features which seem to bear a dialog function (Ward 1997; Ward & Tsukahara submitted). Specifically, we found that these features seem to function as signals by which the speaker allows or encourages the listener to produce back-channel feedback. Attempting to maximize performance as a predictor of back-channel feedback by the listener, we quantified the feature as a region of pitch less than the 26th-percentile pitch level for the speaker, and continuing for 110 milliseconds or more. The presence of such features correlates well with the presence of back-channel feedback 200 to 1200 milliseconds later, in casual English conversations (with a similar correlation also present in Japanese).

However, these findings can be accounted for in more than one way. One is the explanation assumed above: that the low-pitch region is a prosodic feature which functions as a signal from the speaker to the listener.

A second possible explanation is that low-pitch regions function as markers, indicating the presence of some syntactic, semantic, or pragmatic event. Specifically, low-pitch regions may mark clause completion, utterance finality, completion of the transmission of some unit of information, the presence of some kind of discourse structure boundary, and so on. Thus, the correlation between low-pitch regions and back-channels could perhaps be accounted for in terms of a hidden variable, some syntactic, semantic, or pragmatic event, that caused both the pres-

ence of a low-pitch marker in the speaker’s utterance and the back-channel response by the listener.

A third possible explanation is that these prosodic features are mere epiphenomena. In general, the presence of stresses or accents (which involve high pitch) indicates the presence of new or important information. It is therefore conceivable that it is the lack of new information which is the hidden variable, causing both the region of low pitch (as an artifact of the lack of high pitch features), and the subsequent back-channel response by the listener.

The purpose of this paper is to describe two investigations seeking to flesh out these alternative accounts of the role of these low-pitch regions in dialog. Section 2 describes an attempt to do so based on subjective impressions of the contexts of occurrence of low-pitch regions. Section 3 describes an attempt to do so by statistical analysis of the lexical items that occur with these low-pitch regions.

An additional motivation for this study, in addition to the intrinsic interest of the phenomenon itself, is the desire to reconcile two approaches to the study of prosody. One approach focuses on the ways in which prosody marks lexical, syntactic, and semantic events; research in this vein is generally conducted using read speech. A second approach focuses on the conversation-control functions of prosody, typically studied in naturalistic dialog. The inescapable question that arises is, how do these two uses of prosody interact? Integrative work addressing this question, bridging the gap between the two schools of prosody research, and aiming for a comprehensive model of prosodic phenomena, has been rare. This paper, as an attempt to explore all sides of one phenomenon, is a preliminary attempt at such an integration.

2 Subjective Analysis

We examined occurrences of low-pitch regions, using the above definition, in 68 minutes of casual English conversation. Here we carefully listened to the contexts of occurrence, noting what syntactic, semantic, and pragmatic functions tended to appear. In general, it seem that low-pitch regions co-occur with four main types of discourse activity. In this section we consider these four categories, and examine the ways in which they can correlate with back-channel feedback.

1. First, low-pitch regions often occur at points where the speaker considers that he has transmitted some infor-

⁰nigel@sanpo.t.u-tokyo.ac.jp I am very grateful to Liz Shriberg for penetrating questions and helpful suggestions, and for providing the Switchboard data. This work was supported by a grant from the Inamori Foundation.

file = marc01:1.au

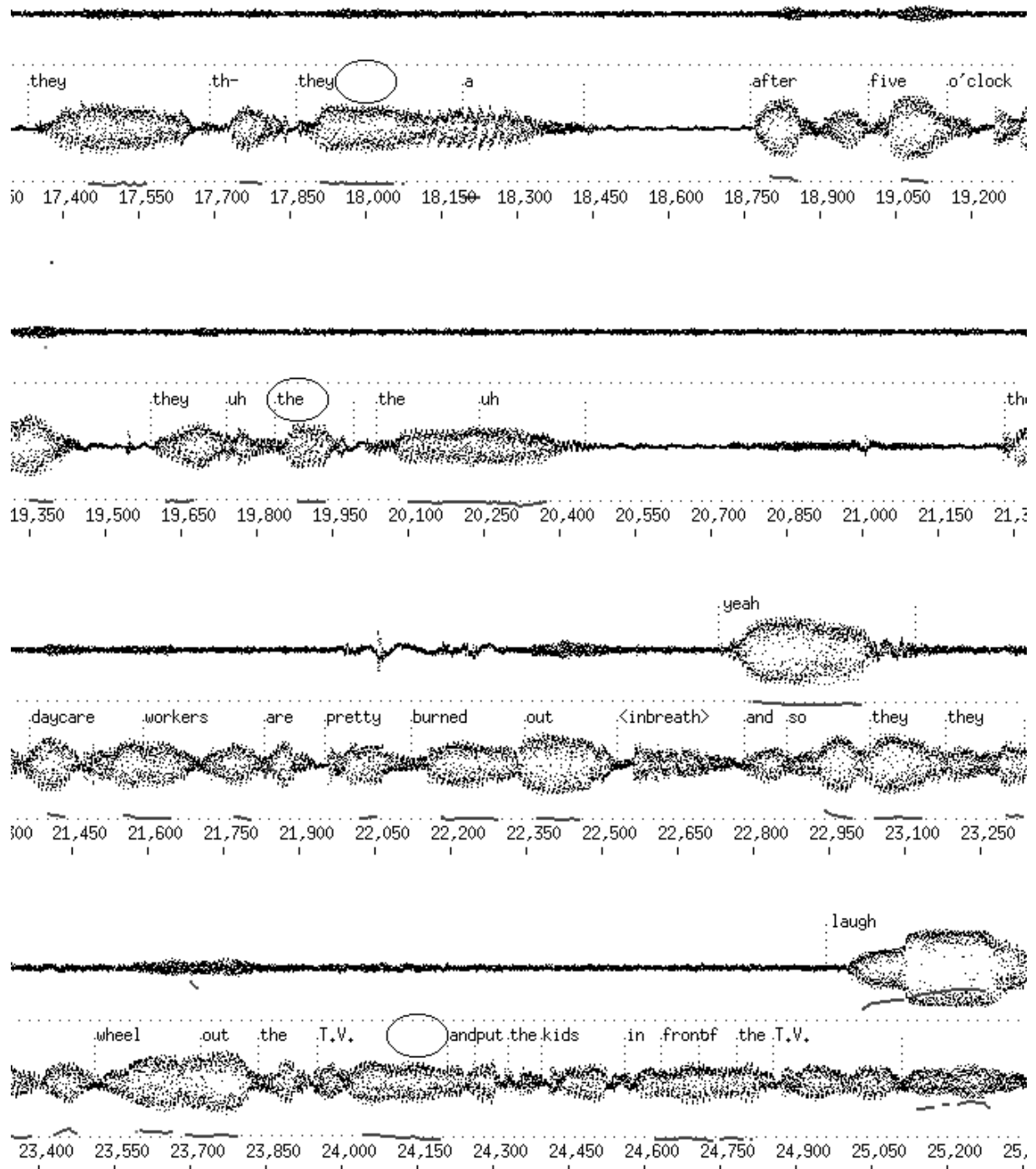


Figure 1: Conversation Fragment. Each of the four strips includes two rows and a timeline. In each strip the top row is one speaker and the bottom row the other. Each row includes: a transcription, the signal, the pitch, and the 30th percentile pitch level (horizontal dotted line). Wide ovals indicate the onset of low-pitch regions (see text). Although the pitch-tracker had a hard time with these speakers, it did well enough for purposes of roughly identifying low-pitch regions. This fragment occurred after some talk about television-watching habits and effects on children.

mation. At these points it is logically appropriate for the hearer to confirm receipt or understanding or interest with a back-channel. We can think of these low-pitch regions as conveying ‘this completes that thought, did you follow?’

Sometimes what has been transmitted is a complete new fact or proposition. Such low-pitch regions often co-occur with completion of a grammatical clause. Concomitantly, they also often co-occur with related lexical items, including clause connectives. (Note that these clause-ending markers and clause-ending low-pitch regions generally do not function as sentence ends. Rather, they indicate, perceptually, some degree of incompleteness, and also the likelihood of the utterance continuing. There is, of course, a closely related prosodic feature indicating finality and/or turn end, namely a sharp pitch drop; this also gives rise to low-pitch regions, but these generally end fairly quickly, generally lasting less than 110ms.)

Other times it seems that the low-pitch region occurs before an information unit is complete; that is, it occurs at a point where the speaker has expressed just enough information for the listener to infer his point, or where it seems that the speaker is ‘hinting’ at the meaning to be conveyed, and inviting the listener to jump ahead and infer the upcoming information. The last low-pitch region in Figure 1 (the oval in the bottom row) seems to be such a case. In such cases back-channel feedback sometimes appears before the speaker has completed a grammatical phrase or full proposition, and sometimes back-channel feedback in such cases takes the form of completing the speaker’s thought or sentence.

Rather less often, a low-pitch region which seems to mark the conveying of information appears with a repetition of a previous content word, produced for emphasis or clarity or when recovering from a false start. In such cases also, it often welcomes back-channel feedback; we can perhaps consider the low-pitch region to convey ‘I said it again, did you get it that time?’

2. Second, low-pitch regions also occur frequently with disfluencies and markers of formulation difficulties. In these cases we can think of the low-pitch region as saying, ‘I’m stuck, but keep listening, something meaningful will come out soon’. Two such low-pitch regions are seen in Figure 1 (at the first two ovals). Typical of such cases is the word *the* appearing in the lengthened, unreduced pronunciation which indicates formulation difficulty (Fox Tree & Clark 1997).

A related communicative function is taking the floor before actually saying anything, where the speaker utters some kind of filler or call for attention, and low-pitch regions occasionally occur at these times.

A disfluency with low pitch sometimes elicits back-channel feedback, presumably functioning as encouragement to continue. Many such disfluencies, however, do not seem to welcome back-channel feedback. In such cases, it seems that the speaker is using the disfluency to hold the floor.

3. Third, a low-pitch region often occurs together with

back-channel feedback itself. In the corpora such cases occasionally elicit a confirmatory word or sigh.

4. Fourth, there are fairly rare cases where 110 ms of low pitch occurs as part of a substantially longer region of low pitch, with a special meaning. These cases include utterances with reduced pitch range, as in parentheticals and ‘self-directed speech’, that is comments said ‘under the breath’, and perhaps groans.

Thus low-pitch regions seem to bear four interrelated functions. Interestingly, some of these can be distinguished without reference to the lexical context. For example, category 2, low-pitch regions involving disfluencies, typically occurs towards at the start of an utterance; roughly within the first 700ms or so. Category 3, low-pitch regions co-occurring with back-channel feedback, are distinguishable because such utterances are generally short and/or spectrally stable (in Switchboard (Jurafsky *et al.* 1998) over half of English back-channels belong to that series of ‘words’ typified by *uh-huh* and *uh*, which are drawn from limited phonetic inventory (Ward 1998), and over half of the rest are variants of *yeah*, which is typically lengthened). Category 4, low-pitch regions in groans and the like, is distinguishable by the unusually long duration of the low pitch region.

3 Statistical Analysis

The analysis of the previous section being based on only a small amount of data, and on one person’s subjective judgements, we set out to find independent evidence regarding the possible functions of low-pitch regions. To do this we measured the propensity of various lexical items to occur together with low-pitch regions.

Specifically, we gathered statistics from 65 conversation sides (38330 words) from an independent corpus of English, Switchboard. These sides were selected because a low-pitch discrimination function, defined for a another purpose, fell between the 25th and 27th percentile pitch levels for these sides. For purposes of identifying low-pitch regions, unvoiced speech regions adjacent to low-pitch regions were treated as having low pitch.

There are four ways in which a word can be involved with a low-pitch region : A. a word being fully inside a region, B. a word overlapping the start but not the end of a region, C. a word overlapping the end but not the start of a region, and D. a word fully containing a low-pitch region, overlapping on both sides.

The first topic explored was disfluencies. Because repetition of a word is often indicative of a disfluency (Bear *et al.* 1992), we investigated whether low-pitch regions occur more often during such repetitions. Specifically, we measured the frequency with which repeated words are involved, in any way, with a low-pitch region. It turns out that the first of a pair of repeated words involved a low-pitch region 27.9% of the time, and the second of a pair

	<i>inside</i> (A) %	<i>start</i> (B) %	<i>end</i> (C) %	<i>cont.</i> (D) %	<i>any</i> %	<i>count</i>	<i>freq.</i> %	<i>len.</i> ms.	<i>i.l.</i> %	
I	9.3	↓2.7	5.3	↓0.8	↓18.1	1540	4.0	172	10.4	I
AND	10.4	8.6	5.5	2.7	↑27.2	1330	3.5	300	11.4	AND
THE	↑13.2	↓3.3	↓2.1	↓1.5	↓20.2	1126	2.9	147	18.7	THE
YOU	7.6	↓5.9	6.8	↓1.2	↓21.5	1103	2.9	197	8.3	YOU
TO	10.9	↓3.7	↓2.5	↓0.7	↓17.7	1009	2.6	165	17.5	TO
A	↑13.6	↓4.3	↓2.8	↓0.7	↓21.4	888	2.3	120	18.2	A
THAT	↑11.5	11.4	5.1	3.2	31.2	783	2.0	312	13.5	THAT
IT	↑14.3	↓4.5	6.1	2.3	27.2	753	2.0	254	17.5	IT
UH	↑13.5	↑13.8	↑10.1	↑8.6	↑46.0	733	1.9	583	14.7	UH
KNOW	8.0	10.5	5.5	3.2	27.1	660	1.7	307	8.0	KNOW
OF	11.6	6.7	↓1.3	↓0.6	↓20.2	639	1.7	143	15.8	OF
SO	8.4	8.0	7.1	1.8	25.3	450	1.2	418	8.7	SO
YEAH	5.8	↑14.3	5.6	↑7.6	33.3	447	1.2	719	6.3	YEAH
IN	↑14.0	6.4	3.8	↓0.7	24.9	421	1.1	189	14.5	IN
THEY	9.9	↓3.6	7.0	1.4	↓22.0	414	1.1	176	9.2	THEY
HAVE	5.3	8.3	3.5	0.8	↓17.9	375	1.0	241	6.4	HAVE
LIKE	7.0	↓4.5	4.5	1.6	↓17.6	374	1.0	284	9.1	LIKE
BUT	9.3	8.5	6.8	2.0	26.5	355	0.9	356	11.3	BUT
WE	7.2	↓2.6	4.6	↓0.6	↓15.1	345	0.9	191	7.8	WE
IT'S	11.4	↓4.0	6.2	0.0	↓21.6	324	0.8	263	15.4	IT'S
WELL	4.4	5.0	4.4	2.8	↓16.6	320	0.8	351	5.0	WELL
WAS	11.7	4.5	7.8	1.6	25.6	309	0.8	245	16.5	WAS
JUST	5.0	5.0	9.0	2.3	↓21.3	300	0.8	338	15.0	JUST
IS	10.9	↑13.9	3.4	1.4	29.6	294	0.8	242	13.9	IS
DO	↓4.1	7.5	3.1	2.0	↓16.7	293	0.8	242	4.4	DO

Table 1: Words and their involvements with low-pitch regions. For each word, the first four columns indicate the percentage of tokens which interact with a low-pitch region in each way (see text). The *any* column is the sum of the first four columns, thus counting all words which overlap in any way with a low-pitch region. ↑ and ↓ designate frequencies which differ significantly from expectation (the aggregate ratios from the last row of the table) at the 1% confidence level by a chi-squared test. The *count* column indicates the total number of tokens of each word in the corpus, and the *freq* column the frequency of each word. The *len* column indicates the average duration, in milliseconds. The *i.l.* column indicates the percentage of tokens which appear entirely in low pitch. The table includes the 25 most frequent words in the corpus plus those words which have significant behavior in any of the first five columns.

of repeated words involved a low-pitch region 22.1% of the time. Words in general are involved with low-pitch regions 24.7% of the time. Thus the first occurrences of the words do appear more often in low-pitch regions (significant at the 5% level by a chi-squared test), as would be expected given that the first occurrence of a repeated word is generally disfluent. Moreover, the second occurrences of the words occur less often with low-pitch regions (although the difference is not significant), as might be expected, in that the second repetition is often the restart or beginning of a correction.

The second exploration was a determination of which lexical items were involved with low-pitch regions more or less than the average lexical item. The data is seen in Table 1. Although it is difficult to draw firm conclusions, some patterns seem to be present:

The pronouns *I*, *you*, *they*, *we*, *he*, and *she* are seldom involved with low-pitch regions at all. This is what would be expected from the existence of category 1 of the previous section, given that pronouns seldom bear new information (although these may also be due in part to the fact that

subject pronouns are typically early in the phrase, before declination sets in). The fact that *it* does not follow the same pattern may be due to the frequency of the various non-referential uses of *it*.

The determiners *the*, *a*, *that* and *an* tend to appear inside low-pitch regions (although not in other positions). This is understandable as a result of disfluencies (category 2) involving problems formulating noun-phrases, which often manifest themselves at the onset of such phrases.

The prepositions *in*, *for*, *with*, and *at* also often appear inside low-pitch regions. Again this may be due to a disfluencies (category 2) in the formulating of upcoming noun phrases. The fact that other prepositions, including *to*, *of* and *up*, do not fall into this pattern, may be due to the higher frequency of these words in positions other than pre-noun-phrase.

The ‘word’ *uh* is involved with low-pitch regions in every way. This again is what one would expect given the correlation between low-pitch regions and disfluencies (category 2).

	<i>inside</i> (A) %	<i>start</i> (B) %	<i>end</i> (C) %	<i>cont.</i> (D) %	<i>any</i> %	<i>count</i>	<i>freq.</i> %	<i>len.</i> ms.	<i>i.l.</i> %	
FOR	↑15.6	5.0	2.7	2.7	26.0	262	0.7	272	19.5	FOR
UHHUH	↓2.0	↑17.4	6.3	↑16.6	42.3	253	0.7	1166	2.0	UHHUH
THINK	4.4	6.0	4.8	0.8	↓16.1	248	0.6	315	4.0	THINK
DON'T	↓3.1	7.2	4.5	0.9	↓15.7	223	0.6	274	3.1	DON'T
THAT'S	7.8	↓2.3	4.6	0.5	↓15.1	219	0.6	300	10.0	THAT'S
MY	6.9	6.0	5.5	0.0	↓18.4	217	0.6	223	6.9	MY
REALLY	↓2.0	11.2	6.8	2.0	22.0	205	0.5	445	2.0	REALLY
NOT	5.2	6.3	4.7	3.1	↓19.4	191	0.5	286	4.7	NOT
BE	9.8	↓2.7	6.0	1.1	↓19.6	184	0.5	230	12.0	BE
RIGHT	3.3	13.3	8.8	↑17.7	43.1	181	0.5	835	3.9	RIGHT
WITH	↑14.9	3.9	3.3	2.2	24.3	181	0.5	226	14.4	WITH
WHAT	3.9	7.2	5.6	1.1	↓17.8	180	0.5	275	5.6	WHAT
UM	13.4	11.7	8.9	↑6.7	40.8	179	0.5	784	13.4	UM
HE	7.9	4.9	6.7	0.6	↓20.1	164	0.4	195	8.5	HE
IF	8.1	3.1	4.3	0.6	↓16.1	161	0.4	190	14.3	IF
AS	11.4	2.9	2.1	0.0	↓16.4	140	0.4	203	14.3	AS
AT	↑19.4	5.4	4.7	2.3	31.8	129	0.3	180	22.5	AT
UP	8.1	4.1	2.4	0.8	↓15.4	123	0.3	250	9.8	UP
HAD	5.7	9.0	0.8	1.6	↓17.2	122	0.3	250	5.7	HAD
WOULD	7.5	3.3	5.0	1.7	↓17.5	120	0.3	201	9.2	WOULD
KIND	4.5	6.3	2.7	0.0	↓13.5	111	0.3	252	6.3	KIND
THEY'RE	11.2	2.0	2.0	0.0	↓15.3	98	0.3	266	14.3	THEY'RE
SHE	10.8	4.3	0.0	0.0	↓15.1	93	0.2	277	12.9	SHE
NO	4.3	3.3	3.3	2.2	↓13.0	92	0.2	391	4.3	NO
SOME	8.0	2.3	2.3	1.1	↓13.6	88	0.2	304	10.2	SOME
BEEN	5.1	2.5	2.5	0.0	↓10.1	79	0.2	241	5.1	BEEN
MUCH	2.8	0.0	4.2	1.4	↓8.5	71	0.2	298	4.2	MUCH
AN	↑21.9	4.7	3.1	0.0	29.7	64	0.2	185	25.0	AN
YEARS	9.5	↑22.2	12.7	9.5	54.0	63	0.2	534	9.5	YEARS
OUR	6.3	4.8	1.6	0.0	↓12.7	63	0.2	179	6.3	OUR
THREE	1.8	3.6	3.6	1.8	↓10.9	55	0.1	332	3.6	THREE
ONLY	0.0	2.1	4.2	0.0	↓6.2	48	0.1	244	0.0	ONLY
PUT	2.2	0.0	2.2	0.0	↓4.4	45	0.1	206	11.1	PUT
FEEL	2.6	0.0	5.3	0.0	↓7.9	38	0.1	302	2.6	FEEL
FIRST	0.0	0.0	2.7	0.0	↓2.7	37	0.1	357	2.7	FIRST
TAKE	0.0	0.0	0.0	2.8	↓2.8	36	0.1	262	5.6	TAKE
DOES	3.3	0.0	0.0	0.0	↓3.3	30	0.1	254	3.3	DOES
avg	8.0	8.0	5.8	2.9	24.7	-	-	335	9.3	avg
total	3055	3048	2236	1116	9455	38330	100.0	-	3579	total

Table 1 (continued): more words, plus averages and totals for the corpus as a whole.

The typical back-channels *yeah*, *right*, *uhhuh* and *uh* tend to start or contain low-pitch regions (category 3).

The word *and* also has some propensity to be involved with low-pitch regions. This may be because filled pauses are often transcribed as *and*, or because *and* tends to appear after a clause or thought is complete (categories 1 and 2).

The above analysis must be regarded as tentative since there are at least two potential confounds. The first is the fact that some words are, lexically, unaccented, and would for that reason alone, tend to fall in low-pitch regions. That is, lexical items which are seldom stressed will seldom have high pitch, and so are a priori more likely to occur in low-pitch regions. The *i.l.* ('intrinsically low') column of the table indicates this as the percentage of tokens which

are completely in a low pitch, regardless of whether or not that low pitch period constitutes part of a 110ms low-pitch region. As expected, there is some correlation between these values and the values for column A, although the correlation is far from perfect. The second confounding factor is word length. Words which are very short will logically tend to have higher values in the A (fully inside) column, whereas words which are very long will tend to have higher values in the D (containing) column, other things being equal. However in the table the correlation between the *len* column and these other columns does not appear to be strong.

4 Summary

First, we considered the possibility that low-pitch regions could be analyzed as ‘discourse markers’, bearing specific pragmatic functions, rather than as ‘signals’. Both subjective judgements and statistical analysis suggest that low-pitch regions are indeed associated with specific pragmatic functions. These functions, detailed above, include: the transmission of information, the occurrence of a disfluency, and the occurrence of back-channel feedback. However, as seen in Section 2, the two accounts are compatible, indeed inextricable: since these pragmatic functions are themselves compatible with the subsequent occurrence of back-channel feedback, it is impossible to say whether low-pitch regions function as markers or signals or both.

Second, we considered the possibility that these low-pitch regions are not prosodic features in their own right, instead merely reflecting the absence of other prosodic features. The present study did not test this proposition directly, but the fact that low-pitch regions do appear to bear specific pragmatic functions is further evidence that they do exist as real prosodic features.

The introduction to this paper mentioned the idea of an attempt to ‘bridge a divide’ in prosody research, by investigating the interactions between various functions of prosody. However, the above results suggest that low-pitch regions occur relatively more frequently in positions that do not involve content words, well-formed syntactic structures, or new information. Thus they tend to occur in locations in which the other functions of prosody are less active. In terms of our original question regarding the relation between lexical/syntactic prosody and dialog prosody, in this case there turns out to be little relation after all. It is interesting to speculate whether this is true in general, and whether it is appropriate, after all, for the two main strands of prosody research — the one focusing on read speech and the other on conversation and dialog — to proceed as two complementary but non-interacting paradigms.

In any case, all of these conclusions are tentative, and will remain so pending the development of more research infrastructure, in the form of large corpora of conversations with rich functional annotations.

5 References

- Bear, John, John Dowding, & Elizabeth Shriberg (1992). Integrating Multiple Knowledge Sources for Detection and Correction of Repairs in Human-Computer Dialog. In *Proceedings of the Association for Computational Linguistics*, pp. 56–63.
- Fox Tree, Jean E. & Herbert H. Clark (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, 62:151–167.
- Jurafsky, Daniel, Elizabeth Shriberg, Barbara Fox, & Traci Curl (1998). Lexical, Prosodic, and Syntactic Cues for Dialog Acts. In *Association for Computational*

Linguistics, Workshop on Discourse Relations and Discourse Markers.

Ward, Nigel (1997). Responsiveness in Dialog and Priorities for Language Research. *Systems and Cybernetics*, 28(6):521–533.

Ward, Nigel (1998). The Relationship between Sound and Meaning in Japanese Back-channel Grunts. In *Proceedings of the 4th Annual Meeting of the (Japanese) Association for Natural Language Processing*, pp. 464–467.

Ward, Nigel & Wataru Tsukahara (submitted). Prosodic Features which Cue Back-Channel Feedback in English and Japanese. *Journal of Pragmatics*.