Register as a variable in prosodic analysis: The case of the English negative

Malcah Yaeger-Dror

Department of Linguistics, University of Arizona, Tucson, AZ 85721, USA

Received 4 April 1993; revised 12 November 1995

Abstract

The next generation of text-to-speech systems will have to be more sensitive to sociolinguistic 'style' variables. In order to assist in the adaptation of synthesis to a wider range of contexts, this article examines several sociolinguistic parameters which have been shown to influence the realization of negatives in actual discourse, analyzing their effects on the realization of negatives in English prose readings. Consistent with the results found in an earlier study, the analysis shows that pitch prominence on negatives is not common in read prose passages, and is even less common in read dialogue. Informational content (and consequently pitch prominence on negatives) is more important in prose addressed to children than in narrative reading for adults, while the more formal the prose for adults, the less likely prominence is to occur. The results show a surprising absence of conformity with 'theoretical' linguistic expectations, highlighting the necessity for consideration of register as an important variable for speech synthesis.

Zusammenfassung


Résumé

La prochaine génération de systèmes de synthèse à partir du texte devra être plus sensible aux variations socio-linguistiques. Dans ce contexte, on a étudié plusieurs paramètres socio-linguistiques qui ont une influence sur la réalisation de la négation dans le langage parlé, en examinant leurs effets sur l'accentuation des négatives lors de la lecture de prose anglaise. Conformément aux observations faites lors d'une étude antérieure, cette analyse a montré que la prominence intonative n'est pas fréquente dans la lecture de prose et encore moins fréquente dans des dialogues lus. Le contenu informationnel (et donc la prominence intonative) est plus important dans la prose lue aux enfants que dans celle lue aux adultes. Dans ce dernier
cas, plus la lecture est formelle, moins la prominence intonative apparaît. Les résultats montrent une étonnante absence de conformité avec les prédictions linguistiques ‘théoriques’.

Keywords: Spontaneous speech; Read speech; Speaking style; Discourse variables; Focus

1. Introduction

In 1988, Rodman made the point that “In the long run, the field of sociolinguistics may make the largest contribution to the field of speech recognition” (Rodman, 1988). An entire issue of this journal (Vol. 11, Nos. 4–5, October 1992) was recently given over to the analysis of ‘style’ variation as it influences speech, with the understanding that future text-to-speech routines will be expected to have a much greater degree of adaptability than has been possible heretofore. The present study will take this growing need for adaptability of text-to-speech systems as a point of departure, and assume that variation should be modeled on human performance (Argente, 1992; Bladon et al., 1987; Granström, 1992). Having determined that one ‘style’ worth isolating is prose read for the visually handicapped (Fant and Kruckenberg, 1989; Granström, 1992), this study will analyze speech read in that style, in order to determine whether it will be important to incorporate a range of pragmatic variables into an algorithm for prose synthesis, or whether a simpler algorithm is likely to be both comprehensible and acceptable to listeners.

The study will concern itself primarily with the analysis of how negatives are realized intonationally in this specific, narrowly defined discourse context, to determine whether negatives should be synthesized with pitch prominence or not. This problem was chosen because earlier studies have shown that negatives carry important information; consequently, the intelligibility of negatives is critical to listener understanding. However, other studies have shown that negatives also ‘carry’ critical pragmatic as well as semantic information, and the paper will show that both pragmatic and semantic variation influence the prosodic realization of negatives in actual discourse.

Section 2 will review in greater detail the information which is already available on the prosodic realization of negatives. Section 3 will describe the corpora chosen for the present analysis, Section 4 will describe the analysis, and Section 5 will draw conclusions, and suggest future areas of research which will benefit from an understanding that both speech recognition and synthesis can ultimately benefit from theoretical and empirical sociolinguistic input.

2. Review of the literature

2.1. On the realization of ‘accentuation’

Most linguists agree that ‘focal’ or ‘new’ information in a discourse should be realized with accent on the ‘focal’ word (e.g., Bruce, 1992; Coker and Umeda, 1971; Engstrand, 1989a,b, 1992; Fowler, 1988; Fowler and Housum, 1987; Hirschberg, 1990; Koopmans-Van Beinum, 1992; Levac, 1993; MacWhinney and Bates, 1978; Nooteboom and Kruyt, 1987; Prince, 1981). In English, accent is generally acknowledged to be realized phonetically as (relative) vowel length, peripherality and intensity, as well as pitch prominence; however, for present purposes (for reasons which will emerge in Section 4), the defining criterion for ‘accent’ will be pitch prominence.

2.2. On the realization of negation – the semantic perspective

Theoretical studies of negation have concluded that (except in a few narrowly circumscribed instances) negatives convey important semantic information in a discourse (Horn, 1990; Tottie, 1991). Consequently,
Table 1
Conflicting claims of the influence of these theories on prosodic parameters

(a) Claims made by the Cognitive Prominence 'Rule', i.e., new/focal information will be phonetically salient (pitch prominent and uncontracted), while old information will be nonsalient

<table>
<thead>
<tr>
<th>Contraction</th>
<th>Pitch</th>
<th>New Information</th>
<th>Old Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Prominent</td>
<td>- Contracted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Prominent</td>
<td>+ Contracted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Claims made by the Agreement 'Rule', i.e., negatives supportive of others' positions will be phonetically salient (pitch prominent and uncontracted) even if they provide no focal information; disagreements will be nonsalient

<table>
<thead>
<tr>
<th>Contraction</th>
<th>Pitch</th>
<th>Agreement/Supportive</th>
<th>Contradiction/Nonsupportive</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Prominent</td>
<td>- Contracted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Prominent</td>
<td>+ Contracted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

despite the fact that negatives are 'closed class' items linguistically (Prince, 1981; Hirschberg, 1990), and closed class items usually do not provide new information (Prince, 1981) or receive cognitive focal prominence (Hirschberg, 1990), analyses of negatives have assumed that they should be semantically regarded as conveying focal information (Horn, 1990; Tottie, 1991) and therefore should be prosodically prominent. This will be referred to as the semantic Cognitive Prominence Rule. Table 1a presents the claims of the Cognitive Prominence Rule more graphically.1

Acoustic studies of read sentences have confirmed that negatives are overwhelmingly pitch prominent in a speech when pragmatic concerns are minimized and only semantic information is conveyed. The best documented acoustic study of the Cognitive Prominence Rule (O'Shaughnessy and Allen, 1983) elicited sentences from 3 MIT linguists, to determine whether pitch was prominent on negative bearing elements; examples can be found in (1). The negatives analyzed are underlined in the text: (1) O'Shaughnessy examples, cited in (O'Shaughnessy and Allen, 1983).

a. Which books was Joe not studying?
a'. Which books wasn't Joe studying?
b. Which books weren't the boys studying?
b'. Which books were the boys not studying?

They found that pitch was most often prominent on the negative in isolated read sentences such as those in (1).

Independently, Hirschberg (1990) based her module for intonational synthesis on her analysis of National

1 Quantitative studies of negation in actual speech have documented vowel reduction (contraction) on negatives, although contraction and vowel reduction have not been considered jointly. This paper will not consider the issue of contraction (as vs. vowel lengthening), but it should be understood that contraction is (at least theoretically) inversely related to pitch prominence, and has been shown to occur preferentially in just those situations where pitch prominence is least likely to occur. This is true in prose (Biber, 1988; Tottie, 1991) as well as conversation (Yaeger-Dror, 1996). However, since the choice of full vowel (not) or deleted vowel (n't) is regulated for the present corpus by genre and other prose conventions, only pitch variation will concern us here.
Public Radio (henceforth NPR) news broadcast readings, which she refers to as ‘newsspeak’. Initially she assumed that negatives, being closed class, should not be pitch prominent. However, like O’Shaughnessy, Hirschberg found that a reader used pitch prominence on negatives. Based on her results, Hirschberg proposed that negatives, or negative bearing elements, should be consistently pitch prominent in synthesized speech, and she proposed an algorithm for synthesis which requires pitch prominence on negatives. Both quantitative studies supported the theoreticians’ understanding that negatives should be realized with pitch prominence, because they provide critical focal information.

2.3. On the realization of negation – the pragmatic perspective

On the other hand, Schegloff, Jefferson and Sacks, in an important Language article (Schegloff et al., 1977), presented evidence that in conversational speech, there is a ‘preference for agreement’, to which speakers adapt their speech. This ‘rule’ of conversation states that all other things being equal, speakers will attempt to agree with coparticipants in a conversation. Agreements will be referred to as ‘supportive’ statements, and examples will be coded ‘[s]’. Examples of negatives used supportively will be found in (5) and (7) below. Schegloff, Jefferson and Sacks presented evidence that when negatives are not supportive, negatives will be downplayed as much as possible (when not avoided altogether). In general, only when a speaker does not pick up on an error does the listener disagree overtly. Where such overt disagreements occur, they will be referred to (following Brown and Levinson, 1978), as ‘face threatening’ statements, and examples will be coded ‘[ft]’. The hypothesis that agreement is preferred over disagreement will be referred to as the CA (Conversational Analysis) Social Agreement Rule. Table 1b presents the claims of the Social Agreement Rule more graphically. An example of a disagreement which conforms to these expectations can be found in (2). The negatives, again, are underlined.(2) Sample sentence cited in (Schegloff et al., 1977).

2 Based primarily on evidence of phonological variation, both Bell (1991a) and Yaeger-Dror (1991) have independently come to the conclusion that radio (as vs. TV) news presents the most self-conscious broadcast speech. In the UK, RP, or ‘BBC English’ (Trudgill, 1986; Bell, 1991a), is understood by sociolinguists to reflect the British ‘Standard’. Fortunately, or unfortunately, a media ‘standard’ such as that maintained in other ex-British colonies like Canada or New Zealand (Bell, 1991b) cannot actually be found for the US; however, NPR English, under its guise as ‘newsspeak’ (Hirschberg, 1990) or ‘news broadcasterese’ (Di Paolo, 1992) has been chosen by some American linguists as reflecting the neo-standard target for American speech. Wightman and Ostendorf (1991, 1992), and Price et al. (1991), like Hirschberg, have taken advantage of the fact that NPR news readers are most likely to use this stereotypical ‘standard’.

3 To quote: “negative article, negative modal, negative do”.

4 This does not imply that only pitch is relevant to the analysis, but that they found fundamental frequency variation to be the most salient acoustic parameter to be analyzed.

5 Other studies have referred to such ‘face threatening’ comments as ‘remedial’ (Goffman, 1971), or ‘confrontational’. I have chosen ‘face threatening’ because it is most commonly used.

6 Transcription conventions include:

• bold text to signify increased (perceived) loudness,
• italic text to signify (perceived) pitch prominence,
• colons to signify increased (perceived) duration on the immediately preceding sound (e.g., “That’s what I: can’t make out”),
• as well as a hyphen to symbolize abrupt ‘cut off’ of speech (e.g., “I’m getting at- at the: at-”).
• Punctuation reflects transcriber perception of intonation rather than sentence structure. That is, a comma signals reduction in pitch and amplitude, and a slow down across the last several syllables, while period signifies a more extensive reduction of these parameters; a question mark signifies a rise in pitch, combined with the slow-down and (generally) reduction in amplitude, while an exclamation mark signifies a rise in pitch, amplitude and duration over the last syllables. The latter represent larger ‘break’s than the former, while the cut-off represents a break without the intonational changes.

• extreme lowered amplitude, as is common in parentheticals, is transcribed with a degree sign at each end of the relevant segment (e.g., “I’m not sure ...”),
• General comments are in double parentheses (e.g., ‘(pause)’ or ‘(unclear)’).
• Inbreaths are transcribed as .h (e.g., “within a .h single dialect”).
• All examples cited are taken directly from the transcribed corpus denoted in single parentheses at the end of a citation. Table 2 provides information concerning these corpora, and further details can be found in (Sacks, 1992).
• In all of the examples cited, the word (or location) at issue is underlined.
Disagreement, to show that Wife-1′s preceding comment does not sound plausible.

\begin{itemize}
  \item [a.] Husband-2: That′s what \textit{I} can′t make out. uh, y′know. If that′s true, some of the birds are ahead of ’im.
  \item [a′] c′n′t even see ’im. \(7\) (CH-4)
\end{itemize}

In example (2), even though the negative provides important new information, it does not receive pitch prominence (as witnessed by the transcription convention of italicising perceptually pitch-prominent items), because it is used to express disagreement.

Yaeger-Dror (1985) pointed out that the Cognitive Prominence Rule and the Social Agreement Rule conflict, and attempted to determine quantitatively which is stronger. The key to the conflict initially appeared to be variation in style and register. A slight digression on the distinction among registers is therefore relevant. Although a full analysis of such variation is not appropriate here, certain parameters must be taken into consideration when determining the material for analysis:

\subsection*{2.4. On the differentiation of speech register parameters}

One key factor is the distinction between task and speech oriented material (Yaeger, 1974; Labov, 1989): Material which is read for no purpose other than analysis of the speech can be referred to as \textit{speech oriented}; Shockey (1995) has referred to such speech as ‘lab speech’, or ‘ideal’ speech. O′Shaughnessy′s material, for example, was purely speech oriented. In contrast, Hirschberg′s data, while carefully read, were used to convey information to listeners, and can be defined as \textit{task oriented}; Shockey (1995) refers to such speech as ‘real’ speech. While this distinction is quite important for the analysis of phonological variation (Yaeger, 1974; Labov, 1986, 1989), it does not appear to be critical for the present research situation, since both O′Shaughnessy and Hirschberg achieved similar results.

\footnote{\textit{The present speaker (Husband-2) does not believe (as was claimed by the previous speaker, Wife-1) that the lead bird in a formation is actually the furthest back. Not only the speaker mitigates the face threatening implications of his disagreement by prosodically neutralizing the negative, but in addition he does a certain amount of interactional work to avoid an overt disagreement. He does not say “That \textit{can′t be}! the other birds can′t even see the lead bird if that is true!” but first implies that the fault here is probably his own “That′s what \textit{I} can′t make out”, then interpolates both filled and unfilled pauses, to permit the previous speaker sufficient time to self correct if possible: “uh, y′know. if that′s true.”, and finally arrives at an actual disagreement when no repair is made by the original speaker: i.e., if the previous speaker is correct, then the so-called leader-bird cannot even be seen by the followers. Even when the negative is actually used, it is contracted and reduced prosodically.

The 1977 paper presented many examples garnered from actual (American English) conversations, showing ways in which even very young speakers will avoid direct disagreements; three such examples, demonstrating listeners’ ability to avoid a negative/direct disagreement at all costs:

\textit{(6.i)} A: Hey, the first time they stopped me from sellin’ cigarettes was this morning. (1.0)
B: from \textit{selling} cigarettes?
A: from \textit{buying} cigarettes, They said, (GTS-3:42: SJS (43), p. 370)

\textit{(6.ii)} Kids playing.
Steven: 1, 2, 3, ((pause)) 4, 5, 6, ((pause)) 11 8 9 10.
Susan: 11? 8 9 10?
Steven: 11, 8, 9, 10.
Nancy: 11?
Steven: 7, 8, 9, 10. (GJ:FN: SJS (58), p. 375)

\textit{(6.iii)} Ken: `e likes that waider over there,
A: Waiter?

It seems that in many cases listeners will exhaust attempts at eliciting corrections from the speaker before directly disagreeing. See also example \textit{(4)b′}. Other examples are cited in (Schegloff et al., 1977) and (Schegloff, 1989).}
Another key factor, initially proposed by Labov (1966), distinguished several ‘style’s which he differentiated on the basis of apparent ‘attention paid to speech’ (with typical points along the continuum taken by word list style (WLS), read sentences (RS) and so-called casual style (CS). (Justification of the relationship between ‘style’ and attention paid to speech is found in (Labov, 1966; Bell, 1984; Prince, 1987).) This continuum has also been shown to influence phonology and syntax (Labov, 1966, 1972; Ochs, 1979), and certainly will prove to be important to the analysis of prosodic variables, although it has not been considered in prosodic analyses to date. The present study will be simplified by choosing a task with a ‘style’ similar to O'Shaughnessy’s and Hirschberg’s.

What is more generally termed ‘style’, is referred to by psycholinguists as ‘types of discourse’ (Levelt, 1989, p. 111), and by British-tradition sociolinguists as ‘genres’ or ‘text types’ (Svartvik, 1990; Biber, 1988, pp. 68ff), while American sociolinguists discuss variation in ‘registers’ (Yaeger-Dror, 1991; Finegan and Biber, 1994). Registers, in turn, vary along several continuous parameters, or to use Biber’s term ‘dimensions’. For example, in Biber’s analysis (Biber, 1988), the informational-interactional ‘Dimension I’ covaried with most of the syntactic and morphological linguistic variables he studied. Although they did not share the same terminology, Yaeger-Dror (1985) and Biber (1988) independently determined the importance of this ‘dimension’. 8

While Hirschberg (1990) studied informational speech, Table 2 lists (on lines begun with “”) the corpora used for (Yaeger-Dror, 1985), which included both informational and interactional data. In that study, Yaeger-Dror checked the original perceptual analysis revealed in Jefferson’s transcription of conversational data against acoustic analysis. To permit comparison with other studies cited, only the acoustic analysis will be discussed here. The 1985 study measured fundamental frequency, relative amplitude and duration for all negatives in the corpora analyzed: over 800 tokens in all. Of those negatives, only 301 could be classed as disagreements. 9

---

8 Yaeger-Dror (1985) initially distinguished between ‘instructional’ and ‘interactional’. Biber’s Dimension I actually contrasts (Biber, 1988) ‘informational’ and ‘involved’. As a choice for a general term, Biber’s ‘informational’ is both more general and more appropriate than ‘instructional’, since one can wish to convey information without being overtly didactic. On the other hand, while Biber’s term ‘involved’ resonates with the older understanding of ‘style’ being defined by the degree of attention paid to speech, it implies a degree of involvement which is not borne out by either Biber’s or Yaeger-Dror’s data, and in the present study the endpoints of the continuum will be termed ‘interactional’ and ‘informational’.

9 Comprising between 33% (8 of 24 negatives used by the therapist) and 60% (87 of 146 used by other adults) of all negatives used.
Yaeger-Dror (1985) found that in a primarily ‘informational’ tutorial setting [‘SCRL’ corpus], DARPA researchers used pitch prominence relatively often on information bearing negations in primarily informational statements; this is consistent with the Cognitive Prominence Rule. However, even in this setting they did so less consistently than either O’Shaughnessy’s or Hirschberg’s speakers.

**Informational** needs took precedence not just in tutorial-instructional register (the SCRL corpus) but also in the conversations analyzed when (occasionally) segments were presented in a marked ‘informational manner’. Turns which occur either in an informational setting, or in an informational-neutral turn within a more interactional conversation, will be referred to as ‘neutral’ statements, and examples will be coded ‘[n]’. An example is found in (3).(3) Informational presentation of a sign read, during a telephone conversation (speaking about a building which has been shut down).

A: An’ the doors are all boarded up an’ it has signs on it
Then
Please - do - not - enter.

An’ there’s no - way to get in the building. (TG)

These informational negatives were more likely to be prominent even when the larger conversational context was interactional.

Yaeger-Dror (1985) concluded that informational settings should be clearly distinguished from interactional settings, since negations are more likely to be pitch prominent in the informational setting.

Within the meta-group of registers considered ‘interactional’, Goffman (1971), Sacks (1992) and Brown and Levinson (1978) all found that a speaker’s intent, whether ‘neutral’ [n], ‘supportive’ [s] or ‘confrontational’ [ft], influences speech behavior. It was discovered that the dichotomy between ‘neutral’ and ‘affected’, and the continuum from ‘polite’/‘supportive’ to ‘confrontational’, all are important factors in the analysis of accentuation on negation. The 1985 data were categorized for these semantic and pragmatic variables which had been isolated in the Schegloff et al. (1977) study on which the Social Agreement Rule was based. As already mentioned, the same coding will be used in the present analysis.

Pitch was more likely to be prominent when a speaker was self-correcting (as in examples in (4)) than when (s)he was merely conveying information (as in the examples in (3)).(4) Self-corrections, in the middle of explanatory discussions of new techniques being advocated:

a. [n] JSH: That’s not necessarily the best way, but...
   (SCRL)

b [ft] DK: We assumed within a single dialect, that everyone has’s same internalized ((cough))
   very abstract representation for a word. ((pause))

[ft] LL: I’m not sure if that assumption is made cross dialects, is ((unclear))

DK: Does English?
   (long pause [for answer])
   All English speakers?
   (longer pause)

b’ [s] I guess not. 10
   (SCRL)

This can be understood when we consider that the Cognitive Prominence Rule is not countered by the Social

---

10 As in (2), we see that direct disagreement is avoided: it is clear that LL’s ‘I’m not sure’ implies ‘A single representation is not appropriate.’ DK then twice attempts to garner support, with the lack of response from listeners implying agreement with LL rather than DK. This leads to a self-correction by DK, with the negative self-correction (which displays agreement with the listener(s)) clearly pitch and amplitude prominent.
Agreement Rule in this context: The negative provides new information (favoring Cognitive Prominence), and cannot be heard as threatening other parties to the conversation (in fact, agrees with their inferred position, and so need not be reduced.) In fact, only one self-correction of 20 used was not pitch prominent, and that one occurred in the tutorial where reduction should be least likely!

Pitch was most likely to be prominent where actual agreement was involved, although in most such cases no new information was conveyed by the negative. Examples are found in (5a), (5b) Supportive interchanges.

a. [n] Therapist: You say you d’wanna be an engineer? (GTS-5)


Roger: You see the fault that I’m getting at-the: at-... wi’th’

Therapist: schools’s, they haven’t got an answer to the problem. (GTS-5)

(5)a and b can be defined as ‘neutral’ interactionally, but are not agreements; although they convey important information, the negatives are both contracted and nonprominent. In the 1985 study, negatives in such neutral contexts were pitch prominent more often than those in disagreements (20% > 3%), but less often than those used to agree with a previous statement, like those in (5)a’ and b’ (20% < 95%). Note that the negatives used in (5)a’ and b’ do not convey any new information, but display emphatic agreement with (5)a and b, respectively. In such cases the Social Agreement Rule favors pitch prominence on the agreement bearing element, even if that element is a negative. Thus, although the Cognitive Prominence Rule is neutralized (the actual new information has been conveyed in (5)a and b), the Social Agreement Rule plays an important part in this context: the negative provides agreement with a preceding statement. As a result, the informative negatives in (5)a and b are contracted and nonprominent, while the supportive but non-informative negatives in (5)a’ and b’ permit the highest percentage of pitch prominence. Consequently, (5) provides not just two examples of a pattern, but paradigmatic examples of that pattern.

Pitch prominence was less likely to occur in a neutral statement (like those in (1) and (3)), and least likely in a statement which could be construed as a ‘face threatening’ correction, like the one found in (2). Thus, Yaeger-Dror (1985) found that the two most important factors influencing pitch prominence were informational-interactional setting and interactional intent of the speaker. Table 3 compares the percentages of pitch prominent negatives, and confirms that while speech in purely informational settings may support the Cognitive Prominence, permitting a high percentage of pitch prominence, speech from more interactional registers supports the Social Agreement Rule, permitting relatively high percentages of pitch prominence in agreements or self-corrections, but not in other-corrections.

<table>
<thead>
<tr>
<th>‘Dimension I’</th>
<th>Disagreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>–</td>
</tr>
<tr>
<td>%</td>
<td>(N)</td>
</tr>
<tr>
<td>Informational [n]</td>
<td>50</td>
</tr>
<tr>
<td>Interational</td>
<td>–</td>
</tr>
<tr>
<td>Supportive [s]</td>
<td>100</td>
</tr>
<tr>
<td>Neutral [a]</td>
<td>20</td>
</tr>
<tr>
<td>Face threat [f/Adult]</td>
<td>–</td>
</tr>
<tr>
<td>Face threat [f/Teenage]</td>
<td>–</td>
</tr>
</tbody>
</table>
This does not imply that only one parameter— or ‘dimension’— is relevant at a time, but (at least for now) assumes a simplified model with data chosen from idealized endpoints of the Informational-Interactional continuum. Note that Biber’s support for this ‘Dimension I’ was determined from a study of many syntactic and morphological variables, as used in the many registers of the London-Lund Corpus (Biber, 1988; Svartvik, 1990).

Comparing all interactions from which data were gathered, informative nondisagreeing negatives buried in a context which was less informational were only pitch prominent approximately 20%. Since this avoidance of focus on the negative was caused by even a very generalized interactive context, the Cognitive Prominence Rule cannot be as central as initially posited.

Tottie (1991) confirmed these results: Although she had initially assumed that pitch prominence would cooccur with negation, her quantitative analysis of data from the London-Lund Corpus of Spoken [British] English, which provides its own perceptual analysis of pitch variation (Svartvik, 1990), confirmed that pitch prominence on negatives in that corpus occurs ‘infrequently’ across all registers for which the prosodic transcription is available.

[s] The interactional-supportive situation, coded ‘[s]’, should permit the highest percentage of pitch prominence, since the Social Agreement Rule either dominates or reinforces the Cognitive Prominence Rule— examples are found in (5)a’ and b’. High percentages should also occur when a speaker self-corrects, and the two rules reinforce each other—as in examples (4)a and b’.

[n] The purely informational situation and the interactional, but neutral situation, coded ‘[n]’, permit the Cognitive Prominence Rule to dominate— examples are found in (1), (3), (4)a, (5)a and (5)b.

[ft] The percentages should be lowest when the two rules are clearly in conflict, when a ‘remedial’ (Goffman, 1967, 1971) or ‘face threatening’ (Brown and Levinson, 1978) statement is being made, and it is in these cases that Yaeger-Dror (1985) found the Social Agreement Rule most consistently outweighs the Cognitive Prominence Rule— examples are found in (2) and (4)b. In such a setting, pitch prominence varied from 0% (for the group therapist, N = 811), and under 3% (in polite adult interactions) to 14.6% for the teenaged group therapy patients, and 19.1% for the tutorial speakers in the informational setting, when they must disagree with other speakers.

Given that the CA Social Agreement Rule appears to outweigh the Cognitive Prominence Rule even in an informational setting, it is no surprise that in actual conversations pragmatic interactional significance often conditioned whether pitch was prominent.

Table 3 shows that speakers were more likely to focus on the negative with pitch prominence at the informational end of the scale than in a more interactional setting; it also reveals that teenagers’ use of pitch prominence in disagreements differed significantly from adult conversational usage, but did not differ significantly from adults’ in the informational setting.

2.5. On the analysis of read prose

Klatt (p.c.) wondered to what degree the registers commonly mimicked for synthesis would follow the CA Social Agreement Rule, and to what degree they would follow the Cognitive Prominence Rule. The corpus for the present study has been construed as narrowly as possible in order to answer this question (at least partially), and only data which can be used as a direct model for one specific register of speech to be synthesized were analyzed: Read prose was chosen for analysis because it has already been funded for speech synthesis (e.g., Fant and Kruckenber, 1989; Granström, 1992), because it includes scripted [pseudo-interactional] dialogue as well as more neutral material, and because data ‘samples’ of the register, with a transcript, could be acquired easily.

11 Note that except for the teenaged group therapy patients, disagreements were not just nonprominent, but totally reduced in the majority of instances (68/87 disagreements in conversations, and 6/8 in the therapist’s speech: 75–78%).
From the library 'Books on Tape' section, 'Books for the Blind' were chosen which appeared to reflect a range of appropriate material. The books themselves were used as transcripts.

The review of the literature would predispose one to expect the Cognitive Prominence Rule to be dominant, under the assumption that neutral read prose is informational. However, only analysis of actual prose reading would provide conclusive evidence. While such 'careful' taped speech is significantly different from conversational speech, the variables which were found to influence interactional prosody cannot be ignored in analysis of read dialogue (or monologue). Four such corpora were chosen for analysis, to determine if actual reading in such a register requires pitch prominence on negatives used, or if prominence on negatives in this register would sound unnatural in a synthesized corpus. Thus, the present study has both theoretical and practical ramifications: If pitch prominence is preferred by the professional readers of this specialized register, this supports the Cognitive Prominence Rule, and synthesis algorithms should include such a rule. However, if pitch is not often prominent on negatives, that disconfirms the Cognitive Prominence Rule, and a more complex algorithm should be devised.

3. Description of the corpora to be analyzed

The works chosen for the present analysis can be found in Table 2 (marked with ' · '). Each was chosen to represent a specific niche of prose-for-synthesis; each included a large amount of dialogue, and the tape had been read by a professional.

- One children's author was chosen (Cleary, 1968, 1981), since one might assume that prose-for-children would be even more emphatically and informationally presented than prose for adults. Furthermore, the teenagers whose speech was analyzed for Yaeger-Dror (1985) used more pitch prominence than adults, even on disagreements, so proponents of both the Cognitive Prominence Rule and the Social Agreement Rule would expect read prose for and about children to permit a greater degree of pitch prominence.

- Two popular American authors for adults were chosen – Tyler (1988) and Atwood (1983, 1979) –, but the Atwood (read by Atwood herself) included almost no negatives and was discarded.\(^{12}\)

- One monologuist was chosen (Keillor, 1985), specifically because his book is a compendium of monologues which originally were presented over National Public Radio, and his monologue style epitomizes my own internalized concept of the 'ideal' NPR 'voice', and would, it was assumed (wrongly, we shall see), come closest to the prosodic patterns found by Hirschberg, who also studied speech used by NPR professionals. All the readings which were chosen were assumed to be fairly careful (in style), fairly similar to NPR 'voice' (in register), as well as speech-oriented and neutral (in interactional intent). In actuality, while the prose passages were fairly neutral, often a reader 'acted out' an intent appropriate to a given dialogue passage, requiring considerable variation in the 'interactional intent' of scripted dialogue, and (consequently) recourse to the categories carried over from (Yaeger-Dror, 1985).

Each of the corpora will be discussed, and conclusions about pitch prominence in these corpora will be used to determine the relative importance of informational ('semantic') and inter-active ('pragmatic') input to the formulation of an algorithm mapping prominence onto read prose.

It was projected that in read dialogue pitch prominence percentages would reflect either the informational context, or (in dialogue) the readers' intuitions for how actual speakers would present such a dialogue. Note that all readers except Keillor were professional actors, and women.

\(^{12}\) Obviously this lack of negatives is worthy of both literary and linguistic analysis, both of which are beyond the bounds of the present study.
4. The analysis

4.1. Why use pitch analysis?

Most phonetic studies have chosen one criterion for the determination of prosodic 'prominence'. Most studies which have compared the importance of various criteria to the perception of prosodic prominence have found that durational and amplitude variation are not insignificant influences on perception. However, in real-life situations both amplitude and duration are influenced by segmental and prosodic factors (cf. Lehiste, 1970), while their affect on the actual pitch or durational realization of a given segment is much smaller relative to the total variation.

- Amplitude is helpful as an analytical tool in only the most careful experimental situations; if it were measurable for this study, the variation caused by prosodic factors might well overwhelm the intrinsic amplitude variation. However, extensive experience with analysis of running casual and read speech of the sort to be used in this study convinced me that careful analysis of amplitude would not be helpful for this analytical study. This does not imply that amplitude variation programmed to co-vary with pitch variation would not lead to more comprehensible and lifelike synthesis; it merely maintains that amplitude tracks garnered from such a corpus, even though it be professionally recorded, would not be helpful in interpreting actual prosodic variation.

- Both Lehiste (1970) and Klatt (1975) showed that in English the variation due to intrinsic vowel duration (much less phonological/environmental conditioning factors!) can neutralize prosodic durational variation – even in quite careful reading. Moreover, O'Shaughnessy and Hirschberg intuitively determined that a negative is considered prosodically prominent even in the case where the negative’s vowel is contracted, and the pitch prominence is ‘carried’ entirely by a reduced Aux vowel, as occurred in examples (2) and (5)a/b. Consequently, it appears that durational variation is not the most important variable to be considered for a study of negation and disagreement in English. (See also the analysis in (Wightman et al., 1992), whose results are consistent with these conclusions.)

- On the other hand, according to Lehiste (1970), intrinsic pitch variation on vowels (with high vowels having higher pitch than lower vowels) is never more than 20 Hz; while in a carrier phrase 20 Hz is certainly above the perceptual (jnd) threshold, pitch variation within a given sentence in the present corpus was generally between 100 and 200 Hz, swamping any possible influence of intrinsic pitch. Moreover, given that the vowel tested for relative prominence is often a low vowel (which has a lower intrinsic pitch) ["not, wasn’t, hadn’t, can’t ..."] – [æ], [a] or [ə] – pitch prominence could never be attributed to intrinsic pitch.

- Those who feel that pitch is more significant a determiner than (say) duration or amplitude generally choose either perceptual or acoustic criteria to determine prominence. At least two groups of researchers are presently developing reproducible perceptual criteria for the designation of prosodic prominence. However, workshop presentations, and email communications (on the Prosody and Fonetiks nets) have demonstrated that these perceptual criteria are not yet considered reliable. Moreover, as already determined, the present study will limit itself to an acoustic analysis of fundamental frequency to permit direct comparison with O'Shaughnessy and Hirschberg and to permit computation of a variable which will be useful for realistic-sounding speech synthesis; pitch variation will be measured acoustically rather than perceptually.

O'Shaughnessy and Hirschberg determined pitch prominence acoustically, but neither study defined the criteria for the determination of ‘pitch prominence’. Subsequently, Wightman and Ostendorf (1992) explained their algorithm for the detection of pitch prominence. They conclude that their results “indicate that we have not yet extracted features appropriate for the detection of these rises and further research will be necessary to

---

13 See (Granström and Nord, 1992).
identify them and incorporate them into the system". 14 The present study will attempt to follow the rule
developed for the Wightman and Ostendorf (1992) study, while recognising that the acoustic criteria chosen
may require modification. 15

Acoustical prominence of the fundamental frequency was the only measure evaluated, and was interpreted
very broadly, to permit even a limited ‘Cognitive Prominence’ to be included. It is also theoretically possible
that cases would occur in which there would be negative prominence (that is, pitch lowering) on a token, as
proposed by Bolinger (1978); however, no cases of this type of prominence were found in either the 1985 study
(N = 812) or the present study (N = 716).

4.2. Choice of pitch analysis techniques

Segments from each of the tapes were run through either MacSpeechLabII or Signalyze 2.1, using a
Macintosh IIci with a MacRecorder interface. Note that the 8 bit A-D interface for the analysis of fundamental
frequency was sufficiently clear for the high quality recordings used here (as can be seen from the pitch tracks
in Appendix A, all of which were stored from (Keillor, 1985)), but it would not be sufficient for the analysis of
corpora like those used in (Yaeger-Dror, 1985). 16

Since the results of the tracking routine could not be automatically entered in the system using MSLII, tokens
were counted by hand. They were considered prominent or not, according to the following ‘rule(s)’:

"(1) If the trackable fundamental frequency was no higher than the immediately preceding vowel’s \( F_0 \), and
did not have any contour, then it was not categorized as pitch prominent. All of the tracks in Fig. 2 are examples
of this non-prominent pitch-type.

"(1’) Tokens which had a higher fundamental than the preceding vowel, but which had a rising fundamental,
and the following vowel had a higher fundamental following the same contour, were not perceived as
prominent. These tokens were termed transitional, and were not categorized as prominent.

"(2) Fig. 3 includes examples of both local pitch prominence and full (pitch and durational) prominence. If
the pitch was raised on the vowel of \( \text{not} \) – or, in a contracted negative, on the vowel of the auxiliary onto which
the negative was contracted – this was considered evidence of pitch prominence, even if there was no contour.
Such a negative was considered pitch prominent if the fundamental frequency of the vowel was raised relative
to the fundamental frequency of immediately preceding vowels, and was not lower than that of the subsequent
vowel.

"(2’) If the pitch was rising, or with rise-fall contour in the vowel, even if the pitch peak was not actually
higher than on the immediately preceding vowel, this was also categorized as ‘pitch prominent’. If such local
pitch prominence occurred, even if that prominence was not the pitch peak for the entire sentence, the negative
was classified as prominent.

---

14 I quite agree that a viable model for prosodic variation is still lacking at this time. The linguistics community would be thrilled if a
model could be proposed, integrating fundamental, amplitude, duration, and ‘break indices’, which could capture some of the patterns of the
speech we study. Unfortunately, with all due respect to the researchers who have been attempting to devise such a system, sociolinguists
have found that much task-oriented speech is so different from the speech for which ToBI was initially proposed that there is no adequate
ToBI transcription possible; for the present corpora, that is especially true for the conversations, group therapy sessions, tutorial and even
the stylized debate material. Footnote 6 describes some aspects of Jefferson’s transcription system, which was developed to transcribe
prosodic variation as efficiently and transparently as possible, and is used in many analyses of various speech registers where prosodic
variation is important. The interested reader will see that the Jefferson transcription is flexible and can deal with actual (i.e., interactively
contexted) prosodic variation. Our field’s difficulty in developing a functional multidimensional analytic transcription will be discussed
further in Section 5.

15 See also work by Ten Bosch (1993).

16 The earlier study used a Rabiner ‘black box’ hardware pitch track, connected to a PDP-11; although duration and amplitude were also
measured and recorded for each negative in the transcripts, only those tokens with actual pitch prominence were so designated.

17 Note, not just ‘not higher than the nearest local maximum’, but ‘not higher than the immediately preceding vowel’.
While this decision criterion clearly ignores other forms of prominence (amplitude and duration), (hopefully) it accurately compares the data with both O’Shaughnessy’s and Hirschberg’s results, and with the template which would be provided by Hirschberg’s synthesis algorithm. For the same reason, both contracted and uncontracted tokens were pooled on the following figures and tables.

4.3. Choice of agreement parameters

Tokens were categorized systematically for all of the pragmatic variables isolated in earlier studies: all prose passages and instructive dialogue were categorized as informational/neutral [n]. All negatives embedded in dialogue which contradicted some earlier speaker’s position were categorized as face threatening [ft], and all negatives embedded in dialogue specifically agreeing with a previous speaker’s position were categorized as supportive [s], as were other supportive moves. Examples of each of these pragmatic categories can be found in (6) and (7). (6) Neutral descriptive prose passage.

a. [n] The Dieners had not been getting full use of the dining room. (Keillor-M)
b. [n] If other towns have it, we don’t know about it. (Keillor-L)
c. [n] as if he hadn’t talked enough, to get over adolescence completely. (Keillor-L)
d. [n] glad I didn’t quite die:.

e. [n?] ’cause you sure could not trust your parent. (Keillor-M)
f. [n?] but didn’t, because Grampa offered them our house, (Keillor-M)

Table 4
Percent pitch prominence for each reading category, for the 4 authors whose work was analyzed [Number of tokens of negatives in each category are in parentheses]

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>(Total)</th>
<th>Prose</th>
<th>Support</th>
<th>Neutral</th>
<th>Remedial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keillor, L.</td>
<td>satire</td>
<td>(150)</td>
<td>50</td>
<td>-</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Keillor, M.</td>
<td>satire</td>
<td>(140)</td>
<td>35</td>
<td>50</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>Tyler</td>
<td>literate</td>
<td>(225)</td>
<td>19</td>
<td>50</td>
<td>32</td>
<td>17</td>
</tr>
<tr>
<td>Cleary</td>
<td>children’s</td>
<td>(201)</td>
<td>49</td>
<td>33</td>
<td>29</td>
<td>60</td>
</tr>
</tbody>
</table>
Dialogue, with interactional coding as 's' = support, 'n' = neutral, 'ft' = face threat.

a. [s] "I know it isn't easy." (Father to daughter) (Cleary)
b. [s] "Shoot, you don't need to lose weight!" (waitress to customer) (Tyler)
b' [s] "You are surely not old enough to have a daughter that grown up!" (waitress to customer)
c. [s] "you're no::: skeptic," (Keillor-M)
d. [n] "We did not know what would become of us ...," wrote Oskar. (Keillor-L)
e. [ft] "You don't get along with other people. You don't make an effort."
   (Keillor-L)
e' [ft] "What are you doing? You can't shoot people here!" (italics in original) (Keillor-L)
f. [ft] "This is no way to behave ..." (Cleary)
g. [ft] "Howie, that's not polite."
   (Cleary)

4.4 Analysis

For each corpus, all negatives found on at least one side of a tape of each read text were labeled following this system, then the tape of those segments was analyzed for pitch prominence on the negatives, following the description defined above. Given the high quality of the tapes, the pitch tracks permitted accurate measurement, even for the female readers.

If the reading-prose prosody reflects the same rules as the data analyzed in earlier studies, we should expect

Hypothesis I. • pitch prominence on neutral [n] prose, and dialogue
   > as both O'Shaughnessy and Hirschberg found.
Hypothesis II. • pitch prominence on emphatic expressions of agreement [s]
   > as Yaeger-Dror found.
Hypothesis III. • variable prominence on disagreements [ft]
   > where Yaeger-Dror found percentages variable, with more pitch prominence on negatives in speech by teenagers.
Hypothesis IV. • relatively more pitch prominence in prose for children than in prose for adults, whether or not disagreement is involved.
   > as Yaeger-Dror found.
Hypothesis V. • relatively more prominence in read than in actual conversations
   > since (presumably) clarity is of the essence, and actual 'agreement' is neutralized.

4.5 Results

Fig. 1 and Table 4 present the data for the readings. As in Table 3, the number of tokens for a given group can be found in parentheses on the table. In Table 5 these results can be compared with the results from the earlier studies.

Note that where there is a distinction between superficial agreement and support of the coparticipant, the relevant parameter is presumed to be support rather than agreement. In b' the protagonist has claimed she should be on a diet, so the waitress's move is superficially a disagreement. However it is not a face threatening act, but a supportive move, and is categorized accordingly. The protagonist speaks of her college-bound daughter to substantiate her claim to 'old' age, so the waitress's reply is (again) superficially a direct disagreement, while ultimately (within an understanding of the culture painted by the author) a supportive comment. Another interchange (from the same dialogue) reflects those in (5) above:

[s] "Well, to me I'm not ordinary", Maggie said.
[s] "No, indeedy!" Mabel said.
Hypothesis I – The most surprising results were for the negatives in the reading of prose passages. It had been hypothesized that these would be the most consistent with the Cognitive Prominence expectations:

- They are ‘informational’ – and therefore should follow the Cognitive Prominence Rule.
- They have no interactional ‘vector’ – and need not be reduced by the Social Agreement Rule.
- They are read by professional readers (e.g., Jill Eikenberry, Garrison Keillor), and so should most closely parallel the materials from which Hirschberg’s algorithm was developed.

However, negatives in neutral read prose were never prominent more than half of the time, and in read neutral dialogue were generally prominent less often. Even the NPR monologist, Garrison Keillor, did not present neutral negatives pitch prominently more than 50% of the time for any category. Although I had projected that neutral prose reading would have a very high percent of pitch prominent negations, neither neutral prose nor dialogue permit high percentages. It is clear that to the degree that we would want synthesized book readings to reflect actual book readings, a synthesis rule which imposes pitch prominence on all negatives is at least 50% wrong, and is no better as a synthesis algorithm for read-prose-passages than for read-dialogue.

There does not appear to be a consistent rule for supportive or face threatening statements; perhaps this reflects Labov’s findings that read speech is erratic – that is less rule governed than more casual speech (Labov, 1986; Yaeger, 1974) – even for phonological variables:

Hypotheses II–III – Is pitch more often prominent on agreement [s] than disagreement [ft]? Pairwise comparisons of face threatening and supportive negations for adult readings show that where comparison was possible, supportive negatives were more likely to be pitch prominent. In 6 of 7 cases, the supportive negatives were consistently more prominent than neutral ‘informative’ negatives.

Hypothesis IV – The evidence presented in Fig. 1 supports Hypothesis IV: As projected, the neutral prose-reading for children (Cleary, permitting almost 50% pitch prominence on neutral-prose negatives) permits more prominence than the readings for adults (20%), as does the Keillor monologue which includes children’s dialogue (50%). In fact, pairwise comparisons of Cleary and other readings shows that in 5 of 6 comparisons, Hypothesis IV is confirmed: Both in [n] prose-reading and in [ft] face threatening dialogue the reader for children used pitch prominence more consistently than readers for adults. However, in both cases where comparison was possible, readers for adults used pitch prominence on supportive negations [s] significantly more often than the reader for children.

Hypothesis V – However, the actual percentages, and comparisons with prominences from actual interactions were quite surprising. On the whole these readings are even less biased toward informational (i.e., pitch prominent) presentation than actual conversations: Supportive negatives were not pitch prominent more than
Fig. 2. (a) Nonprominent pitch, contracted negative. (b) Nonprominent pitch, reduced negative. (c) Nonprominent pitch, unreduced negative. (d) Nonprominent pitch, unreduced negative.
Fig. 3. (a) Prominent pitch, contracted negative. (b) Prominent pitch, contracted negative. (c) Prominent pitch, uncontracted negative. (d) Prominent pitch, uncontracted negative.
half the time. Only face threatening negatives were more systematically pitch prominent than actual 'conversational' negatives, but even in readings for children, pitch was only prominent 60% of the time.

5. Discussion

In a theoretical domain, we could say the study was initiated with the avowed intent of determining the relative importance of the **Cognitive Prominence Rule** to speech registers which are likely to be used for speech synthesis. An underlying purpose was to determine the relative importance of this **Cognitive Prominence Rule** and the **Social Agreement Rule** in different speech situations in read prose. The speech situations were carefully chosen to permit tentative conclusions to be drawn about appropriate algorithms for speech synthesis.

Like Wightman and Ostendorf, and Hirschberg in the cited publications, my primary task here was to determine the variation within one specific parameter (Fundamental Frequency), rather than to integrate several parameters. I agree with Wightman and Ostendorf that in the long run a one dimensional analysis is not intellectually satisfying. However, I also agree (see citation above) that a detailed one dimensional analysis which permits comparison across registers, as well as permitting practical algorithms to be devised, is more practical at our present stage of knowledge, and that at this time it is premature and counterproductive to attempt to present results for this type of study using an integrated system. There are a number of papers from recent meetings in Japan which appear to follow this philosophy, and rely on documenting fundamental frequency changes alone, as a practical alternative to a multidimensional analysis which we cannot yet agree on, much less implement. The present paper ascribes to that philosophy.

Hirschberg (1993) has recently discussed the fact that the vast majority of pitch variation can now be accounted for 'by rule', but that researchers will have to spend the next few years analyzing specific narrow problems to ‘fine tune’ the rules for those prosodic patterns (like the negative) which have so far eluded simple rule algorithms. The present paper has conducted an analysis of the English negative, which proves to be one such locus for fine tuning. The study found that while pitch prominence on negatives might not sound unnatural for a proof-reading algorithm, further study is needed to determine a much more complex algorithm for a more varied synthesis corpus. Even synthesis of prose fiction will require analysis of other writers (and readers), to determine appropriate prosody for negatives in different registers and in different pseudo-pragmatic situations. However, even with the limited corpora used here, I hope to have shown that ‘fine-tuning’ of the system is not only possible but imperative, and that an accurate ‘fine-tuning’ requires a nuanced understanding of register variation.

It would have made sense to assume that in a register used for synthesis, where informational needs are high, we should expect focal prominence on negatives to be used more often in prose passages or scripted dialogue than in conversations. However, Table 5 showed that in a neutral, purely informational, setting, none of the readers used pitch prominence on the negatives more than 50% of the time. In dialogue, the readers used pitch prominence more often than conversationalists documented in earlier studies, but not more often than tutorial lecturers, and much less often than political debaters.

The pitch prominence percentages for read prose match those for actual conversational material more closely than they match the projected percentages and the synthesis routine described in the literature. In fact, only one

---

corpus analyzed to date approaches the synthesis routine’s prescription – televised political debates. Debates have by far the most pitch prominence since the need for clarity is reinforced by the ritualized obligation to disagree, so both the CA Social Agreement Rule and the Cognitive Prominence Rule should reinforce the tendency to use pitch prominence on negatives in this setting. Oehrle and Yaeger-Dror (1992) showed that in political debates speakers will use pitch prominence on negatives more liberally than in other interactions analyzed. In fact, the three male politicians whose speech was analyzed used pitch prominence on negatives between 60 and 80% of the time in disagreements. Note that in this maximally emphatic situation, where both the informational and interactional dynamics combine to predispose the speaker to use pitch prominence, it is still less likely to occur than in isolated sentences, or in the materials studied by Hirschberg, but it is certainly much more likely to occur than in the readings used for the present study. An algorithm for synthesis should take these facts into consideration, lest the synthesized prose sound more like a political debate than like a conversation.

One important conclusion of this study is that a nuanced analysis of register variation is critical to our understanding of the variable being studied, and that analysis of actual variation clarifies for us how fine the distinction among registers should be. The practical consequences are simple to implement. Presumably the algorithm as formulated by Hirschberg would be ideal for synthesizing a proof-reading program, or a news broadcast. However, if we were to formulate an algorithm for read prose, we would alter the algorithm: adult and children’s literature would be differentiated from each other, quotation marks would be used to distinguish dialogue from descriptive prose, with higher percentages of pitch prominence for children’s dialogue, and lowest percentages for adult dialogue, with non-dialogue midway between. On the other hand, perhaps because of the paucity of examples of supportive negatives in dialogue, one infers from the evidence that the coding of the distinction between face threatening and supportive dialogue may not be necessary.

While only a few texts have been studied in detail, it seems clear that even in neutral prose reading, the readers do not consistently use pitch prominence on negatives. Pitch prominence is most likely to occur in the descriptive prose passages of texts read to (perhaps ‘acted out for’ would be a more appropriate locution) children, especially in conflicts acted out for such an audience, but conflicts in adult readings are only pitch prominent half the time or less. Pitch prominence is even less likely to occur in prose passages or neutral dialogue read by a ‘Standard American’ NPR-type reader. Thus, in story-texts which could be ‘read’ by a synthesizer, it appears that ‘informational’ presentation should not be used categorically, even for descriptive prose passages. Given that the only material found to use a higher percentage of prominence on negatives was political debates, one can project that even 60% pitch prominence on negatives would give an inappropriately aggressive, combatative, impression to the listener. In passages enclosed in quotes, pitch prominence on negatives is perhaps somewhat more likely than in other circumstances, but even here should not exceed 50%, to avoid leaving an aggressive impression.

For further reading, see (Biber and Finegan, 1994; Brazil, 1984; Finegan, 1994; Goffman, 1981; Jefferson, 1975; Nagabuchi et al., 1993; Yaeger-Dror and Nunamaker, 1992).

---

20 Recently, studies have been published which demonstrate that there are specific social situations in which confrontational ‘footing’ is preferred; in such settings – news interviews (Clayman, 1988, 1993; Greatbatch, 1993; Heritage, 1985; Heritage and Greatbatch, 1991; Pomerantz, 1989, Schegloff, 1989) and call-in formats (Hutchby, 1992a,b), as well as courtroom discussions (e.g., Drew, 1993) and political debates (Adams, 1991a,b; Adams and Edelsky, 1988, 1990; Oehrle and Yaeger-Dror, 1992) – a speaker is expected to disagree with coparticipants.

21 Anyone interested in the practical detail work necessary to implement the present coding scheme for synthesis of literature-on-tape would be grateful if this particular distinction need not be coded: The distinction between prose and dialogue is clearly labeled in the text, and the distinction between adult and children’s prose need only be coded once per text, but the distinction between supportive and remedial interchanges is not simply coded, and certainly such coding cannot be automated.
Acknowledgements

This paper would not have been conceived without the supportive questioning of Dennis Klatt, to whose memory the paper is dedicated. The analysis was carried out in the Speech Research Lab of the University of Arizona, and thanks are due to Kerry Green for access to the facilities, and Tom Bourgeois for reconfiguring the system, as well as to NSF for funding for acoustical equipment. Thanks also to Henrietta Cedergren, Scott Jacobs, Dick Oehrle and Patti Price, for helpful suggestions while the paper was being prepared for the Acoustical Society Convention where it was first presented, and to Molly Diesing and an anonymous reviewer, for their comments thereafter. Any remaining limitations are clearly to be laid at the door of the author (e-mail: malcah@ccit.arizona.edu).

Appendix A

The pitch tracks shown in Fig. 2 and Fig. 3 were made from an audio recording of Keillor (1985), using MSL II; 100 Hz is marked on the pitch tracks by a dotted line.

References

K. Adams (1991a), “Must a politician 'Talk like a lady' too?”.
M. Atwood (1979), *Life before Man* (Fawcett, New York).
M. Atwood (1983), *BlueBeard's Egg* (Fawcett, New York) [1990. Read by M. Atwood; Random House, Toronto].


Y. Sagisaka, N. Campbell and N. Higuchi, Eds. (Forthcoming), *Computing Prosody* (Springer, Bonn).


L. Shockley, Ed. (1995), *University of Reading Speech Research Laboratory Work in Progress 8* (Department of Linguistic Science, University of Reading, Whiteknights).


