The Effect of Contour Type and Epistemic Modality
on the Assessment of Speaker Certainty

Agustín Gravano [1], Stefan Benus [2], Julia Hirschberg [1], Elisa Sneed German [3], Gregory Ward [3]

[1] Department of Computer Science, Columbia University, New York, NY, USA
[2] Dept. of English and American Studies, Constantine the Philosopher University, Nitra, Slovakia
[3] Department of Linguistics, Northwestern University, Evanston, IL, USA

{agus,julia}@cs.columbia.edu, sb513@nyu.edu, {elisa,gw}@northwestern.edu

Abstract

In an empirical study participants were asked to rate the perceived degree of certainty of utterances that contained either the modal *would* or main verb *be* (e.g. *That would be me* vs. *That's me*), and which were also variously produced with one of three intonational contours (downstepped, declarative, and yes-no-question). We found that both downstepped contour and epistemic *would* made a significant and independent contribution to the assessment of speaker certainty. That is, participants rated utterances with the downstepped contour as most certain, followed by those with the declarative contour, while the yes-no-question contour was perceived as highly uncertain. Similarly, participants rated speakers' responses with epistemic *would* as significantly more certain than those without it.

1. Introduction

It has long been noted that a speaker's epistemic disposition toward a proposition can be conveyed linguistically via different modules of the grammar: morphologically (via affixes), syntactically (via word order), semantically (via verb choice), and prosodically (via intonational contour). One commonly expressed type of epistemic disposition is speaker certainty: the degree to which a speaker is committed to a proposition expressed in a particular context. In this paper, we examine two constructions that have been claimed to be associated with the conveyance of speaker certainty: the downstepped intonational contour [3, 9, 10] and epistemic *would* [6, 7, 8].

1.1. Previous studies

The so-called downstepped contour [9] is characterized by a series of pitch accents uttered in an increasingly compressed pitch range, producing the impression of a flight of stairs in the F0 contour[1]. In the ToBI system of intonational transcription, the most common version of this contour is represented as H* !H* (!H*) L- L% (illustrated in Figure 1), where each subsequent high tone (H*) is measurably lower than the preceding one.

Downstepped contours are quite common in Standard American English. For example, in the AT&T Communicator Corpus of read speech [2, 5], the H* !H* L- contours represent the most frequent pattern of the 2888 intermediate phrases in this corpus, comprising about 40% of all contours. They occur almost twice as often as the standard declarative contours ((H*) H* L-; illustrated in Figure 2) in this corpus [4]. In the

Boston Directions Corpus, downstepped contours occur in over 49% of intermediate phrases in read speech and in 37% of intermediate phrases in spontaneous speech, with H* !H* L- contours representing 21% of read and 15% of spontaneous intermediate phrases. Yet, despite its relative frequency in naturally
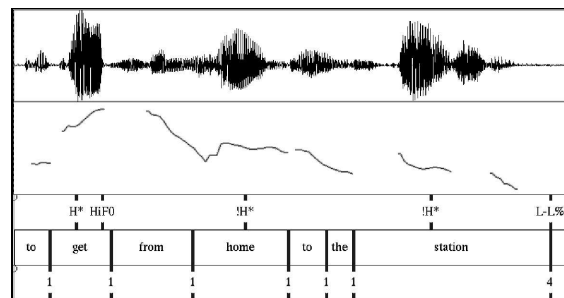

Figure 1: *Pitch track showing a downstepped contour.*

occurring speech, very little empirical work has been done on the meaning of the contour. In their pioneering survey of the meaning of various intonational contours, [10] suggested that downstepped contours mark discourse topic structure, occurring frequently in phrases which signal topic beginnings and endings. Moreover, they claimed that the interpretation of sequences of downstepped pitch accents might be characterized as conveying that the Hearer should be able to infer, from the shared beliefs of Hearer and Speaker, the existence of discourse entities realized with such accents. A possibly related observation is that downstep serves as an alternative to deaccenting, when information being expressed represents given information in the discourse [2, 11, 12]. Thus, to the extent that given information represents information about which the speaker is certain, we can hypothesize that the speaker's use of downstep will convey an epistemic disposition of certainty.

As for the epistemic *would* construction, illustrated in (1), previous researchers have disagreed about the contribution of epistemic *would* to utterance interpretation.

(1) **A:** *Who has the best on-time-arrival record this month?*
    B: *That would be Continental.* [*Chicago Tribune*, 01/10/05]

[6, 8], for example, characterize epistemic *would* as conveying tentativeness. More recently, however, [13] argue against these accounts: On the basis of a large corpus of naturally occurring data, they claim that epistemic *would* is a focus marker indicating the presence of a salient open proposition. Moreover, according to [13], use of epistemic *would* conventionally

---

[1] The type of downstep that occurs in English is sometimes called *catathesis* in order to distinguish it from downstep in African tone languages, where tone is morphemic.

implicates "that the speaker believes she or he has conclusive objective (that is, empirical or logical) evidence for the truth of the proposition encoded in the utterance" ([13], page 75). This rules out, among other things, decisions, predictions, and wild guesses by the speaker. In (1), the open proposition corresponding to A's question is *x has the best on-time-arrival record this month*, with *Continental* instantiating the variable in this open proposition for whose truth the speaker has "verifiable evidence". Thus, to the extent that a true, verifiable proposition represents information about which the speaker is certain, we can hypothesize that epistemic *would* also conveys an epistemic disposition of certainty.
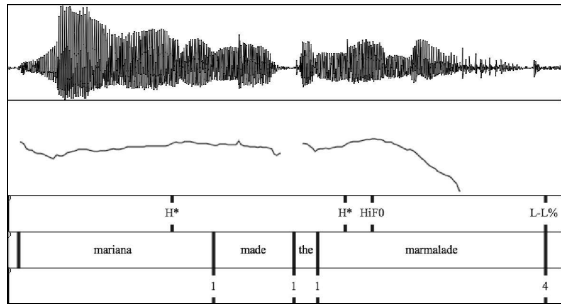

Figure 2: *Pitch track showing a declarative contour.*

1.2. The present study

We designed an empirical study to examine whether and to what extent these two constructions – downstepped contour and epistemic *would* – affect speaker certainty. We asked participants to rate utterances which contained either the modal *would* or main verb *be*, and which were variously uttered with one of three intonational contours: downstepped, declarative, and yes-no-question. The downstepped and declarative contours were described above; the yes-no question contour, illustrated in 3, is transcribed as (L*) L* H- H% in the ToBI system.
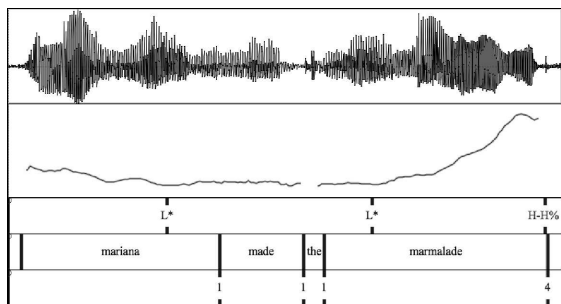

Figure 3: *Pitch track showing a yes-no-question contour.*

To preview the main findings, we found that both constructions made a significant and independent contribution to the assessment of speaker certainty. That is, participants rated utterances with the downstepped contour as the most certain ones, followed by those with the declarative contour, while the yes-no-question contour was perceived as highly uncertain. Similarly, participants rated speakers' responses with epistemic *would* as significantly more certain than those without epistemic *would*.

2. Perception study I: Textual condition

Our initial goal was to assess the effect of epistemic modality alone on the degree of perceived certainty. To achieve this, we performed a perception experiment in which participants read short conversations and rated the certainty of target sentences with or without epistemic *would*. This textual condition allowed us to study the effect of modality on the assessment of speaker certainty, in isolation from the effect of intonation.

2.1. Materials and method

The materials for this study were based on 20 short dialogues containing a target sentence of the form *That would be x*, selected from a larger corpus collected at NU from printed and overheard conversations. An example dialogue is provided in (2), with the target sentence underlined.

(2) **Jennie:** *What a great party!*
David: *Yeah, but we're stuck cleaning up all the crap.*
Jennie: *Hey, somebody left their iPod out on the floor.*
David: *That would be my roommate.*

For each original target sentence we created a matching sentence by replacing *would be* with *is*. Thus, the matching sentence for (2) was, *That's my roommate.* Additionally, we selected 40 separate short dialogues as fillers, in which the target sentences did not contain either of the target constructions, as illustrated in (3).

(3) **David:** *Have they posted the results of the swim meet yet?*
Ronnie: *Some of them. You already know who won the 100m Butterfly.*
David: *The winner of that event was Chris Jesperson.*

In the first perception study, participants were shown transcripts of the 20 original dialogues, in which the target sentence had either *would be* or *is* (i.e., they saw only one version of the target constructions), and the 40 fillers. These 60 tokens were presented in a different random order to each participant, and the epistemic modality of each stimulus was also chosen at random. Participants were asked to rate the perceived certainty for each token's target sentence, using a 5-degree Likert scale (Very uncertain, Somewhat uncertain, Neither certain nor uncertain, Somewhat certain, Very certain).

12 undergraduate students (8 female, 4 male; mean age: 20.3) from Northwestern University participated in this study; all were native speakers of American English. They completed the study in a quiet lab, indicating their classification decisions in a computer interface on a lab workstation.

2.2. Results and discussion

The only independent variable in the statistical analysis of our first perception study was *Modality*, with two levels: *would be* and *is*. To compute the dependent variable, *Certainty*, we first assigned a numeric value to each degree in the Likert scale (Very uncertain = -2, Somewhat uncertain = -1, Neither = 0, Somewhat certain = 1, Very certain = 2). We next normalized each rating by participant, using z-scores[2], to account for participant variation in use of the scale. The mean *Certainty* for stimuli with modal *would* was -0.13 (st.dev. = 1.11), and for stimuli without modal *would* was -0.03 (st.dev.

---

[2] $z = (X - mean)/st.dev$, where X is a rating by participant P, and $mean$ and $st.dev$ are calculated over all ratings by participant P.

= 1.04). There were 120 data points in each group. A one-way ANOVA reported no significant difference between the two means ($F(1, 238) = 0.58$, $p = 0.45$). After doubling the data to simulate a larger sample, the p-value was still non-significant at 0.28, suggesting that increasing the sample size would probably not lead to finding a significant difference. So, contra [6, 7, 8], it appears that modal *would* alone is no more uncertain than main verb *be*.

3. Perception study II: Spoken condition

We performed a second perception experiment, where participants now listened to the same target sentences, produced with three different intonational contours, and rated the perceived degree of certainty. This spoken condition allowed us to study the effect of both epistemic modality and intonational contour, combined or in isolation, on the assessment of speaker certainty.

3.1. Materials and method

Each of the 40 stimuli used in the first perception study was recorded using three different intonational contours: declarative, downstepped, and yes-no-question. Thus, for each of the 20 original dialogues there were six recorded stimuli (2 modalities × 3 intonational contours), for a total of 120 recorded stimuli. Additionally, we divided the 40 fillers into three groups at random, and recorded 13 of them with a declarative contour, 13 with a downstepped contour, and 14 with a yes-no-question contour.

The 120 stimulus sentences and the 40 filler sentences were recorded by a male speaker (an author of this paper) in a sound-proof booth using a close-talking head-mounted microphone. The Praat software [1] was used for capturing, digitizing, and analyzing the sentences. We verified the intonational contour, and checked the average amplitude and pitch range of each sentence; the sentences with undesirable intonation or outlier values for pitch range and amplitude were re-recorded.

In this second perception study, a different group of participants was presented with 60 tokens: the 20 original dialogues, in which the target sentences were produced with one of the three intonational contours and had either modal *would* or main verb *be*, and the 40 fillers. These 60 tokens were presented in a different random order to each participant, and modality and contour were also chosen at random for each stimulus. Participants were shown a transcript of the dialogue context, although they did not see the target sentence itself, as illustrated in (4):

(4) **Jennie:** *What a great party!*
David: *Yeah, but we're stuck cleaning up all the crap.*
Jennie: *Hey, somebody left their iPod out on the floor.*
David: 🔊

Participants were then asked to rate the perceived certainty of the target sentence, using the same 5-degree Likert scale described in Section 2.1. They could listen to the target sentence as many times as they wished, by clicking a button in the computer interface.

A different pool of 30 undergraduate students (24 female, 6 male; mean age: 21.2) from Northwestern University participated in this study, all native speakers of American English and reporting no hearing problems. They completed the study in a quiet lab using headphones to listen to the tokens, and indicating their classification decisions in a computer interface on a lab workstation.

3.2. Results and discussion

In the analysis of our second perception study, we wanted to test the hypothesis that both contour and epistemic modality influenced participant perceptions of certainty. Our independent variables were *Modality* (with two levels: *would be* and *is*) and *Contour* (with three levels: declarative, downstepped and yes-no-question). For the dependent variable, *Certainty*, we used the same conversion and the same participant normalization described in Section 2.2.

Table 1: *Certainty mean ± standard deviation for each of the six stimulus types.*

	declarative	downstepped	yn-question	·
would be	**0.57** ±.50	**0.73** ±.43	**−0.84** ±.72	**0.15** ±.90
is	**0.39** ±.62	**0.67** ±.54	**−1.02** ±.66	**0.01** ±.96
·	**0.48** ±.57	**0.70** ±.49	**−0.93** ±.70	**0.08** ±.93

Table 1 shows the mean and standard deviation of *Certainty* judgments for each of the six types of stimuli (there were 100 data points in each cell). An initial two-way ANOVA assuming the full model ($Certainty = Modality + Contour + Modality \times Contour$) revealed no significant interaction between the *Modality* and *Contour* factors ($F(2, 594) = 0.698$, $p = 0.498$). In other words, intonation and epistemic modality made independent contributions to the assessment of speaker certainty. The almost parallel lines of Figure 4 illustrate this lack of interaction.
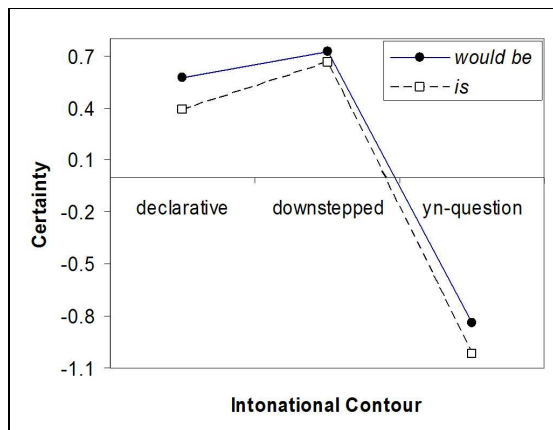

Figure 4: *Interaction plot for Modality and Contour.*

In consequence, we ran a two-way ANOVA test assuming the additive model ($Certainty = Modality + Contour$), and found both factors to have a significant effect on *Certainty* (*Modality*: $F(1, 596) = 8.8$, $p = 0.003$; *Contour*: $F(2, 596) = 456.32$, $p \approx 0$). For *Modality*, this indicates that sentences with the *would be* construction (observed mean = 0.15, see Table 1) tended to be perceived as significantly more certain than sentences with the *is* construction (mean = 0.01). For *Contour*, downstepped utterances were perceived as most certain (mean = 0.7), followed by declarative utterances (mean = 0.48); the yes-no-question contour was perceived as strongly uncertain (mean = −0.93). A Tukey test revealed that the pairwise differences in means between all three contours were significant at the 95% level. It is also worth noting that the standard deviations in the textual condition were much higher than in the spoken condition, which presumably shows the important disambiguating

role of intonation in the assessment of speaker certainty.

Next, we studied the effect of *Modality* separately for each intonational contour. For stimuli uttered with a declarative contour, the perceived certainty of sentences with modal *would* was significantly higher than that of sentences without it (ANOVA, $F(1, 198) = 5.29$, $p = 0.02$). For the other two contours, despite the fact that *would be* sentences had a higher *Certainty* mean than *is* sentences, such differences were not statistically significant, although they approached significance for the yes-no-question contour (downstepped contour: ANOVA, $F(1, 198) = 0.81$, $p = 0.37$; yes-no-question contour: ANOVA, $F(1, 198) = 3.42$, $p = 0.066$).

We also studied the effect of *Contour* separately for each epistemic modality type. For stimuli containing the *is* construction, ANOVA ($F(2, 297) = 222.51$, $p \approx 0$) and Tukey tests (at the 95% level) revealed, again, that downstepped productions were perceived as significantly more certain than declarative ones, and declarative productions, in turn, as more certain than yes-no-question ones. Finally, for *would be* stimuli, ANOVA ($F(2, 297) = 234.96$, $p \approx 0$) and Tukey tests (95%) showed both downstepped and declarative productions as significantly more certain than yes-no-question ones, although no significant difference was found between these two. Table 2 summarizes these findings.

Table 2: *Summary of findings of Perception Study II;* '$>$' *means 'perceived as significantly more certain than'.*

General results in the spoken condition
would be $>$ *is*
downstepped $>$ declarative $>$ yn-question
Given a particular intonational contour
declarative: *would be* $>$ *is*
downstepped: –
yn-question: *would be* $>$ *is*   (approaching significance)
Given a particular epistemic modality
would be: downstepped / declarative $>$ yn-question
is: downstepped $>$ declarative $>$ yn-question

Our initial hypotheses that both epistemic modality and intonational contour influence hearer perceptions of certainty are borne out by this perception study. The presence of epistemic *would* does indeed convey more certainty than *is* in the spoken condition. Downstepped contours do indeed convey more certainty than either declarative or yes-no-question contours. Note however that, when we examine the effect of *Modality* on contour interpretation, we find a significant effect only for declarative contours, although we approach significance for yes-no-question contours. That is, the perceived certainty associated with epistemic *would* is clearer in declarative and yes-no-question contours than in downstepped contours. Also, for either construction, the yes-no-question contour is interpreted as the most uncertain one; and, for sentences not containing epistemic *would*, the downstepped contour is perceived as the most certain of the three contours.

4. Conclusions

Empirical studies in which participants were asked to rate the perceived degree of certainty of utterances produced with either the modal *would* or main verb *be*, and with one of three intonational contours (downstepped, declarative, and yes-no-question) found effects for both contour and epistemic *would* on certainty ratings. Both the use of a downstepped contour and the use of epistemic *would* made significant and independent contributions to participants' assessment of speaker certainty. Participants rated utterances with the downstepped contour as most certain, followed by those with the declarative contour, while the yes-no-question contour was perceived as highly uncertain. Similarly, participants rated speakers' responses with epistemic *would* as significantly more certain than those without it. These findings support our initial hypotheses that both epistemic *would* and the downstepped contour can be employed to convey speaker certainty.

6. References

[1] Boersma, P.; Weenink, D.; 2001. Praat: Doing phonetics by computer. http://www.praat.org

[2] Hastie, H.W.; Prasad, R.; Walker, M.; 2002. What's the trouble: Automatically identifying problematic dialogues in DARPA Communicator Dialogue Systems. In *Proc. of the 40th Annual Meeting of the Assoc. for Computational Linguistics*, 384-391.

[3] Hirschberg, J.; Pierrehumbert, J.; 1986. The intonational structuring of discourse. In *Proc. of the 24th Annual Meeting of the Assoc. for Computational Linguistics*, 136-144.

[4] Hirschberg, J; Gravano, A.; Nenkova, A.; Sneed, E.; Ward, G.; 2007. Intonational overload: Uses of the H* !H* L- L% contour in read and spontaneous speech. In *Laboratory Phonology 9*, J. Cole & J.I. Hualde (ed.). Berlin: Mouton de Gruyter. 455-482.

[5] Hirschberg, J.; Rambow. O.; 2001. Learning prosodic features using a tree representation. In *Proc. of Eurospeech 2001*, Aalborg, Denmark. 1175-1180.

[6] Palmer, F.R., 1990. *Modality and the English Modals.* 2nd ed. London: Longman. Original edition, 1979.

[7] Palmer, F.R., 2001. *Mood and Modality.* 2nd ed. Cambridge: Cambridge Univ. Press. Original edition, 1986.

[8] Perkins, M.R., 1983. *Modal Expressions in English.* London: Frances Pinter.

[9] Pierrehumbert, J., 1980. *The Phonology and Phonetics of English Intonation.* Ph.D. Thesis, MIT. Cambridge, MA.

[10] Pierrehumbert, J.; Hirschberg, J.; 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*, P. R. Cohen, J. Morgan & M.E. Pollack (ed.). Cambridge, MA: MIT Press, 271-311.

[11] Prince, E.F., 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*, P. Cole (ed.). New York: Academic Press, 223-255.

[12] Prince, E.F., 1992. The ZPG letter: Subjects, definiteness, and information status. In *Discourse description: Diverse analyses of a fund raising text*, S. Thompson & W. Mann (ed.). Philadelphia/Amsterdam: J. Benjamins, 295-325.

[13] Ward, G.; Birner; B.J.; Kaplan J.P.; 2003. A pragmatic analysis of the epistemic *would* construction in English. In *Modality in Contemporary English*, R. Facchinetti, M. Krug & F. Palmer (ed.). [*Topics in English Linguistics 44*, General Editors: B. Kortmann & E. Closs Traugott.] Berlin/New York: Mouton de Gruyter, 71-79.