ON THE DISCOURSE FUNCTION OF INTONATION

Diev Huber
Chalmers University of Technology
Department of Information Theory
S-412 96 Gothenburg, Sweden

ABSTRACT
This study explores the differences between discourse intonation and the kind of pitch contours typically found in isolated sentences. Three kinds of material are evaluated systematically: (1) orally read lists of semantically unrelated sentences, (2) orally read narrative texts, and (3) dialogues. The material consists of equivalent samples of Swedish, English and Japanese speech, produced by native speakers (both males and females) of the respective languages. It will be shown that discourse intonation differs from intonation in semantically unrelated sentences with respect to practically all $F_0$ parameters investigated in this study.

1. INTRODUCTION

Human speakers typically associate their verbal speech utterances with intrinsic patterns of voice fundamental frequency. This phenomenon has been widely attested, and is acknowledged as a universal, innate quality of speech, common to all speakers, in all languages, and in all kinds of spoken utterances. Numerous scientific studies within a variety of disciplines have been undertaken to investigate the form and function of these fundamental frequency patterns, to establish their communicative status, and to disentangle the seemingly infinite variety of linguistic and paralinguistic conditioning factors that human speakers so aptly and automatically employ in their speech. A key aspect in this regard is the ability of human speakers to use variations in pitch consistently in different settings, and distinguish them systematically: (1) orally read lists of semantically unrelated sentences, (2) orally read narrative texts, and (3) dialogues. The purpose of this study is to investigate these differences, i.e. between discourse intonation and the kind of pitch contours typically found in isolated sentences.

2. DATA

Three kinds of material are evaluated systematically: (1) orally read lists of semantically unrelated sentences, (2) orally read narrative texts, and (3) dialogues. These materials were selected from the CTH [2] speech databases and comprises equivalent samples of Swedish, English and Japanese speech. The English and Japanese dialogues consist of simulated telephone conversations conducted within the applications domain of conference registration, whereas the Swedish dialogues were conducted spontaneously.

Ten native speakers of the respective languages participated in the recordings selecting for this study: 3 speakers of Standard Swedish (2 males, 1 female), 3 speakers of American English (2 males, 1 female) and 4 speakers of Standard Japanese (3 males, 1 female). Registration of the speech samples was conducted in speechoachic, sound-insulated recording studios both at ATR in Kyoto (Japan) and at CTH in Gothenburg (Sweden), using high-quality digital recording equipment.

3. ANALYSES

Approximately one minute of recorded speech per speaker and speech style was analysed for this study. Pitch extraction was performed using the DWAPIT pitch determination algorithm presented earlier in [3]. Pitch estimates were obtained at 16-ms intervals for both periodic and aperiodic (laryngealized) stretches of speech. Tracings into intonation units (IU) was performed following the approach published in [4]. To this approach, two global declination lines which approximate the trends in time of the peaks (tope) and valleys (baseline) of $F_0$ across the utterance, are computed by the linear regression method. Computation is reiterated every time the Pearson correlation coefficient drops below a preset level of significance. Segmentation is thus performed without prior knowledge of higher level linguistic information, with the termination of an intonation unit being determined by the general resensing of the intonation contour wherever in the utterance it may occur. The $F_0$ onsets (intercepts) and offsets (end points), durations, declination line slopes and key values of these intonation units, as well as their time-alignment with features of linguistic structure were established individually for each of the speakers participating in this study.

4. RESULTS

4.1 Number of Intonation Units

A total of 586 intonation units has been established in the accumulated material for all ten speakers. The distribution of these intonation units per language and speech style is summarized below in figure 1.

Figure 1. Intonation units per language and speech style. The bar heights depict the ratio between the number of intonation units and the number of sentences contained in the respective material.

These distributions reveal a clear and consistent tendency, observable in each of the three languages, to subdivide orally read texts into a larger number of prosodically cued chunks than both the list and the dialogue material. All ten speakers produced predominantly one intonation unit per sentence in the list reading task, as predicted by most studies of sentence intonation, whereas in the text reading task the individual sentences were processed on the average in terms of between 2 and 3 intonation units.

Quite obviously, differences in sentence structure and informational content need to be taken into account for a comprehensive assessment of these ratios. This is particularly relevant with regard to the relatively few values obtained for the dialogues which clearly reflect (1) the comparatively larger proportion of short and incomplete sentences included in the material, and (2) the more frequent use of intonation units that stretch over the time extent of several consecutive sentences (cf. [6] for a more detailed discussion). Also, the dialogue material investigated in this study contains significantly less subordination than the texts and sentences, and only few examples of it-clefts and wh-clefts that typically occur as separate, prosodically cued chunks in read narrative.

Considering the higher degree of interlanguage variability found in the
dialogues, it must also be appreciated that the Swedish material consists of hesitations, false starts, fragmentary spontaneous conversations, i.e. constructions, etc. than the simulated dialogues in the English and Japanese samples.

The overwhelming majority (84.6%) of intonation units correspond in a clearly defined way with units of syntactic structure. This regular syntax-phonetic correspondence, however, is significantly more prevalent in the Japanese (98.2%) than in English (82.2%) and Swedish (79.9%) material. It is also slightly more prominent in the orally read texts (85.5%) as compared with the dialogues (83.8%).

Most commonly in our accumulated dialogue material, intonation units correspond in a regular fashion with single sentences (40.3%), whereas in the text material the results are more inconsistent between the three languages investigated in this study. In 36.6% of the English and 32.4% of the Swedish texts, intonation units start time-align clauses. In the Japanese text material, on the other hand, only about one tenth (10.1%) of the intonation units position to the clause correspondence class, thus indicating a markedly different prosodic processing behavior.

Larger structures beyond the sentence domain (i.e. stretching over two or three consecutive sentences) are almost exclusively found in the dialogues, with only 1.1% 3-sentence occurrences in the English and 2.1% 2-sentence occurrences in the Swedish texts. Conversely, intonation units corresponding to single constituents in the sub-sentence domain (i.e. nounphrase-subjects, verbphrases, adverbs, gerundial clauses, etc.) occur more often in the text (41.9%) than in the dialogue (24.9%) material, with a significant prevalence in the Japanese (60.3%) as compared with both the English (35.9%) and Swedish (29.5%) speech samples.

One of the dialogue materials has been scrutinized at such a detailed level of linguistic analysis. For the speech samples produced in the list reading task, a predominant one-to-one relationship between isolated sentences and single, coherent intonation units has already been established in the previous section.

4.2 Prosody-Syntax Alignment

The overwhelming majority (84.6%) of intonation units identified by the segmentation algorithm correspond in a clearly defined way with units of syntactic structure. This regular syntax-phonetic correspondence, however, is significantly more prevalent in the Japanese (98.2%) than in English (82.2%) and Swedish (79.9%) material. It is also slightly more prominent in the orally read texts (85.5%) as compared with the dialogues (83.8%).

Most commonly in our accumulated dialogue material, intonation units correspond in a regular fashion with single sentences (40.3%), whereas in the text material the results are more inconsistent between the three languages investigated in this study. In 36.6% of the English and 32.4% of the Swedish texts, intonation units start time-align clauses. In the Japanese text material, on the other hand, only about one tenth (10.1%) of the intonation units position to the clause correspondence class, thus indicating a markedly different prosodic processing behavior.

Larger structures beyond the sentence domain (i.e. stretching over two or three consecutive sentences) are almost exclusively found in the dialogues, with only 1.1% 3-sentence occurrences in the English and 2.1% 2-sentence occurrences in the Swedish texts. Conversely, intonation units corresponding to single constituents in the sub-sentence domain (i.e. nounphrase-subjects, verbphrases, adverbs, gerundial clauses, etc.) occur more often in the text (41.9%) than in the dialogue (24.9%) material, with a significant prevalence in the Japanese (60.3%) as compared with both the English (35.9%) and Swedish (29.5%) speech samples.

One of the dialogue materials has been scrutinized at such a detailed level of linguistic analysis. For the speech samples produced in the list reading task, a predominant one-to-one relationship between isolated sentences and single, coherent intonation units has already been established in the previous section.

4.3 Declination Line Parameters

The declination line parameters onset (intercept), offset (endpoint), duration, slope and key were calculated separately for each of the 586 intonation units investigated in this study. Statistical evaluation of these data revealed the following tendencies:

(1) Intonation units aligning with the isolated sentences from the list reading task are on the average shorter, steeper, less varied, and substantially lower topline intercepts in the discourse material.

(2) Important features of prosodic variation such as for instance rising baselines, "bi-modal" toplines, and narrow versus wide key (cf. [5]) do not occur in the list material at all, but are frequently used in discourse.

(3) The only parameter for which no statistically significant differences could be established between the different kinds of material is the baseline endpoint, which thus appears independent of context.

There are, however, significant differences in the frequency of occurrence of these patterns between the three languages, as reflected in the following percentages:

**SWEDISH**: 26.8% **ENGLISH**: 33.4% **JAPANESE**: 39.8%

Even more importantly, the use of laryngealization as a boundary marker differs markedly between the three languages, where it occurs least frequently in the lists of semantically unrelated sentences (Swedish 7.3%; English 10.1%; Japanese 15.3%) and most frequently in the narrative texts (Swedish 60.4%; Japanese 49.3%). The respective figures for the discourse material (Swedish 32.1%; English 35.7%; Japanese 37.2%) reveal a somewhat intermediary status for the

**REFERENCES**


