Semi-automatic Synthesis of Intonation for Dutch and British English

H.W. Zelle, J.R. de Pijper and J. 't Hart
Eindhoven, the Netherlands

1. Aim

The aim of the work reported here is to develop computer programs with which arbitrary utterances of restricted length in Dutch and English could be provided with acceptable pitch contours. The full melodic range of the language should be available. An additional requirement is that the amount of input data should be restricted to a minimum. This implies that the program should have built-in 'knowledge' of the rules and restrictions governing the internal structure of the pitch contours, and of accepted and forbidden combinations in each of the languages.

2. Background

We must make a distinction between Dutch and British English in view of differences in progress in our analysis of the intonation of the respective languages. Our earlier analysis of Dutch intonation has incorporated a corpus of considerable extension, which makes us confident enough to state that we have, in our 'grammar of Dutch intonation', incorporated practically all melodic possibilities. These results have been described in 't Hart and Collier (1975).

De Pijper's analysis of British English intonation has primarily concentrated on the most frequently occurring pattern of that language; in addition, it appeared possible to give explicit recipes for the construction of each of Halliday's (1970) seven Primary Tones (cf. 't Hart and de Pijper, 1983). There are reasons to doubt the adequacy of Halliday's analysis of the main features into precisely these seven categories. However, this is not the issue here. The examples recorded on Halliday's tape remain examples of fully acceptably intonated British English; this has been verified by de Pijper in listening experiments with native British subjects, (de Pijper, 1983).

The results of these analyses is a representation of both Dutch and English intonation in fully standardized form. In this only the perceptually relevant pitch movements come into play, in the shape of straight lines, with standard specifications of the size, slope and position of these movements in the syllable. On perceptual grounds, we have found no reason to apply a smoothing of these straight line contours.

3. Input data

It is assumed that an arbitrary speech file is stored on disk, in the form of formant, LPC, or channel vocoder parameters. These data may be derived from analysis of natural speech or obtained by diphone concatenation or allophone synthesis by rule. The experimenter has to provide the following additional data:

1. The temporal location of the vowel onsets of the syllables that should receive a pitch accent;
2. The location of syntactic boundaries to be marked intonationally;
   2a. In cases where this marking is required to be done by means of a continuation rise: the location of the last voiced frame before the boundary;
3. A code indicating the type of basis intonation pattern underlying the desired contour:
4. The choice made between a great number of optional variation possibilities.

Ad 1.
We do not have a sufficiently complete set of rules from which we can derive which words should be pitch-accented. Neither do we have rules for lexical stress of polysyllabic words. In the case of speech produced by diphone concatenation, the vowel onsets of all syllables are automatically available.

Ad 2.
In Dutch, the two most frequently occurring intonational markings for major syntactic boundaries are the postponed, or non-final fall, and the continuation rise followed by a (silent) fall. For the former, an indication of the frame number of the word boundary is sufficient. For the latter, the occurrence of a voiceless stretch immediately before the boundary would prevent its audibility. Therefore, the program also requires the location of the end of voicing.

The program for British English does not yet incorporate the feature of joining two sub-contours together by means of boundary markers. But since a continuation rise is one of the possibilities of that language as well, the same measure will have to be taken.

Ad 3.
For British English, there are Halliday's seven Primary Tones, numbered from 1 through 7, where 6 means Tone 13, and 7 means Tone 53 (in Halliday's terminology).

For Dutch, there are nine basic patterns in the program's menu. Pattern 1 is the 'hat-pattern' with choice of one to four pitch accents, but without continuation rise. Pattern 2 is the same, but with the option of one or more continuation rises at so-called primary boundaries. In no. 3 the final pitch

accent is given by a rise, and in no. 4 the only one by a fall, the contour beginning high. No. 5 is the 'cap-pattern', with the typically 'scooped' rise, no. 6 is the 'valley-pattern' with gradual rise of 'inclination', nos. 7-9 are patterns with one or more half-falls ('call contours', 'terrace contours', or contours in which the half-fall serves as a boundary marker).

Ad 4.
Examples of optional variations are: postponing non-final falls to mark (secondary) boundaries, final rise, gradual fall, high onset. The program offers these options only if the intonation grammar allows these variations in the chosen pattern. Other options are: excursions other than standard (= 6 semitones), final frequency other than standard (= male 75 Hz, female 180 Hz).

4. Generating the pitch contour

All pitch movements called for by the choice of the pattern and by the options are to be superimposed on the declination line. The slope of this line is calculated from the automatically derived utterance duration with (D in semitones per second and t in seconds):

$D = -8.5/t$ for $t > 5$ s and $D = -11/(t+1.5)$ for $t \leq 5$ s (Cohen, Collier and 't Hart, 1982).

The program detects pauses within utterances (operationally defined as stretches of more than 250 ms during which the amplitude remains below a given threshold). During such pauses, declination is stopped and non-final falls are made 1 semitone smaller, such that a kind of 'declination resetting' takes place. Together with the standard (or chosen) final frequency, the slope defines the initial frequency. If a high onset is chosen, the initial frequency is 6 semitones higher (in the case of standard excursions). If pauses have been detected the initial frequency is adapted accordingly.

Figs. 1-3 show some examples. Fig. 1 refers to an English sentence (from Halliday's tape), Figs. 2 and 3 to a Dutch sentence, spoken without and with a pause, to show the effect of automatic declination resetting.

For each pattern, the grammar dictates the nature and order of the pitch movements, as well as their position with respect to the vowel onsets (or voicing offsets) of the syllables involved. As a consequence, the input data are sufficient for the program to generate the entire pitch contour.

5. Results and Applications

For Dutch intonation, the adequacy of the grammar has been tested on a large corpus earlier, and that of the fully standardized stylizations has been demonstrated on many occasions. For British intonation, the development

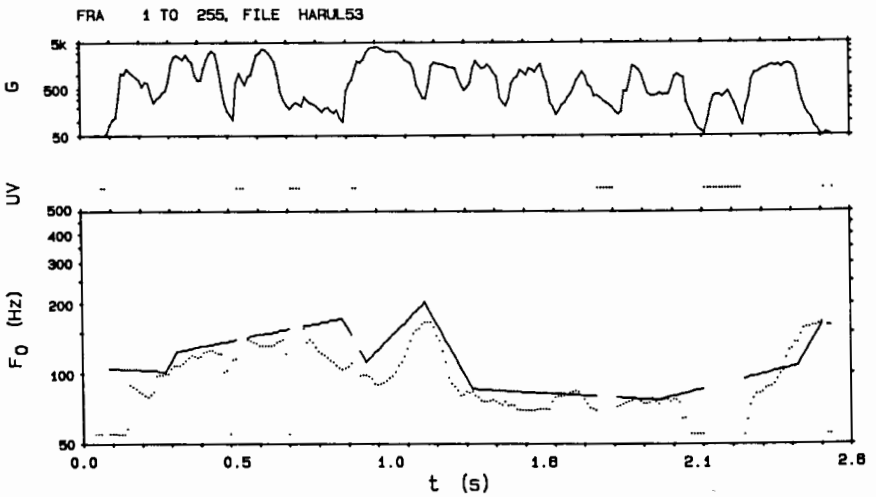

Fig. 1. Example of the sentence 'He's never taken **Jane** on any of his visits though', as spoken with Tone 53 (dashed), and as intonated artificially by means of the program (solid line). Necessary input data were: Tone: 7; Tonic (vowel onset of 'Jane'): 96; Pretonic (vowel onset of 'never'): 32; Secondary Tonic (vowel onset of 'visits'): 187; End of voicing: 250.

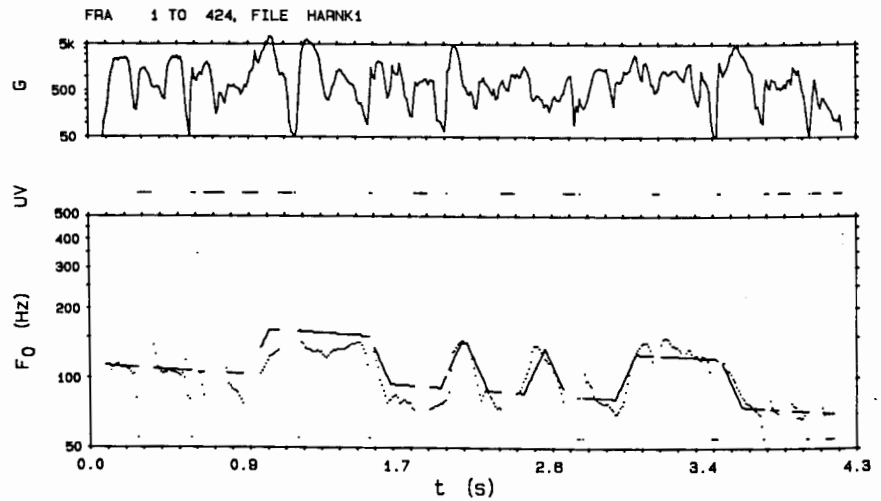

Fig. 2. Example of the Dutch sentence 'Wie het Fonetisch Con**gres** bezoekt / en Amster**dam** wil zien // hoeft maar een **half** uur in de **trein** te zitten' (Who the Phonetic Congress attends / and Amsterdam wants (to) see // needs only a half hour in the train to sit) spoken (dashed) and intonated artificially (solid line). Input data were (free options omitted): Pattern: 2; No. of accents: 4; Flat hat between accents 1 and 2: no; Flat hat between accents 2 and 3: no; Flat hat between accents 3 and 4: yes; Frame no. accent 1: 96 (vowel onset of 'Congres'); Frame no. boundary: 157 (end of 'bijwoont'); Primary boundary?: no; Frame no. accent 2: 204 (vowel onset of '(Amster-)dam'); Frame no. boundary : 256 (end of voicing of 'zien'); Primary boundary?: yes; Frame no. accent 3: 302 (vowel onset of 'half'); Frame no. accent 4: 356 (vowel onset of 'trein').

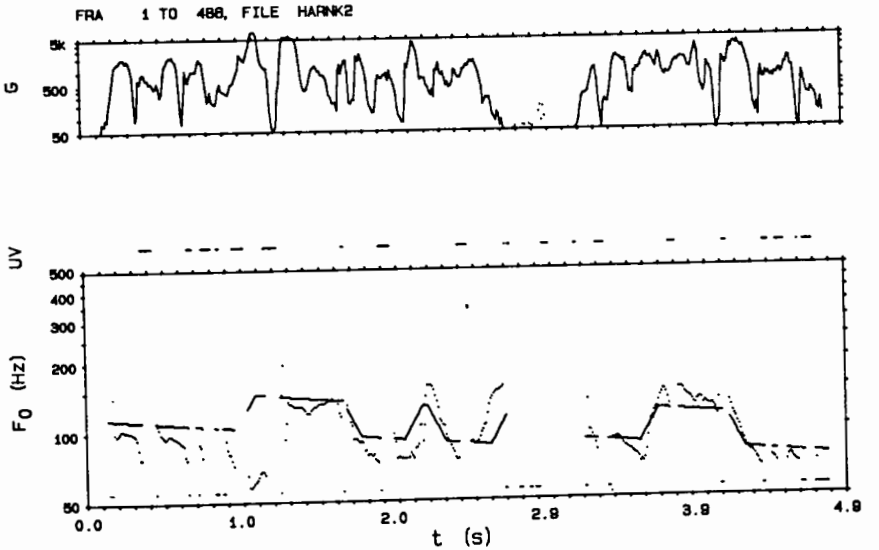

Fig. 3. The same sentence as in Fig. 2, now spoken with a pause, giving rise to automatic declination resetting.

of a grammar is still under way; on the other hand, testing of the acceptability of the stylized contours has been carried out very thoroughly with native users of the language ('t Hart and de Pijper, 1983).

Everyday application in the laboratory has been in speech (re-)synthesis activities where no original F_0 curves are available, such as in diphone synthesis, or in experiments in which the course of F_0 is a controlled experimental variable. The obvious advantage of the program is that experimenters no longer have to be instructed in the grammar of intonation before being able to actually produce acceptable synthetic pitch contours.

References

Cohen, A., Collier, R. and 't Hart, J. (1982). Declination: construct or intrinsic feature of speech pitch? *Phonetica* **39**, pp. 254-273.

Halliday, M.A.K. (1970). *A course in spoken English: intonation*, Oxford University Press, Oxford.

't Hart, J. and Collier, R. (1975). Integrating different levels of intonation analysis, *J. of Phonetics* **3**, pp. 235-255.

't Hart, J. and de Pijper, J.R. (1983). *Proc. Xth Int. Congress of Phonetic Sciences.*

de Pijper, J.R. (1983), *Modelling British English intonation. An analysis by resynthesis of British English intonation.* Foris Publications, Dordrecht.