SYNTHESIS OF LOGICAL ACCENTUATION IN DIFFERENT LANGUAGES

Antti Sovijärvi & Reijo Aulanko

Department of Phonetics, University of Helsinki,
Vironkatu 1, SF-00170 Helsinki, Finland

ABSTRACT

The phonetic realization of logical emphasis in different languages (Finnish, Hungarian, Estonian, Swedish spoken in Finland, German and Italian) is studied with speech synthesis by creating short sentences where different words in different sentence positions are accented one by one, using the possibilities of our method for changing the intensity, fundamental frequency, and duration parameters. In our system, four quantity degrees can currently be produced with the prosodic rules, and the general pitch level can be lowered or raised from the neutral level, in addition to slowly rising or falling pitch contours and several types of local modifications in pitch. The synthesis rules for logical accentuation in these languages are presented, and the differences among them are considered.

INTRODUCTION

In a normal, neutral utterance there is generally thought to be one nuclear point, which has the main sentence stress. We use the term **logical accentuation** for utterances where one constituent that expresses a paradigmatic opposition with some implicit alternative(s) gets a special, additional stress. Other terms often used for this kind of accent are e.g. contrastive sentence stress and emphatic accentuation [4, 5]. A speaker has at his/her disposal certain phonetic means for actualizing the necessary logical contrast in the acoustic speech signal. The phenomenon is generally assumed to be universal in its phonetic realization, but different languages may, nevertheless, use different phonetic systems for the realization of this kind of logical accentuation. For instance, the quantity of a logically accented syllable is often lengthened, in some languages especially the vowel and in Finnish the syllable-final consonant.

Our task is to find out (for a number of languages) the pitch and stress contours (and their changes) of logically accented words in short sentences and to formulate these results as synthesis rules. One aim of the project is to improve the quality of synthetic speech by making the prosodic rules correspond more closely to the prosody of natural speech. The synthesis rules for producing logical accentuation could be implemented in various rule synthesis programs.

METHOD

Recordings

As a basis for synthesizing accented variants of a sentence, we have recorded examples of logical accentuation in six different languages: Finnish (speaker M.L.), Hungarian (I.S.), Estonian (A.K.), Swedish spoken in Finland (L.N.), German (M.R.), and Italian (G.G.). The affirmative test sentence (see the Appendix) was semantically identical in all languages, meaning approximately '**Quite soon, I shall be going to a conference abroad.**' The first variant was a sentence with neutral sentence stress distribution. In the other variants a word or word pair was logically accented. We have analyzed the prosody of the recorded sentences and then synthesized them. The (male) speakers were all barytones, with the exception of the Hungarian who was a bass and the Italian who was a tenor. The synthesis, however, is realized in the pitch range of a bass in all languages, and the pitch movements found in the informants' utterances have been adapted for the synthesis accordingly.

Synthesis equipment

In our synthesis project we have been using the speech synthesizer OVE IIIb controlled by an HP 21 MX real-time computer [11]. Developed starting out from Sovijärvi's **beat phase theory** for word stress [7], our system is based on the use of **four-phase diphones** ("keys") stored in the computer's memory [8]. In our system each diphone has four matrix rows, and each matrix row contains 16 parameter values controlling the OVE IIIb synthesizer. Thus, each diphone with its four phases has a total of 64 parameter values. The rule synthesis system was designed primarily for synthesizing **Hungarian**, and the whole 'library' contains now of about 1330 keys, 80 of which represent sound combinations specially constructed for this paper, departing from the original 'library'.

In the diphones, the fundamental frequency (F0) of each unstressed and unmodified vowel varies within a range of one semitone (82-87-82 Hz = E-F-E). The contours of A0 (= amplitude of the first harmonic) are typically different for stressed and unstressed vowels [11], and they have been imitated in the sentence examples of all the six languages. In consonants the values of the parameters A0, AH, AN, FN, K0, K1 and K2 depend on the essential manner of articulation. The value of F0 is constantly 82 Hz in all unstressed, unmodified and unvoiced consonants [9].

Synthesizing Intonational Variants

For regulating the optional variations in intonation and/or stress contours we use special prosodic control symbols [9]. The slightly rising-falling basic pattern of the F0 contour of a vowel is always preserved, regardless of the application of any prosodic symbols. The prosodic control symbols are the following [10, 12]:

An overall **raised F0 level** can be produced with the symbols **#** (5 semitones) and **^** (2 semitones) in a sequence delimited by brackets (<>).

An overall **lowering of the F0 level** (by 2 semitones) is produced with the symbol **=**.

The minus sign (-) produces a **falling intonation:** F0 falls until the slash (/) or the first pause sign (_) gradually during the three next sounds (1st sound 2 semitones lower, 2nd sound 3 semitones lower, 3rd and subsequent sounds 4 semitones lower).

The plus sign (+) causes a **rising progredient intonation:** F0 increases until the slash or the first pause sign by 1 semitone for each sound until the fifth sound. In subsequent sounds F0 remains at the level of the fifth sound. (This symbol is not represented in the examples of this paper.)

A special **interrogative intonation** typical of Hungarian interrogative sentences that presuppose a *yes-no* answer and contain more than two syllables, is produced with the question mark (?): F0 rises by 7 semitones on one syllable. (This symbol is not used in the examples of this paper.)

The symbols " and ' are used for producing the **stronger and weaker degrees of stress.** The former symbol produces a 10% increase in duration, an increase in F0 by 4 semitones, and a rise in A0 by the dB values 3, 8, 6, and 2 in the successive subphases of the stressed vowel; furthermore, in the vowel of the following syllable, F0 increases by 2 semitones, and, in the consonant(s) appearing between the syllables, by 3 semitones above the level of 82 Hz. The latter symbol (i.e. weaker stress) produces an increase in F0 by 2 semitones and a slight rise in A0 by 0, 4, 3, and 1 dB in the successive subphases of the vowel after the symbol. That is, these two symbols are used to vary both the musical and dynamic aspects of intonation, whereas all the symbols mentioned previously only deal with the musical aspect.

The colon (:) is used for **lengthening** a sound: the duration parameter (DR) is increased by 150% in the third phase of a vowel and in the second phase of a consonant. An **overlong** degree of a sound can be produced by a double colon (::), which causes the DR to be increased by 300% in the corresponding phases.

The semicolon (;) is used for **shortening** a sound: the DR of the third phase of a vowel is reduced to 30% of the original duration.

The comma (,) is used for shortening a sound to a **half-long** degree: the duration of the third phase of a vowel is shortened to 70% of the original.

The **rate of speech** is controlled by a special coefficient placed before the input sequence. (In all the sentence examples of this paper the speed coefficient was set at 38.)

In order to investigate the quality of intonation in our four-phase synthesis system with material from the six languages analyzed, we constructed accentual variants of a sentence on the basis of the recordings described above, and experimented with several ways of synthesizing their intonation and stress contours. In our synthesis system we can change and vary very quickly the pitch, intensity, and/or duration of the accented words (or word pairs) in the written string, which has, at first, rather arbitrary symbols of prosody. Examples are listed in the Appendix in three different manners: (a) the normal orthographic form of the test sentence as translated into the six languages, (italics are used to denote the words with logical emphasis) (b) the input strings for synthesizing these sentences (with the diphone names and all the necessary prosodic control symbols), (c) the (highest) pitch levels of syllable nuclei displayed as semitones (zero meaning the rest level, negative numbers a lower pitch level, and positive numbers a higher pitch level).

RESULTS

About the Sentence Examples

Based on our completed 'library' of Hungarian speech-sounds we used in the sentence examples about the same vowel and consonant qualities [1-3, 6]. In the Finnish sentences (both neutral and logical variants), for instance, the symbols for synthesizing the short vowels [a] and [e] are written as **A2;** and **E2;**. However, for the relatively dark Finnish [s] (denoted with **S3** in the input strings) we had to construct new keys. The difference in duration between single and geminate tenuis plosives [p t k] is far greater in Finnish than it is in Hungarian, where the geminate plosives are relatively much shorter [2]. For this reason we have used the symbol **::** in the word *piakkoin* 'quite soon'.

The Hungarian (and other) examples demand, according to our judgment, sometimes either the symbol **^** for the rise of two semitones or the symbol **'** for the same rise and the weaker stress. Therefore we have synthesized the important words *hamarosan* 'quite soon' and *külföldi* 'abroad', changing these symbols.

The Estonian sentences presuppose the vowel quality [ɔɾ] in the words *sõidan* 'I travel' and *õige* 'quite' which have the corresponding marks of **O3**. The symbols **P3 T3 K3** correspond to the letters b d g.

In the sentences representing Swedish spoken in Finland we have very often used the slash (/) which acts as an end point for the pitch-lowering effect of the minus sign (-). The symbol **U2** represents the vowel [ʉ], which differs from the more [y]-like corresponding sound in Swedish spoken in Sweden.

In the German examples (for instance in the words *werde, einer,* and *ausländischen*) there are the special vowel symbols **E3 I3 U3** which correspond to the sounds [ə ɪ̯ ʊ̯], respectively, and the special symbol **R2** for the German variants of [ʀ ʁ ʔ].

In the Italian language we have to employ the half-long duration mainly in the cases of a stressed vowel and of a syllable ending in a consonant. For this reason we use the comma (,) for

instance in the word *andrò* 'I shall be going': A2,NDRO2;.

The Synthetic Realization of Logical Accentuation

The logically accented nuclei have almost always the synthesized height of five semitones above the rest level [13]. In only one case in the examples, namely the Italian words *fra poco* 'quite soon', we find the higher intonation of 2 → 7 → 5 on the nucleus of the syllable *po-*; the string used to produce this is ^<FRA2;>#<P'02K02;>. The other alternatives for the peak height - 4 or 7 semitones - sound rather unnatural, according to our auditory experiments with various synthetic versions.

The difference of the highest tonal point between the usually and logically accented affirmative sentences is 3-5 semitones in all six languages. This means that the whole intonational contour in the neutral sentences is definitely narrower with a phonetic certainty in a given language than in the same sentences with any logically accented word or word pair (5-8 against 8-11 semitones, depending on the language).

We have not yet thoroughly investigated the interrogative sentences from the logical accentuation point of view. However, we have recently described [13] altogether three cases of logical accentuation, two sentences of which represent different questions (with the highest pitch of 4 or 5 semitones) and only one represents an affirmative sentence (with the corresponding height of 7 semitones).

At the present stage the controlling system does not allow the synthesis of intonational phenomena of those languages which are characterized by an alternating "winding" tonal patterning and often by its relatively variable durative dimensions. The languages of this kind include for instance American and British English, French, and Swedish spoken in Sweden. In contrast, the language type in which the intonational patterns are largely based on the fixed or nearly fixed patterning of quantity and stress in the concatenation of syllables and/or words is suitable for the synthesis system presented in this paper.

REFERENCES

[1] Bolla, K. "Magyar hangalbum (A magyar beszédhangok artikulációs és akusztikus sajátságai)." Magyar fonetikai füzetek (Hungarian Papers in Phonetics) 6, 1980, 1-167.

[2] Bolla, K. "A finn beszédhangok atlasza (A beszédhangok képzési és hangzási jellemzői)." Magyar fonetikai füzetek (Hungarian Papers in Phonetics) 14, 1985, 1-249.

[3] Bolla, K. & Valaczkai, L. "Német beszédhangok atlasza (A beszédhangok képzési és hangzási sajátságai)." Magyar fonetikai füzetek (Hungarian Papers in Phonetics) 16, 1986, 1-210.

[4] Carlson, R. & Granström, B. "Word accent, emphatic stress, and syntax in a synthesis by rule scheme for Swedish. STL-QPSR 2-3, 1973, 31-36.

[5] Garde, P. "Contraste accentuel et contraste intonationnel." Word 23, 1967, 187-195.

[6] Olaszy, G. "A magyar beszéd leggyakoribb hangsorépitő elemeinek szerkezete és szintézise. A számitógépes beszédelőállítás néhány kérdése." Nyelvtudományi Értekezések 121. sz. Budapest: Akadémiai Kiadó, 1985.

[7] Sovijärvi, A. "Alustavia mittaushavaintoja suomen yleiskielen sanapainosta. (Vorläufige Messungsbeobachtungen über den Wortakzent der finnischen Hochsprache.)" Virittäjä 62, 1958, 351-365.

[8] Sovijärvi, A. "Examples of some synthesized Hungarian sentences." Sprache und Sprechen. Festschrift für Eberhard Zwirner zum 80. Geburtstag. Tübingen: Max Niemeyer, 1979, 113-120.

[9] Sovijärvi, A. "On the realization of Hungarian phonemic combinations and prosody in the light of experiments in speech synthesis." Congressus Quintus Internationalis Fenno-Ugristarum, Turku 20.-27.VIII.1980, Pars VI, 1981, 273-278.

[10] Sovijärvi, A. & Aulanko, R. "Unkarinkielisten puhunnosten sääntösynteesijärjestelmästä. (Introducing Speech Synthesis by Rule on Hungarian)." 11th Meeting of Finnish Phoneticians - Helsinki 1982 (ed. A. Iivonen & H. Kaskinen). Publications of the Department of Phonetics, University of Helsinki 35, 1982, 175-193.

[11] Sovijärvi, A. & Aulanko, R. "Rule synthesis of intonation and stress based on four-phase diphones." XIII Meeting of Finnish Phoneticians - Turku 1985 (ed. O. Aaltonen & T. Hulkko). Publications of the Department of Finnish and General Linguistics of the University of Turku 26, 1985, 217-231.

[12] Sovijärvi, A. & Aulanko, R. "Rule synthesis of variable intonation." Proceedings of Nordic Acoustical Meeting 20-22 August 1986 in Aalborg, Denmark (ed. H. Møller & P. Rubak), Aalborg University Press, 1986, 211-214.

[13] Sovijärvi, A. & Aulanko, R. "Die automatische Regelung der Intonations- und Betonungsgestalten beim Synthetisieren des Ungarischen." Gedächtniskolloquium für Eberhard Zwirner (ed. H. Bluhme). Publications of the Department Germaanse Filologie, Universität Antwerpen. (In press)

Appendix. Sentence examples

Finnish. Matkustan piakkoin erääseen ulkomaiseen kokoukseen.

```
=<_M'A2;TKUS3TA2;M>^<PI>A2;=<K::OI2N_>E2;RE:S3E2N_^<UL>KOMA2;I2S3E2N3=<K'OKO-UKS3E2N_>
    0-2-2           20-2          000         2000               0-2-4-6    [range: 8 semitones]
Matkustan piakkoin erääseen ulkomaiseen kokoukseen.
=<_MA2;TKUS3TA2;M>^<PI>A2;=<K::OI2N>E2;RE:S3E2N_#<UL>KO=<MA2;I2S3E2N3>KO=<KO-UKS3E2N_>
   -2-2-2          20-2          000        50-2-2               0-2-4-6    [11 st]
Matkustan piakkoin.erääseen ulkomaiseen kokoukseen.
=<_MA2;TKUS3TA2;M>#<PI>A2;=<K::OI2N_>^<E2;>=<RE:S3E2N_>^<UL>KO=<MA2;I2S3E2N3>=<KOKO-UKS3E2N_>
   -2-2-2          50-2          2-2-2          20-2-3                 -2-2-4-6  [11 st]
```

Hungarian. Hamarosan megyek egy külföldi összejövetelre.

```
^<_HA>MA=<ROS2AN>^<ME>=<G2EK_ET2>K'YL=<FQLDI>^<QS:>=<-EJQVETELRE_>
   20-2-2        2-2     -2    2-2-2      2-4-6-6-6-6       [8 st]
Hamarosan megyek egy külföldi összejövetelre.
^<_HA>MA=<ROS2AN>^<ME>=<G2EKET2>#<KYL>=<FQL-DI/Q-S:EJQVETELRE_>
   20-2-2        2-2     -2    5-2-5      -2-5-6-6-6-6      [11 st]
Hamarosan megyek egy külföldi összejövetelre.
#<_HA>MA=<ROS2ANMEG2EKET2>^<KYL>=<FQL-DI/-QS:EJQVETELRE_>
   50-2-2       -2-2 -2    2-2-5       -4-6-6-6-6-6         [11 st]
```

Estonian. Ma sõidan õige varsti välismaale ühele konverentsile.

```
_MA2;S'O3I2T3-A2,N_=<O3I2K3E2,>V'A2;RST-I_V'E=<LISMA2LE2;_>'YH2=<E2;LE2;K'ON:VE2;>-RE2;NCI=<LE2;_>
  0   2-2          -2-2      2-2        2-2-2-2        2-2-2          0-2-3-4-6       [8 st]
Ma sõidan õige varsti välismaale ühele konverentsile.
_MA2;S'O3I2T3A2,NO3I2K3E2,^<VA2;RS>TI_#<VE>^<LIS>MA2-LE2;_^<YH2>E2;LE2;^<KON:>=<-VE2;RE2;NCILE2;_>
  0   20        00        20        520-3            200           2-5-6-6-6       [11 st]
Ma sõidan õige varsti välismaale ühele konverentsile.
_MA2;SO3I2T3A2,N_O3I2K3E2,#<VA2;RS>^<TI_VE>LIS=<-MA2LE2;/>^<YH2>E2;LE2;^<KON:>VE2;-RE2;N=<CILE2;_>
  0   00        00        52       20-5-6            200           20-3-6-6        [11 st]
```

Swedish spoken in Finland. Jag skall resa ganska snart till en utländsk sammankomst.

```
_JA2;SKA2;R'E2SA2;GA2;NSK-A2;/SN'A2RT:IE2N_'U2:TLE2;NSKS'A2;M:A2;N3-KOMST_
  0    0    20    0-2          2    0 0   20          20-3          [5 st]
Jag skall resa ganska snart till en utländsk sammankomst.
_JA2;SKA2;R'E2^<SA2;>GA2;NSKA2;SN'A2RT:I-E2N_#<U2:T>LE2;NSK=<S'A2;M:A2;N3>-KO;MST_
  0    0    22          00        2    0 -2   50           0-2-3      [8 st]
Jag skall resa ganska snart till en utländsk sammankomst.
_JA2;SKA2;RE2S-A2;^<GA2;NS>KA2;#<SNA2RT:>IE2N_'U2:TLE2;NSK=<S'A2;M:A2;N3>-KO;MST_
  0    0   0-2       20          5      0 0   20           0-2-3      [8 st]
```

German. Ich werde ziemlich bald zu einer ausländischen Versammlung abreisen.

```
_IJ2=<VE2R2D-E3/ŠCI:MLIJ2^<PA2;L>C:UA2;I3-NE3R2_'A2;U3S-LE2;NDIS2:E3N_FE2;R2Z'A2;M-LUN3/=<A2;-PR2A2;I3ZE3N_>
 0     -2-4      00        2       0 0-3       2-3-4-4              02-3                 -2-6-6   [8 st]
Ich werde ziemlich bald zu einer ausländischen Versammlung abreisen.
_IJ2V-E2R2DE3/CI:M-LIJ2/^<PA2;L>C:UA2;I3N-E3R2_#<A2;U3S>-LE2;NDIS2:E3N_=<FE2;R2Z'A2;MLUN3_-A2;PR2A2;I3ZE3N_>
 0    -2-4     0-3        2       0 0-2       5-3-4-4             -20-2                  -4-6-6   [11 st]
Ich werde ziemlich bald zu einer ausländischen Versammlung abreisen.
_IJ2=<VE2R2D-E3/CI:MLIJ2>#<PA2;L>C:UA2;I3N-E3R2_=<'A2;U3SLE2;NDIS2:E3N_FE2;R2Z'AMLUN3_-A2;PR2A2;I3ZE3N_>
 0     -2-4     -2-2       5       0 0-2        0-2-2-2              -20-2             -4-6-6     [11 st]
```

Italian. Fra poco andrò ad una conferenza all'estero.

```
_FRA2;P'O2^<KO2;_>A2,ND^<RO2;>A2,D^<U:NA2;>KO2,NFE2;R'E2,NCA2;^<A2,L:'E2,ST>-E2;RO2;_
  0    22      02        0    22     0020              2     4-2-4             [8 st]
Fra poco andrò ad una conferenza all'estero.
^<_FRA2;>=<P'O2>^<KO2;_>A2,ND^<RO2;>A2,DU:NA2;=<KO2,NFE2;R'E2,N>^<CA2;>_=<A2,L:>#<E2,ST>-E2;RO2;_
    2      02         02        0   00      -2-202                -2       5-2-4  [9 st]
Fra poco andrò ad una conferenza all'estero.
^<_FRA2;>#<P'O2KO2;_>=<A2,NDRO2;A2,D>U:NA2;=<KO2,NFE2;R'E2,N>-CA2;_=<A2,L:>^<E2,ST>-E2;RO2;_
    2        75         -2-2      -2   00      -2-20-3             -2      2-2-4     [11 st]
```