Study on the phonetic pitch movement of the accentual phrase in Korean read speech

Hyongsil Cho, Stéphane Rauzy

To cite this version:

HAL Id: hal-00285529
https://hal.archives-ouvertes.fr/hal-00285529
Submitted on 5 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Phonetic pitch movements
of accentual phrases in Korean read speech

Hyongsil Cho & Stéphane Rauzy

Laboratoire Parole et Langage, CNRS
University of Provence, FRANCE
{hyongsil.cho;stephane.rauzy}@lpl.univ-aix.fr

Abstract
The minor prosodic unit in Korean language, generally called
an Accentual Phrase, is usually defined by its syntactic or
phonological characteristics. This article looks at the
correlation between phonetic pitch movements and accentual
phrase boundaries using a technique of pattern extraction and
prediction by a probabilistic grammar.

1. Introduction
Some recent studies have concluded that silent pauses, which
are usually considered the most important cue for the
definition of major prosodic boundaries, are a redundant
factor in perception and that the sentence boundary can be
predicted by other acoustic cues without taking the silent
pause into consideration [5][6]. This might mean that the
usual definition of prosodic units needs to be revised.
The aim of this study is to examine the role of phonetic
pitch movements in the definition of minor prosodic units.
The main methodology is based on the use of a probabilistic
grammar trained on a semi-automatically annotated corpus for
modelling the correlation between the phonetic pitch
movement and the accentual phrase boundary.

2. The Korean language
2.1. Prosodic units and boundaries in the Korean language
For the hierarchy of prosodic units in Korean, the framework
of K-ToBI, based on intonational phonology, has been quite
widely adopted assuming a hierarchical phonological structure
as illustrated in Figure 1.

![Figure 1. The intonational structure of the Korean language as described in the K-ToBI framework.](image)

Here, an Accentual Phrase (AP) is smaller than an
Intonational Phrase (IP) and larger than a phonological word
(W), which is defined as a lexical item plus a case marker or
postposition. An IP is marked by a boundary tone (%) and
final lengthening. An AP is marked by a phrasal tone, THLH
(where T=H if the AP initial segment is aspirated or tense,
T=L otherwise), but not by final lengthening [14].

2.2. The phonetics and phonology of accentual phrases
The underlying patterns of AP phrasal tone, LHLH or HHLH,
are realized on the surface differing across dialects. The most
frequent tonal patterns of AP in Seoul dialect, which is the
standard dialect of Korean, are illustrated in Figure 2.

![Figure 2. Seoul Korean AP tonal patterns illustrated in the K-ToBI.](image)

K-ToBI has a special tier called ‘phonetic tone tier’ to
mark the surface realization of AP tones [15]. However, all
the forms suggested in Figure 2 are still described only with L
and H, without considering the pitch variation range. For
this reason, tonal patterns in K-ToBI are difficult to apply directly
to speech synthesis or recognition, even though the concept of
K-ToBI has been widely adopted as the basic prosodic rule in
speech technology. In fact, it would be more reasonable to
consider the K-ToBI phonetic tier as a broad phonetic
transcription approaching a phonological description rather
than a narrow phonetic one. Because its phonological status is
sometimes misunderstood, K-ToBI has even been considered
as inadequate to describe the Korean language [18].

In the light of this complexity, researchers in speech
technology prefer to define the Korean accentual phrase by its
morpho-syntactic cues such as POS (Part Of Speech) or other
information obtained from graphic text [16][17][25].

This study attempts to outline a basis for the phonetic
characteristics of the accentual phrase with a narrower
transcription of pitch movements.

3. Corpus
3.1. Composition
The forty continuous passages from the Euromil corpus [4]
were freely translated and adapted to the Korean language by
the first author. All of the forty passages were each recorded
by ten standard Korean native speakers (5 males, 5 females) in an anechoic chamber and digitized in wav files. The complete recordings of the 400 passages last a total of just over 2 hours 7 minutes and the half of the recordings (20 passages for each of 10 speakers) were adopted as the data for this article.

3.2. Basic annotation
The data, composed of 200 sound files, was at first annotated using INTSINT. INTSINT (INternational Transcription System for INTonation) is a theory-independent annotation system for intonation, developed in the LPL in Aix-en-Provence over the last twenty years. It has been used for the phonetic modelling and symbolic coding of the intonation patterns of a number of languages [11], including English [1], French [23], Italian [10], Catalan [8], Brazilian Portuguese [9], Venezuelan Spanish [20], Russian [22], Arabic [21] and isi Zulu [19].

In the INTSINT framework, intonation patterns are represented as a sequence of tonal segments using an alphabet of 7 tonal symbols: T(op), M(id), B(ottom), H(igher), S(ame), L(ower), U(pstepped) and D(wonstepped). These tonal segments are aligned directly with the acoustic signal although it is assumed that at a more abstract level the alignment is determined by the prosodic structure of the utterance.

A phonetic interpretation of the INTSINT tonal segments can be carried out using two speaker dependent (or even utterance dependent) parameters of the pitch domain.

- **key**: like a musical key, this establishes an absolute point of reference defined by a fundamental frequency value (in Hertz).
- **range**: this determines the interval (in octaves) between the highest and lowest pitches of the utterance.

The targets T, M and B are defined 'absolutely' without regard to the preceding targets as below.

- T = key*√(2^range)
- M = key
- B = key/√(2^range)

The T and B are thus at equal (log) distance from the key (= M) and the interval between them corresponds to the speaker's range in octaves.

Other targets are defined with respect to the preceding target as follows:

- H = √((Pi-1*T))
- L = √((Pi-1*B))

These are thus situated half way (on a log scale) between the preceding target and the top or bottom of the range.

- S = Pi-1

This tone has the same value as the preceding target.

- U = √((Pi-1*√((Pi-1*T)))
- D = √((Pi-1*√((Pi-1*T)))

These are thus situated one quarter of the way between the preceding target and the top or bottom of the range.

The position of each target compared to the others is shown in Figure 3.

Figure 3. INTSINT, a system for annotation of intonation.

Thus, the final annotation provides a sequence like [M S H L U T D B]. To optimize the automatic INTSINT annotation, it is preferable to extract target points by MOMEL (MOdelling of the MELody) which filters out micromelodic components from macromelodic ones. The combination of MOMEL and INTSINT provides the possibility of two way conversion between the acoustic measurement and the phonetic analysis.

In this study, all the steps of Momel-Intsint were carried out in a semi-automatic way using the Momel-Intsint Plug-In developed for Praat [3][13].

3.3. Boundary marking
After automatic annotation of the intonation with the INTSINT alphabet, prosodic boundaries for AP and IP were added manually by the first author. To avoid the influence of IP final pitch movement on the prediction of the accentual phrase, we put an AP boundary marker only after a pure AP (IP initial and medial AP). The number of pure APs amounted to 2 122 for the whole of the data.

An example of the final annotation is shown in the figure 4.

Figure 4. Final Annotation with INTSINT alphabet and boundary markers (# for IP and / for AP).

The final annotation of each sound file was saved as Praat TextGrid file.

4. Probabilistic grammar and AP boundary prediction

4.1. Probabilistic grammar
Probabilistic grammar is a data modelling technique playing an increasing role in the fields of computational linguistic and
machine learning theory. The approach allows learning from a data training corpus, the regularities appearing as a sequence of symbols. The underlying statistical model makes use of the mathematical apparatus of conditional probability to capture the contextual dependencies between symbols and to identify pattern regularity.

In this study, the probabilistic grammar is built on the patterns model extracted by data training. The Patterns Model belongs to the family of probabilistic finite state automation approaches [2] (for Hidden Markov Model, see also [24]). The approach is characterized by an optimal extraction of the information content contained in the training sample. In this approach, unlike the n-gram model, the left context is not limited to a fixed number of symbols but rather takes into account the regularities of the training corpus. As a result, if any sequence of symbols occurs frequently in the data sample, the pattern is included in the model without considering the number of the symbols composing the sequence. This feature is particularly interesting for analysing a small size corpus like the data of the present study.

4.2. Accental phrase boundary prediction

For the preliminary step of the prediction, we built a patterns model by extracting statistically significant left context patterns from the corpus. The 200 TextGrid files containing the final annotation were put in one single data file. The pattern model is then used to investigate the sequential structure of INTSINT targets and boundary markers in the full data.

In a second step, we made a prediction of AP boundaries, using the predictive power of the grammar. For this step, the manually annotated AP boundaries were removed from the corpus to keep only the information concerning the sequences of INTSINT tones and IP boundaries. The model was then used as a predictive tool to insert AP boundaries between tones at the most probable location in the sequence.

5. Result

5.1. Patterns model construction

From the whole corpus containing 2 122 manually annotated AP boundary markers, 573 statistically significant left context patterns were extracted which characterize the model. The power of prediction of the model is evaluated by comparing the raw entropy of the distribution of the tones and boundaries markers (E=2.16) with the entropy of the symbols distribution when accounting for the model information (E=1.54). The mutual information brought by the model is thus of 0.62.

Among the 573 patterns, 72 of them concerned pre-AP boundary movements and each of the AP tonal patterns proposed by K-ToBI was demonstrated in a number of phonetic forms, which confirms the detail observed in [16].

For example, as we may see in the following table, 13 different kinds of phonetic patterns could be interpreted as the simple rising form described as LHa in K-ToBI system.

<table>
<thead>
<tr>
<th>Reference N.</th>
<th>Occurrences</th>
<th>Intsint annotation</th>
<th>K-ToBI interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>75</td>
<td>BH</td>
<td>LHa</td>
</tr>
<tr>
<td>235</td>
<td>65</td>
<td>BU</td>
<td>LHa</td>
</tr>
<tr>
<td>556</td>
<td>112</td>
<td>DU</td>
<td>LHa</td>
</tr>
</tbody>
</table>

### Table 1. Phonetic forms of LH tonal pattern.

For a static AP tonal pattern which is described as HHa by K-ToBI, the phonetic pitch movements were - as expected - more diversified depending on the position of the starting point of the static pattern.

<table>
<thead>
<tr>
<th>Reference N.</th>
<th>Occurrences</th>
<th>Intsint annotation</th>
<th>K-ToBI interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>495</td>
<td>28</td>
<td>US</td>
<td>HHa</td>
</tr>
<tr>
<td>159</td>
<td>39</td>
<td>TS</td>
<td>HHa</td>
</tr>
<tr>
<td>297</td>
<td>56</td>
<td>HS</td>
<td>HHa</td>
</tr>
<tr>
<td>130</td>
<td>136</td>
<td>T</td>
<td>HHa</td>
</tr>
<tr>
<td>247</td>
<td>403</td>
<td>H</td>
<td>HHa</td>
</tr>
</tbody>
</table>

### Table 2. Phonetic forms of HH tonal pattern.

This result, easily predictable but often neglected, confirms for us the necessity of a multi-directional approach for an optimal analysis of the prosody.

5.2. Prediction of AP boundaries

The results of the prediction can be summarized in terms of measure of precision and recall.

In the original corpus containing 2 122 pure accental phrases, all the AP boundary marks were removed automatically. Then, 1 654 AP boundaries were inserted by the prediction based on the patterns model. Among the 1 654 inserted boundaries, only 981 were found to be positioned at the correct location, which leads to an estimate of the precision and recall measures given table 2.

| Precision | 0.59310762 |
| Recall    | 0.46229972 |

### Table 3. Precision and recall of the predicted AP boundaries.

Several experiments were conducted in order to test the robustness of these results. The analysis was performed individually on each speaker without showing any significant deviation from the mean values of table 2, and a grouping by gender of speakers does not reveal any specific trend.

Given that the result was not significantly different even when we applied the model built from one speaker’s data to an others’, we may consider this result as a reference on the correlation between the pitch movement and the identification of the accental phrase.
Conclusion

In this study, we observed the character of phonetic pitch movements in a minor prosodic unit, by examining the possibility of predicting its boundary from the pitch movement. Accental phrases, which are usually defined by their phonological tonal form, could also be identified to a certain degree in this study only by their phonetic pitch movement. In further work it is planned to extend this experimental procedure to a larger corpus. It would be also interesting to look at the distribution of the "errors" of prediction, which will help us to get an optimal way to move from the acoustics to the phonology for the description of prosody.

Acknowledgement

The recording and labelling of the ten-speaker version of the Eurom1-KR corpus was carried out in cooperation with our partners in Seoul National University in the framework of the Franco-Korean Aanvis-KR project which was supported by the Egide PAI Star. We thank our Korean partners for this fruitful collaboration. We should also like to thank Daniel HIRST in LPL Aix-en-Provence for his kind help.

References