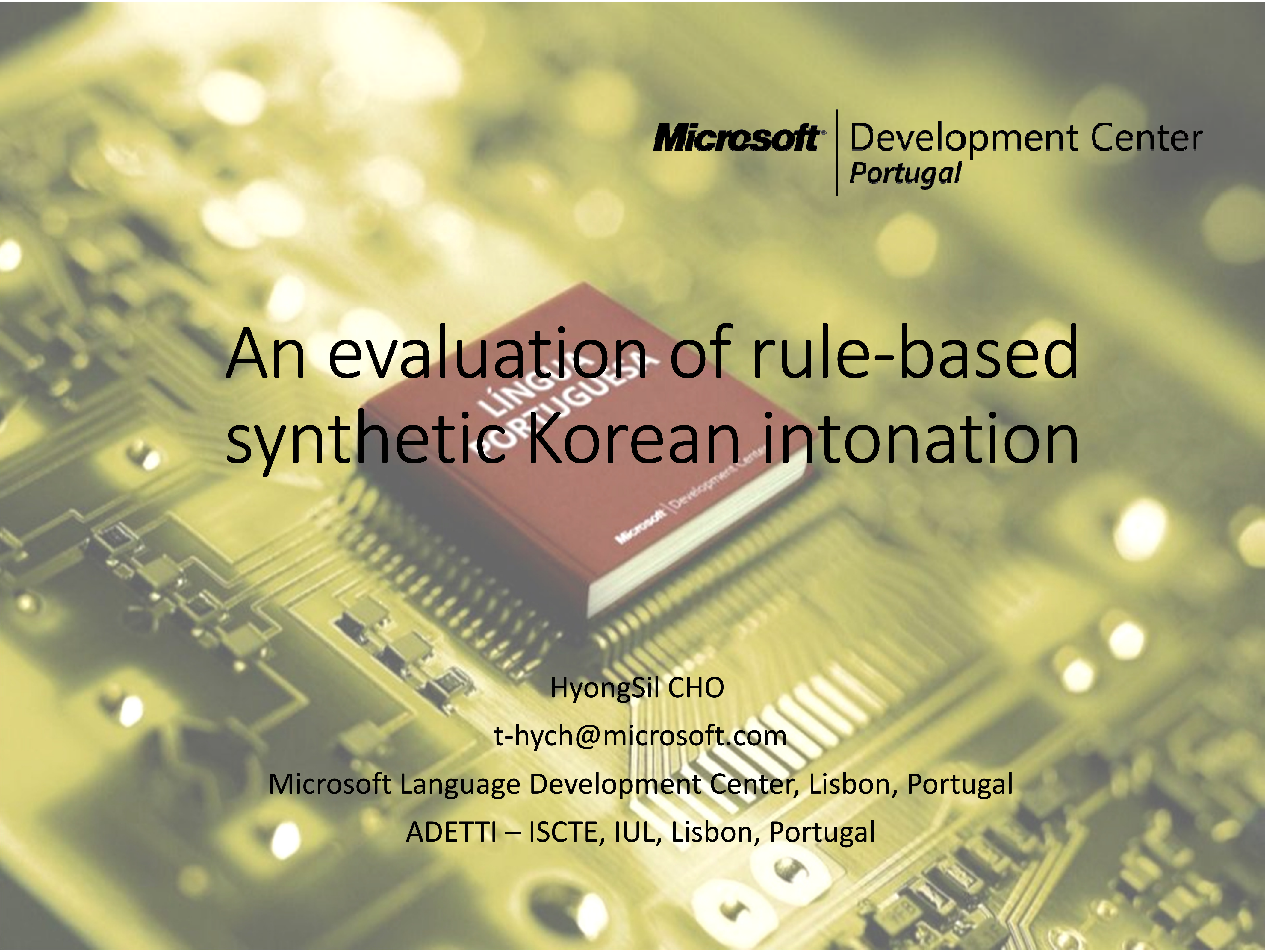


Microsoft | Development Center
Portugal

An evaluation of rule-based synthetic Korean intonation

HyongSil CHO

t-hych@microsoft.com

Microsoft Language Development Center, Lisbon, Portugal

ADETTI – ISCTE, IUL, Lisbon, Portugal

Introduction

- Speech synthesis technology has progressed remarkably over the past few years, especially with regard to segmental naturalness.
- Although the sound quality of segmental aspects has improved, a definition of an adequate model for the generation of prosody is largely still an unsolved problem.
- This is a matter of some concern, because prosody, and in particular intonation, plays a key role in the perceived naturalness of synthetic speech [2].
- In this study, we examine the possibility of a simplified rule-based synthesis system for Korean. Given that the language has a variety of boundary tones at the end of the sentence and that these boundary tones contain important linguistic and paralinguistic information, we decided in a first step to keep the original intonation intact for the end of the sentence and apply a simple algorithm to generate the intonation of the rest.
- We then made an MOS scale evaluation by Korean native speakers to compare the naturalness of synthesized sentences to the original ones.

Materials

- Korean Multext [8]: the Korean version of Eurom1 corpus
  - ➢40 passages localized into the Korean language (and culture) from the English text of Eurom 1
  - ➢Read by 5 males and 5 females native speakers of the Standard Korean language
  - ➢The total duration is 2 hours 7 minutes
  - ➢In this study,
    - One half of this corpus (20 passages by 10 speakers) was used for data analysis,
    - One female speaker’s files were used as the resource for synthesis

- Momel-Instint: In this study, we extracted, by using the Momel-Intsint Plug-In for Praat, the most frequent AP tonal pattern from our corpus and the average pitch rate of each of seven values from a female speaker’s data

Data analysis

- Extraction of AP tonal patterns

INTSINT annotation	Occurrence (%)
U	10.86
H	9.02
D	8.13
L	6.70
S	6.53
T	3.45
LH	3.05
DU	2.93
HL	2.35
Total	53.02

- ✓ the most frequent AP tonal pattern in our corpus is “U” (a simple rising contour)
- ✓ some other patterns like LH or DU are also from rising contours
- ✓ almost one half of APs (44.3%)in our corpus were pronounced in rising contour

- Calculate

the average pitch rate

for 7 Intsint values

Intsint value	**Original pitch value (Hz)**	**Normalised pitch value (Hz)**
T	316	**305**
M	219	**216**
B	161	**153**
U	216	**207**
D	215	**210**
H	230	**242**
L	**186**	**192**

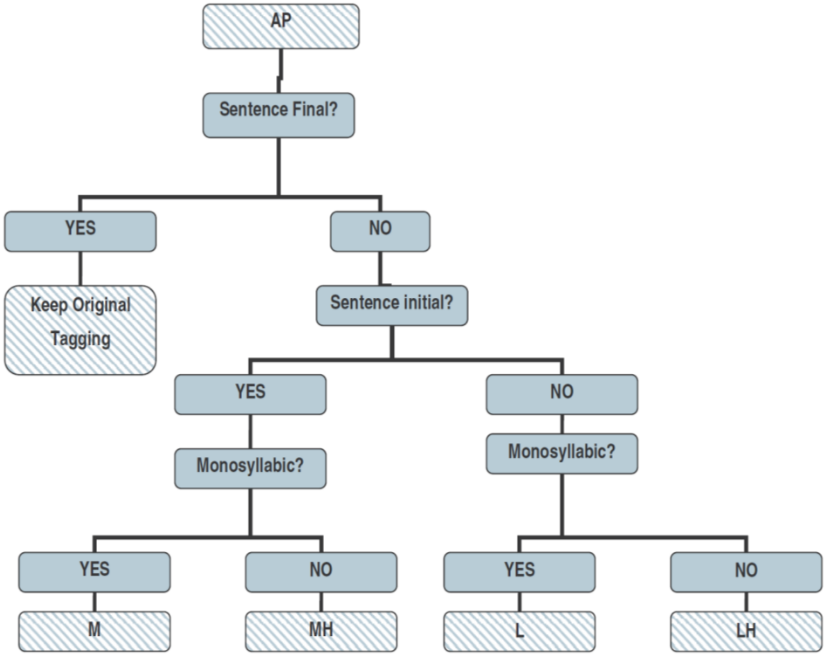
Synthesis by PSOLA

Microsoft | Development Center Portugal

- Specifying target points: Since we need two target points per AP to apply the "LH" sequence, we first removed the entire target points from the data and re-distributed two target points on the first and the last syllable of each AP.

- F0 curve generation:
  - each sentence initial AP is tagged "MH"
  - all sentence medial APs are tagged "LH"
  - a sentence final AP keeps its original tagging

- After this tree, a sentence is transcribed as #(MH)+(LH)…(LH)+(LHB)#



Evaluation

- 20 original sentences
- 20 re-synthesized sentences
- Listened in random order
- By 10 native speakers of Korean
- Rated by the mean opinion score (MOS): a way of measuring the acoustic quality of speech sound. Originally developed to evaluate compressor/decompressor (CODEC) systems and digital signal processing (DSP), MOS is largely adopted to the evaluation of synthesized speech. In this study, 10 Korean native speakers were invited to give a rating among :
    1) Very unnatural
    2) Unnatural
    3) Acceptable
    4) Natural
    5) Very natural

Result and conclusion

- The average score rated by ten participants is

  - ➢ 3.9 for twenty original sentences
  - ➢ and 3.4 for twenty re-synthesized sentences.
  - ➢ In some cases, native speakers even preferred the synthesized sentences to the original recording.
  - ➢ A clear preference for the natural speech when AP initial syllable was pronounced by “H” in the original recording and re-synthesized by “L”.

⇒ Even though the score is not so high, given that the difference between two groups of sentences is not significant, we may conclude that we can reach an acceptable level of naturalness with one single AP tonal pattern (if we preserve diverse patterns of IP boundary tones).

References

1) Boersma, P. & Weenink, D. 2006. Praat: doing phoneticsby computer. (Version 4.5.08) freely downloadable from http://www. praat.org
2) Bunnell, H T, Hoskins, S R, Yarrington, D. 1998. Prosodic vs. Segmental Contributions to Naturalness in a Diphone Synthesizer. In: Proceedings of the third ESCA/COCOSDA Workshop on Speech Synthesis, Jenolan, Autralia.
3) Chan, D., Fourcin, A., Gibbon, D., Granström, B., Huckvale, M., Kokkinas, G., Kvale, L., Lamel, L., Lindberg, L., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., & Zeiliger, J. 1995. EUROM: a spoken language resource for the EU. Proceedings of the 4th European Conference on Speech Communication and Speech Tecnology, Eurospeech '95, (Madrid) 1, 867-880.
4) Cho, H. & Rauzy, S. 2008 Phonetic pitch movements of accentual phrases in Korean read speech. Proceedings of Speech Prosody 2008. Campinas, Brazil.
5) Hirst, D.J. 2007 A Praat plugin for the Momel and INTSINT with improved algorithms for modelling and coding intonation. Proseedings of ICPhS 2007, Saarbrucken, Germany.
6) Jun, Sun-Ah 2005. Prosodic Typology in Sun-Ah Jun (ed.) Prosodic Typology: The Phonology of Intonation and Phrasing. pp. 430-458. Oxford University Press.
7) Jun, Sun-Ah. 2000. K-ToBI Labelling conventions. http://www.linguistics.ucla.edu/people/jun/ktobi/ktobi3-2.pdf (UCLA.)
8) Kim, S., Hirst, D., Cho, H., Lee, H., & Chung, M. 2008. Korean MULTEXT: A Korean Prosody Corpus. Proceedings of Speech Prosody 2008. Campinas, Brazil.
9) Kim, Y., Byeon, H. and Oh, Y. 1999. Prosodic Phrasing in Korean; Determine Governor, and then Split or Not. Proceedings of Eurospeech99, 539-542, 1999.
10) Lee, H. & Son, M. 2007. Perception of phrasal tones in Korean. Hangeul 2007 vol 3.
11) Lee, H.Y. 2004. H and L are not enough in intonational phonology. Eoneohag 39. The Linguistic Society of Korea. 200408, 71-79.
12) Lindstrom, A., Bretan, I. and Ljungqvist, M. 1996. Prosody Generation in Text-to-Speech Conversion Using Dependency Graphs. Proceedings of The Fourth International Conference on Spoken Language Processing. Philadelphia. USA.
13) Natvig, J. E. & Heggtveit, O. 2003. Prosodic Unit Selection for Text-to-Speech Synthesis. Telektronikk volume 2.
14) Yoon, K. 2006. A Prosodic Phrasing Model for a Korean Text-to-speech Synthesis System. Computer Speech and Language, 20(1):69-79, 2006.

COMPETE
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento regional