Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody*

Kyuchul Yoon
(Kyungnam University)

Yoon, Kyuchul. 2007. Imposing Native Speakers' Prosody on Non-native Speakers' Utterances: The technique of cloning prosody. ***The Journal of Modern British & American Language & Literature. ??,*** **???-???.** This paper introduces a technique for cloning the prosodic features of a native speaker's utterance onto the same sentence spoken by a non-native speaker. Three acoustic aspects of prosodic features were examined: the fundamental frequency (F0) contour, segmental durations, and the intensity contour. The fundamental frequency contour and the segmental durations of the native speaker's utterance were imposed on the non-native speaker's utterance using the PSOLA (pitch-synchronous overlap-add) algorithm (Mouline & Charpentier, 1990) implemented in Praat (Boersma, 2005). The intensity contour cloning was also done in Praat. The technique used for cloning one or more of these prosodic features was elaborated, and its implications were discussed.

Key words: Korean, prosody cloning, native speaker, non-native speaker, PSOLA, Praat

I. Introduction

One of the most critical tasks in the acquisition of a foreign language is the acquisition of the prosodic features of the language. The definition of prosodic features can vary, but for the purposes of this study, they were operationally identified as the intonation contour, assignment of phrase breaks, durations of the speech segments and the intensity contour of an utterance.

* This work was supported by Kyungnam University Foundation Grant, 2007.

In the traditional classroom environment, foreign or second language teachers explicitly taught the prosodic features of the target language. For example, Korean teachers teaching English to Korean students put much emphasis on word stress, intonation pattern, etc. In most cases, students learn these prosodic features of a foreign language by listening to or watching and attempting to imitate their teachers' or a native English speaker's speaking style as they observe it played on a cassette or video tape or CD-ROM/DVD educational software.

In such cases, the feedback the students may receive when they make mistakes is limited. In a classroom environment, teachers give their students specific instructions, making them repeat the target utterance. It is not uncommon for the teacher to draw the intonation contour of the target utterance on the chalkboard to help students understand the point the teacher is making.

In a self-study environment with CD-ROM/DVD education software, however, students get less feedback than they do in the classroom environment. There is also software, such as Dr.Speaking (언어과학, 2002), that provides visual feedback. This type of software records what the learner produces following a native speaker's utterance, displays the learner's intonation contour, as well as the exemplary intonation contour of the native speaker to show how closely the two match. Studies have shown that a visual display of the learner's intonation contour along with a native model provide valuable feedback to students (Abberton & Fourcin, 1975; Stibbard, 1996).

In both the classroom and the self-study environment, the major form of feedback is largely visual, which may not be optimal, given the circumstance that the student is attempting to learn a spoken form of the target language. Even if the feedback is presented in an entirely audio format, it is merely presented in the form of other people repeating the same target utterance.

In an alternate case, however, such as one wherein feedback is presented in the voice of the student, but with the prosodic features of the native speaker (especially for a student who may be less talented in learning foreign languages), this new type of audio feedback could provide

a new degree of motivation that none of the traditional feedback methods could have offered. This new type of audio feedback might work as follows: The software equipped with the capability to allow its user to apply the new technique plays the target sentence uttered by a native speaker, records what the language learner repeats, imposes only the prosodic features of the native speaker onto the learner's utterance, and plays back the learner's utterance with the native speaker's prosody, demonstrating to the second language learners that they could "speak" like the native speaker.

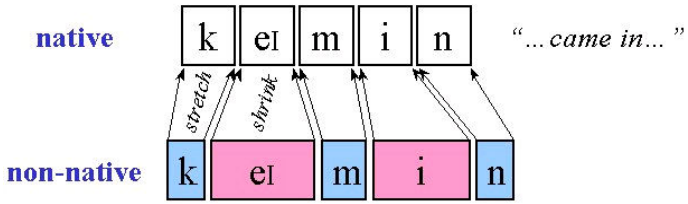
This paper describes, in detail, this new technique for super-imposing or cloning some, or all of the prosodic features of a native speaker's utterance onto the same utterance produced by a non-native language learner. The phrase breaks, segmental durations and the intonation contour were manipulated using the PSOLA (pitch-synchronous overlap-add) algorithm (Mouline & Charpentier, 1990) implemented in Praat (Boersma, 2005). The intensity contour was also manipulated in Praat.

II. Methods

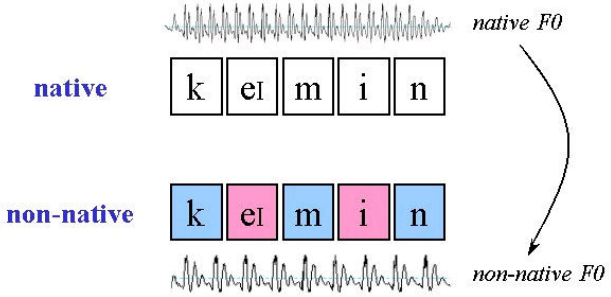
For the purposes of this study, a male native speaker of Korean in his late thirties read aloud the English question sentence "What did you say before that?", which was also repeated and recorded by a male native speaker of English (Ladefoged, 2006). The Korean speaker was a high school graduate who had not received any college education. His level of English proficiency was low.

Using the target sentence, the procedure for transferring the prosodic features of a native speaker's utterance to a non-native speaker's utterance proceeded in three stages. For illustrative purposes, however, a sample phrase"came in" is presented below. In the first step, the speech segments of the non-native version were aligned with those of the native version (See Figure 1). The segment alignment step is the most important of all, because the quality of subsequent manipulations depends on it. This alignment was followed by the stretching or shrinking of the non-native segments in conformity with the native segments, using the PSOLA algorithm (Mouline & Charpentier, 1990) implemented in Praat (Boersma, 2005). As a result, the non-native segments took on the same durations and proportions as the native segments. As an added benefit, the location of the phrase breaks were the same for the two versions of the target utterance.

<Figure 1> Illustration of step 1: alignment of speech segments. Given the sample phrase "came in", the segments of the non-native version are aligned and manipulated, i.e. stretched or shrunken, with respect to those of the native version.



<Figure 2> Illustration of step 2: native F0 imposition. After the segmental durations are adjusted, the non-native F0 contour is replaced with the native F0 contour.

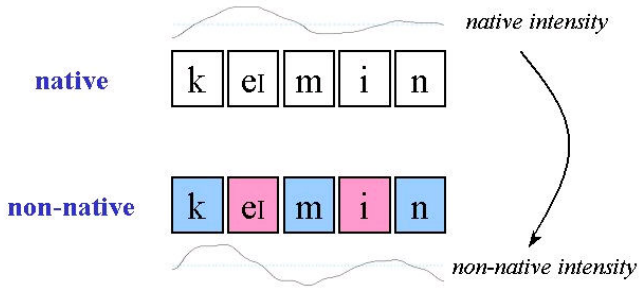


Note that the durational adjustment was performed uniformly, which means that no sub-segmental consideration was made. If, for example, the formant transition at the beginning of the vowel [eɪ] in the native sample phrase was longer than that contained in the non-native sample phrase before the duration manipulation, the formant transition in the non-native

utterance after the manipulation will be much shorter because of the uniform shortening. One way to get around this problem would be to fine-tune the alignment process. If the alignment were done by sub-segment, e.g. the formant transition versus the steady-state part of a vowel, or the gap versus the burst part of a stop, the performance should be improved.

In the second step, the fundamental frequency (F0) contour of the native version was super-imposed on the non-native version (See Figure 2). This was done in Praat by replacing the F0 contour of the duration-treated non-native version with that of the native version. This was possible, because in the previous step, the durations of the native and non-native version were adjusted. This step was based on the assumption that the duration manipulation of step 1 was perfect. Therefore, it is possible that the relative position of the F0 peak in the vowels of the non-native utterance may be slightly different.

<Figure 3> Illustration of step 3: native intensity contour imposition. After the adjustments of the segmental durations and the F0 contour, the non-native intensity contour is replaced with the native intensity contour.



In the third and final step, the intensity contour of the native version was imposed on the non-native version (See Figure 3). In Praat, this was done by mathematically "neutralizing" the intensity contour of the non-native version and importing the intensity contour of the native version. As pointed out for the second step of F0 manipulation, this step also depends on how well the segments have been aligned during the previous step. If one were to proceed from step 1 all the way through 3, the process will have replaced all the prosodic features of the non-native utterance with those of the native utterance. If you stop after the second step, you will have replaced the durations and the F0 contour only.

1. Selective cloning of prosodic features

It is also possible to selectively clone some of the prosodic features of the native speaker's utterance. There are a number of possible ways to do this. Of the three prosodic features, i.e. the segmental durations, F0 contours and intensity contour, it is possible to clone only one feature in isolation from the native utterance onto the non-native utterance. Conversely, however, two prosodic features can also be simultaneously applied. Depending on the purposes and parameters of the work, different sets of features can be manipulated.

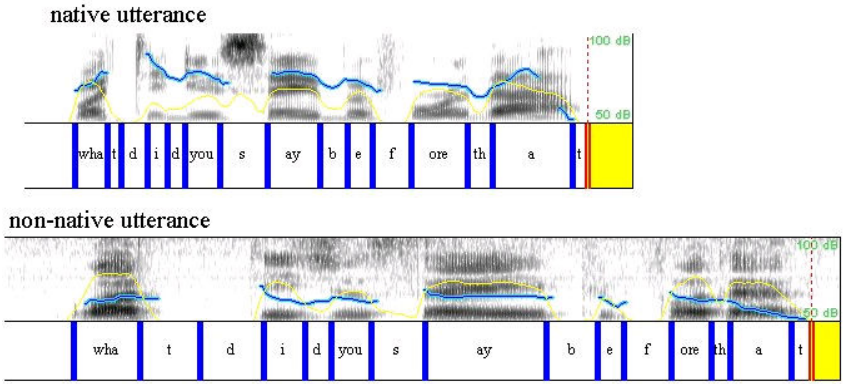
If the only prosodic feature that needs to be copied from the native utterance is its segmental duration proportions, then step 1 of the Method section is sufficient. However, if either the F0 contour or the intensity contour need to be imitated, an additional step will be necessary. Recall that in step 1 the segmental durations of the non-native utterance were adjusted with respect to those of the native utterance (Figure 1). This may be called for operational purposes, the "normal order". In the additional step, the reverse order of step 1 needs to be performed, i.e. the durations of the native utterance need to be adjusted with respect to those of the non-native utterance. For example, the [k] segment of the native utterance in Figure 1 will have to be shortened.

During the process, additional frames are added to the original sound signal or excess frames are deleted from the original sound signal (Mouline & Charpentier, 1990), which also affects the F0 and intensity contour of the native utterance. After this additional step, the new version of either the F0 contour or the intensity contour of the native utterance can be imposed on the non-native utterance. As the procedure involves an additional step which modifies the original native utterance, the resulting non-native utterance cannot be said to contain the original F0 and intensity contour in the strictest sense.

If two of the prosodic features need to be copied, either the normal or reverse durational manipulation can be combined with either the F0 or the intensity contour manipulation. For example, if it is the segmental durations and the F0 contour of the native utterance that need to be imposed on the non-native utterance, then the normal durational manipulation as shown in Figure 1 can be applied with a subsequent F0 contour replacement. If, however, it is the segmental durations and the intensity contour of the native utterance which needs to be imposed, then the subsequent replacement can be done with the intensity contour of the native utterance. The super-imposition of the F0 and intensity contour of the native utterance onto the non-native utterance can begin with a reverse durational manipulation, followed by the transfer of the F0 and intensity contour of the native utterance onto the non-native utterance.

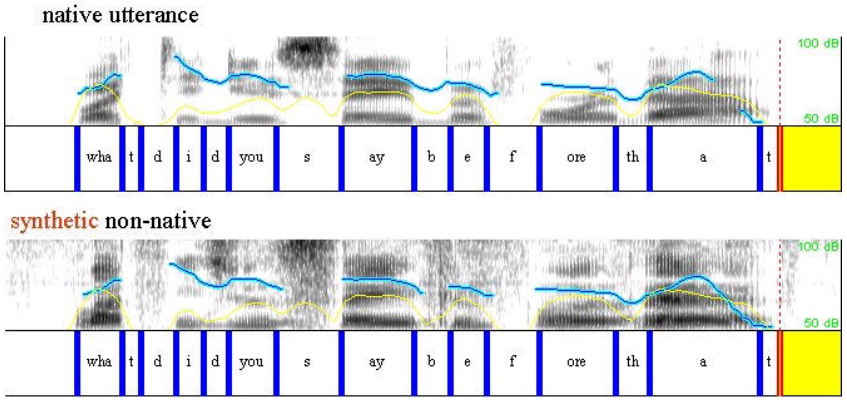
III. Results

<Figure 4> A spectrographic comparison of the native and non-native utterances before the application of the technique. The target sentence was "What did you say before that?". The thicker line represents the F0 contour, and the thinner line represents the intensity contour.



A spectrographic comparison of the native versus the non-native utterance before and after the application of the technique is shown in Figures 4 and 5. The technique of super-imposing all of the prosodic features was employed for the target utterance: "What did you say before that?". As shown in Figure 4, the non-native utterance is different from its native counterpart in terms of every aspect of the prosodic features. Although both speakers were male, the native speaker was generally higher in its F0 contour.

<Figure 5> A spectrographic comparison of the native and non-native utterances after the application of the technique. The thicker line represents the F0 contour and the thinner line represents the intensity contour.

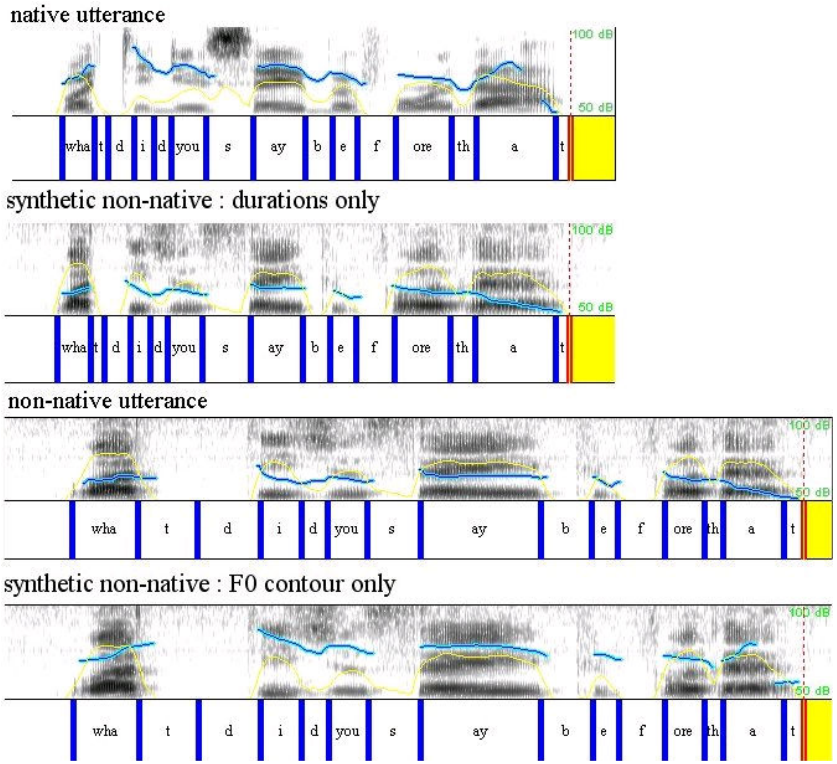


However, after the application of the technique (See Figure 5), the two utterances became almost identical in all aspects of their respective prosodic features. The durations of the matching segments were the same, although, as pointed out earlier, their precision depended on the accuracy of the segment alignment process. The F0 and intensity contour appeared to be almost identical, although slight sub-segmental variations may have been present. Differences in segmental quality were observed, for example, in the [s] segment. This may have been partly due to the difference in the formant characteristic of the [s] segment of the non-native speaker, and/or partly due to the weakness of the PSOLA algorithm itself (Mouline & Charpentier, 1990).

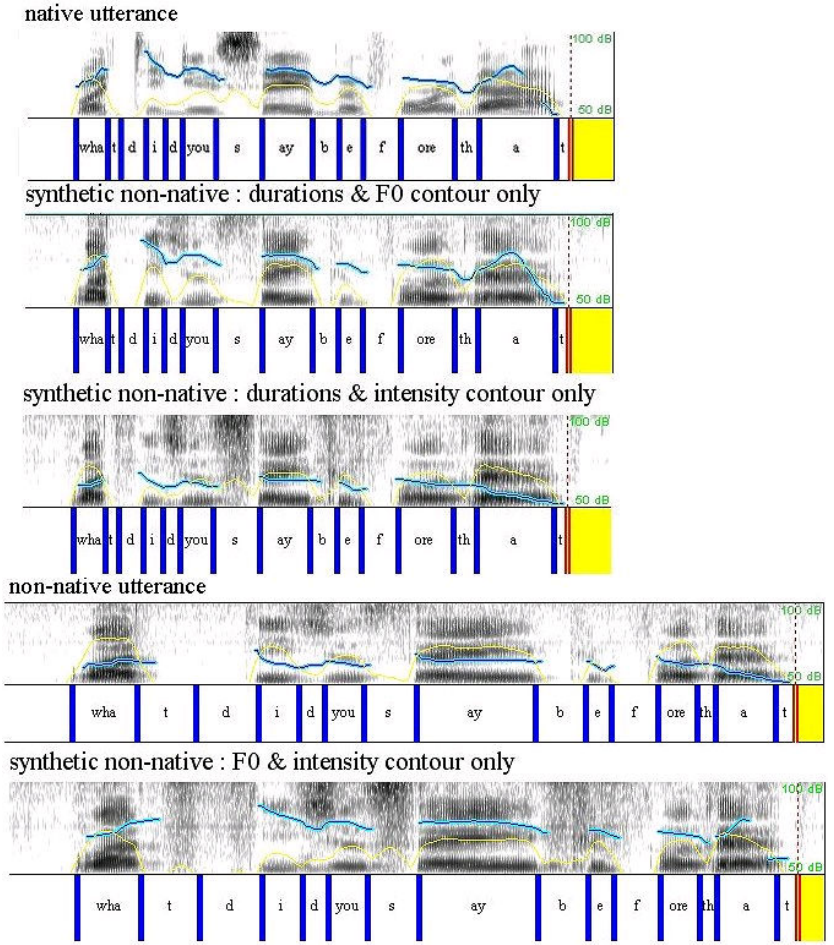
Figures 6 and 7 show the result of manipulating only one of the prosodic features, i.e. either the durations or the F0 contour, or two of the prosodic features, i.e. the durations and F0 contour, the durations and

intensity contour, or the F0 and intensity contour. The “durations only” panel in Figure 6 shows the native speaker’s durations, but the relatively flat, although “compressed” by the PSOLA algorithm, intonation contour of its longer original version (Lower panel in Figure 4) was observed. In the “F0 contour only” panel of Figure 6, the F0 contour, although “extended” by the PSOLA algorithm, closely resembled that of the native speaker (Upper panel in Figure 5).

<Figure 6> A spectrographic comparison of the non-native utterances after imposition of either the durations, or the F0 contour of the native utterance. The thicker line and the thinner line represent the F0 contour and the intensity contour respectively. The original native and non-native utterance were given for comparison.



<Figure 7> A spectrographic comparison of the non-native utterances after imposition of two of the prosodic features of the native utterance. The thicker line represents the F0 contour and the thinner line represents the intensity contour. The original native and non-native utterance were given for comparison.



The “durations & F0 contour only” panel in Figure 7 appears very similar to the synthetic non-native utterance in Figure 5. However, a close examination of its intensity contour shows that it is not the same as that of the original native utterance. For example, in the original native

utterance, the intensity contour for “you” is higher than that for “did”. The pattern is reversed in the synthetic non-native utterance. When the intensity contour of the native utterance is imposed onto the non-native utterance, the pattern is observed as shown in the “durations & intensity contour only” panel.

The results showed that it is possible to manipulate some, or all of the prosodic features involved in this study. The manipulation started with the adjustment of the segmental durations, followed by the swapping of both the F0 and intensity contour or either of the two features. Despite some degradations in the quality of the synthesized utterances, and discrepancies in sub-segmental alignment, the technique appears to be potentially useful for various purposes in many areas.

IV. Conclusion

This paper has presented a technique for cloning the prosodic features of a native speaker's utterance onto the same sentence uttered by a non-native speaker. The technique of cloning some of the prosodic features selectively was also presented. The spectrographic comparison of the synthetic utterances shows that this technique can be a useful tool for a variety of purposes.

In terms of the second language education, this technique can be used to give higher quality audio feedback to learners. By having the learners listen to their own voice with the prosodic features of the targeted example of a native speaker, an educator could motivate them from an entirely new perspective. This new technology could be integrated as an additional audio feedback source into existing language education software which provides visual feedback in the form of F0 contour matching. In order for this to work, segmental alignment should be done automatically with the help of accurate automatic speech recognition technology.

As illustrated above, the selective application of prosodic features can give different levels of audio feedback. For example, the learner could be given back his or her utterance with either the native speaker's F0 contour alone, or alternately, with the segmental durations alone. This could make learners more aware of each prosodic feature being manipulated. By having the learners pay more attention to a particular aspect of the prosodic features of the target language, learners may be able to acquire the prosody of the target language with increased efficiency.

This technique could also be used for correcting the pronunciation of patients with a vocal disorder. Given a target utterance made by a normal speaker, the patient could be motivated by listening to his or her pronunciation with all the normal prosodic features of a target speaker.

Since the essence of this technique is swapping prosodic features between speakers, it could also be used in relevant perception experiments. With the help of a sophisticated automatic speech aligner or an automatic speech recognition system, the whole process of the technique can be automatized and incorporated into the existing second language education software.

References

언어과학(주). Dr. Speaking . CD-ROM. URL http://www.eoneo.co.kr. Version 1.02. 2002

Abberton, E. & A. Fourcin. "Visual feedback and acquisition of intonation". In E. H. Lenneberg and E. Lenneberg, editors, *Foundations of Language Development (2nd ed.)*, 157-165. Academic Press, 1975.

Boersma, P. “Praat, a system for doing phonetics by computer”, *Glot International*, 5(9/10)(2005): 341-345.

Ladefoged, P. *A Course in Phonetics*. Thomson Wadsworth, 2006.

Moulines, E. & F. Charpentier. "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones". *Speech Communication*, 9:n(1990): 5-6.

Stibbard, R. "Teaching English intonation with a visual display of fundamental frequency". *The Internet TESOL Journal*, 2(8), 1996.

Appendix (Praat Script)

```
################################################################################
# prosody-cloning.praat (Written by Kyuchul Yoon, kyoon@kyungnam.ac.kr)
# Given two utterances with corresponding textgrids (aligned by the segment),
# this script copies the durations and/or the F0 and/or intensity contour
# from the native utterance, using the PSOLA algorithm. The two textgrids
# should have the same number of intervals, otherwise, an error message pops up.
# The section below illustrates how to clone one or two or all three prosodic
# parameters from the native utterance to its corresponding nonnative utterance.
# (D for duration, F for F0 contour, I for intensity contour)
# List of output files:           D-nonnative.wav, F-nonnative.wav, I-nonnative.wav,
#                                 DF-nonnative.wav, DI-nonnative.wav, FI-nonnative.wav
#                                 DFI-nonnative.wav
################################################################################
# (1) native D                    PSOLA, modify nonnative D wrt/ native D
# (2) native F          (a)       PSOLA, modify native D wrt/ nonnative D
#                       (b)       PSOLA, copy native F to nonnative utterance
# (3) native I          (a)=(2a)  PSOLA, modify native D wrt/ nonnative D
#                       (b)       NON-PSOLA,copy native I to nonnative utterance
# (4) native D + F      (a)=(1)   PSOLA, modify nonnative D wrt/ native D
#                       (b)=(2b)  PSOLA, copy native F to nonnative utterance
# (5) native D + I      (a)=(1)   PSOLA, modify nonnative D wrt/ native D
#                       (b)=(3b)  NON-PSOLA, copy native I to nonnative utterance
# (6) native F + I      (a)=(2a)  PSOLA, modify native D wrt/ nonnative D
#                       (b)=(2b)  PSOLA, copy native F to nonnative utterance
#                       (c)=(3b)  NON-PSOLA, copy native I to nonnative utterance
# (7) native D + F + I  (a)=(1)   PSOLA, modify nonnative D wrt/ native D
#                       (b)=(2b)  PSOLA, copy native F to nonnative utterance
#                       (c)=(3b)  NON-PSOLA, copy native I to nonnative utterance
################################################################################
form Specify files and folders
  comment NATIVE utterance
  word natFolder native
  word natSound_(with_dot_wav) native.wav
  word natTextgrid_(with_dot_textgrid) native.TextGrid
  natural natTierOfSegment 1
  comment NONNATIVE utterance
  word nonnatFolder nonnative
  word nonnatSound_(with_dot_wav) nonnative.wav
  word nonnatTextgrid_(with_dot_textgrid) nonnative.TextGrid
  natural nonnatTierOfSegment 1
  comment New synthetic NONNATIVE utterances will be created here
  word outFolder_(to_be_created) clonedUtterances
endform
#######################################
# CREATE FOLDER AND READ TEXTGRID FILES
#######################################
system_nocheck mkdir 'outFolder$'
Read from file... 'natFolder$'/'natSound$'
Rename... natSoundObj
durNatSound = Get total duration
Read from file... 'natFolder$'/'natTextgrid$'
Rename... natTextgridObj
natNumIntervals = Get number of intervals... natTierOfSegment
Read from file... 'nonnatFolder$'/'nonnatSound$'
Rename... nonnatSoundObj
durNonnatSound = Get total duration
Read from file... 'nonnatFolder$'/'nonnatTextgrid$'
```

```
Rename... nonnatTextgridObj
nonnatNumIntervals = Get number of intervals... nonnatTierOfSegment
################################################
# CHECK IF THE NUMBER OF INTERVALS ARE THE SAME
################################################
if (natNumIntervals <> nonnatNumIntervals)
  exit WARNING! The number of intervals of the two textgrids are not the same!
endif
################################################
# GET MANIPULATION OBJECTS FOR LATER DURATION SWAP
################################################
select Sound natSoundObj
To Manipulation... 0.01 60 400
#noprogress To Manipulation... 0.01 60 400
Rename... natSoundManipObj
select Sound nonnatSoundObj
To Manipulation... 0.01 60 400
# noprogress To Manipulation... 0.01 60 400
Rename... nonnatSoundManipObj
################################################
# GET PITCH TIER OF NATURAL FOR LATER PITCH SWAP
################################################
select Manipulation natSoundManipObj
Extract pitch tier
Rename... natPitchObj
############################################################
# CREATE AN EMPTY DURATION TIER OBJECT FOR TWO UTTERANCES
############################################################
Create DurationTier... natDurTier 0 durNatSound
Create DurationTier... nonnatDurTier 0 durNonnatSound
####################################################
# GET DURATION RATIO OF EACH NONNATIVE INTERVAL WRT/
# CORRESPONDING NATIVE INTERVAL AND ADD POINTS TO
# NONNATIVE DURATION TIER OBJECT
####################################################
nonnatStartTime = 0
natStartTime = 0
for iInterval to nonnatNumIntervals
  select TextGrid nonnatTextgridObj
  nonnatEndTime = Get end point... nonnatTierOfSegment iInterval
  nonnatIntervalDur = nonnatEndTime - nonnatStartTime
  storeNonnatStartTime = nonnatStartTime + 0.00000000001
  nonnatStartTime = nonnatEndTime
  # Get the ratio
  select TextGrid natTextgridObj
  natEndTime = Get end point... natTierOfSegment iInterval
  natIntervalDur = natEndTime - natStartTime
  natStartTime = natEndTime
  ratioOfInterval = natIntervalDur / nonnatIntervalDur
  # Insert a point to the duration tier
  select DurationTier nonnatDurTier
  Add point... storeNonnatStartTime ratioOfInterval
  Add point... nonnatEndTime ratioOfInterval
endfor
####################################################
# GET DURATION RATIO OF EACH NATIVE INTERVAL WRT/
# CORRESPONDING NONNATIVE INTERVAL AND ADD POINTS TO
# NATIVE DURATION TIER OBJECT
####################################################
nonnatStartTime = 0
natStartTime = 0
for iInterval to nonnatNumIntervals
  select TextGrid natTextgridObj
  natEndTime = Get end point... natTierOfSegment iInterval
  natIntervalDur = natEndTime - natStartTime
  storeNatStartTime = natStartTime + 0.00000000001
  natStartTime = natEndTime
  # Get the ratio
  select TextGrid nonnatTextgridObj
  nonnatEndTime = Get end point... nonnatTierOfSegment iInterval
  nonnatIntervalDur = nonnatEndTime - nonnatStartTime
  nonnatStartTime = nonnatEndTime
  ratioOfInterval = nonnatIntervalDur / natIntervalDur
  # Insert a point to the duration tier
  select DurationTier natDurTier
  Add point... storeNatStartTime ratioOfInterval
  Add point... natEndTime ratioOfInterval
endfor
#################################
# CLONE NATIVE DURATIONS ONLY (1)
#################################
select Manipulation nonnatSoundManipObj
# Store the original nonnatSoundManipObj
Copy... copyNonnatSoundManipObj
plus DurationTier nonnatDurTier
Replace duration tier
select Manipulation copyNonnatSoundManipObj
Get resynthesis (PSOLA)
# Then we get a new synthetic nonnative sound (1): D only from native
Rename... synNonnatSoundObjD
# Save the sound file in the output folder
synSoundD$ = "D-" + nonnatSound$
Write to WAV file... 'outFolder$'/'synSoundD$'
# The manipulation object can be used later on: D only from native
To Manipulation... 0.01 60 400
# noprogress To Manipulation... 0.01 60 400
Rename... synNonnatManipObjD
############################################################
# CLONE NONNATIVE DURATIONS ONLY (2a/3a) AND
# GET PITCH TIER OF MODIFIED NATURAL FOR LATER PITCH SWAP
############################################################
select DurationTier natDurTier
plus Manipulation natSoundManipObj
Replace duration tier
select Manipulation natSoundManipObj
Get resynthesis (PSOLA)
# Then we get a new synthetic native sound (2a): D only from nonnative
Rename... synNatSoundObjD
# The manipulation object can be used later on: D only from nonnative
To Manipulation... 0.01 60 400
# noprogress To Manipulation... 0.01 60 400
Rename... synNatManipObjD
Extract pitch tier
Rename... synNatPitchObjD
#############################
# CLONE F0 CONTOUR ONLY (2b)
#############################
synSoundF$ = "F-" + nonnatSound$
```

```
select Manipulation nonnatSoundManipObj
plus PitchTier synNatPitchObjD
Replace pitch tier
select Manipulation nonnatSoundManipObj
Get resynthesis (PSOLA)
# Then we get a new synthetic nonnative sound (2b): F only from native
Rename... synNonnatSoundObjF
Write to WAV file... 'outFolder$'/'synSoundF$'
####################################
# CLONE INTENSITY CONTOUR ONLY (3b)
####################################
synSoundI$ = "I-" + nonnatSound$
# Intensity object of the duration-copied native utterance (3a)
select Sound synNatSoundObjD
# Get the intensity value in dB
synNatSoundObjDIntensityValue = Get intensity (dB)
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... synNatSoundDIntensityObj
Down to IntensityTier
Rename... synNatSoundDIntensityTierObj
# Intensity object of the original nonnative utterance
select Sound nonnatSoundObj
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... nonnatSoundIntensityObj
# Inverse the intensity object by getting the maximum and subtracting self
maxNonnat = Get maximum... 0 0 Parabolic
Formula... 'maxNonnat' - self
# And make IntensityTier object
Down to IntensityTier
Rename... nonnatSoundIntensityTierObj
# Multiply the nonnative utterance with its own inverse IntensityTier and then
# by the IntensityTier of the synthetic natural D utterance
select Sound nonnatSoundObj
plus IntensityTier nonnatSoundIntensityTierObj
Multiply
Rename... nonnatSoundInverseObj
plus IntensityTier synNatSoundDIntensityTierObj
Multiply
# Before writing, adjust the average intensity value in dB
Scale intensity... 'synNatSoundObjDIntensityValue'
# Another new synthetic nonnative sound (3b): I only from native
Write to WAV file... 'outFolder$'/'synSoundI$'
################################################
# CLONE DURATIONS + F0 CONTOUR (4a+4b)=(1+2b)
################################################
synSoundDF$ = "DF-" + nonnatSound$
select Manipulation synNonnatManipObjD
plus PitchTier natPitchObj
Replace pitch tier
select Manipulation synNonnatManipObjD
Get resynthesis (PSOLA)
Rename... synNonnatSoundObjDF
Write to WAV file... 'outFolder$'/'synSoundDF$'
#####################################################
# CLONE DURATIONS + INTENSITY CONTOUR (5a+5b)=(1+3b)
#####################################################
synSoundDI$ = "DI-" + nonnatSound$
# Intensity object of the original natural utterance
select Sound natSoundObj
# Get the intensity value in dB
natIntensityValue = Get intensity (dB)
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... natIntensityObj
Down to IntensityTier
Rename... natIntensityTierObj
# Intensity object of the duration-copied nonnative utterance
select Sound synNonnatSoundObjD
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... synNonnatSoundIntensityObj
# Inverse the intensity object by getting the maximum and subtracting self
maxNonnat = Get maximum... 0 0 Parabolic
Formula... 'maxNonnat' - self
# And make IntensityTier object
Down to IntensityTier
Rename... synNonnatSoundIntensityTierObj
# Multiply the nonnative utterance with its own inverse IntensityTier and then
# by the IntensityTier of the natural utterance
select Sound synNonnatSoundObjD
plus IntensityTier synNonnatSoundIntensityTierObj
Multiply
Rename... synNonnatSoundInverseObj
plus IntensityTier natIntensityTierObj
Multiply
# Before writing, adjust the average intensity value in dB
Scale intensity... 'natIntensityValue'
Write to WAV file... 'outFolder$'/'synSoundDI$'
#############################################################
# CLONE F0 CONTOUR + INTENSITY CONTOUR (6a+6b+6c)=(2a+2b+3b)
#############################################################
synSoundFI$ = "FI-" + nonnatSound$
# Intensity object of the duration-copied native utterance
select Sound synNatSoundObjD
# Get the intensity value in dB
synNatIntensityValue = Get intensity (dB)
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... synNatIntensityObj
Down to IntensityTier
Rename... synNatIntensityTierObj
# Intensity object of the F0-copied nonnative utterance
select Sound synNonnatSoundObjF
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... synNonnatSoundIntensityObj
# Inverse the intensity object by getting the maximum and subtracting self
maxNonnat = Get maximum... 0 0 Parabolic
Formula... 'maxNonnat' - self
# And make IntensityTier object
Down to IntensityTier
Rename... synNonnatSoundIntensityTierObj
# Multiply the nonnative utterance with its own inverse IntensityTier and then
# by the IntensityTier of the natural utterance
select Sound synNonnatSoundObjF
plus IntensityTier synNonnatSoundIntensityTierObj
```

```
Multiply
Rename... synNonnatSoundInverseObj
plus IntensityTier synNatIntensityTierObj
Multiply
# Before writing, adjust the average intensity value in dB
Scale intensity... 'synNatIntensityValue'
Write to WAV file... 'outFolder$'/'synSoundFI$'
################################################################################
# CLONE DURATIONS + F0 CONTOUR + INTENSITY CONTOUR (7a+7b+7c)=(1+2b+3b)=(4+5)
################################################################################
synSoundDFI$ = "DFI-" + nonnatSound$
# Intensity object of the native utterance
select Sound natSoundObj
# Get the intensity value in dB
natIntensityValue = Get intensity (dB)
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... natIntensityObj
Down to IntensityTier
Rename... natIntensityTierObj
# Intensity object of the DF-copied nonnative utterance
select Sound synNonnatSoundObjDF
To Intensity... 70 0
# noprogress To Intensity... 70 0
Rename... synNonnatSoundIntensityObj
# Inverse the intensity object by getting the maximum and subtracting self
maxNonnat = Get maximum... 0 0 Parabolic
Formula... 'maxNonnat' - self
# And make IntensityTier object
Down to IntensityTier
Rename... synNonnatSoundIntensityTierObj
# Multiply the nonnative utterance with its own inverse IntensityTier and then
# by the IntensityTier of the native utterance
select Sound synNonnatSoundObjDF
plus IntensityTier synNonnatSoundIntensityTierObj
Multiply
Rename... synNonnatSoundInverseObj
plus IntensityTier natIntensityTierObj
Multiply
# Before writing, adjust the average intensity value in dB
Scale intensity... 'natIntensityValue'
Write to WAV file... 'outFolder$'/'synSoundDFI$'
select all
nocheck Remove
############### END OF SCRIPT ##################
```

윤 규철
경남 마산시 월영동 449
경남대학교 영어학부
kyoon@kyungnam.ac.kr

Received:
Revised: