

Identifying Frication and Aspiration Noise in the Frequency Domain: The Case of Korean Alveolar Lax Fricatives

Yoon, Kyuchul¹⁾

ABSTRACT

This paper introduces the technique of semi-automatically identifying different types of noise in the frequency domain. Given the lower cutoff frequency of the frication noise, and a user-specified constant, the technique identifies the boundary between the frication and aspiration noise in a Korean lax fricative followed by the vowel /a/ by comparing the upper and lower sums of energy with respect to the cutoff frequency. The user-specified constant can be adjusted for different speakers. When the technique was applied to distinguish the two types of noise of Korean lax fricatives from the same speaker, the average and standard deviation of the difference between the manually inserted boundaries and the automatically inserted boundaries were 2.67ms and 1.80ms respectively.

Keywords: Frication, aspiration, frequency domain, Korean, alveolar lax fricatives, Praat

1. Introduction

The Korean lax fricatives are known to have two different types of noise when they are followed by a vowel. When a Korean lax fricative is followed by a low vowel such as /a/, the frication noise is usually followed by a longer aspiration noise [2]. As <Figure 1> shows, this relatively long aspiration noise is one of the features that distinguish the lax version of Korean alveolar fricatives from the tense version in a low vowel environment.

The boundary between the two types of noise is usually determined by hand. This of course introduces errors. As suggested in [2], the manual work can be improved by having a computer do the dividing on the basis of energy distribution in the frequency domain of the spectrogram. As can be seen in <Figure 1>, the energy distribution in the frequency domain of

the fricative segment is unique in that the frication portion of the fricative has energy concentrated in the upper frequency region compared to the energy of the aspiration portion.

It is possible to think that a trained phonetician manually draws the line between the two types of noise based on this energy pattern distributed in the frequency domain over time. A human annotator may depend on the change of visual patterns. However, in order to make a computer do the same thing on the spectrogram, it may be necessary to make it aware of the change of energy distribution in the frequency domain over time. The energy distribution at a particular frame in the spectrogram can be seen as the sums of energy across a certain reference frequency. For the Korean lax alveolar fricatives, the low cutoff frequency of the frication portion could serve as the reference frequency.

The frame-by-frame change of energy distribution pattern can be checked to identify the frame where there is the transition from the frication to the aspiration noise. That frame can be said to be the boundary between the two types of noise. This algorithm was used in [3]. He used a Praat [1] script to implement the algorithm and used it to measure the duration of two types of noise semi-automatically. The algorithm appears to have worked fine in his work. However, the algorithm was never

1) Yeungnam University kyoony@ynu.ac.kr
(This research was supported by the Yeungnam University Research Grants in 2008 (208-A-235-095))

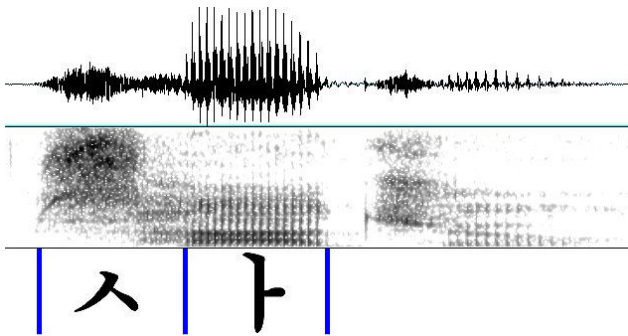


Figure 1. The Korean syllable /sa/ consists of a lax alveolar fricative /s/ and a vowel /a/. Each segment is labeled with a Korean hangul symbol.

tested thoroughly in a separate experiment.

According to [3], the script needs two user-defined variables for each speaker. One is the reference frequency. The script adds the energy above and below this frequency for each frame along the time axis, comparing the energy difference between the upper sum and the lower sum of energy. When the difference of energy reaches a certain point, i.e. the criterion value, the script determines the frame as the boundary between the two types of noise. The user needs to specify this variable for each speaker.

Since the user cannot be expected to know the exact criterion value for a particular speaker before running the script, she needs to run the script just one time to identify the value. This is possible because the script, after the first execution, shows the user the list of candidate criterion values along the time axis. The user just needs to pick up the appropriate criterion value that corresponds to the time point where a human annotator would have inserted a boundary manually between the two types of noise. Once this value is set, it can be used for all the lax fricatives produced by that particular speaker. The frame advances every 5 msec, which is also user-settable. The frame advance can then be said to determine the precision of the annotation done by the script.

An important implicit assumption that was made in [3] was that once the reference frequency is set, the criterion value is the same for the same speaker. He did his work with this assumption, but never verified it using an experiment. If this assumption had been wrong, his findings would have been flawed. One of the goals of this paper is to verify his assumption by comparing the manual annotation for a particular speaker with the automatic annotation done by a script. If the script performs as well as a phonetically trained annotator for one speaker, we can conclude that the assumption holds true and that the semi-automatic annotation using a script can replace a human annotator.

2. Methods

2.1 Algorithm

The algorithm proceeds as follows. Given the beginnings of the lax fricative and the following vowel, which can be regarded as the target area for scanning the energy pattern, the script extracts a windowed selection of the target area at a specified interval. The reference frequency can be set to the low cutoff frequency of the frication portion. In <Figure 2>, it was set to 3800Hz. The criterion value can be set heuristically. Here it was set to 300. The value will be adjusted later after the first execution of the script. Thus it can be set arbitrarily at this stage.

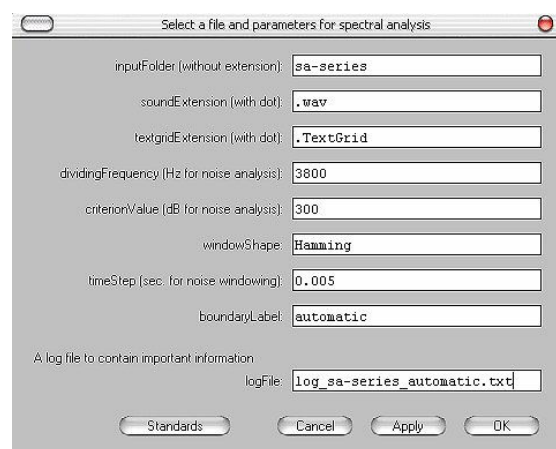


Figure 2. The pop-up dialog box after the execution of the script. The reference frequency (*dividingFrequency*) was set to 3800 and the criterion value (*criterionValue*) to 300.

The type of window (Hamming by default) and the time step (5 msec by default) can be specified by the user. In other words, the scanning proceeds every 5 msec and each frame is also 5 msec long. Each frame is then converted into a long-term average spectrum. With respect to the reference frequency, the sums of energy above and below it, called the sum of high band energy and the sum of low band energy respectively, are calculated and their subtracted values are logged in a separate text file. The script checks the series of this logged criterion values and inserts a boundary when the value falls below the user-specified criterion value (See <Figure 3>). The boundary label can also be specified by the user.

The candidate criterion values along with their time stamps are displayed in a pop-up window after the first execution of the script (See <Figure 4>). At the bottom of this pop-up window, the user-specified criterion value of 300 is shown with the time stamp 0.8347 of the wrong boundary inserted by the script during

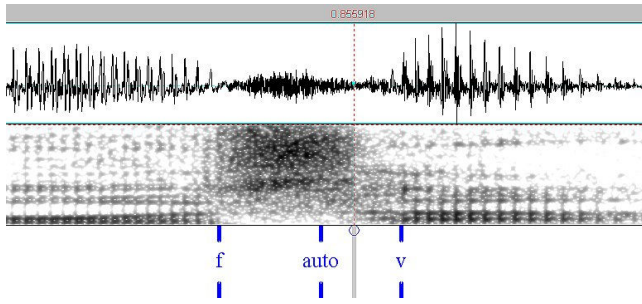


Figure 3. The script automatically inserted a wrong boundary (labeled *auto*) in between the fricative onset (labeled *f*) and the vowel onset (labeled *v*) after the first execution. The dotted vertical line is where a phonetically trained annotator would have inserted the correct boundary.

the first execution. As shown in <Figure 3>, a phonetically trained annotator would have inserted the correct boundary in a different frame. The highlighted frame in <Figure 4> represents this frame and its criterion value, i.e. sum of high band energy - sum of low band energy, is around -460. By making a note of the cursor location of the dotted vertical line in <Figure 3>, one can identify the “correct” criterion value for that particular speaker in the pop-up window. Since the correct criterion value is set, the user can use it for the rest of the files in the specified folder. <Figure 5> shows a correct boundary inserted automatically by the same script (See <Figures 3> and <Figure 5> for comparison).

| time_sec | H_band_E_Sum | L_band_E_Sum | criterionValue |
|----------|--------------|--------------|----------------|
| 0.7697 | -34 | 164 | -197.83 |
| 0.7747 | 73 | 141 | -68.01 |
| 0.7797 | 225 | 166 | 59.17 |
| 0.7847 | 325 | 112 | 212.44 |
| 0.7897 | 506 | 150 | 356.48 |
| 0.7947 | 531 | 130 | 400.84 |
| 0.7997 | 509 | 59 | 449.37 |
| 0.8047 | 671 | 142 | 528.97 |
| 0.8097 | 622 | 92 | 529.46 |
| 0.8147 | 591 | 112 | 478.66 |
| 0.8197 | 535 | 168 | 366.64 |
| 0.8247 | 567 | 153 | 413.58 |
| 0.8297 | 438 | 70 | 368.60 |
| 0.8347 | 400 | 110 | 289.19 |
| 0.8397 | 388 | 188 | 199.91 |
| 0.8447 | 225 | 243 | -17.08 |
| 0.8497 | 222 | 235 | -12.65 |
| 0.8547 | -240 | 321 | -460.40 |
| 0.8597 | -327 | 190 | -517.59 |
| 0.8647 | -448 | 271 | -718.68 |
| 0.8697 | -443 | 241 | -684.22 |
| 0.8747 | -511 | 291 | -802.41 |
| 0.8797 | -526 | 230 | -755.57 |
| 0.8847 | -254 | 439 | -693.14 |

criterionValue 300 at 0.8347 seconds

Figure 4. The pop-up Info window after the first execution of the script. The correct criterion value can be found by checking the time steps in the first column. User can identify the time where she would have manually inserted the correct boundary, and use the criterion value in the same row.

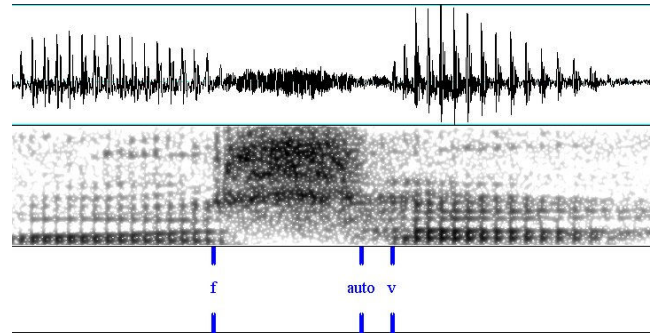


Figure 5. With the adjusted criterion value, the script can now correctly insert the boundary (labeled *auto*) between the fricative onset (labeled *f*) and the vowel onset (labeled *v*).

2.2 Experiment

Once the reference frequency and the criterion value are set, we can test the assumption that each speaker has a single criterion value. We can do this by comparing the same set of files annotated manually by a trained phonetician and automatically by the algorithm implemented in a Praat script.

The experiments were performed in two parts: the repeated series and the non-repeated series. In the first repeated series experiment, a trained phonetician (the author) manually identified the boundary between the two types of noise in three Korean words containing the lax alveolar fricatives. The three words, numbered 18, 24, and 43 and marked with an asterisk in <Table 1>, were repeated by a speaker (the author) nine, twelve and ten times respectively. The recording was done in a quiet room using a Shure SM10A head-worn microphone and the recording function of Praat. The sampling rate was 22050 Hz.

Table 1 The list of words used in the experiments. The words marked with an asterisk was also used in the repeated series experiment. The numbers following the asterisk represent the number of repetition during the recording. The lower panel represents its romanized version.

| Word-initial /s/, no coda | Word-initial /s/, with coda | Word-medial /s/, no coda |
|---------------------------|-----------------------------|--------------------------|
| 01. 사강 | 13. 사부 | 24. 사강 (*12) |
| 02. 사강 | 14. 사상 | 25. 사제 |
| 03. 사견 | 15. 사심 | 26. 사중 |
| 04. 사견 | 16. 사염 | 27. 산길 |
| 05. 사교 | 17. 사사 | 28. 산맹 |
| 06. 사과 | 18. 사주 (*9) | 29. 산불 |
| 07. 사내 | 19. 사찰 | 30. 산소 |
| 08. 사다 | 20. 사방 | 31. 살뎡 |
| 09. 사돈 | 21. 사외 | 32. 살딱 |
| 10. 사람 | 22. 사표 | 33. 삼베 |
| 11. 사병 | 23. 사회 | 34. 삼성 |
| 12. 사법 | | 35. 삼중 |
| | | 40. 아사 |
| | | 41. 제사 |
| | | 42. 지사 |
| | | 43. 고사 (*10) |
| | | 44. 주사 |
| | | 45. 지사 |
| | | 46. 이사 |
| | | 47. 어사 |
| | | 48. 유사 |
| | | 49. 풍사 |
| | | 50. 순사 |
| | | 51. 린사 |

| Word-initial /s/, no coda | Word-initial /s/, with coda | Word-medial /s/, no coda |
|---------------------------|-----------------------------|--------------------------|
| 01. sagag | 13. sabu | 24. saggam (*12) |
| 02. sagam | 14. sasang | 25. sagje |
| 03. saggeon | 15. sasim | 26. saggung |
| 04. sagyeon | 16. saeob | 27. sangil |
| 05. sago | 17. saja | 28. sanmaeg |
| 06. sagwa | 18. saju (*9) | 29. sanbul |
| 07. sae | 19. sachal | 30. sanso |
| 08. sada | 20. satang | 31. sallim |
| 09. sardon | 21. satwe | 32. salljag |
| 10. saram | 22. sapyo | 33. sambe |
| 11. sabyeong | 23. sahwe | 34. samseong |
| 12. sabeob | | 35. samjung |
| | | 40. asa |
| | | 41. jesa |
| | | 42. sisa |
| | | 43. gosa |
| | | 44. jusa |
| | | 45. seosa |
| | | 46. isa |
| | | 47. eosa |
| | | 48. jusa |
| | | 49. songsa |
| | | 50. sunsa |
| | | 51. seonsa |

The thirty-one repeated tokens were manually annotated by the phonetician for the boundary between the frication and aspiration noise in the fricative segment. The same set of tokens were also automatically annotated by the Praat script as explained in the algorithm section. The accuracy of the annotations between the human phonetician and the Praat script was examined by measuring the differences in the location of the boundaries.

In the second experiment, the non-repeated series of words was used to compare the performance of the human annotators and the Praat script. The fifty-three words in Table 1 were manually labeled by two trained phoneticians (including the author). The same set of tokens were also labeled automatically by the script. The performance between the human labeler and the script was compared. In addition, the inter-annotator differences were also checked, that is, between the two phoneticians. The inter-script differences, that is, with different criterion values, were measured too.

3. Results

The histogram of differences between the manually inserted boundaries and the automatically inserted boundaries for the repeated series experiment are given in <Figure 6>. The mean and the standard deviation of the thirty-one differences were 2.19 msec and 1.63 msec respectively. Most of the differences lie below 5 msec except for one token, which was 6.4 msec. The token is given in <Figure 7>.

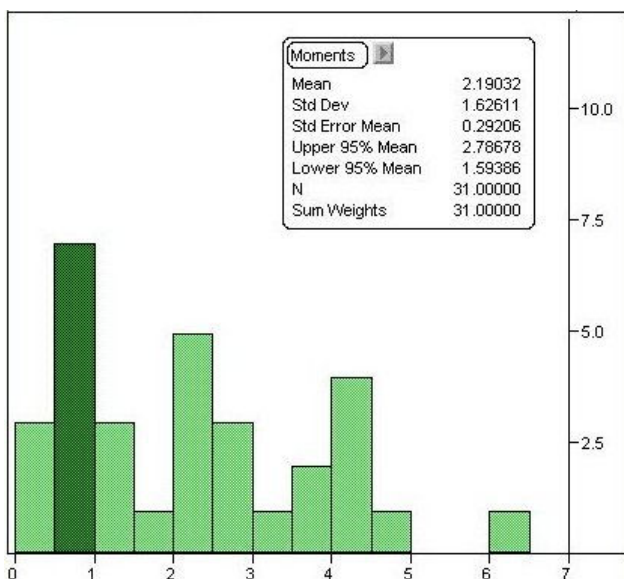


Figure 6. The histogram of differences between the manually inserted and automatically inserted boundaries for the repeated series of words. The horizontal axis represents differences in msec.

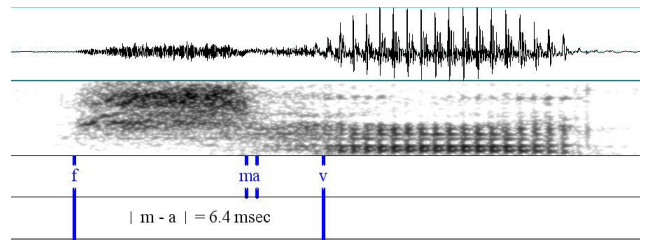


Figure 7. The word with the biggest discrepancy between the manual and the automatic boundaries in the repeated series experiment. The difference was 6.4 msec. The m and a represents manual and automatic respectively.

As can be seen from the spectrogram in <Figure 7>, the boundary inserted by the script appears to be a valid one, although different from the choice the human annotator made. If this was the worst case, doing the annotation automatically using a script can be said to be a viable approach. This also suggests that there is indeed a single criterion value for the speaker participated in the experiment. The result from the repeated series experiment appears to justify the use of this approach with non-repeated series experiment.

The histograms from the non-repeated series experiment (with fifty-three non-repeated words) are given in <Figure 8>. The differences between the boundaries inserted manually and automatically by the trained phonetician (the author) are given in the left panel of <Figure 8>. The mean and the standard deviation were 2.52 msec and 1.85 msec respectively. The right panel is from another trained phonetician. The mean and the standard deviations were 2.67 msec and 1.80 msec respectively. The histograms in <Figure 8> are not very different from that in <Figure 6>, except that the difference in the worst case is 7.8 msec for both histograms. However, the patterns we see in the histograms seems encouraging. If what the script does is not very

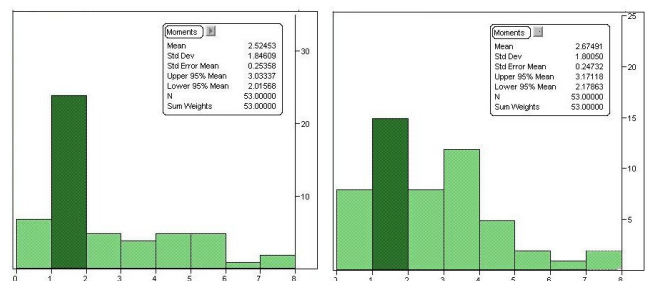


Figure 8. The histogram of differences between the manually inserted and automatically inserted boundaries for the non-repeated series of fifty-three words. The left panel is from the phonetician (the author) participated in the repeated series experiment and the right panel is from another phonetician. The horizontal axis represents differences in msec.

different from what a phonetically trained annotator does, it appears that we can safely hand over the job to the script with appropriate reference frequencies and criterion values.

The histograms in <Figure 9> represent differences between human annotators and between automated scripts with different criterion values. The histogram on the left panel shows differences between the two trained phoneticians and the differences lie mostly below 6 msec. The mean and the standard deviations were 3.24 msec and 1.81 msec respectively. The histogram on the right shows differences between the two automated scripts customized to represent the phoneticians who did the annotations. The mean and the standard deviation were 1.23 msec and 2.17 msec respectively.

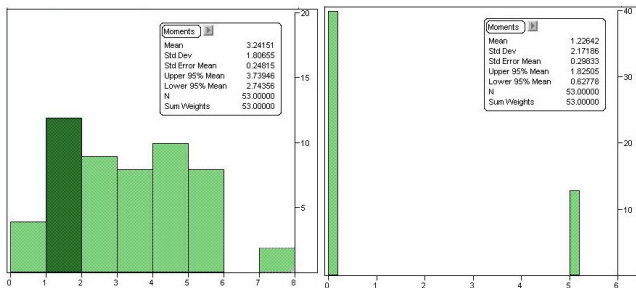


Figure 9. The histogram of differences between the two phoneticians for the non-repeated series of words (left). The panel on the right is the histogram of differences between the two scripts with different criterion values. The horizontal axis represents differences in msec.

Despite the differences in criterion values, i.e. one was -400 and the other was -300, there were no differences in the location of the boundaries for the forty words. This was expected because the scripts were executed onto the same set of stimuli and the purpose of the scripts was to do what trained phoneticians would do. In an ideal situation, the trained phoneticians would insert the boundaries at the same place, and the cases of forty words can be a reflection of this situation. For the other thirteen words, the differences were 5 msec. Note that the differences would be in multiples of 5 msec because the frame size and frame advance were set to 5 msec by default.

The means and standard deviations of the differences in msec between the manually inserted boundaries and the automatically inserted boundaries for the two experiments are summarized in <Table 2>.

Table 2. The summary of the means and the standard deviations of the differences from the two experiments. The numbers are given in msec. Phoneticians A and B represent the two human annotators. Scripts A and B represent the same script with different criterion values.

| | Experiment with repeated words | |
|---------------------------------|---|--------------------|
| | Mean | Standard deviation |
| Phonetician A vs. Script A | 2.19 | 1.63 |
| | | |
| | Experiment with non-repeated words | |
| | Mean | Standard deviation |
| Phonetician A vs. Script A | 2.52 | 1.85 |
| Phonetician B vs. Script B | 2.67 | 1.80 |
| Phonetician A vs. Phonetician B | 3.24 | 1.81 |
| Script A vs. Script B | 1.23 | 2.17 |

4. Conclusion

This paper introduced a technique of identifying different types of noise in the frequency domain and applied it to isolated words containing Korean alveolar lax fricatives produced by the same speaker. The technique was implemented in a Praat script. With a criterion value customized to each speaker, the script was able to identify the boundaries between different types of noise. As the results from the experiments show, the differences of boundaries inserted by the human annotators and the automated scripts are minimal. Even in the worst cases, the differences were less than 8 msec.

From the means and the standard deviations obtained from the experiments, it appears that the automated scripts can replace human annotators. The approach can be valuable if the annotation involves hundreds of tokens from a speech corpus. The assumption that there is a single criterion value for each speaker can be said to hold true for the Korean lax alveolar fricatives followed by the vowel /a/. The criterion value in this case might reflect the consistency of a speaker in the production of the fricatives. In the same vein, the speaker variability in the production of the same fricatives may be expected to show up in the form of different criterion values.

Although this approach proved to be useful in annotating words containing Korean lax fricatives followed by the vowel /a/, future study needs to verify the viability of this technique in the words containing other vowels produced by different speakers. <Figure 10> shows the application of this technique in the identification of the boundary between the /s/ and /h/ in the phrase Miss Henry produced by a female native speaker of

English. The successful application of this approach implies that the technique can be employed for any task involving the identification of the boundaries between fricatives with different energy patterns in the frequency domain.

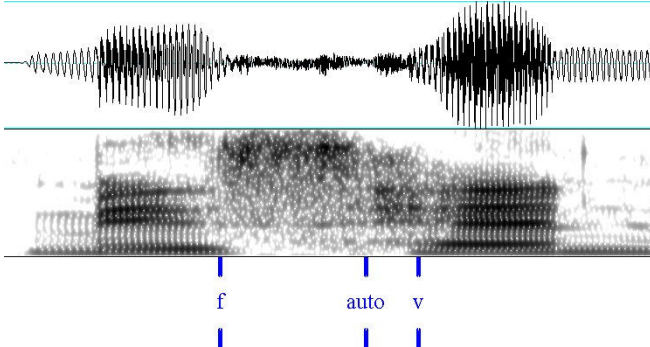


Figure 10. The automated identification of the boundary (labeled auto) between /s/ and /h/ in the phrase Miss Henry produced by a female native speaker of English. The f and v represent the beginnings of /s/ and the vowel following /h/.

References

[1] Boersma, Paul. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5(9/10). pp. 341-345.
 [2] Yoon, Kyuchul. (2002). A production and perception experiment of Korean alveolar fricatives. *Speech Sciences* 9(3). pp. 169-184.
 [3] Yoon, Kyuchul. (2005). Durational correlates of prosodic categories: The case of two Korean voiceless coronal fricatives. *Speech Sciences* 12(1). pp. 89-105.

Appendix

```
Praat script
#####
# divideNoise4-automatic.praat (Written by Kyuchul Yoon,
# kyoon@ynu.ac.kr)
#####
form Select a file and parameters for spectral analysis
word inputFolder (without extension) sa-series
word soundExtension (with dot) .wav
word textgridExtension (with dot) .TextGrid
positive dividingFrequency (Hz for noise analysis) 3800
real criterionValue (dB for noise analysis) -400
word windowShape Hamming
positive timeStep (sec. for noise windowing) 0.005
word boundaryLabel automatic
comment A log file to contain important information
word logFile log_sa-series_automatic.txt
endform
# Clears the Info window.
clearinfo
# Make a list of all sound files in the folder.
Create Strings as file list... fileListObj
'inputFolder$/*'soundExtension$'
Sort
numFiles = Get number of strings
pause 'numFiles' sound files identified. Continue?

printline
time sec'tab$'H_band_E_Sum'tab$L_band_E_Sum'tab$criterionValue
fileappend 'logFile$'
fileName'tab$'specBegin'tab$'H_band_E_Sum'tab$L_band_E_Sum
... 'tab$criterionValue'newline$'
# Loop through each file.
for iFile to numFiles
select Strings fileListObj
soundName$ = Get string... iFile
prefix$ = soundName$ - soundExtension$
textgridName$ = prefix$ + textgridExtension$
Read from file... 'inputFolder$/'soundName$'
Rename... soundObj
Read from file... 'inputFolder$/'textgridName$'
Rename... textgridObj
# Temporary matrix file for picking out aspiration location
tempOut$ = "tempOut"
filedelete 'tempOut$'
# Flag for presence of aspiration point in TextGrid
aspBegin = 0
```

```
select TextGrid textgridObj
numLabels = Get number of points... 1
for i from 1 to numLabels
label$ = Get label of point... 1 i
if label$ = "fricative"
noiseBegin = Get time of point... 1 i
elseif label$ = "boundaryLabel"
aspBegin = Get time of point... 1 i
elseif label$ = "vowelBegin"
noiseEnd = Get time of point... 1 i
endif
endfor
# If there's already an aspiration point, delete it so we can try again
if aspBegin > 0
select TextGrid textgridObj
Edit
editor TextGrid textgridObj
Move cursor to... aspBegin
Remove
Close
endeditor
endif
specNum = 0
repeat
select Sound soundObj
Edit
editor Sound soundObj
specBegin = noiseBegin + timeStep * specNum
specEnd = specBegin + timeStep
Select... specBegin specEnd
Extract windowed selection... slice 'windowShape$' 1 yes
Close
endeditor
# Band energy querying across supplied dividing frequency
To Spectrum
To Ltas (1-to-1)
numBands = Get number of bands
divBand = Get band from frequency... dividingFrequency
divBand = floor (divBand)
lowBandEnergySum = 0
for i from 1 to divBand
lowEnergy = Get value in band... i
lowBandEnergySum = lowBandEnergySum + lowEnergy
endifor
highBandEnergySum = 0
for k from (divBand + 1) to numBands
highEnergy = Get value in band... k
highBandEnergySum = highBandEnergySum + highEnergy
endifor
subtracted = highBandEnergySum - lowBandEnergySum
fileappend 'tempOut$' 'specBegin:4' 'tab$' 'subtracted:4'
'newline$'
fileappend 'logFile$' 'textgridName$' 'tab$' 'specBegin:4' 'tab$'
... 'highBandEnergySum:0' 'tab$' 'lowBandEnergySum:0' 'tab$' 'subtracted:
d:2'
... 'newline$'
# Print the same info in the Info window
printline
'specBegin:4' 'tab$' 'tab$' 'highBandEnergySum:0' 'tab$' 'tab$'
... 'lowBandEnergySum:0' 'tab$' 'tab$' 'subtracted:2'
select Sound slice
plus Ltas slice
plus Spectrum slice
Remove
specNum = specNum + 1
# Stop if windowed selection crosses over the end of noise segment.
until (specBegin > noiseEnd - timeStep)
# Read in the values from the temporary output file and determine the
point
# where criterion value has been reached.
Read Matrix from raw text file... 'tempOut$'
sliceNum = 1
skipNoise = (noiseEnd - noiseBegin) / 5 + noiseBegin
repeat
sliceNum = sliceNum + 1
sliceTime = Get value in cell... sliceNum 1
sliceRatio = Get value in cell... sliceNum 2
until ((sliceRatio <= criterionValue) and (skipNoise <= sliceTime))
select Matrix 'tempOut$'
Remove
# Write a label there.
select TextGrid textgridObj
Insert point... 1 sliceTime 'boundaryLabel$'
Write to text file... 'inputFolder$/'textgridName$'
# And in the Info window, tell the user what you did
printline
printline criterionValue 'criterionValue' at 'sliceTime' seconds
printline
fileappend 'logFile$' 'newline$' 'criterionValue' 'criterionValue'
... at 'sliceTime' seconds'newline$' 'newline$'
# Delete the temporary matrix file
filedelete 'tempOut$'
# Get time of vowelEnd for TextGrid confirmation
# noiseBegin is already defined earlier
vowelEnd = Get time of point... 1 4
select Sound soundObj
Edit
editor Sound soundObj
Select... noiseBegin-0.1 vowelEnd+0.1
Zoom to selection
Close
endeditor
plus TextGrid textgridObj
Edit
# Comment the following line to make this script run continuously
#pause Check 'textgridName$' aspiration point!
Remove
endifor
select Strings fileListObj
Remove
##### END OF SCRIPT #####
```

• **윤규철 (Yoon, Kyuchul)**
 영남대학교 영어영문학부
 경상북도 경산시 대동 214-1
 Tel: 053-810-2145 Fax: 053-810-4607
 Email: kyoon@ynu.ac.kr
 관심분야: 음성학, 음운론
 현재 영어영문학부 교수