Spectral Characteristics of the Release Bursts in Korean Alveolar Stops

Hansang Park

Faculty of Foreign Language and Literature
Daegu Haany University, Korea
phants@dhu.ac.kr

Abstract

This study investigates spectral characteristics of the release bursts in Korean alveolar stops in terms of intensity, center of gravity, and skewness of the spectra of the release burst across phonation types and speakers. The results showed that there was no significant difference in intensity, center of gravity, or skewness across phonation types but a significant difference across speakers. This means that difference in phonation type does not lead to any significant difference in the spectra of the release burst. This study suggests that the difference in the spectral characteristics of the release burst across phonation types can be ignored in speech synthesis or speech recognition.

1. Introduction

Stops are articulated as a sequence of occlusion, transient, fricative segment, aspirative segment, and the initial part of a following voiced sound.[1] In other words, in a CV sequence, the pre-release phase involves closure and release burst while the post-release one comprises frication noise, aspiration, and the earlier part of the vowel accompanied by vocal fold vibration.[1] The transient, or release burst, is a response to the pressure release of the vocal tract exclusive of the effect of air turbulence.[1] Duration of the release burst ranges 2-30 ms and is, in general, less than 10 ms.[1] More often than not, formants are not represented clearly since they change over time in a major way and are also affected by zero function caused by the following frication noise.[1]

Spectral characteristics of the release burst have been employed in the studies of place of articulation as it is a response to the pressure release of the vocal tract.[2] Bilabial stops are characterized by a flat or falling spectral tilt, alveolar ones by an increasing spectral tilt, and velar ones by a rise-and-fall spectral tilt.[2] On the other hand, place of articulation of word-initial obstruents has been studied in terms of statistical moments, such as mean, center of gravity, skewness, and kurtosis of the spectrum.[3]

Spectral characteristics of the release burst have also been employed in the studies of phonation types.[4][5] Aspirated and fortis stops have greater energy in the release burst than lenis ones in Korean.[4][5] The present study investigates spectral characteristics of the release burst in Korean alveolar stops in terms of intensity, center of gravity, and skewness, under the hypothesis that there is no significant difference in spectral moments across phonation types or speakers. The current study also provides center of gravity and skewness for the Hamming-windowed and pre-emphasized speech samples, so that the results from the Hamming-windowed and pre-emphasized speech samples can be compared to those from the raw speech samples.

2. Methods

2.1. Subjects

Three male native speakers of Standard Korean in their mid-twenties participated in the experiment. All subjects were students at the University of Texas at Austin who had been in the US for less than 6 months. None of the subjects had any history of speech disorders.

2.2. Materials

The materials used in the present study is represented below:

\[
\begin{align*}
\hat{t}^h \text{ada} & \quad \text{“to ride"} \\
\text{tada} & \quad \text{“to be ‘ta’"} \\
\text{t*ada} & \quad \text{“to pick“}
\end{align*}
\]

All tokens provided above are infinitival forms of real Korean verbs except [tada]. However, [tada] is also spoken in real speech, meaning “(The answer) is ‘ta’.” Vowels were set to /a/. Tokens are different from each other only in their initial consonant. Homorganic coronal stops were selected to minimize the effects of place of articulation. Tokens were embedded in a carrier sentence, ‘[______ga ani a _____da]’

“It is not _____ but ______.” Utterance-initial CV sequences were available in the first slot of the carrier sentence. The two syllables of the token in the first slot were metathesized in the second slot of the sentence to obtain VCV sequences. This study handles only utterance-initial stops.

2.3. Recording and Digitization

Subjects were asked to read a list of randomly ordered sentences to the effect that each sentence was repeated 10 times. Subjects were instructed to separate each sentence by a long pause. Recording was conducted in the Phonetics Lab in the Department of Linguistics at the University of Texas at Austin. Speech data were recorded onto analogue tapes via a Marantz SuperScope. The analogue signals were digitized on SoundScope at a sampling rate of 22,050 Hz. Combination of 3 speakers x 3 phonation types x 10 repetitions produces a total of 90 tokens.

2.4. Measurements

As was stated above, duration of the release burst ranges 2-30 ms.[1] Speech samples with a duration of 6.4 ms has been extracted to obtain spectra of the release burst.[6] However, a
comprehensive examination of the speech samples at hand demonstrated that the release burst is immediately followed by vocal fold vibration in fortis stops, where duration of the release burst was sometimes less than 6.4 ms. This study extracted only initial 5 ms of the release burst, so that no vocalic segment is included in the extracted speech samples. The onset of the release burst was set at the onset of a substantial change in the waveform. Samples of the release burst are illustrated in Figure 1, 2, and 3.

Praat 4.1.5 was used in both the extraction of the release burst and the measurement of intensity, center of gravity, and skewness. The waveform and spectrum of an extracted speech sample are illustrated in Figure 4 and 5.

Figure 1. Release burst in [tʰada]

Figure 2. Release burst in [tada]

Figure 3. Release burst in [t*ada]

Figure 4. Waveform of an extracted speech sample

Figure 5. Spectrum of an extracted speech sample

The spectrum of a Hamming-windowed and pre-emphasized speech sample is given in Figure 6. The spectrum was drawn from the same speech sample that was used for Figure.

Figure 6. Spectrum of a Hamming-windowed and pre-emphasized speech sample
As shown in Figure 6, the spectrum was boosted at higher frequencies. Intensity is calculated by formula (1).
\[
10 \log_{10} \frac{1}{n P_0} \sum |x|^2
\]  
(1)
where \( n \) stands for the number of samples, \( P_0 \) for the normative auditory threshold for a 1000 Hz sine wave, that is \( 2 \times 10^{-5} \) Pa, and \( x \) for the amplitude of a sample.
Center of gravity is computed by formula (2).[7]
\[
\frac{\int_{f_1}^{f_2} S(f) df}{\int S(f)^2 df}
\]  
(2)
where \( f \) stands for frequency, \( S(f) \) for the complex spectrum of speech samples, and \( p \) for power. The center of gravity is the average of \( f \) over the entire frequency domain, weighted by \( |S(f)|^p \). The denominator represents "energy."[7] For \( p = 1 \), it is the absolute spectrum derived from FFT. For \( p = 2 \), it indicates a power spectrum. The power was set to 2 in this study. The spectral center of gravity is a measure for how high the frequencies in a spectrum are on average.[7] The center of gravity of the spectrum given in Figure 5 is 988 Hz.

Skewness is computed using the central spectral moment given in formula (3).[7]
\[
\frac{\int_{-\infty}^{\infty} (f - f_c)^3 |S(f)|^p df}{\int |S(f)|^2 df}
\]  
(3)
where \( f_c \) stands for the spectral center of gravity. The \( n \)-th central spectral moment is the average of \( (f - f_c)^n \) over the entire frequency domain, weighted by \( |S(f)|^p \). The power was set to 2 in this study. The normalized spectral skewness is obtained by the third spectral moment divided by the 1.5 power of the second moment.[7] The skewness is a measure for how much the shape of the spectrum below the center of gravity is different from the shape above the mean frequency.[7] The skewness of the spectrum given in Figure 5 is 2.6.

3. Results
Mean of intensity, center of gravity, and skewness across phonation types and speakers for both the raw speech samples and Hamming-windowed and pre-emphasized samples are given in Table 1. It should be noted that WP represents "for the Hamming-windowed and pre-emphasized samples" in Table 1.

Table 1. Mean of intensity, center of gravity, and skewness across phonation types and speakers for both the raw speech samples and Hamming-windowed and pre-emphasized samples

<table>
<thead>
<tr>
<th>Speaker Type</th>
<th>Phonation Type</th>
<th>Aspirated</th>
<th>Lenis</th>
<th>Fortis</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGS</td>
<td>Intensity</td>
<td>98.42</td>
<td>56.58</td>
<td>57.30</td>
</tr>
<tr>
<td></td>
<td>Center of Gravity</td>
<td>1161</td>
<td>1107</td>
<td>14.00</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>2.24</td>
<td>2.42</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>Center of Gravity (WP)</td>
<td>4853.55</td>
<td>4832.47</td>
<td>4837.64</td>
</tr>
<tr>
<td></td>
<td>Skewness (WP)</td>
<td>-.198722</td>
<td>-.652724</td>
<td>-.235732</td>
</tr>
<tr>
<td>LSS</td>
<td>Intensity</td>
<td>61.32</td>
<td>60.97</td>
<td>62.72</td>
</tr>
<tr>
<td></td>
<td>Center of Gravity</td>
<td>12.24</td>
<td>12.31</td>
<td>11.09</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>2.04</td>
<td>2.50</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>Center of Gravity (WP)</td>
<td>4912.40</td>
<td>4962.04</td>
<td>4612.94</td>
</tr>
<tr>
<td></td>
<td>Skewness (WP)</td>
<td>-.069014</td>
<td>-.059917</td>
<td>-.119938</td>
</tr>
<tr>
<td>SHU</td>
<td>Intensity</td>
<td>58.07</td>
<td>58.84</td>
<td>59.05</td>
</tr>
<tr>
<td></td>
<td>Center of Gravity</td>
<td>10.94</td>
<td>7.05</td>
<td>8.30</td>
</tr>
<tr>
<td></td>
<td>Skewness</td>
<td>2.27</td>
<td>3.49</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td>Center of Gravity (WP)</td>
<td>3678.66</td>
<td>3465.36</td>
<td>3655.40</td>
</tr>
<tr>
<td></td>
<td>Skewness (WP)</td>
<td>3751.66</td>
<td>3554.06</td>
<td>2889.10</td>
</tr>
</tbody>
</table>

A two-way ANOVA was conducted to test whether there is a significant difference in intensity, center of gravity, and skewness across phonation types and speakers. First, for intensity, there was no significant interaction effect between phonation type and speaker.(\( F(4,81) = 0.814 \), \( p > 0.05 \)) There was no significant difference in intensity across phonation types.(\( F(2,81) = 0.719 \), \( p > 0.05 \)), while there was a significant difference in intensity across speakers.(\( F(4,81) = 16.776 \), \( p < 0.0001 \)). Post hoc comparison by Tukey’s HSD showed that LSS has a significantly greater intensity than SHJ or LGS. Next, for center of gravity, there was no significant interaction effect between phonation type and speaker.(\( F(4,81) = 1.294 \), \( p > 0.05 \)) There was no significant difference in center of gravity across phonation types.(\( F(2,81) = 0.820 \), \( p > 0.05 \)), while there was a significant difference in center of gravity across speakers.(\( F(2,81) = 5.447 \), \( p < 0.01 \)). Tukey’s HSD demonstrated that SHJ has a significantly less center of gravity than LSS or LGS. Finally, for skewness, there was no significant interaction effect between phonation type and speaker.(\( F(4,81) = 1.639 \), \( p > 0.05 \)) There was no significant difference in skewness across phonation types.(\( F(2,81) = 3.028 \), \( p > 0.05 \)), while there was a significant difference in skewness across speakers.(\( F(2,81) = 6.882 \), \( p < 0.01 \)). Tukey’s HSD showed that SHJ has a significantly greater skewness than LSS or LGS.

For the Hamming-windowed and pre-emphasized speech samples, a two-way ANOVA was conducted to test whether there is a significant difference in center of gravity and skewness across phonation types and speakers. For center of
gravity, there was no significant interaction effect between phonation type and speaker. (F(4,81) = 0.600, p > 0.05) There was no significant difference in center of gravity across phonation types. (F(2,81) = 0.290, p > 0.05), while there was a significant difference in center of gravity across speakers. (F(2,81) = 45.834, p < 0.0001) Tukey’s HSD demonstrated that SHJ has a significantly less center of gravity than LSS or LGS. For skewness, there was no significant interaction effect between phonation type and speaker. (F(4,81) = 1.643, p > 0.05) There was no significant difference in skewness across phonation types. (F(2,81) = 0.236, p > 0.05), while there was a significant difference in skewness across speakers. (F(2,81) = 19.266, p < 0.01) Tukey’s HSD showed that SHJ has a significantly greater skewness than LSS or LGS.

In summary, there was no significant difference in intensity across phonation types, while there was a significant difference in intensity across speakers. For both the raw speech samples and Hamming-windowed and pre-emphasized samples, there was no significant difference in center of gravity or skewness across phonation types, while there was a significant difference in center of gravity across speakers. There was no significant difference in statistical moments between the raw speech samples and Hamming-windowed and pre-emphasized samples. This means that windowing or pre-emphasis does not affect significance of the difference in statistical moments of the spectral characteristics of the release burst. This also means that there is no significant difference in the spectra of the release burst across phonation types. Therefore, it can be interpreted that difference in spectral characteristics of the release burst can be ignored in speech synthesis or speech recognition.

The results provided in this study are different from the previous study. The previous study claims that aspirated and fortis stops have greater energy in the release burst than lenis ones in Korean [4][5] while this study demonstrated that there is no significant difference in intensity, center of gravity, or skewness across phonation types. The source of the difference between the current and previous studies cannot be clarified due to the absence of a detailed description of the methodology in the previous study.

4. Conclusions
This study investigated spectral characteristics of the release burst in Korean alveolar stops in terms of intensity, center of gravity, and skewness, under the hypothesis that there is no significant difference in spectral moments across phonation types or speakers. From the results of this study, we accept the hypothesis that there is no significant difference in intensity, center of gravity, or skewness across phonation types, while we reject the hypothesis that there is no significant difference in intensity, center of gravity, or skewness across speakers. This lead us to claim that difference in phonation type does not lead to any significant difference in the spectra of the release burst. The present study suggests that difference in the spectral characteristics of the release burst across phonation types can be ignored in speech synthesis or speech recognition.

5. References