The Amplitudes of the Peaks in the Spectrum: Data from [a] Context

Anna Esposito
International Institute for Advanced Scientific Studies (IIASS)
Via G. Pellegrino 19, I84019 Vietri sul Mare (SA), Italy
e-mail:annesp@vaxsa.csied.unisa.it

Abstract[1]

This work is devoted to the study of the properties of the sound spectrum at the release of Italian stop consonants in vocalic contexts. The aim is to check if the amplitudes of the peaks in the spectrum can be used as acoustic attributes of the place of articulation of the consonants. This information is useful for defining an automatic algorithm which can discriminate among different place of articulation using simple data such as the values, in dB, of the maximum peaks in different frequency ranges. Moreover, different measurements have been performed (the spectra are computed at the release, averaged over 10 msec after the release, and using a smoothed spectrum) in order to define which measure retains more information about peak amplitudes.

Materials and procedures

The recording and measurements were made at the Research Laboratory of Electronics, Speech Communication Group, MIT, Cambridge, USA. The materials consisted in VCVC utterances produced by seven adult Italian speakers (three females and four males) in a sound-treated room and recorded on a high-quality magnetic tape recording system. The utterances were embedded in a carrier phrase. The measurements were made for the intervocalic consonant. Data were collected for all Italian vowels embedded in stop contexts. However, the results reported in the present paper are derived from the analysis of the stop consonants in the [a] context. The spectral representations used include a DFT spectrum, a smoothed DFT, a spectral averaging. The analysis window (Hamming window) was set to 3.1 msec. The spectrum at the consonant release, the averaged spectrum over the first 4 msec (for [b, d, g]) and over 10 msec (for [p, t, k]) after the release and, the k-averaged[2] spectrum were computed using a software program developed by Klatt (1984). All spectra were preemphasized, and the spectral amplitudes were enhanced by modifying an overall spectral gain control parameter. The amplitudes of the maximum peaks in different frequency ranges were measured by visual examination.

The amplitude attributes

The peaks amplitudes measured in the different frequency ranges described above were compared in order to identify properties that can be useful to discriminate the place of articulation of each consonant. Initially averages of the maximum peak amplitudes in different frequency ranges were computed. However, even though some of these averages differ significantly from one consonant to another, the standard deviations were high and they overlapped. This effect is mostly due to the variability of the peak amplitudes among the speakers. For this reason we decided to exclude these measures and we start to look to the amplitudes of the maximum peaks in specified frequency ranges compared to the amplitudes of the maximum peaks in other

[1]Supported by IIASS, CNR, and INFM Salerno University. Acknoledgements goes to M. Grabriella Di Benedetto for her useful comments and suggestions

[2]The k-averaged spectrum was computed by measuring the VOT length of the voiceless consonant. The cursor was then placed on the waveform at the temporal sampling point corresponding to half the VOT length, and the spectrum was averaged over 5 msec to the left and 5 msec to the right of this sampling point.

Table 1: *Amplitude feature-matching results for labial consonants. The entries give the mean percentage of consonants (based on 21 utterances of each consonant, occurring in [a] context, and obtained from seven speakers) that were correctly accepted or rejected by the set of acoustic features defined above.*

Spectrum at release	
Correct Acceptance	Correct Rejection
[p] 19	[k] 100 [t] 100
[b] 100	[g] 100 [d] 61.9
Averaged Spectrum	
Correct Acceptance	Correct Rejection
[p] 90.5	[k] 100 [t] 95.2
[b] 95.2	[g] 76.9 [d] 5
***k*-Averaged Spectrum**	
Correct Acceptance	Correct Rejection
[p] 95.2	[k] 100 [t] 85.7

frequency ranges. This comparison seemed more reasonable to us because it is possible to reduce the amplitude variability among speakers and repetitions. We carried out several attempts, comparing the maximum peak amplitudes in some frequency ranges with the maximum peak amplitudes in some other frequency ranges or comparing the differences between the maximum peak amplitudes in the different frequency ranges examined. In each attempt we defined a set of acoustic features based on these comparisons. We tested this set of features on the consonants in order to verify if it accepts the consonant under examination and rejects the others. The final results of this trial and error process are the following set of acoustic attributes for each place of articulation:

Labial amplitude attributes:

a1) The differences between the maximum peak in the 0-1kHz and the maximum peak in the 2-3kHz frequency ranges must be greater than 2dB;

b1) The differences between the maximum peak in the 0-1.2kHz and the maximum peak in the 5-7kHz frequency ranges must be greater than 6dB;

c1) The differences between the maximum peak in the 0-2kHz and the maximum peak in the 5-7kHz frequency ranges must be greater than 6dB.

Table 2: *Amplitude feature-matching results for alveolar consonants.*

Spectrum at release	
Correct Acceptance	Correct Rejection
[t] 90.4	[k] 95.2 [p] 19
[d] 80.9	[g] 100 [b] 90.4
Averaged Spectrum	
Correct Acceptance	Correct Rejection
[t] 100	[k] 100 [p] 90.4
[d] 33.3	[g] 100 [b] 100
***k*-Averaged Spectrum**	
Correct Acceptance	Correct Rejection
[t] 90.4	[k] 85.7 [p] 85.7

Table 3: *Amplitude feature-matching results for velar consonants.*

Spectrum at release	
Correct Acceptance	Correct Rejection
[k] 100	[p] 85.7 [t] 85.7
[g] 90.5	[b] 100 [d] 95.2
Averaged Spectrum	
Correct Acceptance	Correct Rejection
[k] 95.2	[p] 100 [t] 90.5
[g] 57	[b] 100 [d] 95.2
***k*-Averaged Spectrum**	
Correct Acceptance	Correct Rejection
[k] 80.9	[p] 100 [t] 95.2

Alveolar amplitude attributes:

a2) The differences between the maximum peak in the 2-3kHz and the maximum peak in the 3-4kHz frequency ranges must be lower than 8dB;

b2) The differences between the maximum peak in the 0-1kHz and the maximum peak in the 1-3kHz frequency ranges must be greater than -10dB;

c2) The differences between the maximum peak in the 0-1.2kHz and the maximum peak in the 4-6kHz frequency ranges must be lower than 13dB;

d2) The differences between the maximum peak in the 0-1kHz and the maximum peak in the 2-3kHz frequency ranges must be lower than 8dB;

Table 4: *New amplitude feature results to distinguish [b] from [d, g]*

Spectrum at release	
Correct Acceptance	Correct Rejection
[b] 80.9	[g] 100 [d] 85.7
Averaged Spectrum	
Correct Acceptance	Correct Rejection
[b] 95.2	[g] 100 [d] 38

e2) The differences between the maximum peak in the 0-1.2kHz and the maximum peak in the 1.5-3.5kHz frequency ranges must be lower than 6dB.

Velar amplitude attributes:

a3) The differences between the maximum peak in the 0-1.2kHz and the maximum peak in the 1.5-3kHz frequency ranges must be lower than -1dB;

b3) The differences between the maximum peak in the 1-2kHz and the maximum peak in the 0-1kHz frequency ranges must be greater than 1dB;

c3) The differences between the maximum peak in the 2-3kHz and the maximum peak in the 3-4kHz frequency ranges must be greater than 1dB;

d3) The differences between the maximum peak in the 1-2kHz and the maximum peak in the 6-7kHz frequency ranges must be greater than 11dB.

The set of amplitude attributes defined above are the same both for voiced and voiceless consonants. However, for voiced consonants the discrimination performances are less good in most of the cases (see tables 1, 2, 3). The voicing, which is always present in Italian, causes pressure fluctuations that leads to variability in the peak amplitudes and causes shifts in the vocal tract resonances. In order to improve the discrimination performances among [b, d, g] we tried to identify, for such consonants, a different set of amplitude attributes. The final result of this process was the following set of amplitude attributes for [b, d, g]. The discrimination performances using this set of amplitude attributes are reported in tables 4, 5, 6.

[b] amplitude attributes:

a11) The differences between the maximum peak in the 0-1kHz and the maximum peak in the 2-3kHz frequency ranges must be greater than 5dB;

Table 5: *New amplitude feature results to distinguish [d] from [b, g].*

Spectrum at release	
Correct Acceptance	Correct Rejection
[d] 100	[g] 100 [b] 52.3
Averaged Spectrum	
Correct Acceptance	Correct Rejection
[d] 85.7	[g] 100 [b] 85.7

Table 6: *New amplitude feature results to distinguish [g] from [b, d].*

Spectrum at release	
Correct Acceptance	Correct Rejection
[g] 95.2	[b] 100 [d] 100
Averaged Spectrum	
Correct Acceptance	Correct Rejection
[g] 95.2	[b] 100 [d] 95.2

b11) The differences between the maximum peak in the 0-1.2kHz and the maximum peak in the 1.5-3kHz frequency ranges must be greater than 0dB;

c11) The differences between the maximum peak in the .8-1.5kHz and the maximum peak in the 2-3kHz frequency ranges must be greater than 1dB.

[d] amplitude attributes:

a22) The differences between the maximum peak in the 0-2kHz and the maximum peak in the 5-7kHz frequency ranges must be lower than 22dB;

b22) The differences between the maximum peak in the .8-1.5kHz and the maximum peak in the 2-3kHz frequency ranges must be greater than -7dB;

c22) The differences between the maximum peak in the 0-1.2kHz and the maximum peak in the 1.5-3.5kHz frequency ranges must be lower than 12dB;

d22) The differences between the maximum peak in the 2-3kHz and the maximum peak in the 3-4kHz frequency ranges must be lower than 8dB;

[g] amplitude attributes:

a33) The differences between the maximum peak in the 2-3kHz and the maximum peak in

the 3-4kHz frequency ranges must be greater than 7dB;
b33) The differences between the maximum peak in the .8-1.5kHz and the maximum peak in the 2-3kHz frequency ranges must be lower than 3dB;

Conclusions

An attempt, even with a different approach, to ceck if the shape of the spectrum can retain information about the place of articulation of the consonants, was made by Blumstein and Stevens (1979). These authors based their considerations on the general shape of the spectrum and defined a set of templates. Overall, about 85% of the utterances are correctly accepted by these templates and about the same percentage of utterances are correctly rejected. They used a fixed time window (26 msec). However, they suggested that it may not be desiderable to postulate a single, fixed time window, and that the gross spectrum shape may be assesed examining, successive spectral samples extending over 10-20 msec, each computed using a relatively short time window.

The preliminary results we reported in this work for the consonants in [a] context, and in a previous work (Esposito 1995) for the consonants in [i] context show that the amplitudes of the peaks in the spectrum examining successive spectral samples over 10 msec (averaged burst and *k*-averaged burst) and computed using a short time window (3 msec) give useful information to discriminate among the voiceless consonants [p, t, k]. (see tables 1, 2, 3). Overall, about 95% of the utterances are correctly accepted by these set of amplitude attributes and about the same percentage of utterances are correctly rejected. This information can also be used to define an automatic algorithm that discriminates successfully among [p, t, k] and it is simpler than the templates defined by Blumstein and Stevens (1979) that are more general and hard to translate in an automatic algorithm. Moreover, the set of amplitude attributes improve the discrimination performances with respect to the templates. Furthermore, the set of amplitude attributes we defined takes into account the vowel contexts. In fact, a similar, but different set of acoustic attributes was defined in a previous work (Esposito 1995) for the same consonants in the [i] environment. The amplitude attributes defined for the consonants in [i] context gave good discrimination performances. However, when applied to the same consonants in [a] context they give very poor discrimination. Then, we modified the sets of amplitude attributes in order to improve the discrimination. Our hypothesis is that it will be necessary to define a set of amplitude attributes for each vowel or for specific vowel classes that share the same distinctive features. We could justify the above consideration on the basis of the conjecture that since articulatory parameters change from one vowel class to another, and there is some anticipatory coarticulation effect which modifies the spectrum shape of the consonant under examination.

In the case of voiced consonants, we obtained some improvements with the new set of amplitude attributes. We can discriminate successfully [g] (95% of correct acceptance both for the spectrum at the release and the spectrum averaged over 4 msec) from [b, d] (about 100% of correct rejection). However, for [b, d] similar information does not work very well: information about formant transitions is required.

With regard to the particular spectra computed it is possible to say that the spectrum during the first 10 msec after the release and the *k* averaged spectrum seem more useful to retain information about amplitude features when the consonants are voiceless. The spectra at the release retain more information about the amplitude attributes of voiced consonants.

References

S.E. Blumstein, K.N. Stevens, 1979, Acoustic invariance in speech production: ... , *JASA*, vol. **64(4)**, 1001-1017.
A. Esposito, 1995, The amplitude of the peaks in the spectrum ... In *Proceeding of ICPhS95*, (K. Elenius and P. Branderud editors), vol **1**, pp. 38-41, Arne Strombergs Grafiska, Stockholm.
D.H Klatt, 1984, *MIT SpeechVax User's Guide*, Copyright 1984 by Dennis H. Klatt.