19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

AUDITORY REPRESENTATION OF SPECTRAL INTENSITY VARIATION IN COARTICULATED VOWELS

PACS: 43.70.Fq

Jacewicz, Ewa; Fox, Robert Allen
Speech Perception and Acoustics Laboratories, The Ohio State University, 1070 Carmack Rd., Columbus, OH, 43210, USA; jacewicz.1@osu.edu; fox.2@osu.edu

ABSTRACT
Variation in the vocal effort of a speaker to signal linguistic stress, accent, or emotion produces significant changes in the overall vowel intensity. Changes are also evident in the vowel's spectrum with the size of intensity variation being more pronounced in the higher frequencies relative to the lower frequency region. A second (and less explored) source of spectral intensity variation results from coarticulation with the immediate consonant environment of a vowel. Despite its smaller size, the general pattern of spectral intensity distribution is similar. The present study is concerned with a possible encoding and auditory representation of this type of variation at early stages of processing by the peripheral auditory system. Auditory spectrograms were obtained by passing the vowel spectrum through three psychoacoustically motivated stages: equal-loudness pre-emphasis, auditory filterbank, and intensity-loudness compression. Overall, the results show greater intensity variation in the higher frequency bands corresponding to the 1.5-4.0 kHz region. This is consistent with findings for the coarticulatory effects across the acoustic frequency bands and suggests that the nature and magnitude of intensity changes found in the acoustic spectrum are carried on through the early stages of auditory processing.

INTRODUCTION
Variation in the vocal effort of a speaker to signal emphasis-related functions such as linguistic stress, accent, or emotion produces significant changes in the overall vowel intensity. As shown in a number of studies, the greater intensity affects also vowel spectrum in an asymmetrical way: the low frequency components are hardly affected and the intensity increase concentrates in the higher spectral regions, usually between 0.5 – 4.0 kHz.

This study examines intensity variation in vowels as a function of coarticulation with the immediate consonantal context. This variation, although less studied, is necessarily present for all speakers at all times and is not related to the intentional emphasis-oriented changes. The question arises as to the nature and magnitude of this type of intensity variation. Specifically, are intensity differences as a function of coarticulation located mainly in the higher spectral regions as in effort-related changes?

Two basic approaches have been developed in the literature to estimate how the spectrum amplitude changes with increased articulatory effort of the speaker. The first, based on models of speech production, is to measure the variation in the tilt of the glottal spectrum taking into account the effects of the vocal tract filter. This has been done by defining either the relation between the amplitude of the first harmonic and the strongest harmonic in the vicinity of the third formant peak with corrections H1*-A3* ([6]) or the dB-difference between a high frequency pre-emphasis and a flat frequency weighting corrected for the influence of vowel-specific formant pattern SPHL-SPL ([2], [3]). The second approach is to measure the energy in the frequency bands across the vowel spectrum. In particular, spectral emphasis measures have been developed to evaluate the contribution of the higher-frequency bands to the overall vowel intensity ([1], [8]).

A specific measure of emphasis called "spectral balance" was proposed by Sluijter and van Heuven [7] as a correlate of linguistic stress. Intensity changes were examined in the four

contiguous frequency bands of 0-0.5, 0.5-1.0, 1.0-2.0, and 2.0-4.0 kHz. [Measuring the intensity distribution in contiguous bands was in part motivated by findings from psychoacoustics.]
This work showed that, in stressed syllables produced with greater physiological effort, the intensity levels increased in the higher parts of vowel spectrum (above 0.5 kHz) but not in the lowest frequency band.

The psychoacoustic basis for the Sluijter and van Heuven study comes from the fact that low-frequency bands do not contribute appreciably to the perceived increase in loudness of a sound while the contribution of the higher frequency bands to loudness is much greater. This is related to the audibility function of the auditory system which determines the relationship between the physical intensity and perceived loudness. That is, to achieve equal loudness across different frequency components in a sound, more intensity is needed for lower frequencies than for higher frequencies, as first shown in [4]. The equal loudness contour and the perceived loudness have been shown to correspond to the acoustic vowel spectrum [5] whose low frequency region is inherently more intense as compared to the higher frequencies. Yet, vowels sounded louder for the listeners as a result of an increase of intensity in the higher frequency regions of the spectrum, which are acoustically enhanced relative to the low-frequency region when spoken with an increased vocal effort.

The spectral balance parameter proposed in [7] was a motivation for the present study. The first aim was to explore the variation in spectral balance as a function of coarticulation which cannot be attributed to the differences in the physiological effort such as to signal linguistic stress or prominence. Although we expected smaller intensity variation across the frequency bands than the prominence-related changes reported in [7], we asked whether the general pattern of such variation in terms of the relative contribution of low and high parts of the spectrum is comparable to that found for vowels spoken with greater physiological effort. The second aim was to address the question of a possible auditory representation of this type of variation at early stages of processing by the peripheral auditory system. In particular, we were interested how the auditory system might encode the patterns of intensity changes in the acoustic spectrum produced by coarticulation.

METHODS

The acoustic data presented in this paper come from a larger study involving twenty speakers, eight vowel categories, and ten consonantal contexts. Only a subset of these data was selected here which pertain directly to the present research focus.

Speakers, materials, and procedure

Twenty speakers (ten men and ten women) of Midwestern American English participated in the study. They were students aged 18-36 years enrolled in a variety of majors at The Ohio State University and were phonetically untrained.

Eight relatively monophthongal American English vowels were selected: /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u/. Each vowel was embedded in a symmetrical C_1VC_1 context. The selected consonant set consisted of ten oral consonants /p, t, k, b, d, g, f, v, s, z/. The CVCs were produced as monosyllables in a stressed position in a short sentence “It’s a … .” Each speaker recorded the stimuli in a single session lasting approximately one hour.

Recordings were controlled by a program in Matlab. The speaker was seated in a sound-attenuated booth. A head-mounted Shure SM10A dynamic microphone was used, positioned at a distance of 2 inches from the speaker’s lips. Speech samples were recorded and digitized at a 44.1-kHz sampling rate directly onto a hard disc drive. A model example word containing the vowel of interest was first displayed on the screen (e.g., “heat”). Next to the word, the stimulus sentence appeared (e.g, “It’s a beeb”). The speaker read the sentence placing stress on the CVC word and producing the vowel quality as in the displayed word. The order of consonantal contexts for each vowel type and the order of vowels themselves were randomized for each speaker. After recording each sentence, the experimenter either accepted and saved the sentence or re-recorded it in the case of mispronunciations.

Acoustic measurements

Prior to acoustic analysis, the tokens were digitally filtered and downsampled to 11.025 kHz. Four contiguous frequency bands B1-B4 were chosen as in [7] B1 (0-0.5), B2 (0.5-1.0), B3 (1.0-2.0), and B4 (2.0-4.0 kHz). These bands were used for each vowel category regardless of whether they included formants and regardless of how many formants fell within one frequency band. This approach was chosen to provide a set of measurements consistent across all vowel categories. The bands in [7] were originally established with specific reference to the vowel /a:/ in Dutch so that B1 included fundamental frequency and B2, B3, and B4 included F1, F2, and F3, respectively. Our selection of bands did not reflect this vowel-specific formant pattern. The present analysis of intensity distribution did not focus on changes to spectral tilt as a function of linguistic stress and our primary interest was in more global effects of consonant environment on intensity variation across frequency bands, including not only formant frequency peaks but also spectral minima. As in [7], the spectrum level of each frequency band was defined as the base-10 log of the summed power. The computed spectrum level was based on 1024-point fft analysis using a 25-ms window (using a Hamming filter) with each fft section representing a range of 10.76 Hz. The intensity values for each frequency band represented the summed values of the squared fft bins across its frequency range. Intensity levels measures were obtained at the center location of rms peak.

Auditory analysis

Auditory spectra were calculated by passing the vowel signal through the following stages. First, vowel spectrograms were used (as a positive time-frequency representation of the signal) to calculate the short-term fft using a 25-ms Hamming window, which served as input to three stages of peripheral auditory processing: (1) equal-loudness pre-emphasis curve (provides a weighting along the frequency axis), (2) auditory filterbank (passes the output of the previous stage through 33 gammatone filters centered at each bin of the fft, center frequencies in ERBs), and (3) intensity-loudness compression (raises the output of the previous stage to the 0.3 power).

RESULTS

The results are presented for the vowels /ɪ/ and /ʊ/ only as representing intensity variation in a high front and a high back vowel, respectively. Consonant contexts included stops only organized by place of articulation: labial (/b, p/), alveolar (/d, t/), velar (/g, k/) and the status of voicing: voiced (/b, d, g/) and voiceless (/p, t, k/). Data in the figures display mean values for 10 male speakers. These data were analyzed using within-subject ANOVAs.

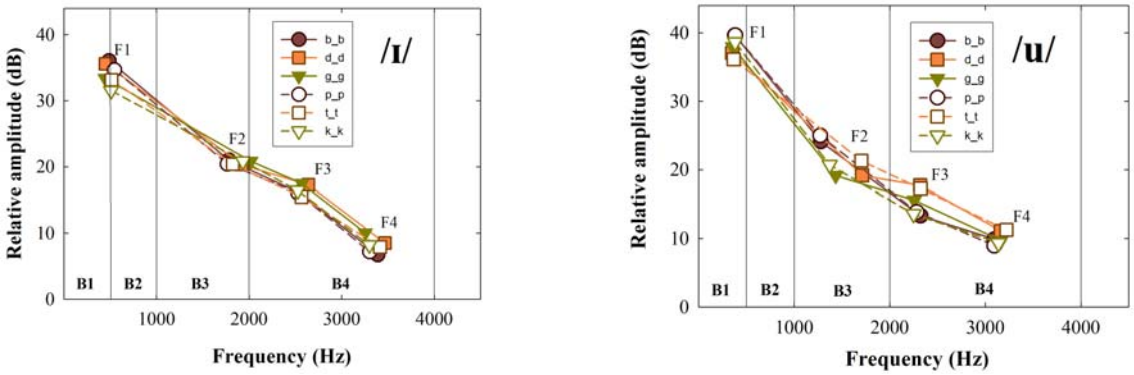

Figure 1. Mean frequencies of formants F1-F4 of the vowels /ɪ/ (left) and /ʊ/ (right) in the context of stops measured at rms peak and the corresponding broad frequency bands B1-B4.

Correspondence between the frequency bands and formants

As a reference for further comparisons, Figure 1 shows the correspondence between the contiguous broad frequency bands B1-B4 (one band per octave) and mean formant frequency values for the first four formants F1-F4 for the vowels /ɪ/ and /ʊ/. As can be seen, the major differences are in the locations of the first two formants, F1 and F2, which fall within bands B1

and B3 for the vowel /u/ and at the boundaries between B1-B2 and B3-B4, respectively, for the vowel /ɪ/.

Effects of consonant voicing on the spectra of /ɪ/ and /u/

Figure 2 (two left panels) shows the distribution of acoustic spectral intensity for the vowel /ɪ/ across four broad frequency bands B1-B4 (upper panel) and the corresponding auditory spectrum of the vowel (lower panel). The auditory spectrum is represented by the output of 33 gammatone filters. In general, a slightly higher intensity value was found for the voiced context in B1, and higher values for the voiceless contexts across the higher frequency bands B2, B3, and B4. Of interest are the greater dB-differences between the voiced and voiceless contexts in bands B2 and B3 as compared to B1. The auditory spectrum shows two peaks, one in the lower frequency region (output of the filters 9 and 10 within band B1) and the second at higher frequencies (filters 21-23) which fall within bands B3 and B4. As can be seen, the effect of consonant voicing on the auditory spectrum is similar to that found in the acoustic broad band analysis: The signal is slightly greater in magnitude in the voiced contexts in the lower frequency region corresponding to band B1 and in the voiceless contexts in the higher parts of the spectrum corresponding to bands B2, B3 and, partially, B4. Consistent with the acoustic pattern for the effect of voicing in B4, the difference in magnitude in auditory spectrum corresponding to the later part of this band is minimal. The observed interaction between voicing and acoustic frequency band was significant ([F(1.6, 14.2) =10.46, p<.001, η^2=.793]).

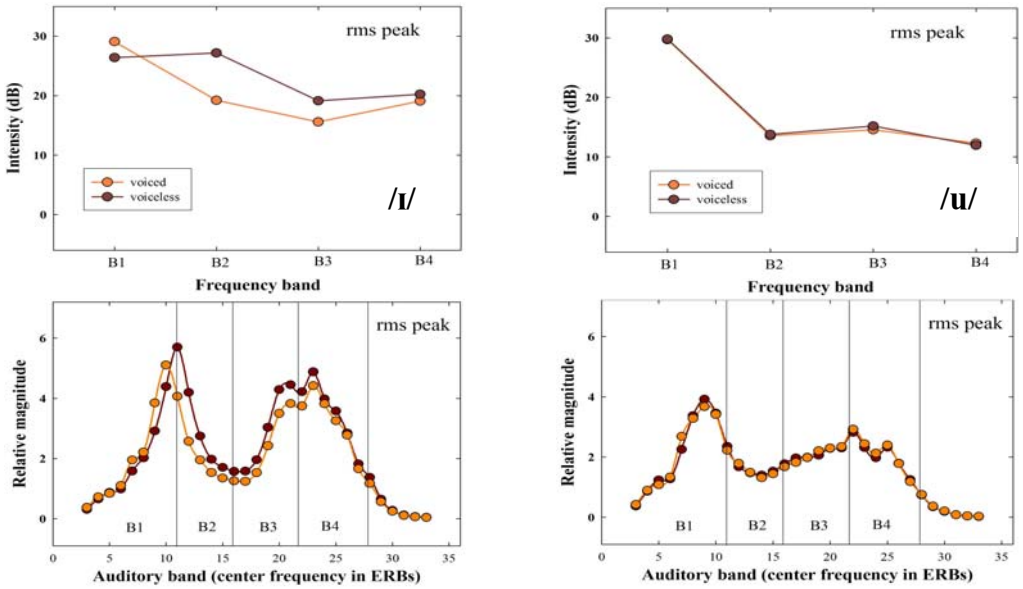

Figure 2. Mean intensity changes across acoustic spectra and the corresponding magnitude changes in the auditory spectra as a function of consonant voicing for /ɪ/ (left) and /u/ (right).

For the vowel /u/ shown in right panels, the correspondence between the intensity distribution across acoustic frequency bands and the auditory spectrum as a function of consonant voicing is again noteworthy. Although no intensity variation across the acoustic spectrum was found as a function of consonant voicing and the interaction between voicing and frequency band was statistically not significant and the same pattern was found across the corresponding auditory spectrum. Another finding for the vowel /u/ was that the second peak in the auditory spectrum corresponding to the higher frequency bands in the acoustic spectrum was much shallower in comparison to the vowel /ɪ/. This is again in line with the intensity distribution pattern for the vowel /u/ across acoustic frequency bands B2, B3, and B4. These results suggest that information about the acoustic spectral intensity variation can be encoded by the peripheral auditory system.

Effects of consonant place of articulation on the auditory spectra of /ɪ/ and /u/

Mean intensity distribution as a function of consonant place of articulation is shown in Figure 3 for the vowels /ɪ/ (two left panels) and /u/ (two right panels). The acoustic intensity distributions across the broad frequency bands are in the upper panels and the corresponding auditory spectra are in the lower panels.

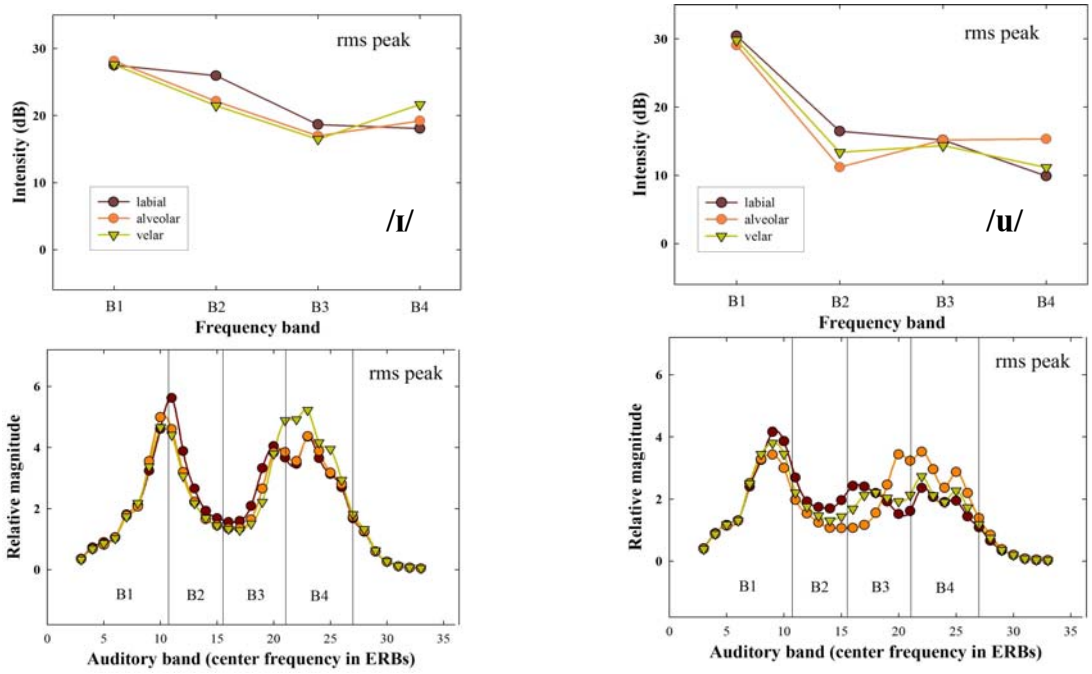

Figure 3. Mean intensity changes across acoustic spectra and the corresponding magnitude changes in the auditory spectra as a function of consonant place for /ɪ/ (left) and /u/ (right).

For the vowel /ɪ/, the interaction between place and frequency band was significant ([F(3.1, 27.4) =12.8, p<.001, η^2=.587]). Considerable intensity variation was observed in the higher frequency bands whereas no such variation occurred in the lowest band B1. The effects of place were remarkably consistent across acoustic frequency bands and the corresponding auditory spectrum. In particular, the spectral intensity was greater in the contexts of labials in bands B2 and B3 and in the context of velars in band B4. The same consonantal context effects were found in the auditory spectrum which was greater in magnitude in the context of labials in the corresponding auditory bands 10-20. Similarly, the "boost" in magnitude in the context of velars in the higher region in the auditory spectrum reflects increased intensity in band B4 as a function of velars.

For the vowel /u/, the intensity changes in B1 were again small and increased considerably in bands B2 and B4. The interaction between place and frequency band was significant ([F(2.9, 26.2) =13.06, p<.001, η^2=.592]). An interesting change in the pattern of consonantal place effects took place in band B3. In the spectral regions spanned by bands B1 and B2, the intensity was greatest in the context of labials and smallest in the context of alveolars. In band B3, the labial and alveolar contexts "crossed" giving rise to greatest intensity in band B4 in the context of alveolars and smallest in labials. This pattern of spectral intensity distribution was well maintained throughout the auditory spectrum shown in Figure 3.

CONCLUSIONS

Overall, the present results demonstrate that the intensity variation across acoustic broad frequency bands as a function of consonantal context conforms to the general pattern of

spectral intensity distribution found in vowels produced with increased articulatory effort. Consistent with the latter, the smallest intensity variation was found in the lowest band B1 and greater differences occurred in the higher bands. This indicates that the spectral balance parameter can also serve as a measure of intensity variation in vowels as a function of coarticulatory context.

Despite this general tendency, there were noticeable differences in the consonantal context effects in specific frequency bands depending on whether the vowel was front or back. The most striking result was obtained for the vowel /u/, where consonant voicing did not introduce any intensity variation across the bands and no differences were found between voiced and voiceless contexts.

The auditory spectra indicate that the nature and magnitude of intensity changes found in the acoustic spectrum are carried on through the early stages of auditory processing. The auditory spectra were remarkably consistent with the patterns of acoustic intensity variation in different vowels and in different coarticulatory contexts. However, they also provided more spectral details about this variation which were hard to detect in the broad-band acoustic analysis.

Based on the present acoustic results, we may tentatively conclude that the contribution of the low-frequency bands relative to higher frequency bands known as the spectral balance is manifested not only in the perceived increase in vowel loudness. The variation in intensity distribution across the frequency bands in coarticulated vowels seems to follow the same general pattern. It remains to be determined whether listener is sensitive to this type of variation across the vowel spectrum.

ACKNOWLEDGMENTS
This study was supported by the research grant No. R03 DC005560 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. The authors wish to thank Dan Hack and Chiung-Yun Chang for assistance at various stages of this research and Jeff Murray for editorial help.

References: [1] D. G. Childers, C. K. Lee: Vocal quality factors: Analysis, synthesis, and perception. Journal of the Acoustical Society of America **90** (1991) 2394-2411.
[2] G. Fant: The voice source in connected speech. Speech Communication **22** (1997) 125-139.
[3] G. Fant, A. Kruckenberg, J. Liljencrants: The source-filter frame of prominence. Phonetica **57** (2000) 113-127.
[4] H. Fletcher, W. A. Munson: Loudness, its definition, measurement, and calculation. Journal of the Acoustical Society of America **5** (1933) 28-105.
[5] R. D. Glave, A. C. M. Rietveld: Is the effort dependence of speech loudness explicable on the basis of acoustical cues? Journal of the Acoustical Society of America **58** (1975) 875-879.
[6] H. M. Hanson: Glottal characteristics of female speakers: acoustic correlates. Journal of the Acoustical Society of America **101** (1997) 466-481.
[7] A. M. C. Sluijter, V. J. van Heuven: Spectral balance as an acoustic correlate of linguistic stress. Journal of the Acoustical Society of America **100** (1996) 2471-2485.
[8] H. Traunmüller, A. Eriksson: Acoustic effects of variation in vocal effort by men, women, and children. Journal of the Acoustical Society of America **107** (2000) 3438-3451.