Tense/lax vowel classification using dynamic spectral cues

Janet Slifka
Speech Communication Group, MIT, U.S.A
E-mail: slifka@speech.mit.edu

ABSTRACT

In English, the feature [tense] or [ATR] (advanced tongue root) has been used to encompass the vowels that are produced on the extreme edges of the acoustic and articulatory spaces. The lax counterparts to these vowels are generally produced closer to the center of those spaces. The present study is based on the hypothesis that the extreme positioning in both articulatory and acoustic space evolves during the course of vowel production. This study uses two measures that attempt to track these changes over time. The dataset consists of 48 citation-form vowels in stressed position, from each of three male speakers, (144 total). The slope of the first formant movement across the vowel distinguished between tense and lax vowels with a 90% accuracy. The location in time of the energy peak in the first formant region as a percentage of vowel duration classified the vowels with an 83% accuracy.

1. INTRODUCTION

Features such as [high], [low], [back], and [rounded] have been used to distinguish vowels. In this case, the word feature refers to binary distinctive features [1], in which the identity of an underlying segment is changed by a change in one of the features that specify that segment. In English, an additional feature, [tense] (for example [2],[3]) encompasses the vowels that are produced on the extreme edges of the acoustic and articulatory spaces. The lax counterparts to these tense vowels are generally produced closer to the center of that space. There has been much discussion of the tense/lax quality and whether or not it is a valid vowel distinction in the way that such contrasts as front-back and high-low are considered to be. (For a discussion see [4]). As a valid vowel distinction, there is the expectation of an articulatory and acoustic correlate of this feature. One such correlate may be in the placement of the tongue root. The tense vowels are sometimes called advanced tongue root (ATR) vowels [5],[2]. Tense vowels are produced with a widening in the cross-sectional area of the pharynx, which can be achieved by moving the tongue root forward.

When the pharynx is widened, the tongue body may be displaced forward and possibly upward, leading to an increase in the constriction in the oral region. For English, the advanced tongue root movement is considered to co-vary with the raised tongue body movement to some degree [6],[7]. These adjustments in the tongue body lead to a lowering of the first formant (F1) frequency for tense vowels as compared to lax vowels. In terms of a Helmholtz resonator approximation, the tense vowels have a greater acoustic compliance in the pharyngeal region and a larger acoustic mass in the oral region leading to a lower F1 than in the lax configuration. The non-low lax vowels are produced with a wider constriction in the oral region and are frequently accompanied by a lowering of the mandible. These actions, according to perturbation theory, should cause F1 to rise. By the same theory, the narrower constriction for tense vowels should lower F1.

The acoustic consequences of the lower F1 as well as the narrower constriction for tense vowels include a reduction in the amplitude in the higher formants (See [8] for a discussion.) The loss of energy in the higher formants also occurs for vowels produced with breathy voicing. It may be that speakers enhance this effect in tense vowels by producing them with breathy voicing [5],[9],[10]. For example, Lotto and colleagues [11] conducted listening tests on the identification of the vowels: /i/, /u/, /ɪ/, and /ʊ/ for a range of the acoustic correlates of breathiness in synthetic vowels. The authors conclude that increased breathiness tends to push the classification toward a tense vowel rather than lax vowel. Halle and Stevens [5] examined acoustic cues for spectral balance in tense/lax pairs and found that lax vowels have more energy in the region of the second and third formants than tense vowels.

Finally, there is a wide body of data supporting vowel duration differences as an acoustic cue between tense and lax vowels. (See [12] for a review.) These are all examples of the wealth of data available to compare such vowel pairs. The majority of these data are acoustic measurements taken at a specific point in time such as at the midpoint of the vowel, at the point of maximal constriction in the oral region, or at the point of maximal constriction in the vocal tract. The present study is based in the hypothesis that advancement of the tongue root and the extreme positioning in both articulatory and acoustic space evolves during the course of the production of tense vowels. If this is the case, then parameters that attempt to track these changes over time should distinguish between tense and lax vowels.

This study uses two measures to evaluate the tense/lax distinction, these are: the slope of the first formant movement across the vowel (F1 slope) and the location in time of the energy peak in the F1 region as a percentage of vowel duration (F1 energy peak).

ISBN 1-876346-48-5 © 2003 UAB

2. METHODS

The current study focuses on non-low vowels only. In English, the set of non-low tense vowels includes: /i/, /e/, /o/, and /u/. Each of these could be considered to be paired with a non-low lax vowel, /ɪ/, /ɛ/, /ɔ/, and /ʊ/. However, /ɔ/ is less uniformly accepted to be a lax vowel. This study uses a limited data set. The consonant-vowel-consonant (CVC) database is a pre-recorded database with three male speakers (mean talker age of 37 years) reading the nonsense word /həCVC/ in isolation. The talkers speak English as their first language and report no hearing loss. The data were originally recorded onto audio tape, and then digitized to 10 kHz. Each speaker produced a slightly different consonant set paired with all eight vowels. Speaker A produced the consonants: /b/, /d/, /m/, /n/, /ʃ/, /s/. Speaker J produced the consonants: /b/, /d/, /f/, /ʃ/, /s/, and /ð/. Speaker K produced the consonants: /b/, /d/, /f/, /k/, /m/, and /n/. The starting point and ending point of each of the 144 vowels were determined by-hand with the aid of spectral information.

A. First Formant Slope

As already stated, the pharyngeal region is widened as the tongue root is advanced, and this causes a narrowing in the oral region. The greater the perturbation from a uniform tube, the more the first three formant frequencies will differ from those of a uniform tube. This difference is shown schematically in Figure 1a for F1 under the assumption that the articulators continue to move toward an extreme position across the course of the vowel. The drop in F1 is accentuated by the narrowing of the constriction in the oral region (/i/ and /e/) or at the lips (/u/ and /o/).

In the lax vowels, the area of the pharyngeal region is not as wide as in the tense vowel counterpart. The constriction in the oral region is wider, and the mandible is often lowered to further increase the cross-sectional area. The lax vowels should show an increase in the first formant frequency (Figure 1b). If these schematics in Figure 1 are a true representation of the physical situation, then a measure of the slope of the F1 movement over time should differ between the two classes of vowels; the slope should be negative for tense vowels and positive for lax vowels.

The neighboring consonants will affect the F1 trajectory depending on the nature of the consonant constriction and the duration of the vowel. A measure of F1 slope could attempt to develop criteria for regions of consonant influence that might, in fact, be different for each consonant or consonant class. F1 slope could then be estimated between the determined limits. However, the expectation is that the movement in F1 is robust for the tense/lax distinction for stressed vowels. For this reason, only the first and last 10 milliseconds (ms) were discarded as being consonantal and to allow for error in the estimation of the vowel limits. The slope of the F1 trajectory was computed across the remainder of the vowel. The F1 track was determined using 14th-order LPC analysis at a 10 ms frame rate. The analysis window was 49 ms, and a mild pre-emphasis filter ($1-0.7z^{-1}$) was applied. A 120 Hz high pass filter was used to reduce the influence of the first harmonic. The poles of the F1 track were selected by hand using spectral information. There was little ambiguity in the F1 selection process.

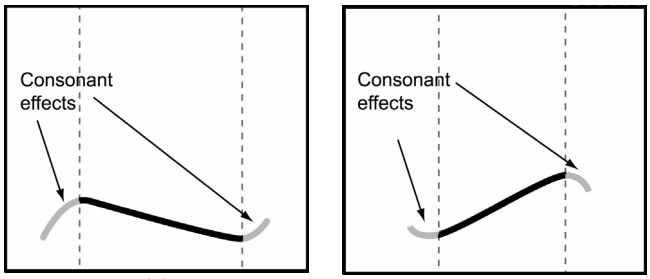

Figure 1: (a) Tense vowels are predicted to have a falling F1 over the course of the vowel. (b) Lax vowels are predicted to have a rising F1 over the course of the vowel. Both predicted trends could be affected by the presence of formant movements due to adjacent consonants as shown in the gray regions.

B. F1-region energy peak

The tense vowels, /i/, /e/, /o/, and /u/ all have a very narrow constriction toward the anterior end of the vocal tract. The vowels /i/ and /e/ have a constriction in the oral region and the vowels /o/ and /u/ have a constriction at the lips. This extreme configuration for the vocal tract shape is characteristic of the non-low tense vowels. This constriction causes increased acoustic losses resulting in a widening of the F1 bandwidth. The result is that the energy in the first formant should weaken as the constriction narrows. This reasoning would imply that the energy in the F1 region should have its peak earlier in the vowel for tense vowels than for lax vowels. A measure of the time to this energy peak as a percentage of the vowel duration is the second measure used in this study.

The results of this study will be used within a model for lexical access [13]. This system attempts to detect the locations of vowel-like regions by filtering the energy in the F1 region (300-900 Hz) and searching for peaks in that energy [14]. The present study uses a simplified version of that process to detect energy peaks in the 300-900 Hz band. The location of the energy peak in the F1 region is expressed as a percentage of the vowel duration. Tense vowels are expected to have the energy peak relatively early in the vowel, and lax vowels are expected to have the energy peak toward the middle or possibly latter part of the vowel.

3. RESULTS

The results for F1 slope are given in Table 1. The lax vowels have a rising slope in 91.7% of the cases, and the tense vowels have a falling slope in 88.9% of the cases. These results agree with the assumption that changes in F1 occur across these vowels as the articulators either move

ISBN 1-876346-48-5 © 2003 UAB

toward an extreme position (tense vowels) or possibly toward a neutral position (lax vowels).

Six lax vowel examples have a falling slope. Four of the six vowels are /ɔ/ as paired with /b/, /f/, /m/, and /ð/. The remaining two are /ɛ/ and /ʊ/ as paired with /b/ and /ð/ respectively. All of the consonant constrictions in these examples are at or near the lips, which lead to the expectation of a lowered F1. Since lax vowels are also expected to be short in duration, the effect of neighboring consonants may play a greater role for these cases. Eight of the tense vowels have a rising slope. Four of the eight vowels are /u/ as paired with /b/, /d/, /k/, and /m/. Two of the eight vowels were /i/ as paired with /b/ and /m/. The remaining vowel is /o/ as paired with /k/. A rise in F1 is not suggested by the locations of all of these consonant constrictions. In this case, it is useful to look at the performance of the individual speakers.

Table 1 also gives the results for F1 slope for each speaker individually. Performance for the lax vowels is fairly consistent as compared to the performance in the tense vowels. All of the tense vowels with rising F1 slope are from speaker K.

Table 1: The slope of the track of the first formant is estimated across the vowel. Results are given for the lax vowels, /ɪ/, /ɛ/, /ɔ/, and /ʊ/ that have a positive slope, and the tense vowels, /i/, /e/, /o/, and /u/ that have a negative slope. There are 72 total vowels in each category.

Lax vowels with positive F1 slope	**Tense vowels with negative F1 slope**
66/72 = 91.7%	64/72 = 88.9%
Results by speaker	

Speaker	**Lax vowels**	**Tense vowels**
A	23/24 = 95.8%	24/24 = 100%
J	22/24 = 91.7%	24/24 = 100%
K	21/24 = 87.5%	16/24 = 66.7%

The other proposed measure is the location of the peak in energy in the F1 region expressed as a percentage of vowel duration. Results for this measure are given in Table 2. For this dataset, the best performance is obtained using a decision line of 38%, giving an overall correct classification of 82.6%. Moving the decision line by +/- 4% reduces the overall performance by 3.4%. The results for each speaker are also included in Table 2, and indicate that speaker differences are greater for the energy peak measure than for the F1 slope measure.

As mentioned above, this study is in support of a model of lexical access. In this model, the acoustic evidence for any feature is expected to be determined by a family of acoustic cues. In the case of tense/lax, these cues may include vowel duration, absolute formant values, some measure of spectral balance, and possibly the two measures discussed here. Figure 2 plots these two measures against each other to visualize the separation. This separation is not complete, but gives an indication that after combination with the evidence from other cues, a reliable separation could be achieved. Figure 3 plots the separation for each speaker individually.

Table 2: The peak in the F1 region (300-900 Hz) was estimated as a percentage of the total vowel duration. Results are given for lax vowels, /ɪ/, /ɛ/, /ɔ/, and /ʊ/, and the tense vowels, /i/, /e/, /o/, and /u/. The decision line is 38% of the vowel duration. There are 72 total vowels in each category.

Lax vowels with F1 energy peak greater than 38% of the vowel duration.	**Tense vowels F1 energy peak less than 38% of the vowel duration.**
60/72 = 83.3%	59/72 = 81.9%
Results by speaker	

Speaker	**Lax vowels**	**Tense vowels**
A	24/24 = 100%	19/24 = 79.2%
J	15/24 = 62.5%	23/24 = 95.8%
K	21/24 = 87.5%	17/24 = 70.8%

There are four examples for which both measures, F1 slope and F1 energy peak, fail to classify the vowel correctly. The one lax vowel example that fails for both measures is /k ɔ k/ as read by speaker K. The three tense vowels that fail for both measures are all spoken by speaker K and are /k u k/, /k o k/, and /m i m/.

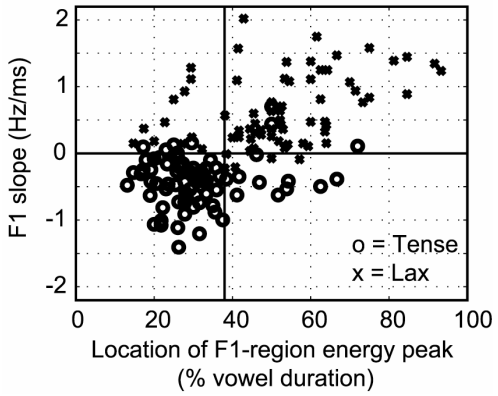

Figure 2: Scatter plot of all 144 vowel tokens for F1 slope versus location of the F1 energy peak in the vowel as a percentage of vowel duration.

4. SUMMARY

This work examines two acoustic measures that were predicted, based on expected movements of the articulators, to vary across the course of the production of tense and lax vowels in English. These measures are the slope of the F1 trajectory and the location of the energy peak in the F1 region. In this study of CVC sequences for stressed vowels in citation format, the slope of the first formant trajectory across the vowel is a robust marker for the tense/lax distinction. Based on the polarity of the slope alone, the overall percent correct classification is 90%. Similar, though not as consistent, results were obtained for the location in time of the F1 energy peak, where the overall percent correct classification is 83%.

ISBN 1-876346-48-5 © 2003 UAB

From an automatic detection point of view, the measure of F1 slope has the advantage that it is an absolute measure. The polarity of the slope alone is the relevant determinant. In the case of many parameters that appear to differ significantly between tense and lax vowels, a comparison must be made. In those cases, quantitative criteria must be developed for such terms as "longer", "breathier," and "earlier." From a linguistic point of view, it is an open question of whether or not the movement in F1 across the course of the vowel is an acoustic correlate of the feature [tense]. A study across languages would have to be made. However, the first step is to expand the English database to include: vowels that are not in stressed position, causal speech, and a greater number of speakers spread across gender.

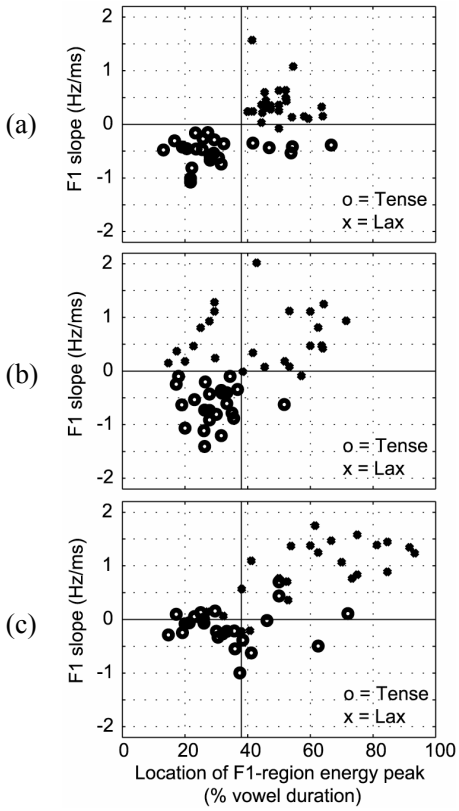

Figure 3: Scatter plot of vowel tokens for F1 slope versus location of the F1 energy peak in the vowel as a percentage of vowel duration as separated by speaker. (a) A (b) J (c) K.

ACKNOWLEDGEMENTS

The author would like to thank Ken Stevens for his time and advice. This work was supported in part by grant DC02978 from the National Institutes of Health.

REFERENCES

[1] R. Jakobson, G. Fant, and M. Halle, "Preliminaries to speech analysis: The distinctive features and their correlates," *Acoustics Laboratory Technical Report 13*, Massachusetts Institute of Technology, Cambridge, MA, 1952. Reprinted by MIT Press: Cambridge, MA, 1967.

[2] B. Bloch and G. Trager, *Outline of Linguistic Analysis*, Waverly Press: Baltimore, MA, 1942.

[3] N. Chomsky and M. Halle, *The Sound Pattern of English*, Harper and Row: New York, NY, 1968.

[4] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*, Blackwell Publishers Ltd: Oxford, UK, 1996.

[5] M. Halle and K.N. Stevens, "On the feature advanced tongue root," In *MIT Research Laboratory of Electronics Quarterly Progress Repor,* vol. 94, pp. 209-215, 1969.

[6] R. Harshman, P. Ladefoged, and L. Goldstein, "Factor analysis of tongue shapes," *Journal of the Acoustical Society of America,* vol. 62, pp. 693-707, 1977.

[7] M. Jackson, "Phonetic theory and cross-linguistic variation in vowel articulation," Ph.D. Thesis, University of California, Los Angeles, CA, 1988.

[8] G. Fant, "On the predictability of formant levels and spectrum envelopes from formant frequencies," in *For Roman Jakobson*, The Hague, Netherlands: Mouton, pp. 109-120, 1956.

[9] K. Denning, "The diachronic development of phonological voice quality", Ph.D. Thesis, Stanford University, San Francisco, CA, 1989.

[10] J. Kingston, N. Macmillian, L. Walsh Dickey, R. Thorburn, and C. Bartels, "Integrality in the perception of tongue root position and voice quality in vowels," *Journal of the Acoustical Society of America,* vol. 101, pp. 1696-1709, 1997.

[11] A.J. Lotto, L.L. Holt, and K.R. Kluender, "Effect of voice quality on perceived height of English vowels," *Phonetica*, vol. 54, pp. 76-93, 1997.

[12] D.H. Klatt, "Linguistic uses of segmental duration in English: acoustic and perceptual evidence," *Journal of the Acoustical Society of America,* vol. 59, pp. 1208-1221, 1976.

[13] K.N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features," *Journal of the Acoustical Society of America*, vol. 111, issue 4, pp. 1872-1891, 2002.

[14] A. Howitt, "Automatic syllable detection for vowel landmarks," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000.