Vowel Height and Duration

Ida Toivonen, Lev Blumenfeld, Andrea Gormley, Leah Hoiting, John Logan, Nalini Ramlakhan, and Adam Stone

1. Introduction

F1 correlates with vowel height: the higher the vowel, the lower the F1. A number of studies have found a positive correlation between F1 and duration in vowels in various languages, such as English (Heffner, 1937; House & Fairbanks, 1953; Peterson & Lehiste, 1960; Scharf, 1962), German (Meyer, 1904; Fischer-Jørgensen, 1940; Maack, 1949), Swedish (Elert, 1964), Inari Saami (Äimä, 1918; Stone, 2014), Thai (Abramson, 1962), and Spanish (Navarro Tomás, 1916). In other words, high vowels are shorter than low vowels. This paper revisits that generalization, as well as the question of whether the generalization is phonetic (mechanical, extrinsic) or phonological (controlled, intrinsic).

The traditional explanation for the positive correlation between F1 and duration appeals to physiology: low vowels take longer to produce because of the extra time it takes for the jaw to open (e.g., Lehiste 1970:19, Lehnert-LeHouillier 2007:80), or because the jaw position of high vowels is close to the jaw position held during the production of most consonants (Catford 1977:197; Maddieson 1997; Gussenhoven 2007). An alternative explanation is that each vowel has a phonologized duration target. Several arguments have been put forward for this view. Lisker (1974) points out that if low vowels are longer because of the time it takes the jaw to move, we would expect the onset and offset formant movements towards low vowels to be longer, not the steady-state of low vowels. However, the steady state is in fact remarkably long (Lehiste & Peterson, 1961). In addition, Tauberer & Evanini (2009) found that duration does not increase as vowels are lowered in language change. Finally, Solé & Ohala (2010) present data from Japanese, Catalan and English, where they investigate the effect of speech rate on vowel duration. For Catalan and English, they find that the duration differences have different size depending on speech rates. This is not expected if the effect is solely mechanical. The results differed for Japanese, where they found a constant change in duration as the speech rate changed. Solé & Ohala (2010) conclude that the positive correlation between vowel duration and F1 is controlled (phonological or high-level phonetic) in English and Catalan, but mechanical (low-level phonetic) in Japanese.

If the duration of vowels depends directly on how much the jaw moves, we would expect a positive correlation *within* categories as well as between categories. In other words, multiple tokens of the same vowel (e.g., [ɪ]) would be expected to display a correlation similar to the correlation between vowels; that is, a slightly lower pronunciation of a given vowel should be slightly longer.

We investigate the vowel duration and height between and within categories in English and Swedish, using F1 as a measure of vowel height. The between-category investigation confirms previous studies: high vowels are shorter than low vowels. However, we did not find the same correlation within categories: a higher instance of the vowel [ɪ] is not shorter than a lower instance of [ɪ].

* Carleton University. E-mail: ida_toivonen@carleton.ca
This paper has benefitted from discussion with the members of the Logic, Language and Information Lab at Carleton University, especially Raj Singh, Erik Anonby, Aysegul Kutlu, and Randall Gess. We also thank the audiences at the CUNY Phonology Forum, NAPhC8 and WCCFL for comments on this work and related work. Many thanks to Paul Melchin for RA support. Ida Toivonen also gratefully acknowledges support from Social Sciences and Humanities Research Council of Canada (SRG #410-2010-1841).

Cascadilla Proceedings Project
Completed October 14, 2014

2. English

The English study considered three different data sets, each considered in turn below.

2.1. English, Study 1

One set of English data were nonce words (tiff tivv keff kevv taff tavv) collected by Andrea Gormley. The words are all monosyllabic and of the form CVC with the vowels [ɪ], [ɛ], or [æ]. A total of 4730 vowels from four different speakers were analyzed. F1 values and durations were extracted with the PRAAT software. For details on data collection and PRAAT analysis for this English study, see Gormley (2010).

In this study and in all other studies we controlled for speaker, since individual speakers have slightly different vowel qualities. For example, male speakers tend to have overall lower formant frequencies than female speakers. The studies also controlled for the voicing of the final consonant, as vowels are on average lower before voiceless consonants than before voiced consonants (Moreton, 2004). We controlled for these factors by specifying them in an ordinary least square regression model. In addition, we eventually fitted separate models for each speaker, and also for voiced and voiceless final consonants to control for potential interactions. For these simpler analyses, we used basic t-tests and Pearson correlation tests. All results reported here reached a significance of $p < .01$ unless otherwise noted. For the statistical analysis, we used the software R.

For all four speakers in Study 1, [ɪ] was shorter than [ɛ], and [ɛ] was shorter than [æ]. In other words, F1 was higher for low vowels than for high vowels, so there was a positive correlation between F1 and duration in the between-category investigation. The table in (1) shows the average results for Speaker A in milliseconds and Hertz. The table shows the results from vowels before voiceless and voiced consonants separately: the numbers in parentheses are the average duration and F1 values of vowels before voiced consonants. Vowels are on average longer and higher (i.e., lower F1) before voiced than voiceless consonants (consistent with previous findings; see above), and this has been taken into account.

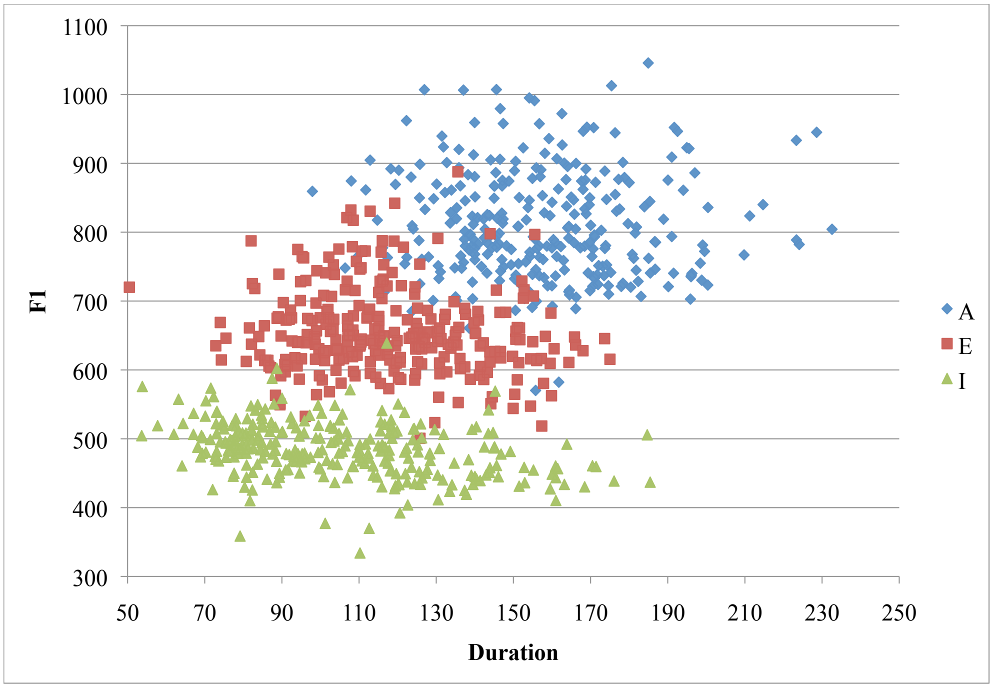
(1)

	Duration	F1
[ɪ]	87 (123) msec	496 (473) Hz
[ɛ]	103 (132) msec	669 (643) Hz
[æ]	147 (169) msec	821 (800) Hz

The results of the study were clear: when comparing vowels of different categories with each other (e.g., when comparing [ɛ] with [æ]), there is a *positive correlation* between F1 and duration: lower vowels (vowels with higher F1) are longer than higher vowels (vowels with lower F1).

This positive correlation is not found within categories. Figure 2 plots the F1 values (y-axis) against duration (x-axis) of all vowels uttered by Speaker A:

(2) English Study 1, Speaker A



The vowels [ɪ, ɛ, æ] are represented in different shades. Pearson correlation tests checking for a possible correlation between F1 values and duration within categories did not reach significance. In other words, *within* categories, we did not find a correlation between F1 and duration. This means, for example, that it is not the case that a slightly lower instance of the vowel [æ] will be slightly longer.[1]

Let us summarize the first English study. The study confirmed the previous finding that high vowels are shorter than low vowels. However, this generalization does not seem to hold within categories. The speakers we examined did not display a positive correlation between F1 and duration within categories. There is a potential problem with using Gormley's data for this study, however, as the study used nonce words and was designed to be a tongue twister. We examined both the intended tongue twisters and the controls and did not find a difference between the two types of stimuli. However, it is possible that actual English words in a different type of experiment would yield different results. We therefore performed two additional studies on English vowels.

2.2. English Study 2

In the second English study, three native speakers of English were asked to read words in a carrier phrase ('say X to me'). The word list consisted of five sets of minimal or near-minimal triplets: *sit, set, sat, bit, bet, bat, big, beg bag, give, Bev, jazz, miss, mess, mass*. Each word was repeated six times and the entire list was randomized differently for each speaker. The subjects were shown one word at a time on a computer screen. The words were presented with even timing (using Power Point). Each session was recorded with a Tascam solid state recorder. The target words were segmented in PRAAT and we used a PRAAT script (a modified version of a script made by Mietta Lennes, http://www.helsinki.fi/~lennes/praat-scripts/) to extract the duration and F1 of the vowels of the target words.

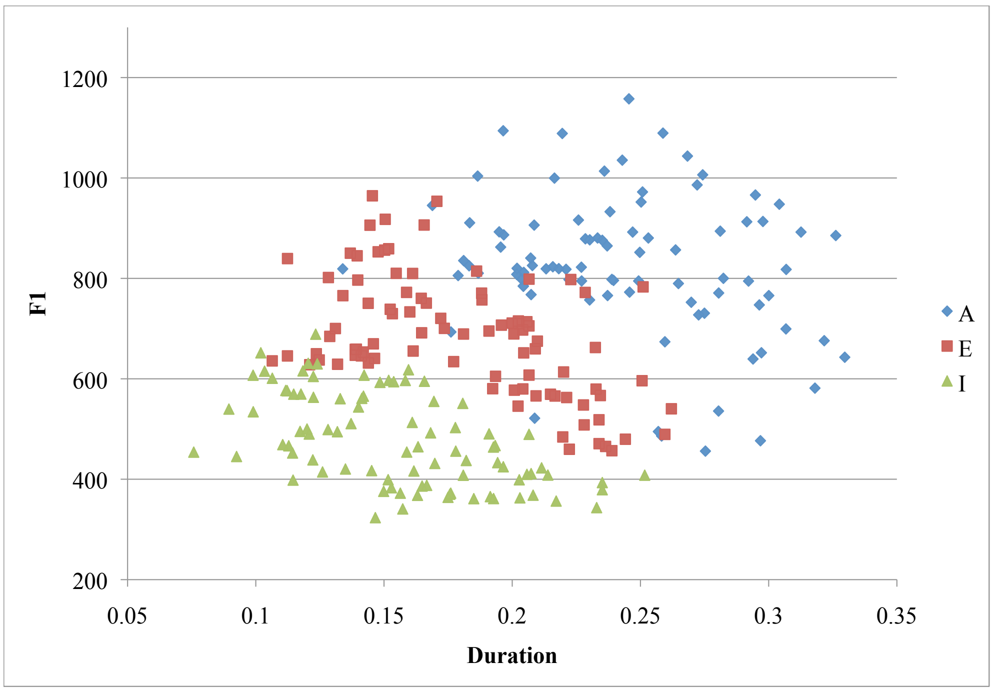
[1] However, there was a weak *negative* correlation between F1 and duration for the [ɪ] vowel for speaker A. In our studies, we occasionally (in individual sounds for individual speakers) find such a negative correlation within a specific category for a specific speaker, but we do not have an explanation for this.

Even though the design and scope for this study were different, the results were overall the same as in Study 1. When comparing between categories, there is a positive correlation between F1 and duration: [ɪ] is shorter than [ɛ], and [ɛ] is shorter than [æ].[2] However, there is no such positive correlation within categories. Duration and F1 values do not covary when different utterances of the same vowel are compared. The average F1 and duration values are given in (3), and the measurements are plotted in (4):

(3)

	Duration	F1
[ɪ]	156 msec	477 Hz
[ɛ]	183 msec	683 Hz
[æ]	241 msec	825 Hz

(4) English Study 2



The within-category negative correlation that is apparent in Figure 2 generally goes away when the voicing of the final consonant is controlled for.[3]

2.3. English study 3

To compare more tokens of exactly the same type (i.e., more instances of the same word) with a smaller set of types (words), we carried out a third study. English study 3 is the same as English study 2, except the stimuli are slightly different. Three native speakers of English were asked to read the words *miss mess mass bit bet bat*, all words with a final voiceless consonant. Each word was read ten times in a carrier phrase. The results were the same as in Study 1 and 2: [ɪ] is shorter than [ɛ], and [ɛ] is shorter than [æ]. However no F1–duration correlation was found in the within-category analysis. The F1 and duration averages are given in (5), and all F1 and duration values are shown in (6):

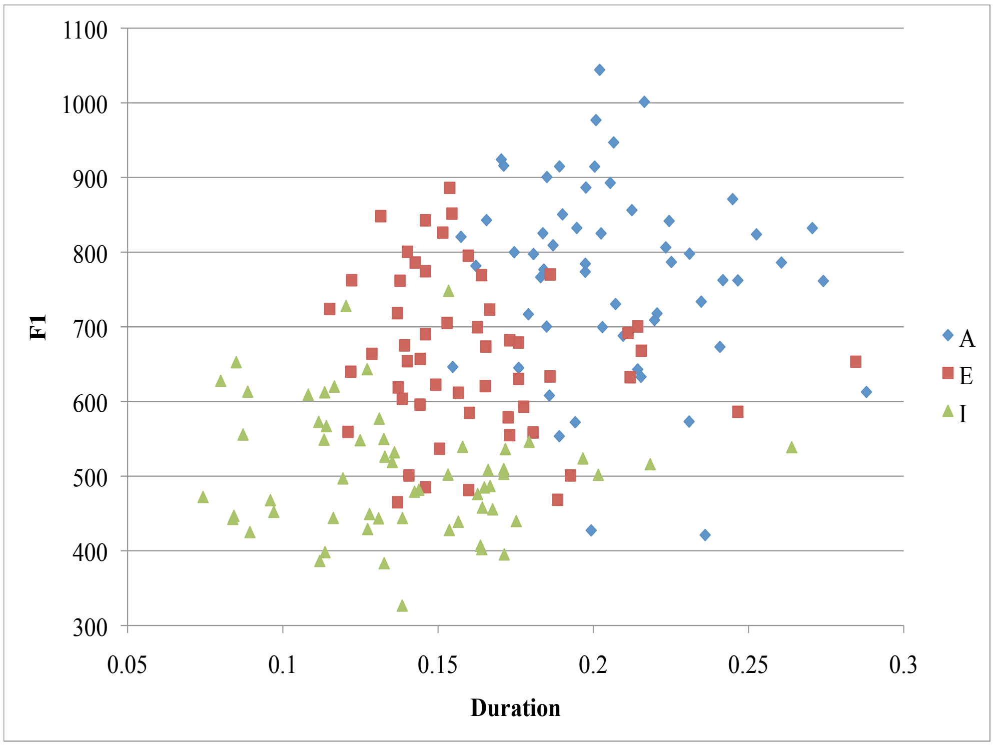
[2] The difference in duration between [ɪ] and [ɛ] was not significant for one of the three speakers. This goes against the widely accepted generalization that high vowels are shorter than lower vowels, but it is consistent with Solé & Ohala's (2010) study, where the results regarding the durations of high and mid vowels were inconclusive.

[3] However, the vowel [æ] is an exception: within this vowel, there is a weak (unexpected) *negative* correlation between F1 and duration for one speaker, even when the voicing of the final consonant is controlled for; see also footnote 1. We return to this in the conclusion.

(5)

	Duration	F1
[ɪ]	138 msec	505 Hz
[ɛ]	162 msec	662 Hz
[æ]	207 msec	772 Hz

(6) English Study 3



2.4. English studies: summary

The English studies confirm the well-known generalization that high vowels are shorter than low vowels. However, there is no correlation between vowel height and duration within categories. That is, a slightly higher instance of a given vowel is not shorter than another instance of the same vowel in English.

3. Swedish

In addition to the English studies, we examined vowel height and duration in Swedish. Swedish has a complementary quantity system (Elert, 1964; Eliasson & LaPelle, 1973; Eliasson, 1985; Schaeffler, 2005), where long vowels are consistently followed by short consonants, and short vowels are followed by long consonants, as illustrated in (7):

(7)

Long V words:		Short V words:	
[haːt]	'hatred'	[hatː]	'hat'
[kaːl]	'bare'	[kalː]	'cold'
[heːta]	'be called'	[hetːa]	'heat'

We examined Swedish vowel height and duration in two separate studies.

3.1. Swedish Study 1

The first study recorded six native speakers of Swedish. Every speaker was asked to pronounce 60 different words five times each in a carrier phrase. The word list was randomized for each speaker. The list consisted of both monosyllabic and disyllabic words, and the monosyllabic words were all closed syllables. The disyllabic words all had initial stress and we considered only the stressed vowels. When examining the data, we considered whether the vowel was inherently long or short. We also considered the quality of the immediately following consonant; whether it was voiced or voiceless and whether it was a fricative or an oral or nasal stop. The table in (8) gives the average F1 and duration for the low central/back unrounded vowel [a/ɑː] and the high front unrounded vowel [ɪ/iː] for speaker KS, and the figure in (9) plots KS's F1 and duration for the same vowels:

(8)

	Duration	F1
[ɪ]	190 ms	404 Hz
[a]	256 ms	720 Hz

(9) Swedish Study 1, Speaker KS

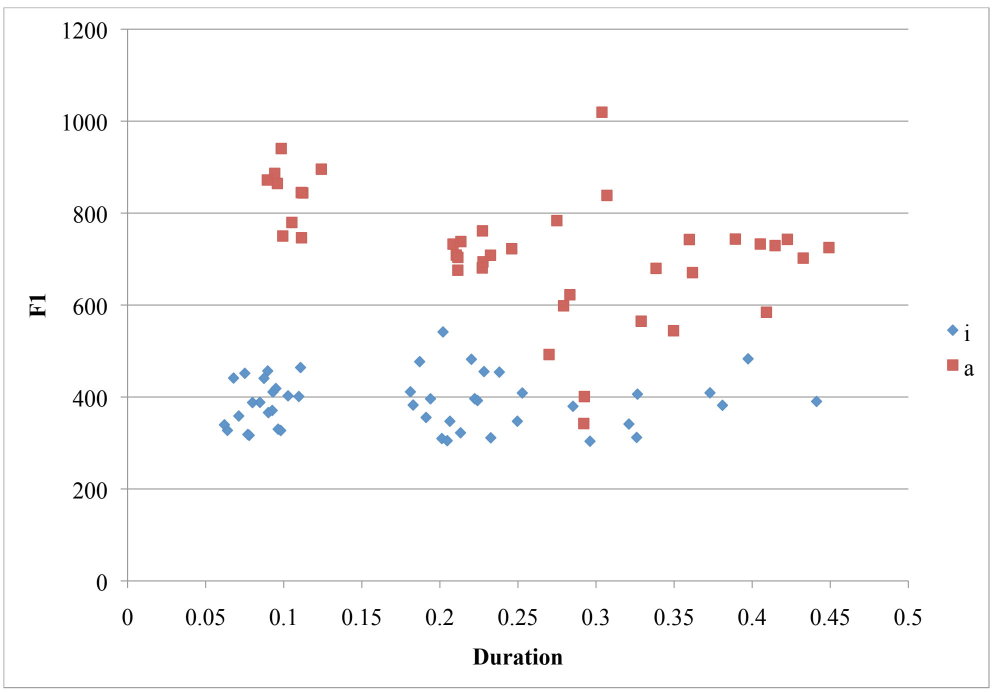


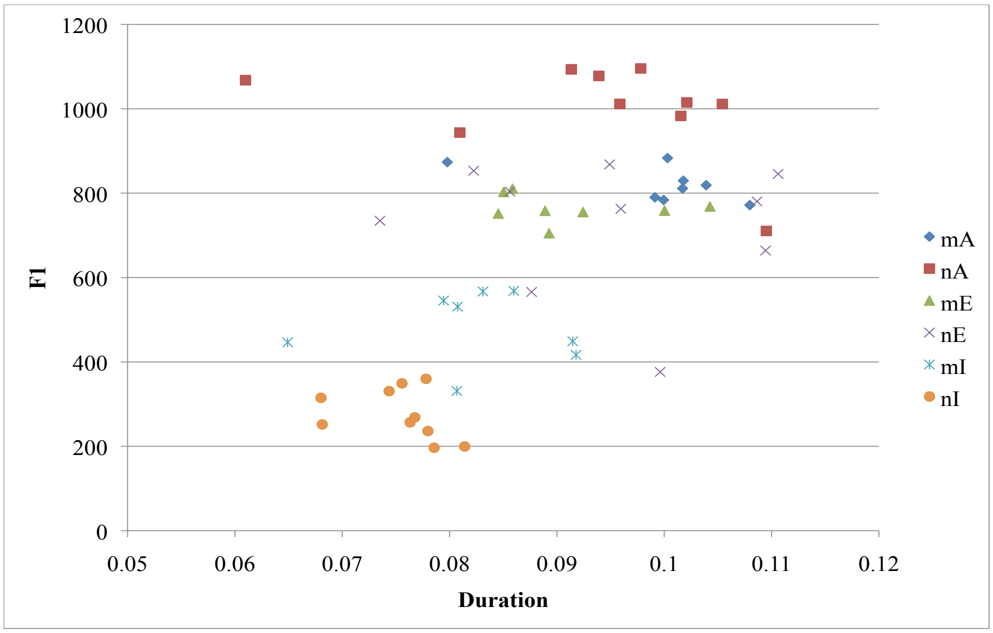
Figure (9) illustrates the difference between inherently short and long vowels: the smaller clusters between 50 and 100 msec are short vowels.

A comparison between categories revealed that high vowels are generally shorter than low vowels in Swedish, as has been previously reported (Elert, 1964). The results were consistent across speakers. That is, there is a positive correlation of F1 and duration between categories in Swedish. The within-category analysis generally showed no F1-duration correlation, but we also found some instances of a weak negative correlation within a category. For example, for speaker KS, we found no within-category correlation between F1 and duration for the vowel [ɪ/iː], and a weak negative correlation within the [a/ɑː] category (cf. footnotes 1 and 3). We return to this unexpected negative correlation in the conclusion.

3.2. Swedish Study 2

Since the stimuli were quite varied in the first Swedish study, we ran a second, smaller study with a more limited word list. In this study, two native speakers of Swedish were asked to read the words *missa, messa, massa* in a carrier phrase (*säg X nu*). One speaker repeated the words eight times, the other speaker repeated the words ten times. We analysed the vowel in the initial, stressed syllable of the words. The figure in (10) plots the F1 and duration values of both speakers ("m" is Speaker 1 and "n" is Speaker 2):

(10) Swedish Study 2



For both speakers, the high *missa* vowel is shorter than the mid *messa* vowel, and the *messa* vowel is in turn shorter than the low *massa* vowel. Within categories, we found no correlation between F1 and duration.

4. Conclusion

Overall, the conclusions of our studies are clear: higher vowels are generally shorter than lower vowels, as has previously been found. However, two different realizations of the same vowel do not display the same tendency. In other words, the within-category results do not mirror the between-category results; at least not in English and Swedish. One might speculate that the within-category effect is so small that it would come out statistically only with a much larger sample. However, we do not find this likely, since we actually found occasional instances of a weak ($p < .05$) *negative* F1-duration correlation within categories. This tendency goes against the expected positive tendency. We do not have an explanation for this (very slight and occasional) negative correlation, but the finding makes it seem unlikely that a larger sample space would result in a positive correlation between F1 and duration. We note again that this negative correlation prevailed even when we controlled for the voicing of the following consonant, which would otherwise exaggerate the effect. We do not have an explanation for the occasional negative correlation, but we find it intriguing and set it aside for future research.

We interpret these data as problematic for the physiological explanation on intrinsic vowel duration. If the effect is purely mechanical, we would expect to find it within as well as between categories. Our results instead support a view where duration is an important cue for distinguishing between vowels in English; it is not a mere side effect of jaw movement. However, this leaves us with an important question: If the jaw opening explanation is incorrect, what explains the universal tendency for high vowels to be shorter than low vowel? Solé & Ohala (2010:647) note: "Duration is one of several distinctive manifestations of vowel identity, but the specific durational targets of vowels may have originated in biomechanical differences. The cross-linguistic tendency then has a physiological explanation, but this tendency has been phonologized in some languages and not others." This reasoning makes sense to us.

However, the results of this study (and also the results of Lisker (1974), Tauberer & Evanini (2009), and Solé & Ohala (2010), mentioned in the introdution) suggest that a straightforward production-based analysis is problematic, at least for Swedish and English. Perhaps a perception-based analysis is possible? Perhaps high vowels somehow "sound" shorter than low vowels and this gets phonologized? This is still an open question, but the results of Gussenhoven (2007) instead seem to suggest that high vowels in fact sound *longer* than short vowels.

We conclude by noting again that higher vowels indeed are shorter than lower vowels and this seems to be a *controlled* feature of the grammar; that is, it is an effect of phonology or at least high-level phonetics. However, it remains to be explained why this generalization should hold cross-linguistically.

References

Abramson, A.S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Indiana U. Research Center in Anthropology, Folklore, and Linguistics, Pub. 20, Bloomington.

Äimä, Frans (1918). *Phonetik und Lautlehre des Inarilappischen. I. Beobachtungsphonetik und deskriptive Lautlehre. II. Instrumentale Versuche und Messungen*, vol. 42–43 of *Mémoires de la Société Finno-Ougrienne*. Suomalais-ugrilainen seura, Helsinki.

Catford, John Cunnison (1977). *Fundamental Problems in Phonetics*. Edinburgh University Press, Edinburgh.

Elert, C.-C (1964). *Phonologic Studies of Quantity in Swedish*. Almqvist & Wiksell, Uppsala.

Eliasson, Stig (1985). Stress alternations and vowel length: New evidence for an underlying nine-vowel system in Swedish. *Nordic Journal of Linguistics* 101–129.

Eliasson, Stig & Nancy LaPelle (1973). Generativa regler för svenskans kvantitet. *Arkiv för nordisk filologi* 133–148.

Fischer-Jørgensen, Eli (1940). Objektive und subjektive Lautdauer deutscher Vokale. *Zeitschrift für die gesamte Phonetik* 4, 1–20.

Gormley, Andrea (2010). *Phonological Mismatch Errors: An Acoustic Analysis*. Ph.D. thesis, Carleton University.

Gussenhoven, Carlos (2007). A vowel height split explained: Compensatory listening and speaker control. Cole, Jennifer & José Hualde (eds.), *Papers in Laboratory Phonology 9: Change in Phonology*, Mouton de Gruyter, Berlin, 145–172.

Heffner, Roe-Merrill Secrist (1937). Notes on the length of vowels. *American Speech* 12:2, 128–134.

House, A.S. & G. Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, 268–277.

Lehiste, Ilse (1970). *Suprasegmentals*. The MIT Press, Cambridge, MA.

Lehiste, Ilse & Gordon E Peterson (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America* 33, 268–277.

Lehnert-LeHouillier, Heike (2007). *The Perception of Vowel Quantity: A Cross-Linguistic Investigation*. Ph.D. thesis, State University of New York, Buffalo.

Lisker, Leigh (1974). On "explaining" vowel duration variation. Tech. Rep. SR-37/38, Haskins Laboratories, New Haven, Connecticut.

Maack, Adalbert (1949). Die spezifische Lautdauer deutscher Sonaten. *Zeitschrift für Phonetik* 3, 190–232.

Maddieson, Ian (1997). Phonetic universals. Hardcastle, W.J. & J. Laver (eds.), *The Handbook of Phonetic Science*, Blackwell, Oxford, 619–639.

Meyer, Ernst A. (1904). Zur Vokaldauer im Deutschen. *Nordiska Studier tillegnade Adolf Noreen*, K.W. Appelberg, Uppsala, 347–356.

Moreton, Elliott (2004). Realization of the English postvocalic [voice] contrast in f1 and f2. *Journal of Phonetics* 32, 1–33.

Navarro Tomás, Tomás (1916). Cantidad de las vocales acentuadas. *Revista de Filología Española* 3, 387–408.

Peterson, Gordon E. & Ilse Lehiste (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32:6, 693–703.

Schaeffler, Felix (2005). *Phonological Quantity in Swedish Dialects: Typological aspects, phonetic variation and diachronic change*. Ph.D. thesis, Umeå University, Umeå.

Scharf, D. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5, 26–30.

Solé, Maria-Josep & John Ohala (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. Fougeron, Cécile, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.), *Papers in Laboratory Phonology 10*, de Gruyter, Berlin, 607–655.

Stone, Adam (2014). Vowel height and duration. *Institute of Cognitive Science Annual Spring Conference*, Carleton University.

Tauberer, Joshua & Keelan Evanini (2009). Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. *Interspeech* 2211–2214.