Russell Marvin

LIN 499

Karthik Durvasula

18 October 2021

The effect of speech rate on the voicing effect in English

Table of Contents

Acknowledgements1

Introduction1-2

Research question2

Background2

Type chapter title (level 3)3

Materials4

Type chapter level (level 2)5

Type chapter title (level 3)6

Measurements4

Type chapter level (level 2)5

Type chapter title (level 3)6

Discussion4

Type chapter level (level 2)5

Type chapter title (level 3)6

Conclusion4

Type chapter level (level 2)5

Type chapter title (level 3)6

Bibliography4

**1. Acknowledgements**

I would like to acknowledge the assistance of Dr. Durvasula in conducting this research project. I also thank the many participants who provided my data, both this semester and last.

**2. Introduction**

The basis for this project is the well-established fact that vowels preceding voiced consonants have longer durations than those which precede voiceless consonants. This voicing effect (hereafter VE) is the subject of my study. My question is whether the rate of speech of the speaker has an impact on the VE, and if so, how?

There are multiple theoretical explanations for the VE. According to Durvasula 2014, the most recent account uses speech perception to explain the existence of the VE. The key to this explanation is that they draw a link between the existence of the VE and “the fact that closure durations of voiced consonants are shorter than those of voiceless consonants.” The idea here is that to increase the contrast for listeners, speakers produce a long vowel which “enhances the perceptual cue of a short closure duration on the following consonant…” In essence, the context provided by the vowel length helps to enhance the perceived shortness of the voiced consonant closure and vice versa.

Given this background information, the central question is how the VE is affected by speech rate. Does the difference between vowels preceding voiced and voiceless consonants change based on the rate of speech at all? Does it change proportionally to the difference in speech rate? I hypothesize that the VE will change in proportion to the change in speech rate to maintain a constant perceptual distance between voiced and voiceless consonants. This is based on research by Solé (2007) which states that “acoustic properties targeted by the speaker adjust to durational variations triggered by changes in speaking rate so as to maintain a constant perceptual distance across rates…” The ratio of the length of vowels preceding voiced consonants to those preceding voiceless consonants has been suggested to be between 1.2 to 1.6, according to Tanner (2019, p. 1). Tanner describes the VE as being “larger than otherwise explainable by purely articulatory properties” (p. 1). This description suggests we should predict the VE to be roughly proportional across speech rates. For example, we might expect a 1.4 voiced to voiceless ratio at slow, normal and fast speech rates, with both vowel lengths increasing or decreasing *together* based on the rate.

Also relevant to my research question, Tanner notes that “the size of the effect has been shown to be at least partially modulated by the effects of… speech rate” (p. 1). Later continuing, “it is possible that VE is neutralized on fast speech, where there may be some upper limit on the extent to which a syllable can be shortened, resulting from some form of ‘maximal compression’ of a vowel.” According to Port and Dalby (1982), “Numerous studies that employ the tempo of a carrier sentence as a variable have demonstrated that the interpretation of temporal cues to segmental contrasts is influenced by the tempo of the context… and generally indicate that perception compensates for changes in the temporal patterns of speech produced at different tempos. “

**3. Materials**

My experiment was conducted using a variety of very basic equipment. For several of the participants, I used my own Macbook Air internal microphone to record. Others used their own internal laptop microphones. This led to some small issues with noise that will be discussed in the final section.

The software I used for the recordings was a web app called Jotform which is an online form builder. Within the application I created a survey to prompt participants to use a microphone recording widget which uses their internal microphones by default. The form first presented each participant with a sample of speech at a given speech rate (fast or slow). They heard a recording of me saying “They say the slogan while they see my boy by the highway.” Then they were asked to begin recording and read a list of sentences: “I will say vat now. I will say cud now. I will say battle now. I will say bicker now.” and so on.

The stimuli used for the experiment were a combination of 13 minimal pairs of words that were contrasted by the voicing of a consonant following a vowel. Each word was placed in a carrier phrase “I will say \_\_\_ again.” In addition to the combination of 26 sentences with the relevant minimal pairs. The experiment also included 13 minimal pairs of words that were contrasted by the voicing of a consonant preceding a vowel. This was done to disguise the feature of interest. Once the participant read and recorded all 52 sentences, they stopped recording and moved on to a “normal speed” recording and recorded themselves reading at that speed. Finally they heard a third speed prompt (slow or fast, depending on which they had not yet done.) Ultimately each participant read all 52 sentences at a slow, normal and fast speed, providing our data. Four of the nine participants completed the experiment in the order “fast, normal, slow” and half completed it in the order “slow, normal, fast.” This was to ensure that the order did not affect the speeds at which they spoke.

The consonants in these pairs were obstruents: stops (e.g. [k]), affricates (e.g. [tʃ]) or fricatives (e.g. [f]). Of the 13 minimal pairs used for the experiment results, the voiced and voiceless words were compared in terms of frequency. The average Lg10 word frequencies between the voiced and voiceless words were similar, 2.55 for the voiceless group and 2.9 for the voiced group. This was done to ensure that the speed of pronunciation of the vowels in question was not the product of a greater familiarity with the words themselves, rather than the speech rate. Research suggests that “high frequency words may both contribute to shortening (leading to maximal compression) and reduce the need for the contrastive effect of VE” (Tanner, 2019, p.1)

The reason that I used a carrier phrase was to get an identical context for the word each time. This avoids creaky voice differences in phrase initial and phrase final contexts, changes in intonation, and lengthening of words at the end of a sentence. The word “now” was used as the following word because a word that started with a vowel, such as “again,” could have caused the final consonant to take onset position as opposed to coda position. This would cause problems for measuring the duration of closure or the length of consonants following our vowels, however this did not end up as part of my calculations. This will be further discussed in the next section.

**4. Measurements**

The phonetic variable that was measured in this experiment was the length of the vowel preceding voiced and voiceless consonants. The use of minimal pairs allows us to examine the effect of voicing discussed above. The beginning of each vowel was identified by the appearance of a very low “voicing bar” in the spectrogram on Praat, several dark formants in the spectrogram, an inverted “U” shape in the waveform, a consistent waveform, and vertical striations in the spectrogram. There is no perfect point, but I used the most consistent combination of indicators to get useful measurements that began and ended at the same points from sample to sample. Often times the end of the vowel is more difficult to pinpoint than the beginning, so for that the solution was mostly the same but with more time spent comparing samples and choosing a consistent endpoint.

ADD INFO ABOUT VOWEL/CONSONANT RATIOS

**5. Tabulations and Charts**

As it is displayed here, the data is not in a tidy format. The data was rearranged in order to use python matplotlib to create plots.

**Table

Description automatically generated**

Table

Description automatically generated with medium confidence

Table 1 and 2: the vowel length measurements preceding voiced consonants (in milliseconds)

Table

Description automatically generated

Table

Description automatically generated with medium confidence

Table 3 and 4: the vowel length measurements preceding voiceless consonants (in milliseconds)

|  |  |  |
| --- | --- | --- |
| **Speed** | **Voiced** | **Voiceless** |
| Fast | 121.7 | 95.69 |
| Normal | 153.81 | 112.38 |
| Slow | 220.05 | 149.62 |

Table 5: averages of vowel length (in milliseconds) for voicing and speech rate, calculated using “avg” from tables 1-4

Chart, bar chart

Description automatically generated

Plot 1: visualization of Table 5, showing avg. vowel length (in milliseconds) for voicing and speech rate

|  |  |
| --- | --- |
| **Speed** | **Difference (Vd-Vl)** |
| Fast | 26.0280342 |
| Normal | 41.4352137 |
| Slow | 70.4233333 |

Table 6: Difference in avg. vowel length (in milliseconds) at different speech rates (voiced – voiceless)

Chart, bar chart

Description automatically generated

Plot 2: visualization of Table 6, showing the difference in milliseconds of average vowel length between vowels that precede voiced vs voiceless consonants (voiced-voiceless)

|  |  |
| --- | --- |
| **Speed** | **avg. dur** |
| Fast | 108.695 |
| Normal | 133.095 |
| Slow | 184.835 |

Table 7: using voiced and voiceless contexts, average duration of vowels calculated in milliseconds for each speed rate

|  |  |
| --- | --- |
| **Speed** | **Proportion** |
| Fast | 1.272064696 |
| Normal | 1.368722001 |
| Slow | 1.470671871 |

Table 8: the proportion of the length of vowels preceding voiced consonants to the length of vowels preceding voiceless consonants – can be interpreted as the “strength” of the VE

**6. Discussion of results and problems**

**6.1 Observations**

These results show that as speech rate increases (“Fast” rate), the VE diminishes and as speech rate decreases (“Slow” rate), the VE increases. This is shown most clearly in Plots 1 and 2, where Plot 2 shows the difference in avg vowel length increasing at a Slow rate of speech and decreasing at a Fast rate of speech. Another way to put this result is that vowels preceding voiced and voiceless consonants were *more* similar than normal in fast speech and *less* similar than normal in slow speech.

In analyzing my data using the values from Table 7, I observed that Fast vowels were 7.06% shorter than Normal vowels and Slow vowels were 7.45% longer than Normal vowels. At the same time, using the values in Table 6, Fast vowels were 37.18% more similar than Normal vowels in duration and Slow vowels were 69.96% more different than Normal vowels. The difference in those percentages (37.18 – 7.06 = 30.12% for Fast vowels) and (69.96 – 7.45 = 62.51% for Slow vowels) is a change in the VE that is *not* accounted for by the change in speech rate, as measured by the length of vowels. This means that the VE changes more than the rate of speech by 30.12% and 62.51% for Fast and Slow vowels respectively. INCLUDE more RESEARCH ABOUT VOWEL/CONSONANT RATIOS.

As Table 8 shows, it is neither the case that VE scales up and down with speech rate to maintain a consistent proportion (of vowel length preceding voiced/voiceless consonants), nor is it the case that this difference is explained by the change in speech rate (previous paragraph). The theory that this change in VE could be explained by the change in speech rate was primarily based on the idea of the compression of vowels at a fast speech rate. My thought was that when given a smaller window of time to articulate a word (fast speech rate), the more similar the vowels would be in a voiced and voiceless context. Given the physical constraints of articulating sounds, when speaking at a faster rate, the speaker has to choose (unconsciously) where to shorten the length of certain sounds. There seems to be a lower bound on the length of time it takes to articulate certain sounds without the speech becoming completely impossible to understand. It seems like it is then a necessity to take away portions of speech disproportionately. For example, if a speaker is speaking at a rate two times the normal rate, perhaps they cannot feasibly make every part of every sound half as fast. We can compare this to an audio recording that is sped up to two times its original rate, in which case we can assume that every part of every sound is taking half the time to play. The consequence of a faster rate of speech then is less perceptual distance between voiced and voiceless consonants via a diminished VE and vice versa for a slow rate. It is ultimately most important that there is a difference in VE between speech rates at all.

Also important to note is the fact that the VE is not consistent between speech rates. This is not a feature that is unaffected by speech rates, rather it does seem to be inversely correlated with speech rate based on Plot 2. Solé 2007 states that “low-level phonetic effects remain constant across rates or should vary in a way that could be explained by physical and physiological principles of speech motor control” (p. 314). This suggests that the VE is *not* one of these “low-level phonetic effects.” So according to Solé, the VE must be one of the acoustic properties “targeted by the speaker [which] adjust to durational variations triggered by changes in speaking rate so as to maintain a constant perceptual distance across rates” (p. 314). In laying out this dichotomy, Solé refers to several proposals on the subject and notes that all of them state that “variations in segment duration… trigger the temporal adaptation of features planned by the speaker, but have no consequence (or merely phonetically predictable consequences) on low-level mechanical effects” (p. 314). “Segment duration” here can be taken to mean speech rate given our context.

According to Solé 2007, which references Cuartero (forthcoming), it was found that “in English, slower speaking rates increased the size of the intended vowel duration contrast before voiced and voiceless stops, that is to say, vowel-duration differences are proportional to the duration of the vowel, larger for longer vowels at slower rates and smaller for shorter vowels at fast rates” (p. 318). This is somewhat similar to my results in that the absolute difference between vowels preceding voiced and voiceless consonants increases at slow speech rates (and vice versa), but the change in my results is not proportional. Rather, the VE is increased in relative terms for slow speech rate and diminished in relative terms for fast speech rate.

This leaves the results of my experiment in a sort of middle-ground, as far as I can tell given what I know right now and have read. There certainly is a change in the absolute difference in length between vowels preceding voiced vs voiceless consonants. However, this change is not proportional. My initial reaction was to call the VE one of the “low-level phonetic effects” that could vary in a way which is “explained by physical and physiological principles of speech motor control”. Yet the inverse correlation between the VE and speech rate indicates that it must indeed be an acoustic feature targeted by speakers. On top of that, it is widely recognized that the VE is larger than can be explained by factors of articulation, as discussed by Tanner (2019). However, the adjustment made to vowel durations at different rates of speech does *not* maintain a constant perceptual distance between voiced and voiceless sounds. This may be the result of a mistake I made by measuring vowel duration only. According to Solé 2007, “the cue to postvocalic obstruent voicing has been characterized as a change in the ratio of the vowel-to-consonant duration in the syllable,” (p. 315) In fact, Port and Dalby assert that “the C/V ratio is the primary factor employed by the subjects to decide on the voicing value of the medial consonants…” in an experiment on a topic similar to this one. This ratio was not taken into account in my experiment. I instead used vowel length as the measured value given that there was so much variation in the pronunciation of the stops due to glottalization. It could be the case that the VE remains proportional at different speech rates when the consonant duration is taken into account. As of now, this would be my assumption, given that Solé 2007 seems to find that the effect of things like the VE varies proportionally to the speech rate in English, as referenced above.

**6.2 Problems**

A potential issue is that I used vowel length as a measurement for speech rate when calculating exactly how fast or slow the “Fast” and “Slow” vowels were compared to normal. This assumes that vowel length changes proportionally with speech rate. This may well be the case, but it could be a better idea to measure the length of each carrier phrase to determine speech rate, which is something that could be done moving forward.

It is possible that the size of the VE in my results was diminished by the frequency of the voiced and voiceless groups being slightly different. Given that the voiced frequency was 2.9 while the voiceless frequency was 2.55, I had to make a judgement that these were sufficiently similar. The purpose of the experiment was not to simply judge the size of the VE so in my opinion this should not have a significant impact on the results with regards to the *change* in VE at different speech rates.

**Bibliography**

Solé, M.J. 2007. [Controlled and mechanical properties in speech: a review of the literature](http://pagines.uab.cat/mariajosepsole/sites/pagines.uab.cat.mariajosepsole/files/8.-18-SoleBeddorOhala-chap18.pdf). In Solé, M.J, Beddor, P. & M. Ohala (eds.). *Experimental Approaches to Phonology.* Oxford: Oxford University Press, 302-321.

Tanner, James & Sonderegger, Morgan & Stuart-Smith, Jane & Consortium, SPADE. (2019). Vowel duration and the voicing effect across English dialects. Toronto Working Papers in Linguistics. 41.

Durvasula, Karthik & Luo, Qian. (2012). Voicing, aspiration, and vowel duration in Hindi. The Journal of the Acoustical Society of America. 132. 2002. 10.1121/1.4755407.

Port, Robert F., and Jonathan Dalby. “Consonant/Vowel Ratio as a Cue for Voicing in English.” *Perception & Psychophysics*, vol. 32, no. 2, 1982, pp. 141–152.