Orthographic Effects on Phonetic Cue Weighting

Grant McGuire

Department of Linguistics, University of California, Santa Cruz
`gmcguir1@ucsc.edu`

Abstract

This paper presents data that tie together two different phenomena; the influence of alphabetic writing on phonemic awareness and the shift in sibilant cue weighting that English-learners undergo during childhood. I present evidence from English listeners' labeling of a non-native sibilant contrast that attention to cues (frication noise vs. formant transitions) can be directed orthographically—i.e. the graphemes given to listeners for labeling determine their cue weighting for the contrast. I further propose that the childhood cue shift is due to learning to read and write English.

Index Terms: cue weighting, orthography, phonemic awareness

1. Introduction

There is considerable evidence that phonemic awareness (PA), the ability to decompose the speech signal into segment sized units, is strongly influenced by one's writing system, such that learning an alphabetic writing system focuses attention on the segment as a unit separate from other larger units such as syllables [1]. However, the articulations that give rise to the acoustic phonetic cues to such contrasts are gradual and do not neatly fit into isolated segmental gestures. The consequence of these coarticulatory effects is that the perceptual cues for identifying a given contrast may be located not only in the segment itself, but also the adjacent segments. For some contrasts, such as plosives, these “adjacent” cues are in fact the most salient [2,3]. This paper demonstrates experimental evidence that orthography can impact such sub-segmental cues and provides evidence that the sibilant cue weighting shift seen in English acquiring children [4] may be due to learning to read and write.

The primary evidence for orthographic influences on speech perception comes from three sources: the correlation between PA and reading ability in children, comparisons of literate and non-literate adults, and comparisons of speakers having different orthographic traditions. As to the first, there is a long history of evidence that children with better reading abilities also show greater PA [5]. While pre-literate children are aware of onset vs. rime characteristics [6], they are generally unable to fully decompose a word into phones [7]. However, children with more schooling show improved abilities at PA, independent of age [8], and children who are better readers overall compared to their peers perform better at PA tasks [9].

As might be expected given the acquisition data, literacy and PA effects are also found in adults. In an important series of studies, Morais and colleagues found that Portuguese illiterates performed much more poorly than literates on a series of PA tests [10,11]. These studies demonstrated that illiterate adults had extreme difficulty in identifying onsets, unlike the literate subjects. While the illiterates were capable of identifying basic syllable and rime components, the literate subjects were considerably better at that task, too. Subjects who acquired reading skills in adulthood showed intermediate abilities.

All of the research above deals with subjects who speak (or are acquiring) a language that uses an alphabetic orthography; other evidence shows that non-alphabetic systems do not necessarily help with PA. For example, [12] found that Chinese speaking adults literate only in Chinese characters (logographic) performed similarly to the illiterates in [10], while those with training in Pinyin (alphabetic) performed similarly to the literates. Such effects are robust and far-reaching; for instance [13] found that these orthographic differences not only affect metalinguistic analysis, but also online speech processing using a primed shadowing task.

While the literature strongly supports the view that many aspects of PA derive from learning an alphabetic orthography, the possibility that sub-segmental cues can be manipulated has been only minimally explored. In [14], 5 and 6-yr old English-learners with different amounts of schooling were given a discrimination task using [b]~[p] and [p]~[p^{h}] stimuli and found strong evidence that phoneme to grapheme learing assists in privileging phonemic over allophonic variation. [15] demonstrated that children with higher PA also show more adult-like cue weighting for sibilant fricatives. Sibilants have cues to their place of articulation both in their frication noise, arising from the constriction made by the tongue in the vocal tract as well as cues in the vocalic transitions into and out of the fricative. As has been amply demonstrated, young children tend to give more weight to the formant transitions than the frication noise, yet, later in childhood, they begin to place more weight on frication noise cues, in much the same way that adults do [15,16,17]. While [14] makes the connection with PA, she does not seem to view orthography as causal, rather the development of PA allows the listener/reader to determine the proper structure.

Given the evidence presented for the development of PA, I believe this is not the case. While children seem to be aware of syllable level information from a very early age, the more detailed knowledge required for segmentation only comes after learning to read using an alphabetic system. Therefore it seems quite possible that learning to read and learning that the frication noise corresponds to the <s> and <sh> orthographic symbols is what drives the shift from formant transitions to frication noise. The following experiment is designed to test this hypothesis by using orthography to change the sibilant cue weighting strategies of listeners.

2. Experiments

The following experiments have the same basic design. Listeners first participated in a same-different discrimination task to assess how well they can identify the cues involved in the Polish post alveolar sibilant contrast, retroflex /ʂ/ and alveolopalatal /ɕ/. Subjects then are asked to classify the sounds using two labels. There were three groups, each were given a different label set: vocalic cues <sha> / <shya>, frication cues <śa> / <ça>, and both <sha>/<xia>. The first set of orthographic symbols is derived from an experiment with the same stimuli where listeners were given example sounds from the contrast and asked to type in on a computer what they heard [18]. The second set was designed to be foreign symbols for a sibilant contrast that English speakers would recognize as consonants and that relied solely on a single symbol changing (i.e. no digraphs). The final set is the Pinyin symbol set for a very similar contrast found in Mandarin Chinese. It was chosen as it shows both a consonantal and vocalic change.

If the orthographic representations do in fact have an effect on cue weighting, then each group should show a different weighting scheme. The vocalic orthography group should preferentially weight the vocalic cues while the frication group should show the opposite effect. The pinyin group, where both symbols are distinct, should show an effect on both cues.

2.1. Methods

2.1.1. Participants

Three groups of native listeners of American English participated. The first group (n = 21) is the vocalic group and were recruited from the XXX campus and payed $10 to participate. The second group (n = 18) is the frication group and was recruited at the XXX campus and received course credit. The final group (n = 24) is the multidimensional group and was also recruited at XXX, receiving course credit. None had any experience with Mandarin Chinese or Polish, though some did have second language experience, primarily Spanish.

2.1.2. Stimuli

The stimuli used were the same as in [18] and are the same used in [18,19]. They consisted of a two-dimensional, 10-step continuum from [ʂa] to [ɕa] varying in frication noise and formant transitions. This continuum was made by taking a representative sample of each sound (produced by native speaker of Polish) in the context of /a/ and separating them at the CV boundary. The frication portions and vocalic portions were separately interpolated to produce two ten-step continua from [ɕ] to [ʂ] (labeled in figures as f0-f9) and [a] excised from [ɕ] to [a] excised from [ʂ] (labeled in figures as v0-v9). These were then recombined into every possible CV combination resulting in 100 stimuli. Perceptual tests in a previous experiment with Polish and Mandarin listeners found them to be sufficiently natural and to be representatives of native categories.

2.1.3. Procedures

In the initial same-different (AX) discrimination task listeners only heard the four stimuli that represented the endpoints of the continuua, i.e. alveolopalatal fricative + alveolopalatal vowel, retroflex fricative + retroflex vowel, alveolopalatal fricative + retroflex vowel, and retroflex fricative + alveolopalatal vowel. These were presented in pairs with a 100ms inter-stimulus interval. There were six possible different pairs (four differing in one dimension and two in both dimensions) and four same pairs, or repetitions of the four stimuli. There were 96 different trials (12 pairs x 2 orders x 8 repetitions) balanced by 96 same trials (24 repetitions of each stimulus). The 192 trials were divided into four blocks of 48 trials each with a self-monitored break between them. Subjects had 1500ms to respond and received accuracy and reaction time feedback after each trial.

In the classification portion of the experiment each of the 100 stimuli were presented to the listeners in a randomized order, one per trial. Subjects were given 1500ms to classify each stimulus by pressing the proper button on a labeled box. Subjects were given feedback as to their reaction time. There were 5 blocks of the stimulus set for a total of 500 trials. The procedure only varied in the labeling of the buttons; the vocalic group was given the labels <sha> / <shya>, the frication group, <śa> / <ça>, and the multidimensional group <sha>/<xia>. All listeners were told that they were hearing sounds from a different language and that they should classify them as best they can. The experiment took approximately 45 minutes to complete.

2.2. Results

In the interest of space, and as it was not the focus of this current experiment, the discrimination portion of the experiment will not be analyzed in detail. For these results the proportion correct for each subject was converted into d', a measure of sensitivity. The d' values for each orthography group were compared using a two level repeated measures ANOVA having the factors Group (orthographic set) and Dimension (frication, vocalic, frication+vocalic).

The Group factor was not significant (f[2,195] = 0.60, p = 0.55), indicating that the groups did not have different sensitivities to the stimuli. The Dimension factor was significant (f[2,130]=24.7,p<0.001) and indicated a better overall sensitivity to the frication+vocalic discrimination (mean = 1.5, sd = 0.65) than the two dimensions singly (frication: mean = 0.95, sd = 0.76; vocalic = 0.97, sd = 0.68). Post-hoc tests confirmed that the multidimensional sensitivity was greater than unidimensional and the two dimensions did not differ in significance from each other.

Before the analysis of the classification results, the data from that part were adjusted to account for variation in label assignment. Because listeners could plausibly assign the <śa> label, for example, to either the /ʂ/ or /ɕ/ stimuli, the following adjustment procedure was devised. First, each subjects mean labeling response to the [ʂa] endpoint stimulus was calculated. If more than half were labeled as <sha> for the vocalic and multidimensional groups, or as <śa> for the frication group, then no changes were made. If fewer than half were labeled as such, then all the responses were flipped for that subject. This was necessary for 6/18 of the frication listeners and 7/24 of the multidimensional listeners. None of the vocalic group participants needed this adjustment.

The adjusted responses were then used in a logistic regression model with the factors fStep, vStep and Group, where fStep refers to fricative step, vStep refers to the vowel step, and the Group factor refers to the orthographic label assignment. The model resulted in significant main effects for all factors and significant interactions between fStep and Group and vStep and Group (see Table 1). Further models were fitted within each group explore the interactions with Group (Tables 2a-c). The results show a significant effect for the vocalic dimension for all groups, but an effect for frication dimension in only the frication and the multidimensional groups.

Table 1. *Analysis of deviance table from model..*

	Df	Deviance	Resid. Df	Resid. Dev.	P value
NULL			6399	16761	
fStep	9	676.13	6398	16085	< 0.001
vStep	9	1830.78	6397	14254	< 0.001
Group	2	72.3	6395	14182	< 0.001
fStep:vStep	19	0	6394	14182	0.95
fStep:Group	2	330.4	6392	13851	< 0.001
vStep:Group	2	659.9	6390	13191	< 0.001
fStep:vStep: Group	2	8.67	6388	13183	< 0.05

Table 2 a-c. *Dimension and Group interactions.*

Vocalic	Df	Deviance	Resid. Df	Resid. Dev.	P value
NULL			2099	6124.5	
fStep	9	0.52	2098	6124	0.47
vStep	9	2028.46	2097	4095.5	< 0.001

Frication	Df	Deviance	Resid. Df	Resid. Dev.	P value
NULL			1799	3544	
fStep	9	491.38	1798	3052.6	<0.001
vStep	9	89.92	1797	2962.7	<0.001

Multi-dimensional	Df	Deviance	Resid. Df	Resid. Dev.	P value
NULL			2499	7025.7	
fStep	9	497.23	2498	6528.5	< 0.001
vStep	9	395.15	2497	6133.4	< 0.001

Figure 1 gives a graphic representation of the data and demonstrates the different Group effects. The results were averaged across all responses for each dimension. Both the vocalic dimension in the vocalic group and the frication dimension in the frication group show effects that resemble classic categorical perception (i.e. “s” curves). The frication curve is weak, largely due to high variance among the subjects. In the vocalic group, there is general tendency to respond “sha”. In the multidimensional group, the results are primarily linear, suggesting categorization, but not categorical perception. Individual subject plots demonstrated that there was a high degree of variance in the use dimensions by this group such that some subjects seem to rely on frication noise, others on the vocalic information, and some used both. None were especially consistent, but as a group they combined to the result above.

3. Discussion

Overall, the expected effects are present; the orthography had a clear effect with very little exposure and no explicit instructions on what cues to attend to. In all groups the vocalic dimension was given some weight, though only the vocalic group showed a classic categorical perception curve. This suggests that there is strong preference to weight the vocalic dimension and that the orthographic labels further reinforced this weighting despite no evidence for this in the discrimination results. While English listeners preferentially weight frication noise for their native sibilant contrast, they use vocalic information for the spectrally similar /f/ ~ /θ/

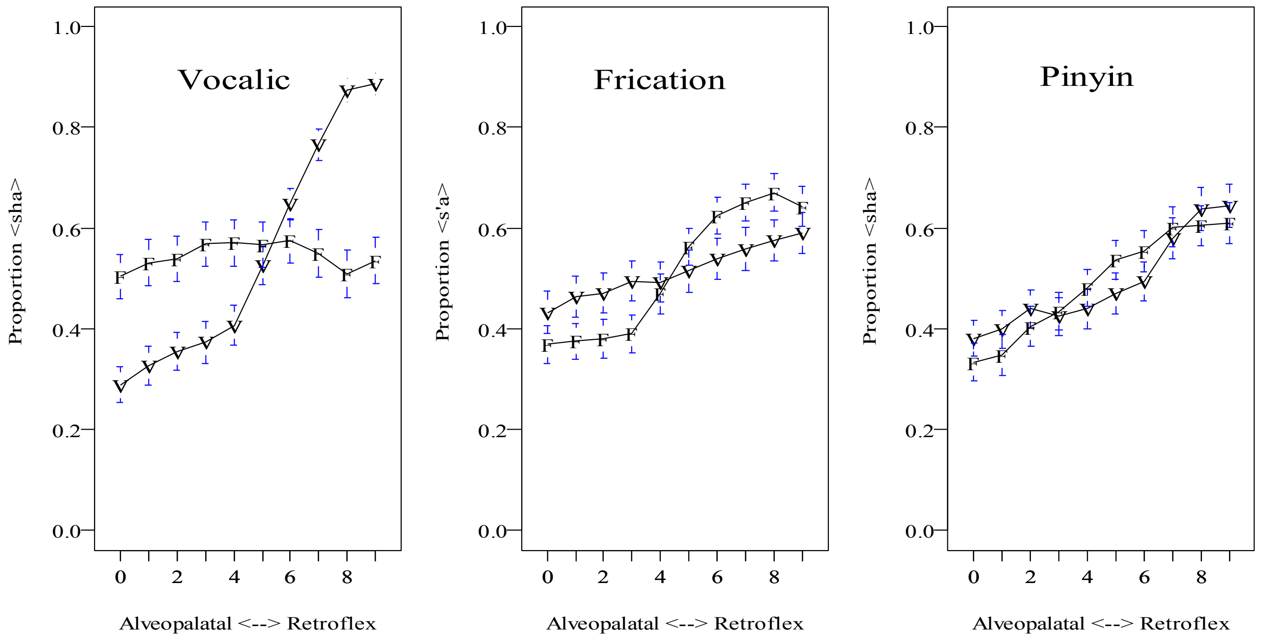

Figure 1: Proportion <sha>/<śa> responses for each group. F = frication dimension, V = vocalic dimension. Leftmost panel represents the vocalic group, middle panel the frication group, rightmost panel the multidimensional group..

contrast [3] and this may be a more general cross-linguistic strategy [20]. Such an effect may be present here.

A secondary orthographic effect is that the multidimensional and vocalic groups both had a symbol for the English post alveolar sibilant <sh>. In the case of the vocalic group this may have led to a general bias to use that label and no subjects using that label for the alveolopalatal fricative (i.e. no need for adjustment).

While the effects were not incredibly dramatic, they were reliable. This likely reflects the extremely limited exposure; listeners only had 45 minutes with the sounds and only 5 opportunities to classify any one stimulus. The multidimensional group showed the most variability and may have had the most difficult task. This is possibly due to having to spread attentional resources across the syllable rather than concentrating on one or the other cue.

4. Conclusions

This study demonstrated orthographic effects on cues below the segment level. Listeners' cue weighting strategies were changed as a result of the label set directing attention to different cues. This supports the view that the orthographic system learned by the reader drives how much internal structure speakers pull from the speech signal. While this may argue against the phoneme as a universal concept, at least from a metalinguistic standpoint, it does not deny that such structure is available to listeners when directed to it. Perhaps more importantly these results instead argue that the extremely discrete, hierarchical view of phonetic structure should not be the default assumption when describing and examining perceptual targets (and presumably articulatory ones, too), regardless of the utility that tools such as the International Phonetic Alphabet provide.

Relatedly, the results of this study call into question the causes of the weighting shift seen in English learning children acquiring its sibilant contrast. That the shift happens in the early stages of learning to read (6-8yrs), that weighting shifts from vocalic to frication (consonantal) cues, that it is related to PA, and that orthography affects both PA and weighting all combine to argue that the weighting shift happens as a result of learning to read. Indeed, the research on fricative cue use, especially that which demonstrates listeners' reliance on fricative noise [20] is focused solely on western languages having alphabetic systems. The primary ramification of this, if true, is that cue weighting may not be solely determined by salience and reliability, but that other factors may conspire to change weighting strategies, see e.g [21].

5. Acknowledgements

Many thanks to Meghan Sumner for initially suggesting that some of my earlier experiment effects were due to orthographic influences. Thanks are also due to Devin Tankersley for running subjects and helping with organizing the data. Finally, I thank the three anonymous reviewers for their very helpful comments.

6. References

[1] Cheung, H., Chen, H.-C., Lai, C.Y., Wong, O.C., & Hills, M. (2001). The development of phonological awareness: Effects of spoken language experience and orthography. Cognition, 81, 227–241.

[2] Liberman, A.M., Delattre P.M., Cooper, F.S., and Gerstman, L.J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 68, 1-13.

[3] Harris, K. (1958). Cues for the Discrimination of American English Fricatives in Spoken Syllables. Language and Speech. 1, 1-7.

[4] Nittrouer, S., and Studdert-Kennedy, M. (1987). The Role Of Coarticulatory Effects In The Perception Of Fricatives By Children And Adults. Journal of Speech and Hearing Research, 30, 319-329.

[5] Bentin, S. (1992). Phonological awareness, reading, and reading acquisition: A survey and appraisal of current knowledge, in R. Frost & L. Katz (eds.), Orthography, Phonology, Morphology, and Meaning, Elsevier Science Publishers B.V., Amsterdam, pp. 193-210.

[6] Kirtley, C., Bryant, P., MacLean, M., & Bradley, L. (1989). Rhyme, rime, and the onset of reading. Journal of Experimental Child Psychology, 78, 243-255.

[7] Bruce, D.J. (1965). The analysis of word sounds. British Journal of Educational Psychology, 34, 158-170.

[8] Bentin S, Hammer R, Cahan S. (1991). The effects of aging and first-grade schooling on the development of phonological awareness. Psychological Science. 2, 271–274.

[9] Lyon, R. (1998). Overview of Reading and Literacy Initiatives: Report to Committee on Labor and Human Resources, United States Congress.

[10] Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 7, 323-331.

[11] Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. Cognition, 24, 45–64.

[12] Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. Cognition, 24, 31-44.

[13] Cheung, H., & Chen, H.-C. (2004). Early orthographic experience modifies both phonological awareness and on-line speech processing. Language and Cognitive Processes, 19, 1-28.

[14] Horlyck, Stephanie, Amanda Reid, and Denis Burnham. "The relationship between learning to read and language-specific speech perception: Maturation versus experience." Scientific Studies of Reading 16.3 (2012): 218-239.

[15] Nittrouer, S. (1996). The relation between speech perception and phonemic awareness: Evidence from low SES children and children with chronic OM. Journal of Speech and Hearing Research. 39, 1059-1070.

[16] Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. The Journal of the Acoustical Society of America. 112(2), 711-719.

[17] Nittrouer, S. & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. The Journal of the Acoustical Society of America, 101(4), 2253-2266.

[18] McGuire, G. (2007). Phonetic category learning. Unpublished dissertation. The Ohio State University.

[19] Cristià, A., McGuire, G., Seidl, A., and Francis, A. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. Journal of Phonetics, 39, 388-402.

[20] Wagner,A., Ernestus., M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. The Journal of the Acoustical Society of America, 120(4), 2267-2277.

[21] Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. Journal of the Acoustical Society of America, 129(2): 999-1011