Explaining a violation of the sonority hierarchy: stop place perception in adjacent [s]

Olle Engstrand and Christine Ericsdotter
Department of Linguistics, Stockholm University

Abstract

Jespersen's sonority hierarchy fails to account for commonly observed syllable structures such as initial [spV], [stV] and [skV] and their final mirror images [Vps], [Vts] and [Vks]. It is likely, however, that stop place of articulation is signaled quite efficiently in these sequences since acoustic place information is present both before and after the stop, i.e., both in [s] and in the vowel. These structures may thus have a perceptual advantage over those predicted by the sonority principle. This was corroborated by the present experiment which indicated that listeners are able to reliably perceive stop place of articulation on the sole basis of an adjacent [s]. It was thus concluded that the observed violation of the sonority principle may have a perceptual basis.

Introduction

One hundred years ago, Jespersen (1897-1899, 1904) presented his famous 'sonority hierarchy', a theory meant to predict the order in which sounds appear in complex syllables. It was claimed, in particular, that consonants tend to appear in order of increasing sonority in syllable onsets, but that the reversed order is true for syllable codas (the 'mirror effect'). This is illustrated by the word *brand* [brænd] in which the relatively sonorous [r]and [n] are adjacent to the syllable nucelus with the less sonorous voiced stops appearing in the word's periphery.

However, Jespersen's theory has been questioned on several grounds. It has been argued, for example, that the theory is in principle untestable since its central term, sonority, is empirically undefined (e.g., Ohala, 1992). Attemtps have been made to cope with this problem in terms of quantifiable dimensions such as degree of jaw opening (e.g., Lindblom, 1983; Björsten, 1997) or amount of modulation in acoustic parameters (Ohala, 1992).

Another obvious objection against the sonority hierarchy is that it does not account for many commonly observed syllable structures. For example, the frequent occurrence of initial [sp], [st] and [sk] clusters runs counter to the prediction that syllable onsets are arranged in order of increasing sonority, and final clusters such as [ps], [ts] and [ks] violate the rule of decreasing sonority towards the end of the syllable. According to the sonority hierarchy, the reversed segment order would be expected.

In the present experiment, clusters of voiceless stops and [s] were considered from a perceptual point of view. More specifically, it was hypothesized that onsets such as [sp], [st] and [sk], and codas such as [ps], [ts] and [ks] have a perceptual advantage over those predicted by the sonority hierarchy. Basically, the idea is that stop place of articulation will be signaled more efficiently in these onsets and codas than in those in which consonants appear in the reversed order, the reason being that place information is then present both before and after the stop, i.e., both in [s] and in the vowel.

It is well known that perception of place of articulation for stops is supported by acoustic information residing in the stop burst as well as in the formant transitions adjacent to the stop consonant. Early confirmation of the role of formant transitions for place perception was provided by the classical Haskins experiments (e.g., Delattre et al., 1955). It has also been

demonstrated (Stevens, 1998, pp. 557 ff.), that spectra for [s] are differentially modulated according to place of articulation of the following stop, and that these effects appear to be substantial enough to be perceptually significant. In combination, these observations imply that a stop consonant intervening between an [s] and a vowel will influence its phonetic context in two ways: it will affect the noise spectrum of the [s] and it will determine the formant transitions present in the vowel segment. In other words, an [s]+stop+vowel sequence and its final mirror image, vowel+stop+[s], will provide the richest possible acoustic information on stop place of articulation. It is thus reasonable to assume that these syllable onsets and codas will entail perceptual advantages over those predicted by the sonority principle. This may explain, at least partly, their relatively common occurrence in the world's languages. Clearly, however, the validity of this hypothesis depends on the extent to which information on stop place of articulation is, in fact, conveyed by acoustic events taking place in an adjacent [s]. The present experiment was carried out to evaluate this claim.

Method

One male native speaker of Icelandic was digitally recorded in a quiet room at 16 kHz while producing 5 tokens each of nonsense test words containing onset and coda clusters of the form [s]+voiceless stop and voiceless stop+[s], respectively. These words are shown in table 1.

Table 1. Nonsense words used in the present experiment.

Onset=[s]+vl stop	Coda=vl stop+[s]
[spaj]	[japs]
[staj]	[jats]
[skaj]	[jaks]

Broadband spectrograms were first inspected for acoustic effects on [s] of the 3 places of stop articulation. For the perceptual test, the [s] portions of the utterances shown in table 1 were excised and presented in isolation. In the onset condition, this segment consisted of the entire [s] friction up to the silent interval corresponding to the stop closure; and in the coda condition, the [s] friction was segmented out to exclude the immediately preceding stop burst. This procedure was applied to all 5 tokens of each test word. Each token was used twice resulting in 60 test items. Of these, 30 items came from onsets, and 30 came from codas.

The onset and the coda blocks were tested separately. In both blocks, the test items were randomized and presented individually to 10 listeners. Half of the listeners heard the onset block first, then the coda block; and half heard the two blocks in the opposite order. Numbers corresponding to the test stimuli were shown on a computer screen. Listeners played the stimuli through headphones by clicking a loudspeaker symbol displayed next to each number, and each stimulus could be repeated as many times as desired. The task was to indicate on a paper form whether the presented [s] represented the beginning of [spaj], [staj] or [skaj] (onset block), or whether it represented the end of [japs], [jats] or [jaks] (coda block). In other words, listeners' ability to perceive stop place of articulation (labial, dental or velar) in adjacent [s] was tested.

The listeners were staff and students of linguistics, speech technology or music acoustics; one listener was a high school student. They had varying language backgrounds: Swedish (5 listeners), Spanish (1), Italian (2), Icelandic (1) and Estonian (1). The non-Swedish listeners were fair to fluent speakers of Swedish.

Results

Acoustic observations

Figure 1 shows amplitude vs. frequency spectra of [s] in the respective labial, dental and contexts. The top row illustrates the onset position with labial, dental and velar context from left to right; and the bottom row shows the analogous information for the coda position. These spectra were averaged over 8-10 ms for the onset position (i.e., the [s] portion nearest to the stop consonant) and over the first 8-10 ms for the coda position (again, the [s] portion nearest to the stop).

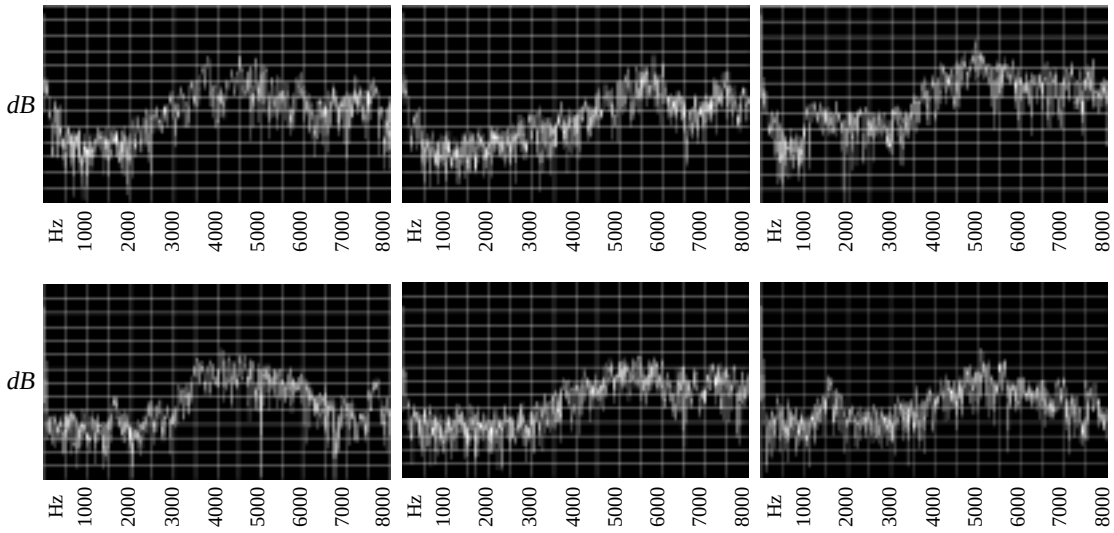

Figure 1. Amplitude vs. frequency spectra of [s] in (from left to right) labial, dental and velar stop context. Top row: onset position; bottom row: coda position. The spectra are averaged over 8-10 ms.

It is evident that the spectra shown in figure 1 a) are differently influenced by the 3 places of articulation for the adjacent stop consonants and b) that these effect are similar in the onset and coda positions. In particular, the energy maximum associated with the labials appears lower on the frequency scale than that for the dental (a maximum plateau appearing from 3500 Hz and 4500 Hz, respectively). For the dentals vs. velars, the upper energy band are in the same general frequency region, but the velars also have an additional energy concentration in the region of approximately 1000 -2000 Hz.

Listener responses

Listener responses are presented as percentages in the confusion matrices in table 2. The leftmost matrix summarizes the data pertaining to onset and coda positions combined, the middle matrix refers to the onset position (i.e., when the voiceless stop follows [s] as in [spaj], [staj] or [skaj]), and the rightmost matrix contains data for the coda position (with the voiceless stop preceding [s], i.e., [japs], [jats] or [jaks]). Stimuli are given in the rows and listener responses appear in the columns. For example, for the [p] stimuli in the combined onset and coda condition, 89 percent of the responses were correct while 5 percent were incorrectly heard as [t] and 4 percent were incorrectly heard as [k].

It is apparent from the high percentages of correct responses appearing in the diagonal cells (shadowed) that listeners were able to reliably perceive place of articulation of an adjacent voiceless stop on the sole basis of acoustic information contained in the fricative [s] segment. With one exception, these percentages were 90 percent or higher in the onset and coda positions. Specifically, the response pattern for the onset position reveals that listeners were able to reliably perceive place of articulation of the following voiceless stop on the sole basis

of acoustic information contained in the preceding [s] segment; and for the coda position, listeners were able to use the acoustic information contained in following [s] to determine place of articulation of the preceding stop (from which, as noted, the burst had been removed). There is, however, one fairly conspicuous asymmetry in these data: [p] in coda position was heard as [t] in 19 percent of the cases, while the reverse was not true, i.e., [t] was heard as [p] in only 5 percent of the cases.

Table 2. Confusion matrices summarizing listener responses (percent) obtained in the present experiment. Left: Onset and coda positions combined; mid: onset position; right: coda position.

	Onset and coda			**Onset**			**Coda**		
	p	t	k	p	t	k	p	t	k
p	89	10	1	99	1	0	79	19	2
t	5	91	4	5	92	3	5	90	5
k	4	3	94	5	3	92	2	2	96

Summary and conclusions

It was hypothesized that onsets such as [sp], [st] and [sk], and codas such as [ps], [ts] and [ks] have a perceptual advantage over those predicted by Jespersen's sonority hierarchy. This hypothesis was based on the assumption that, in such structures, place information is present both before and after the stop, i.e., both in [s] and in the adjacent vowel. Thus, stop place of articulation was assumed to be signaled more efficiently in these onsets and codas than in those in which the consonants appear in the reversed order.

It is well known that place perception for stops is supported by vowel formant transitions; the present experiment was designed to determine whether context-dependent dynamic events in [s] can play a similar perceptual role. Acoustic data suggested that place information for voiceless stops is present in [s]. The perceptual significance of this effect was largely confirmed in a listening test: listeners were able to reliably perceive stop place of articulation on the sole basis of an adjacent [s]. This result thus indicated that the commonly encountered initial [s]+stop and final stop+[s] clusters, in which the onset and coda sequences are mirror images of each other, may have a perceptual basis.

Acknowledgments

One speaker (Pétur Helgason) and 10 listeners have served as subjects in this experiment. Many thanks to you all!

References

Björsten, S. 1997. *Fonemet /s/ och dess inverkan på käkrörelsemönstret i svenskan.* [The phoneme /s/ and its effect on jaw movement patterns in Swedish.] Masters thesis, Department of Linguistics, Stockholm University (with a summary in English).

Delattre, P.C., Liberman, A.M. and Cooper, F.S. 1955. Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27, 769-773. Reprinted in Lehiste, I. (ed.), *Readings in acoustic phonetics* (fourth printing 1974). Cambridge, Massachusetts: The M.I.T. Press, pp. 283-287.

Jespersen, O. 1897-1899. *Fonetik. En systematisk fremstilling af læren om sproglyd.* Copenhagen: Det Schubotheske Forlag.

Jespersen, O. 1904. *Phonetische Grundfragen.* Leipzig and Berlin: Teubner.

Lindblom, B. 1983. Economy of speech gestures. In P.F. MacNeilage (ed.), *The production of speech.* New York: Springer-Verlag, pp. 217-245.

Ohala, J.J. 1992. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Parasession on the syllable*. Chicago: Chicago Linguistic Society, pp. 319-338.

Stevens, K.N. 1998. *Acoustic phonetics.* Cambridge, Massachusetts: The M.I.T. Press.