A Phonetic Basis for the Sonority of [χ]

Sarah Bakst and Jonah Katz

1 Background: French rhotics and sonority

Sonority is a property of segments that relates to phonotactics and syllable shape but whose nature remains in dispute. Clements (1990) uses this property to define the order that segments may occur in a syllable in what he calls the sonority sequencing principle (SSP), which says that "between any member of a syllable and the syllable peak, only sounds of higher sonority rank are permitted." His principle relies on a sonority scale, which ranks segments by their natural class, with obstruents registering low on the scale, followed by nasals, liquids, glides, and finally vowels. The classes are defined by the features vocoid, approximant, sonorant, and what Clements calls "syllabic" (only vowels are [+syllabic]). The sum of the positive values of features for a given class yields the sonority value.

Syllables are usually arranged with the sonority peak at the nucleus with falling sonority towards either syllable edge, so that low-sonority segments tend to precede higher-sonority segments in onsets but the opposite is usually true in codas. Some languages allow sonority plateaux, or multiple consecutive segments of the same sonority, such as a sequence of two stops or a stop plus a fricative. The implication is that languages that allow plateaux, which violate the SSP, also allow sequences that do not violate the SSP, but the reverse does not hold: there should not be languages that only allow SSP violations.

Most accounts agree on the segment ranking above (or a close variant), though the basis for the hierarchy varies. Some definitions are based in phonological features, as in Clements, but others make more reference to phonetics. Parker (2002) creates a hierarchy that is mainly based on acoustics and aerodynamics and defines sonority as a function of intensity. This is later extended to create a universal phonological scale. Wright (2004) maintains that a sonority scale is not a phonological entity but rather a natural organization of sound patterns that falls out from how well place, manner, and voicing cues can be recovered by the listener from the acoustic signal. In contrast to the other definitions mentioned here, Wright does not relate sonority to phonology at all.

French rhotics present a problem for the sonority hierarchy. The rhotic is usually realized as a voiced uvular approximant or occasionally a trill, both of which are considered liquids, but after a voiceless obstruent, the rhotic surfaces as a voiceless uvular fricative. The rhotic patterns like the other liquid in French [l]; both are permitted to be the second member of a complex onset. In theory, French does not allow sonority plateaux in onsets, yet the fricative realization of the rhotic, to the exclusion of all other fricatives, occurs as the second member of a complex onset, such as in /kʁem/ [kχɛm] 'cream'. The sonority scale for French is thus internally inconsistent: other fricatives are not allowed in French complex onsets, but there seems to be an exception for this realization of the rhotic.

One possible phonological treatment would be to make reference to the fact that [χ] is the allophone of a high-sonority rhotic. The underlying representation of [kχɛm] is /kʁem/, which conforms to French sonority rules; /ʁ/ has a sonority value that lies between /k/ and /ɛ/. In fact, Clements (1990) suggests that if the SSP applies at a surface level, it is difficult to account for the many plateaux and reversals observed in language, and he supports the idea that the SSP applies at an abstract level, that is, to the underlying form of a word or syllable. The problem with this analysis is that it makes no predictions about what other segments in French, or any language for that matter, may also exhibit dual behavior with respect to sonority; under this treatment any segment could receive a rhotic label. If such a phonological explanation is required, then the SSP loses the power to predict preferred syllable structure.

Further, application of the SSP at the level of underlying representation would not be feasible in a modern framework like Optimality Theory (OT). The SSP would be modeled in terms of markedness constraints; for French, tautosyllabic obstruent-obstruent clusters would be prohibited. In OT, markedness constraints may only apply to surface forms. Thus the form /kʁem/ would not violate the SSP, but the surface form [kχɛm] would. There would be no way to satisfy the constraint barring adjacent obstruents to allow the fricative surface form of the rhotic without also allowing other tautosyllabic obstruent-obstruent clusters.

There may be acoustic and articulatory reasons why uvular fricatives are phonetically fricatives but are able to behave phonologically as high-sonority liquids. In a fricative modeling study, Yeou & Maeda (1995) show that uvular fricatives are best described as approximants. The area of constriction in a uvular fricative is greater than in other fricatives, and uvular fricatives show less frication noise than other fricatives. Further, uvular fricatives show an approximant-like formant structure due to the long front cavity filter in front of the fricative constriction.

The experiment here tests whether the acoustic characteristics of the uvular fricative translate into Wright (2004)'s "perceptual robustness" that can explain the phonotactics of [χ]. If the fricative is in fact approximant-like, then its occurrence after stops in complex onsets could be explained through a licensing-by-cue approach. This fricative's formant-like structures resulting from the long front cavity in front of the constriction may provide place and manner cues to an adjacent consonant that are recoverable by the listener.

The cue-bearing abilities of [χ] in relation to other fricatives can be tested in a percpetion experiment. The spectrograms in Figure 1 show examples of the syllables [tχa] and [tfa] spoken by a native speaker of French. Formant-like structures are readily visible in the fricative in [tχa] but not in [tfa], suggesting that speakers may be able to recover cues to a preceding stop when the following fricative is [χ] better than when it is [f].

2   Experiment

The experiment compared different continuants in French by testing whether and how well listeners are able to use co-articulatory and transition information about a preceding stop contained in these continuants. Open monosyllables with stop-continuant onsets were presented, where the stop was always [t] and the continuant was [f, χ, l]. Listeners' ability to infer the presence of the stop if the burst information has been removed was assessed. If the following continuant has a clearer formant structure, it is more likely that listeners will be able to recover the presence of the preceding stop; the ability of a segment to provide cues to an adjacent segment may license its phonotactic

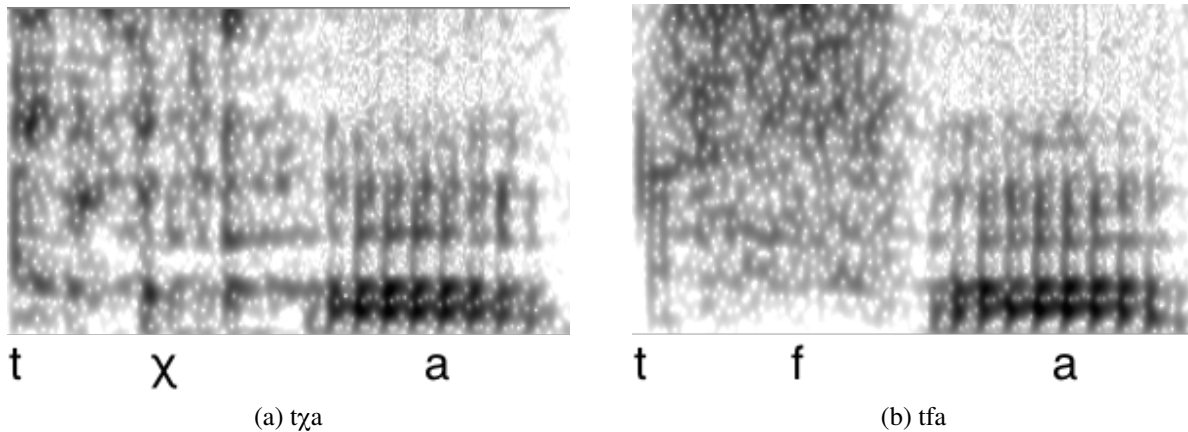(a) tχa   (b) tfa

Figure 1: Spectrograms of [tχa] and [tfa]. Formants are visible in (a) but not (b).

position. This experiment tested the claim that the phonotactic position of [χ] can be explained by its exceptional bearing of transitional cues. If the claim is true, then listeners should be better able to use information provided by [χ] than information provided by [f], and they should perform similarly for [χ] and [l].

If phonotactics or sonority are governed by robustness of perceptual cues, then we expect permitted sequences to have more robust perceptual cues than illicit sequences. This experiment tests how available stop information is recoverable in the following fricative. If uvular fricatives provide usable redundant information about the presence of a preceding stop better than other fricatives, listeners should be more likely to perceive stop-uvular fricative sequences where the stop burst has been removed as the same as sequences where the stop burst has not been removed. If this is the case, listeners should be more likely to respond "same" on "different" trials for uvular fricatives than for labiodental fricatives.

2.1   Stimuli

Recordings for the stimuli were done using a AKG C535 EB condenser microphone sampling at 44 kHz. A native speaker of Parisian French was recorded speaking nonce words realized as [ləCXV][1], where the consonant was {p, t, k}, X was {f, χ , l} and V was {a, e, i, o, u}. These words were spoken in the carrier phrase "Retape ___ pour moi." ("Retype ___ for me.")

The sentences were recorded 10 times each, and the seven best examples were selected. A good example was defined as a token with a clear release burst. Although words with [p] and [k] as the initial consonant were also recorded, the bursts were often compromised before [f] and [χ] respectively, so only stimuli with [t] were used in the experiment. Syllables of the form tfV, tχV, and tlV were extracted from the syncopated words. From the resulting syllables, two stimuli were

---

[1]The sequences were written <l'eCeXV>. A phonological process present in casual French resulted in the deletion of the second [e]. Although Cχ sequences occur naturally in French, all stimuli were created through deletion for consistency. Fougeron and Steriade (1997) found that French speakers have difficulty telling the difference between sequences that are formed through schwa deletion and those that are not despite some articulatory and acoustic differences; these differences have little to no effect on the present study, especially given that all the stimuli were formed in the same way.

created: for one set, the syllables were left intact, and for the other set, the bursts (first 5ms of the token) were removed.
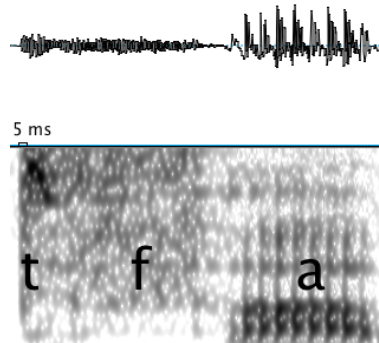

Figure 2: Example of a test syllable [tfa].

2.2 Methods

Twenty-two (21 female) native English speakers participated in the experiment. They were either undergraduate students at UC Berkeley who received extra credit in an introductory linguistics class to participate or were other students or members of the community who were paid $5 in compensation. There were no reports of past or current hearing difficulties. Participants listened to the stimuli through AKG K 240 headphones in a sound-attenuated booth.

In total 210 stimuli were presented in an AX burst-detection discrimination task using the program OpenSesame. Half of the trials were "same" trials, and half were "different", for a total of 420 trials. On "same" trials, the same stimulus was played twice. On "different" trials, the intact and burst-removed versions of the same recorded token were presented. Trials were randomized and the two stimuli of "different" trials were presented in both orders. Stimuli were presented with a 500ms interstimulus interval. Participants used the "1" and "0" keys on the keyboard to indicate whether the syllables they heard were the same or different. They were asked to focus on the beginning of the sound to make their choice, and they were given feedback as to whether they were correct.

2.3 Results

The task measured how sensitive listeners are to a stop burst in different continuant environments. The following analysis compares how often listeners replied "same" to trials where the stimuli were different to how often they replied "same" on trials were the stimuli really were the same.

2.3.1 D-prime

D-prime measures of sensitivity were performed on the responses. Instead of measuring only responses on "different" trials, responses from "same" trials are taken into account. If a participant

responds "different" to every trial, the total accuracy is artificially higher when only "different" trials are analyzed and false alarms are not taken into account.

Most participants had very low sensitivity measures, which means that they were not sensitive to the presence or absence of a burst; the task was indeed difficult. D-prime is measured on a scale of 0-6, and most participants had scores of less than 1 in each condition. All had scores less than 2. Although listeners were not very sensitive to the presence or absence of a burst generally, trends emerge from the data.

Figure 3 compares listeners' sensitivity in the three continuant contexts; the figure gives a qualitative overview of the results. Each point represents a single subject. The first plot compares sensitivity to bursts before [χ] (x-axis) to sensitivity to bursts before [f] (y-axis). More subjects fall above the $y = x$ line than below; these listeners are more sensitive to bursts before [f] than before [χ], which falls in line with the hypothesis. In fact, there are considerably more listeners who have a non-zero D-prime exclusively in the [f] condition than in the [χ] condition. The second plot shows that people are slightly more sensitive in the [f] context than in the [l] context, which does not support the hypothesis. The third graph shows a fairly bimodal distribution: some people are more sensitive to burst presence before [χ] than before [l], and others are more sensitive to bursts before [l] than before [χ].
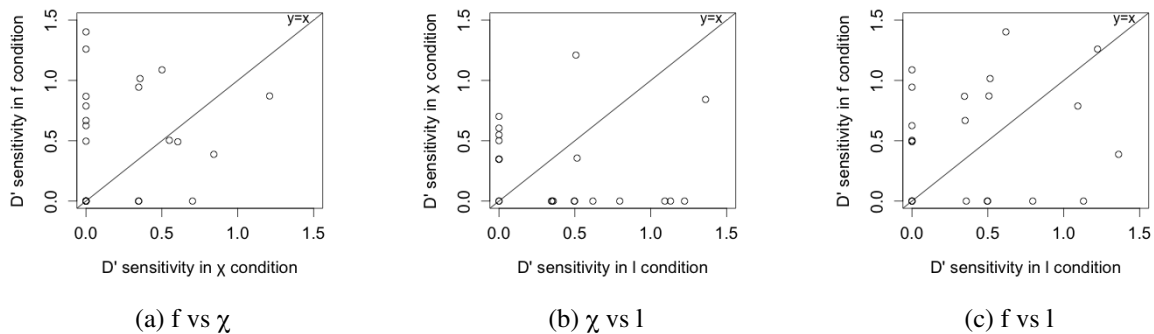

(a) f vs χ	(b) χ vs l	(c) f vs l

Figure 3: Plots comparing listeners' sensitivity to absence or presence of a burst in each continuant condition; each point represents a single participant. In (3a) the y-axis charts sensitivity to the presence of a burst in the f context, and the x-axis charts sensitivity in the χ context; (3b) shows sensitivity in the χ context on the y-axis and in the l context on the x-axis; (3c) charts sensitivity to f on the y-axis and to l on the x-axis. The $y = x$ line is overlaid. Points lying on the $y = x$ line indicate subjects who are equally sensitive in both conditions.

2.3.2 Logistic Regression

A regression analysis was performed to quantify the effect shown above. The response variable was binary (responses were either "same" or "different"); results were analyzed with a mixed effects logistic regression model with subject number and vowel quality set as random variables. Vowel quality was not expected to be a predictor in this experiment. Multiple vowels were included to vary the stimuli and because the results should not vary much across vowel type.

The baseline models the log likelihood of responding "different" to a "same" trial in the [χ] condition. The variables tested as random slopes were the slope for accuracy (corr below), the

difference in false alarms between χ and f (fvar below), the interaction between fvar and the correct response, the difference in false alarms between χ and l (lvar below), and the interaction between lvar and the correct response; these were added to the model in that order and kept if the model was found to be significantly different to the previous model. The final model included fvar and lvar as random slopes of subject number, but neither accuracy (corr) nor any interaction with corr improved the fit of the model. Adding random slopes to vowel also did not improve the fit of the model.

There was a significant effect of context; there was a significant difference in sensitivity in the f condition relative to the other conditions ($p < .05$), but there were no other significant effects. Subjects were more sensitive to the presence of a burst in the f condition than in the other conditions. In the context of this study, on "different" trials participants were probably less able to use transition cues from the f to infer the presence of the preceding t.

```
Formula: binresp ~ corr * (lvar + fvar) + (1 + fvar + lvar | subject_nr) +
   Data: d
   AIC   BIC logLik deviance
 11534 11627  -5754    11508
Random effects:
 Groups      Name          Variance  Std.Dev. Corr
 subject_nr (Intercept) 0.2290261 0.478567
            fvarTRUE     0.1603103 0.400388 -0.289
            lvarTRUE     0.0849097 0.291393 -0.333  0.218
 vowel      (Intercept) 0.0048647 0.069748
Number of obs: 9187, groups: subject_nr, 22; vowel, 5


Fixed effects:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.65345    0.12000  -5.446 5.17e-08 ***
corrTRUE           -0.09517    0.07797  -1.221   0.2222
lvarTRUE           -0.13869    0.10012  -1.385   0.1660
fvarTRUE            0.06119    0.11531   0.531   0.5957
corrTRUE:lvarTRUE   0.17276    0.11076   1.560   0.1188
corrTRUE:fvarTRUE   0.22459    0.10924   2.056   0.0398 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Correlation of Fixed Effects:
             (Intr) crTRUE lvTRUE fvTRUE crrTRUE:lTRUE
corrTRUE     -0.320
lvarTRUE     -0.426  0.383
fvarTRUE     -0.400  0.333  0.361
crrTRUE:lTRUE  0.225 -0.704 -0.552 -0.234
crrTRUE:fTRUE  0.228 -0.714 -0.273 -0.473  0.503
> print(fric2.5.lmer, corr=FALSE)
Generalized linear mixed model fit by the Laplace approximation
```

```
Formula: binresp ~ corr * (lvar + fvar) + (1 + fvar + lvar | subject_nr) +
   Data: d
   AIC   BIC logLik deviance
 11534 11627  -5754    11508
Random effects:
 Groups      Name          Variance  Std.Dev. Corr
 subject_nr (Intercept) 0.2290261 0.478567
            fvarTRUE     0.1603103 0.400388 -0.289
            lvarTRUE     0.0849097 0.291393 -0.333  0.218
 vowel      (Intercept) 0.0048647 0.069748
Number of obs: 9187, groups: subject_nr, 22; vowel, 5

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.65345    0.12000  -5.446 5.17e-08 ***
corrTRUE         -0.09517    0.07797  -1.221   0.2222
lvarTRUE         -0.13869    0.10012  -1.385   0.1660
fvarTRUE          0.06119    0.11531   0.531   0.5957
corrTRUE:lvarTRUE 0.17276    0.11076   1.560   0.1188
corrTRUE:fvarTRUE 0.22459    0.10924   2.056   0.0398 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3   Discussion and Conclusions

The results of the study show that listeners show more sensitivity to the presence of bursts before [f] than before [χ]. Because transition cues to the preceding stop are available in [χ] but not [f] contexts, listeners are less sensitive to whether the burst cues are actually present or not. If the burst is absent in the [f] condition, the lack of transition cues into the fricative means that listeners are unable to recover information that indicates that a stop had been present. If the burst is absent in the [χ] condition, there is sufficient available information leading into the fricative to cue the presence of a preceding stop, even if the burst information has been removed.

Although it is clear that [χ] does not pattern in the same way as [f], there are not definitive conclusions about whether it functions more similar to [l] than to [f]. It is difficult to tell whether the data from the [l] condition is informative because of coarticulation complications. The consonant [t] was selected to the exclusion of other voiceless fricatives because there was little to no burst of [p] before [f] or of [k] before [χ]. In the tokens before [l], many [t] tokens had a lateral release. Tokens selected for the experiment had as little of this lateral release as possible. It is possible that the tokens lacking the release had an unnaturally small amount of co-articulation and therefore less transition information, and significantly less coarticulation than in the [f] or [χ] conditions. This means that the [l] condition might not be comparable to the [f] or [χ] conditions. Regardless of the informativity of the [l] data, we can still draw conclusions about [χ] in relation to other fricatives.

A licensing by cue approach suggests that the reason we find [χ] in consonant clusters (and

less often other fricatives) is that cues to adjacent consonants are more perceptible before [χ] than other fricatives like [f]. This could be in part due to the long cavity that is created in front of the constriction for [χ]. This predicts that if a language's fricative inventory includes back fricatives, these fricatives will be more likely to occur in clusters than other fricatives, and that if any fricative is excluded from this position, it will be less likely to be a back fricative. This prediction crucially does not rely on allophony but rather on qualities that are intrinsic to the segments in question. More broadly, the results of this experiment provide support of the role of Wright (2004)'s cue robustness in phonotactics and some measure of sonority. The results show how the acoustics of these stop-fricative sequences map onto perceptual properties and provide a possible explanation for observed phonotactic patterns.

Further, the results of this experiment provide some insight into the sound change in French from uvular trill to fricative. While it does not explain how or why the change happened, and it is entirely possible that the change from approximant/trill to voiceless fricative was itself an accident of history, it helps to explain the viability of the change. Support for the idea that this might be a natural sound change and not an accident of history would come from languages that also have a [χ] phoneme which may license more preceding stops than other fricatives can.

While the actual differences in sensitivity to burst presence are statistically significant, the differences themselves are very small; the greatest sensitivity for any individual in any context was a D-prime score of 1.4. Further research, both through investigation of the phonotactics of other languages and through more experimentation, is required to more definitively determine whether this result holds.

References

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I*, ed. John Kingston and Mary Beckmann. Cambridge: Cambridge University.

Clements, Nick. 2005. Does sonority have a phonetic basis? Comments on the chapter by Vaux. In *Contemporary views on architecture and representations in phonological theory*, ed. Eric Raimy and Charles Cairns. Cambridge, MA: MIT Press.

Fougeron, Cécile, and Donca Steriade. 1997. Does deletion of French schwa lead to neutralization of lexical distinctions? In *Eurospeech*, volume 2, 943–946.

Gordon, Matthew, Paul Barthmaier, and Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* .

Jesus, Luis M. T., and Christine H. Shadle. 2005. Acoustic analysis of European Portuguese uvular [χ, ʁ] and voiceless tapped alveolar [ɾ̥] fricatives. *Journal of the International Phonetic Association* .

Lindau, Mona. 1985. The story of /r/. In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged.*, ed. Victoria A. Fromkin. Academic Press, Inc.

Parker, Stephen G. 2002. Quantifying the sonority hierarchy. Doctoral Dissertation, University of Massachusetts, Amherst.

Scheer, Tobias. 2007. On the status of word-initial clusters in Slavic (and elsewhere). In *Annual Workshop on Formal Approaches to Slavic Linguistics: The Toronto Meeting 2006*, ed. Richard Compton, Magdalena Goledzinowska, and Ulyana Savchenko. Ann Arbor: Michigan Slavic Publications.

Vaux, Bert, and Andrew Wolfe. 2009. The appendix. In *Contemporary views on architecture and representations in phonological theory*, ed. Eric Raimy and Charles Cairns. MIT Press.

Wiese, Richard. 2001. The phonology of /r/. In *Distinctive feature theory*, ed. T. Alan Hall. Mouton De Gruyter.

Wright, R. 2004. A review of perceptual cues and cue robustness. In *Phonetically based phonology*, ed. Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge: Cambridge University.

Yeou, M., and S. Maeda. 1995. Pharyngeal and uvular consonants are approximants: an acoustic modeling study. In *International Congress of Phonetic Sciences*, volume 2, 586–589.