Phonotactically well-formed onset clusters as processing units in word recognition

Tom Lentz

Utrecht Institute of Linguistics OTS, University Utrecht, Janskerkhof 13a, Utrecht, Netherlands
`Tom.Lentz@let.uu.nl`

Abstract

Phonotactic well-formedness has an effect on speech processing. This is likely due to an independent sub-lexical representation of phonotactics. Researching that knowledge requires isolating it from indirect effects. A prominent indirect effect comes via lexical neighbourhood. The better phonotactically a word is, the more neighbours it has, the harder it is to recognise it.

The present study examined the sublexical effect for phonotactically good word onsets with auditory priming. Word recognition was facilitated for good clusters, in spite of the larger number of lexical competitors. Word recognition latency is corrected for the effect of lexical neighbourhood, additional effects have their origin in the processing differences of the auditory primes. We found that words with good phonotactic onsets are recognised quicker, but that (destructive) manipulation of the prime onset destroys the benefit of good onsets, and can even revert it.

Index Terms: speech perception, phonotactics, language specificity, phonology, word recognition, auditory priming, speech processing.

1. Introduction

1.1. Separating phonotactic and lexical neighbourhood effects

Recognition of spoken words is shown to be affected by components of these words. There is not only a role for phonemes, but also for larger sublexical chunks, like consonant clusters. Phonotactic well-formedness has a strong influence on word recognition. Vitevitch and Luce [1, 2] (henceforth V&L) show that this is an effect that can be modulated by different tasks, in as far as it is obscured by a stronger and opposite effect of lexical neighbourhood density. As good phonotactic forms are in denser lexical neighbourhoods, the direct benefits of easier processing of good phonotactic combinations are cancelled out by the indirect disadvantage of a larger set of lexical competitors.

V&L modulated the phonotactic effect by using different tasks. The strongest facilitating phonotactic influence was found in a mixed word/non-word same/difference task. The authors explain this with the assumption that the non-words force the participants to process the sub-lexical chunks and not focus on the lexical neighbourhood. The lexical effect was still present for words, though. In a lexical decision task, this effect wins for both words and non-words: altogether, good phonotactics properties slowed down processing.

The V&L experiments show the existence of independent, sub-lexical phonotactic effects, and were not intended to look into details of this effect. The measure of phonotactic well-formedness used is phoneme and biphone frequency, taken over the whole word. The present study focusses on phonotactic effects, and is therefore controlled for lexical effects. This allows to address the details of the type of phonotactic well-formedness that facilitates processing on a sub-lexical level.

In order to do so, two types of phonotactic combinations are contrasted. One type is dubbed ‘good’, and is hypothesised to have a special status in the phonotactic grammar. We do not wish to claim that this grammar is of a special cognitive type on the basis of behavioural data only. However, processing good clusters as units *might* be beneficial in speech recognition, as they occur frequently; the experiments show that human listeners indeed process frequent combinations differently.

1.2. Word initial clusters of /s/ and consonant

This study investigated Dutch words starting with s and one other consonant after the s, /sC/-onsets. The phonotactic well-formedness is operationalised as the O/E ratio of the s and the consonant as found in continuous speech, in this case the CGN Corpus of Spoken Dutch [3]. O/E ratios are calculated by dividing the observed frequency O by the estimated frequency E. E can be calculated by multiplying the probability of both components and the total number of tokens. This number reflects how often two phonemes would be observed next to each other if there was no preference for combinations at all.

The actual observed frequencies of the /sC/-clusters can deviate from the expected value. In the case the O/E ratio is below 1, there seems to be a preference against combining the phonemes. If it is larger, the combination can be considered preferred (for brevity’s sake, the term ‘good’ will be used in the rest of the paper). Values around 1 seem to indicate that a language has no preference for or against the combination.

O/E ratios are given in table 1 for all non-marginal Dutch /sC/ clusters that occur word-initially. The clusters come in two groups: O/E above 2, or around 1. Note that statistical measures such as transitional probability are unidirectional.

The good clusters’ overrepresentation might be explained in part by their phonetic properties, but the point here is that they have a special status, which is a phonotactic property of Dutch. It is language specific; many languages do not allow consonant clusters at all and have zero O (and therefore O/E) values for all combinations. The neutral clusters, however, have observed values that can be explained by the frequency of their parts (phoneme frequencies), i.e. the expected value.

This study proposes to look into the extra knowledge that the good clusters are indeed good. This knowledge does not follow from the parts and is therefore connected to the cluster as a unit. Below the possible effects on processing are discussed, after which they are tested.

Table 1: *sC-clusters, ranked on O/E as found in the phonetically annotated part of the corpus of spoken Dutch.*

Cluster	Observed	Expected	O/E
[sp]	5573	1783	3.12
[sx]	10762	3596	2.99
[st]	32173	15840	2.03
good	48508	21219	2.28
[sm]	3548	3541	1.00
[sl]	3667	3304	1.11
[sn]	3742	3136	1.19
[sk]	2703	3241	0.83
neutral	13660	13222	1.03

1.3. Segmentation and epenthesis effects

The good/neutral dichotomy found for /sC/-clusters is possibly relevant to two speech processing issues. First, it could affect segmentation of words. We choose to calculate the O/E ratios over continuous speech to make sure the property is in principle available to infant learners before they construct a lexicon or have access to segmentation, as infants seem to have knowledge of phonotactics early, before they acquire a lexicon [4, 5]. In that sense sub-lexical phonotactics could be the origin of lexical phonotactic effects.

Illegal phonotactics impede certain segmentations and facilitate others, as McQueen showed [6]. To test for that effect of good clusters, clusters were epenthesised in two ways: to [ɛsC] and to [suC]. The effect of the manipulations could differ; if good clusters are recognised faster because the segmentation keeps them together, they should suffer less from the [ɛsC] manipulation than neutral ones.

However, McQueen's findings cannot be directly extrapolated to the present materials, as the benefit of keeping the good clusters together is in conflict with the bad phonotactic well-formedness of the residue [ɛ], a vowel that is not a possible syllable in Dutch on its own. Under a null hypothesis that both clusters are nothing more than the sum of their parts, the [suC] manipulation cannot have a different effect, as both phonemes are still there. However, if clusters are processed as units on some level, the insertion of [u] would cause more problems.

The second important aspect of phonotactic status of clusters on processing is perceptual epenthesis; illegal phonotactic structures can be corrected to legal ones in perception [7, 8]. In Dutch the clusters are legal and therefore not necessarily broken by epenthesis (sC → sVC) . However, they might be preceded by vowels in continuous speech, which matches the epenthesis that best fits fricatives, as Fleischhacker found [9]. That context might therefore be less damaging to the word recognition process.

Both [ɛsC] and [suC] manipulations are phonotactic improvements (clusters are marked in general); both are attested cross-linguistically; they occur in Spanish and Japanese for loanwords, respectively. but they were expected not to match the representations of the words and therefore harm speech processing / word recognition. This is important, as it means the phenomenon does not tap into easier mapping of the acoustic signal to a phonetic perception, but to the mapping of the phonetic signal to a phonological form (that is under the influence of phonotactics).

2. Experiment

To asses to what extent a manipulated version of a word can activate the word lemma, a visual word recognition task with auditory priming was designed. Visual word recognition gives a risk of orthographic influences; these normally correspond to the phonotactics (see [10, 11]). Dutch ortography mirrors the phonemic structure, with the exception of /sx/, which is written 'sch', ('ch' does correspond to /x/ in other contexts). The auditory priming allows the manipulations; the cross-modality ensures the measurements are tightly related to lexical forms.

By using a baseline of unrelated primes, the possible influence of orthography and any other word-specific effects are factored out. Auditory primes that were equal to the visual targets shorten the response time, as the target is already activated by the prime. As this priming effect is related to the word recognition of the prime, it gives an indication of the auditory (phonological) representation of the target words.

Next to primes that are equal to the targets, two epenthetic manipulations of the primes mentioned above were used. The null hypothesis is that priming is the same for both conditions and both manipulations. The first hypothesis to replace it is that good clusters are processed as units and respond differently to the manipulations. The second hypothesis is that the destruction of the adjacency of the cluster by a vowel ([suC]) is more harmful than destruction by a syllable boundary ([esC]).

2.1. Participants

66 people from the university subject pool participated, all native speakers of Dutch. Everyone reported not to have hearing or reading problems.

2.2. Materials

Participants had to respond to 60 experimental items (words starting with /sC/), 60 filler words not starting with /sC/, 60 filler non-words starting with /sC/ and a 60 filler non-words not starting with /sC/, assuring that no yes or no bias was induced and that the prime did not give a clue about the correct answer.

The 60 experimental items came in two conditions; 30 were words starting with good /sC/ clusters and 30 with neutral ones. Most words were monosyllabic, none had more than two syllables. They were selected from the most frequent words fulfilling the condition, based on a weighed average of the frequency in the CGN corpus of spoken Dutch and the CELEX lexical database [12]. As the most frequent 30 words of the neutral condition are less frequent that those of the good condition, the words from the good condition were picked such as to match the other one-by-one as close as possible as far as frequency is concerned.

The primes were recorded by a male speaker of standard Dutch who was not informed about the purpose of the experiment. Manipulated primes came in two versions: either the onset was cross-spliced to become [ɛs] or to become [su]. To avoid results due to cross-splicing alone, the faithful primes were also cross-spliced. Cross-spliced items were made as close to natural Dutch as possible. Three native speakers of Dutch, linguists, checked the naturalness. Two items had to be discarded due to programming errors.

2.3. Procedure

The test was presented in a sound isolated cabin, using an in-house experiment program (FEP) running under Ubuntu LINUX 6 with Xenomai real-time support on a normal desktop computer with an Intel Celeron processor (outside the cabin). Visual targets were presented on a 17" CRT screen. Auditory primes were presented over Beyerdynamic DT250/80 headphones. After the sound was played, there was an 500 ms interval, after which a fixation point (+) was shown for 500 ms. The target was presented for 750 ms after the fixation point; participants had 2500 ms to respond from the start of the target presentation. After every trail, feedback was presented on screen ('correct', 'wrong', or 'too late').

Participants were instructed to respond as quickly as possible to the words presented on the screen, by pressing the yes or the no button on a button box. The yes button was on the side of the dominant hand. Before each word on the screen, they heard an auditory prime over headphones. Participants were asked to write down the last word they heard once every 60 trials, in order to avoid them tuning out and to check their listening. No participants made errors in writing down words.

To avoid learning effects, every participant had one trial per target. The type of prime (unrelated, manipulated or faithful) was rotated per target in a Latin square design. Of each condition, 10 items were presented with an unrelated prime, 10 with a manipulated prime and 10 with a faithful prime. Trials were randomised.

The experiment as described above was run twice. In one group, the manipulated items were of the [ɛs] type, in the other they were of the [su] type.

2.4. Analysis

To avoid the problem of repeated measures and the sphericity assumption common to repeated measures ANOVA, the data were analysed in a mixed-effects linear model with crossed random effects for participants and items [13, 14, 15].[1] The response times for correct responses were used, transformed to their logarithms to unskew them. No outliers were removed, although of course reactions later than 2500 ms had already been discarded as incorrect responses.

Factors were the type of cluster (neutral or good), the type of prime (faithful, manipulated or unrelated) and the group ([ɛsC] and [suC]). The random effect of cluster condition and prime manipulation was nested under item; the random effect of group was nested under participant. To be conservative, the lowest number of degrees of freedom was used to convert t-values to p-values, namely (58 items$-1 =$) 57. Outliers were found by using the interquartile range (cut-off at 3 IQR), for each item per prime type; without outlier removal the same significant effects were found as reported below but slightly less profound.

A model containing type of prime, phonotactic condition, type of manipulation, their interactions, and random effects for participants (including manipulation type group) and targets (including phonotactic condition and prime type) was calculated.

The group effect was small (the [suC] group was slower, 0.03 log(ms)) and not significant ($p > 0.34$, n.s.), showing no overall difference between the runs of the experiment with different prime manipulations. More expected results were for type of prime, both manipulated ($p = 0.0018$**) and unrelated ($p < 0.001$ ***) primes leading to significantly slower responses compared to faithful primes (0.07 and 0.22 log(ms), respectively).

The model incorporated a faster response for good cluster targets, but not significantly so ($p = 0.147$, n.s.). Interaction was present between the type of manipulation and the prime type, because /suC/ manipulation was significantly worse ($p < 0.001$***). Looking at the data, we see that the effect of cluster type is attenuated by an interaction with prime type and group. That interaction is not significant in itself. The /ɛsC/-manipulated primes for good cluster words could ($p > 0.17$, n.s.) show some cancellation of the priming effect (reaction times going up), as can be seen by visual inspection of the data (figure 1 and **??**).

As an effect of good phonotactics was expected (following V&L), a model with simple main effects was constructed including the interaction between cluster type and type of manipulation. Now $p < 0.05*^2$, showing that a simple main effect for condition is lost in a crossed interaction, even though the interaction is not significant.

However, we feel this model to be wrong; although this model is better than an empty model ($p < 0.0136$), the full model shows that the phonotactic effects are explained better by the other factors (the full model is better than the condition model, $p < 2.2e^{-16}$).

Bonferroni correction for the number of contrasts (11) indicates that t-values over 3.1 are acceptable. An insightful model includes the interaction with prime type, excluding its main effect. This allows us to look at the interaction of condition at the different prime types. All four interactions (with as a baseline the faithful primes corrected for the possible condition main effect) now have a t-value well above the threshold: (good cluster + manipulated prime is 0.12 log(ms) slower, $t = 8.3$, neutral cluster + manipulated prime 0.10 log(ms) slower, $t = 7.1$, good cluster and unrelated prime are 0.21 log(ms) slower, $t = 14.37$, and neutral clusters and unrelated primes are 0.23 log(ms) slower, $t = 16.3$). Together, there is a disordinal interaction: the manipulations make the priming for good clusters less effective, while it has better priming in words with unmanipulated primes.

However, this effect can only be attributed to the prime type part, given the full model. On the one hand, as usual, a p value above 0.1 cannot be used to suggest a trend, but on the other hand, it cannot be used to discard the idea that the cluster well-formedness has an effect. The solution for the disappearing phonotactic effect can be found when looking at the prime type / condition interaction for each manipulation separately. A condition main effect in the [ɛsC] group ($t = 3.2$, unrounded value above unrounded threshold) was foun. We also find a trend of an interaction: good clusters suffer more from manipulation (i.e., loose their benefit, see fig. 1). However, in the [suC] group, no such effects can be found (fig.
refchart2). As in that group the effect of manipulated primes is much stronger, we believe this underlies the 'loss' of the phonotactic effect in a floor effect (hardly any priming is left anyway).

The model contains a very strong interaction for manipulated primes of the [suC] type ($t = 3.6$); these primed much less than expected, close to cancelling the whole priming effect.

3. Discussion

The results indicate that the data follow existing predictions: phonotactics help speech processing in general, if the clusters are presented as clusters. As we corrected for word recognition times with a baseline, we isolated them from lexical neigh-

---

[1] However, the crossed random effect issues that Quené and van den Bergh address did not turn out to be too severe in these data.

[2] Not corrected for multiple comparisons.

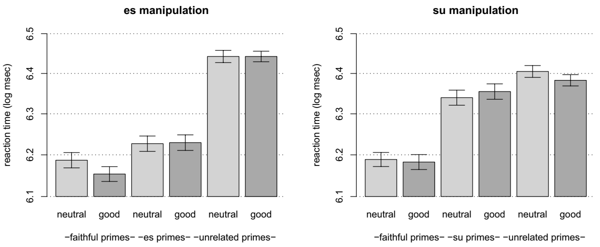

Figure 1: *Group means of log(reaction time) for the ɛsC and the suC manipulation run. Phonotactic conditions are shown in light gray (neutral) and dark grey (good).*

bourhood effects, even though there probably were such effects. The experimental design and the model analysis correct for the manipulation, but the V&L findings that phonotactic well-formedness play a role cannot be supported by these data for the present phonotactic contrast.

The interesting finding is that splitting a cluster with a vowel is significantly worse than a prothesis type of epenthesis [ɛs]. Visual inspection suggest that the good clusters suffer and even more so for being broken, but the data only suggests tendencies in these directions. The simple main effects found for the interaction were included for purposes of comparison with earlier findings, mainly V&L. More research needs to be done to independently show the effect.

4. Conclusions

The phonotactic well-formedness of the stimuli seem to have an effect on their recognition, as it shows up in priming. Destroying clusters reduces priming effects, showing that the clusters are represented as such. The different effects for the two manipulations show that just phoneme similarity is not a sufficient predictor for priming and by inference, for word recognition: both manipulations added just one vowel.

The effect on both good and neutral clusters seems to reside in the fact that clusters are easier to recognise if they are presented as clusters. This is corroborated by the fact that the [suC] manipulation was much worse than the [ɛsC] manipulation.

The clusters cannot be reduced to their parts, indicating that we have to assume at least some phonological property of the combination that is not derived from its parts. Still, the phonological property itself has not been identified.

5. Acknowledgements

This research is funded by the Netherlands Organisation for Scientific Research NWO, under project grant 277-70-001. The author would like to thank René Kager for helping me formulate my hypothesis. Together with Keren Shatzman he pushed me to execute the experiment. I am indebted to Keren for her help with the design of the experiment. Hugo Quené deserves many thanks for putting me on track with the statistical analysis. I thank the three anonymous reviewers for pointing out flaws. Finally, I would like to thank the other members of our research group at the time of designing the experiment, Frans Adriaans and Natalie Boll-Avetisyan, for discussing my work, which was of great value to force me to organise my thoughts and mold them into proper hypotheses. Last but not least I thank Frits van Brenk for his work during the stimulus creation.

6. References

[1] M. S. Vitevitch and P. A. Luce, “When words compete: Levels of processing in perception of spoken words,” *Psychological Science*, vol. 9, no. 4, pp. 325–329, 1998.

[2] ——, “Probabilistic phonotactics and neighbourhood activation in spoken word recognition,” *Journal of Memory and Language*, vol. 40, pp. 374–408, 1999.

[3] N. Oostdijk, “The Spoken Dutch Corpus. overview and first evaluation,” in *LREC-2000 Workshop Proceedings of the workshop on meta-descriptions and annotation schemes for multi-modal/multi-media language resources.*, D. Broeder, H. Cunningham, N. Ide, D. Roy, H. Thompson, and P. Wittenburg, Eds., 2000, pp. 21–25.

[4] P. Jusczyk, A. Friederici, J. Wessels, V. Svenkerud, and M. Jusczyk, “Infants’ sensitivity to the sound patterns of native language words,” *Journal of Memory and Language*, vol. 32, pp. 402–420, 1993.

[5] P. Jusczyk, P. Luce, and J. Charles-Luce, “Infants’ sensitivity to phonotactic patterns in the native language,” *Journal of Memory and Language*, vol. 33, pp. 640–645, 1994.

[6] J. M. McQueen, “Segmentation of continuous speech using phonotactics,” *Journal of Memory and Language*, vol. 39, pp. 21–46, 1998.

[7] E. Dupoux, K. Kakehi, Y. Hirose, C. Pallier, and J. Mehler, “Epenthetic Vowels in Japanese: A Perceptual Illusion?” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, no. 6, pp. 1568–1578, 1999.

[8] E. Dupoux, Ed., *Language, Brain, and Cognitive Development. Essays in Honor of Jacques Mehler.* Cambridge, MA: MIT Press, 2001.

[9] H. Fleischhacker, “Cluster-dependent epenthesis asymmetries,” in *Papers in Phonology 5*, ser. UCLA Working Papers in Linguistics, A. Albright and T. Cho, Eds. UCLA, 2001, vol. 7, pp. 71–116.

[10] T. M. Bailey and U. Hahn, “Determinants of wordlikeness: Phonotactics or lexical neighborhoods,” *Journal of Memory and Language*, vol. 44, pp. 568–591, 2001.

[11] P. Hallé, A. Dominguez, F. Cuetos, and J. Segui, “Phonological mediation in visual masked priming: Evidence from phonotactic repair,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 34, no. 1, pp. 177–192, 2008.

[12] R. Baayen, R. Piepenbrock, and L. Gulikers, “The CELEX lexical database (release 2),” CD-ROM, Philadelphia, PA, 1995.

[13] H. Quené and H. van den Bergh, “On multi-level modeling of data from repeated measures designs: a tutorial,” *Speech Communication*, vol. 43, pp. 103–121, 2004.

[14] ——, “Examples of mixed-effects modeling with crossed random effects and with binomial data”,” *Journal of Memory and Language*, vol. doi:10.1016/j.jml.2008.02.002, in press.

[15] R. Baayen, D. Davidson, and D. Bates, “Mixed effects modelling with crossed random effects for subjects and items,” *Journal of Memory and Language*, to appear.