The role of phonotactics and lexicality on the perception of intrusive vowels

Ela Portnoy and Elinor Payne
University of Oxford

It is well-established that listeners' knowledge of phonotactics influences speech perception in a variety of ways [1,2,3]. This includes processing of consonant clusters, where it has been shown that listeners are biased towards hearing intrusive vowels in illegal clusters [4]. Less well- understood is listeners' susceptibility to hearing vocalic intrusion in *legal* clusters, i.e. where the auditory input can map onto either CC or CVC representations. It is hypothesised that the perception of such sequences is mediated by different types of linguistic information, such as lexicality and frequency, in conjunction with relevant acoustic cues. In addition, knowledge of the *phonetic implementation* of clusters, which varies cross-linguistically, may also influence perception. [5,6] show that clusters in Norwegian are characterised by an open release of C1 more frequently than their equivalent clusters in SSB English, which in turn are more characterised by reduction and overlap. This pattern is even more pronounced in early child productions, such that the dominant acquisition strategy is cluster reduction in English, and epenthesis in Norwegian. This suggests the perceptual threshold for vocalic intrusion differs cross-linguistically, even when the phonological structure is ostensibly the same.

In this paper, we investigate the influence of lexicality, lexical frequency and phonotactic legality on the perception of clusters and intrusive vowels in SSB English. The study is also a first step towards a cross-linguistic comparison with Norwegian. 22 native speakers of SSBE were asked to identity a series of auditory stimuli, in a forced choice task presented orthographically in PsychoPy. The stimuli were taken from recordings of word and non-word items with an original CVCV(C) structure, manipulated to vary V1 duration (6 different intervals: 0, 15, 30, 45 and 60 ms, plus a control 0ms token taken from a CCVC(C) production). For each stimulus, participants chose between items with CVCV(C) or CCV(C) structure, thus testing the perception or otherwise of the intrusive vowel (V1). The clusters investigated were (legal) /bl/ and /gl/, and (illegal) /dl/, and the paired items formed 13 unique lexical frames, such as "blow"/"below", "blieve"/"believe", "glow/gelow", "dlow"/"delow", "blard"/"belard". Participants heard each variant of each lexical frame 4 times, in randomised order, and their responses and reaction times were recorded.

Preliminary results (see Figures 1-5) show a strong effect of both lexicality and phonotactic legality on the perceptual threshold for intrusive vowels, as well as on degree of certainty (response number and speed). Overall, the longer the intrusive vowel, the more likely it would be perceived. However, as hypothesised, the durational threshold varied according to the linguistic constraints of each lexical frame. Where *CVCV(C)* was the only real word in the frame (e.g. "blieve"/"believe"), a vocalic interval of just 15ms was needed for a clear majority response of CVCV(C) ("believe"), and RTs were quicker for CVCV(C) responses. Even where stimuli had no intrusive vowel (0ms), the CVCV(C) response was given 5-23% of the time, suggesting listeners were processing these as reduced variants of the word "believe". By comparison, where *CCV(C)* was the only real word (e.g."glow"/gelow"), intrusive vowels needed to be even longer to be perceived (at least 30ms, and 45ms for CVCV(C) ("gelow") to be the preferred response) and RTs were slower. Up to half of all stimuli with a vocalic interval of 30-60ms were heard as CCV(C) (which suggests they were mapped onto the real word "glow"), whereas stimuli with no vocalic interval very rarely yielded a response of CVCV(C) ("gelow"). Where both items were non-words, (e.g. "blard"/"belard"), a shorter interval was required for perception: with 15ms of vocalic interval, "belard" was heard with some regularity, although 30ms were needed before "belard" became the preferred response, and 45ms before being strongly preferred. However, even with 60ms there were some responses of "blard", and, analogously, at 0ms there were some responses of "belard", suggesting greater symmetry of ambiguity. When both items were non-words but one phonotactically illegal ("dlow"/"delow"), the illegal CCV(C) response ("dlow") was more strongly dispreferred overall.

We discuss the significance of these findings for our understanding of phonological representation and in particular of phonotactic knowledge, and look ahead to possible implications for acquisition.

[1]Jusczyk, P. W. 1995, « Language acquisition : speech sounds and the beginnings of phonology » In J. L Miller and P. D. Eimas (Eds), *Handbook of perception and cognition, Vol 11 : Speech, language and communication*, 263-301, San Diego, Academic Press

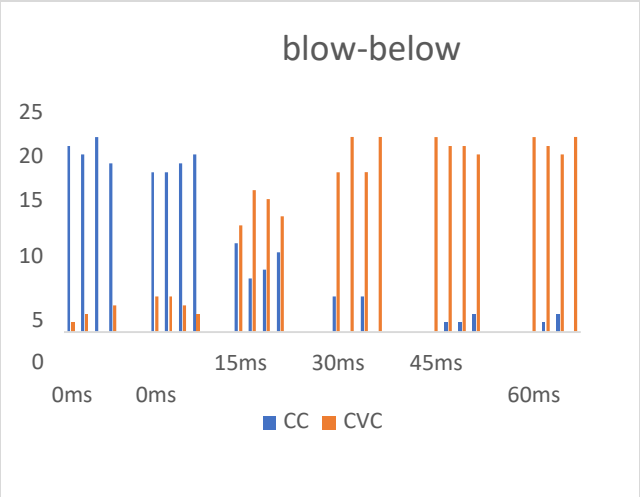
[2] Massaro, D. M. & M. M. Cohen, 1983, "Phonological context in speech perception" in *Perception and Psychophysics*, 34:4, 338-348

[3] Davidson, L. & J. Shaw, 2012, "Sources of illusion in consonant cluste perception" in *Journal of Phonetics*, 40:2, 234-248

[4] Pitt, M. A. 1998, Phonological processes and the perception of phonotactically illegal consonant clusters, Perception and Psychophysics, 60; 6, 941-951

[5] Simonsen, H. N. Garmann, E. Payne, B. Post, E. Holm & P. Hansen, 2015, "Cross-linguistic microvariation in cluster production", oral presentation at *ISBMS*, Chania, Sept 2015

[6] Simonsen, H., N Garmann, P. Hansen, B. Post & E. Payne (*in prep*) "Vowel epenthesis in Norwegian and English clusters: a comparative study in adult and child speech"



Figures 1-5: total number of responses (over 22 speakers) for repetitions 1-4, for each vocalic interval, over 5 lexical frames