Interactive and autonomous modes of speech perception: 
Phonological knowledge and discrimination in English and French listeners

Michael Key

Department of Linguistics, University of Massachusetts, Amherst

The processing of the incoming speech signal can occur in two distinct general modes: the autonomous mode and the interactive mode. The autonomous mode of processing is characterized by an apparent lack of influence from sources of linguistic knowledge or bias (e.g. the lexicon) in certain perceptual tasks in which an influence is expected. For example, recent data from Kingston, Mash, Chambless, Kawahara, Katz, & Key (submitted) shows that wordhood biases do not make pairs drawn from a word↔non-word continuum more cumulatively discriminable than pairs drawn from a non-word↔non-word continuum. In contrast, the interactive mode is characterized by the influence of linguistic knowledge (i.e. feedback) in a perceptual task in which an influence is expected. A classic example of the interactive mode are the lexical biases on phoneme identification reported by Ganong (1980). While it is true that the literature is full of studies that report interactive processing (i.e. some linguistic bias) and relatively devoid of studies reporting autonomous processing (i.e. the lack of a bias where one would be expected), it is also the case that the possible combinations of task type and source of linguistic influence has not nearly been exhausted. In particular, the large majority of these studies have relied on recognition tasks (e.g. phoneme identification, word detection, etc.) and have tested for various influences directly from the lexicon (e.g. wordhood, lexical statistics). As a result, we do not yet have a good understanding of what determines which mode of processing is observed.

The study reported here fills in one of these missing combinations by using discrimination tasks to test for the influence of phonological knowledge using non-word stimuli, thus factoring out possible lexical influences. English has a well-known pattern of regressive place assimilation of coronal stops to labials and velars when the following context is labial or velar (e.g. Wells, 1982): e.g. good girl → goo[g] girl. Previous studies using recognition tasks have shown that English listeners recover the identity of a phoneme that matches the lexical representation (Gaskell & Marslen-Wilson, 1996), with a replication in non-word stimuli (Gaskell & Marslen-Wilson, 1998). This study tested the hypothesis that discrimination tasks would contrast with the recognition results in producing an autonomous mode of processing. Specifically, English listeners should have no more difficulty discriminating coronal from non-coronal stops more when the following context is assimilation-viable (i.e. non-coronal), then when it is unviable (i.e. coronal). In this experiment, the voiced oral stops [b, d, g] were presented in the context preceding heterosyllabic [b, d, g] in non-word stimuli of shape VC.CV. On each trial, listeners’ task was to decide whether the pair of stimuli heard were the same or different; the format of discrimination was AX (‘same-different’).

Figure 1 shows mean discrimination performance when C1 varies between a coronal and a non-coronal. The left pair of bars represent the pairs in which both a coronal was present and the following stop context was viable ([b] or [g]). Discrimination of [b] and [d] in the viable labial assimilation context is compared with discrimination of [b] and [d] in the unviable coronal context (gray bars), and discrimination of [g] and [d] in the viable velar context is compared with that in the unviable coronal context (white bars). Discrimination in both viable contexts is significantly poorer, as confirmed by paired one-tailed t-tests (t(15) = 2.73, p = 0.016 in the labial context, t(15) = 2.28, p = 0.037 in the velar context).

![Figure 1](image1.png)

*Figures 1 (left) and 2 (right): Compared d’ values in English listeners (Fig. 1) and French listeners (Fig.2) between pairs in which C2 is a viable assimilation context and pairs in which C2 is not a viable assimilation context (95% CIs).*
In contrast, French lacks this pattern of place assimilation, and thus no difference in discriminability is predicted for French listeners (Figure 2). Darcy, Ramus, Christophe, Kinzler, & Dupoux (to appear) found that English listeners compensated for place assimilation significantly more often than French listeners in a phoneme monitoring task. In contrast, French listeners perform better in the viable labial context than in the unviable coronal context ($t(7) = 3.01, p = 0.019$), but perform worse in the viable velar context ($t(7) = 7.45, p < 0.001$). The difference between the performance of English and French listeners in the viable labial context is most telling for the predictions of our hypothesis because it shows that AX discrimination apparently induces an interactive mode of processing whose source is language-particular phonological knowledge.

In a follow-up of this experiment, the stimuli remained the same but the format was changed to 4IAX, which has previously been shown to encourage an 'auditory' (i.e. non-phonologically-biased) mode of processing (Gerrits & Schouten, 2004). Figure 3 compares discrimination of coronals from non-coronals in both viable (non-coronal) and unviable (coronal) contexts by English listeners.

Figure 3: Mean discrimination of coronals from non-coronals in viable and non-viable contexts (95% CIs).

Discrimination is no worse in viable contexts than in unviable contexts and discrimination of coronals from non-coronals is no worse than discrimination of two non-coronals. discrimination of [b] from [d] is no worse in a viable context (before [b]) than in an unviable context (before [d]) ($t(15) = 0.78, p > 0.10$); discrimination of [b] from [d] is no worse than [b] from [g] with viable context held constant ($t(15) = 1.27, p > 0.10$). Thus, the results from the 4IAX discrimination task show an autonomous mode of processing.

The autonomy-interaction debate has often been cast in terms of models of speech perception rather than modes, as in this paper. There are two reasons I avoid the debate over models. First, arguments for one type of model over another are often fueled by the claim that effects are ‘early’ or ‘late’. Although I have not done a comprehensive survey of these arguments, it is far from clear what point in the course of processing divides ‘early’ from ‘late’; these are terms of art rather than well-defined stages in a model. Second, the choice of model is ultimately a question of parsimony; autonomous models are often claimed to be more complex because of the additional level of processing they posit. However, autonomous and interactive models should not be compared in this way unless specific implementations (i.e. mechanisms) of each are discussed. In short, invoking Ockham’s Razor is appropriate when choosing between fully explicit models, but inappropriate when choosing between classes of models.

**Selected References**


