Accessing psycho-acoustic perception and language-specific perception with speech sounds

Molly Babel & Keith Johnson

Department of Linguistics, University of California, Berkeley
mbabel@berkeley.edu; keithjohnson@berkeley.edu

In this paper, we reexamine the results of four experiments that explore the difference between language-specific perception and psycho-acoustic auditory perception of speech sounds. These experiments examine this processing difference through two types of tasks: speeded discrimination and similarity rating. These tasks are preferred over labeling tasks as they allow listeners to judge the perceptual similarity of two tokens without having to make explicit judgments about the proper orthographic representation for the sound or whether such a symbol even exists within their repertoire (cf. Best et al. 2001). For the similarity rating task, listeners are asked to compare two stimuli that are presented with a 100 ms ISI on a 5-point equal interval scale ranging from “very similar” to “very different”. This task allows for more leisurely processing of the tokens and, therefore, language-specific perception is evaluated. For the speeded AX discrimination tasks, two stimuli are presented consecutively with a 100 ms ISI. Listeners are instructed to press assigned buttons to indicate if the two sounds are “same” or “different”. Listeners are instructed to respond within 500 ms; their response time and accuracy are presented on the screen between trials to motivate their performance. The fast-paced nature of this task is designed to bypass linguistic processing and hone in on pure acoustic similarity (Pisoni, 1973; Pisoni & Tash, 1974; Fox, 1984).

The first set of experiments (Johnson & Babel, 2007) investigated the perception of voiceless fricatives [θ θ s θ x h] in three vowel environments [i a u] in V.CV₁ sequences by Dutch and English listeners using the two tasks described above. In the similarity rating task, the listener rating scores were analyzed in a repeated measures ANOVA. There was a significant main effect of vowel context (F[2,52] = 65.3, p < 0.01) and consonant pair (F[14, 364] = 94.5, p < 0.01). The consonant pair and vowel interaction was significant (F[28, 728] = 15.5, p < 0.01). Crucially, a consonant pair by language interaction was also found (F[14,364]=3.8, p < 0.01), along with a three-way consonant pair by vowel by language interaction (F[28,728]=1.6, p <0.05). For the speeded discrimination task only correct “different” responses were analyzed in a repeated measures ANOVA. There were no significant language effects. A main effect of consonant pair (F[14, 393] = 15.8, p < 0.001) and vowel environment (F[2, 11] = 18.9, p < 0.001) were found. There was also an interaction between consonant pair and vowel (F[28, 888] = 18.8, p < 0.001).

In the second set of experiments (Babel & Johnson, 2007), the perception of Russian palatalization was compared across Russian and English listeners. Four degrees of Russian palatalization taken from Russian words were implemented in the task design: CV [mat] ‘checkmate’, CjV [mat] ‘rumpled’, CjV [sad] ‘judge’, and CjjV [zmju] ‘snake’. The stimuli were formatted as one of the four aforementioned sequences with a vowel /a u i/ and one of six possible onsets: /m v b d l r/. In the similarity rating task, a repeated measures ANOVA found main effects for language (F[1, 21]=9, p < 0.01), palatalization pair (F[5, 105]=369.2, p < 0.001) and vowel context (F[2, 42]=21, p < 0.001). There were also significant interactions between palatalization pair and vowel (F[10, 210]=18, p < 0.001) and palatalization and language (F[5, 105]=3.4, p < 0.01). Again, “same” pairs of stimuli and incorrect responses were removed prior to the analysis for the speeded discrimination task. A repeated measures ANOVA revealed a main effect of palatalization pair (F[5, 16] = 257.28, p < 0.001) and consonant (F[5, 4] = 13.63, p < 0.05). There were significant interactions between language and palatalization (F[5,16] = 4.52, p < 0.01), palatalization and consonant (F[25, 460] = 2.28, p <0.001), and palatalization and vowel (F[10, 230]=6.83, p < 0.001). The ANOVA also returned a three-way interaction between palatalization, consonant, and vowel (F[50, 1105] = 2.7, p < 0.001). Planned comparisons of the language and palatalization interaction revealed that there were no significant differences between listener groups with the palatalization pairs.
In both sets of experiments, the similarity rating tasks found strong language effects that reflect the phonological organization of the languages. For example, Dutch listeners do not contrast /s/ and /ʃ/ and they rated these sounds as more similar-sounding than English listeners in the rating tasks. The speeded AX discrimination experiments did not reveal language effects; all listeners in these tasks performed comparably, regardless of language background. The results of these experiments suggest that psycho-acoustic perception can be evaluated apart from linguistic perception.

Other work using this paradigm found language effects in AX discrimination as well as rating tasks (Boomershine, Hall, Hume, & Johnson, in press; McGuire, 2007). To reconcile these results with our work, we adopted a binned reaction time analysis with the Boomershine et al. data and the McGuire data. The analysis examined the language effects in the AX discrimination tasks in their experiments through three binned range groups. Range 1 was composed of responses made faster than 500 ms, Range 2 were those responses logged between 500 and 800 ms, and Range 3 were those made after 800 ms. It was predicted that language effects would be found only in slower responses that allowed for linguistic processing. This analysis and its predictions follow from Fox (1984). Fox demonstrated that the lexical effects of the Ganong effect disappear in responses made prior to 800 ms. Our analysis found the language effect in the McGuire data was significant only in the response latencies greater than 800 ms. In the binned range analysis of the Boomershine et al. data the language effects remained, but were the result of consistently longer response latencies by the Spanish listeners and not in any linguistically interpretable way. We argue that the language effect in their data is retained due to delayed reaction times and the bilingualism of the Spanish participants. The work reported argues for multiple stages in perception. Specifically, we provide evidence that it is possible to experimentally probe a stage of auditory perception that is not affected by language specific perceptual warping.

Acknowledgements

This research has been supported by NIH grant R01 DC004421-05 awarded to the second author. Thanks to Grant McGuire, Amanda Boomershine, Kathleen Currie Hall, and Elizabeth Hume for sharing their raw data with us.

References


