Journal of Phonetics 46 (2014) 147–160

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

The role of abstraction in non-native speech perception

Bozena Pajak *, Roger Levy

University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0108, USA

ARTICLE INFO

Article history:
Received 16 July 2013
Received in revised form
25 June 2014
Accepted 2 July 2014

Keywords:
Non-native speech perception
Sound discrimination
Perceptual reorganization
Naïve listeners
Cross-linguistic influence
Inductive inference

ABSTRACT

The end-result of perceptual reorganization in infancy is currently viewed as a reconfigured perceptual space, "warped" around native-language phonetic categories, which then acts as a direct perceptual filter on any non-native sounds: naïve-listener discrimination of non-native-sounds is determined by their mapping onto native-language phonetic categories that are acoustically/articulatorily most similar. We report results that suggest another factor in non-native speech perception: some perceptual sensitivities cannot be attributed to listeners' warped perceptual space alone, but rather to enhanced general sensitivity along phonetic dimensions that the listeners' native language employs to distinguish between categories. Specifically, we show that the knowledge of a language with short and long *vowel* categories leads to enhanced discrimination of non-native *consonant* length contrasts. We argue that these results support a view of perceptual reorganization as the consequence of learners' hierarchical inductive inferences about the structure of the language's sound system: infants not only acquire the specific phonetic category inventory, but also draw higher-order generalizations over the set of those categories, such as the overall informativity of phonetic dimensions for sound categorization. Non-native sound perception is then also determined by sensitivities that emerge from these generalizations, rather than only by mappings of non-native sounds onto native-language phonetic categories.

1. Introduction

The development of speech perception in the first year of life provides a critical foundation for future language learning. Infants undergo profound *perceptual reorganization* (Eimas, 1978): they transition from discriminating almost any speech sound distinction (including those absent from their ambient language) to a state of enhanced sensitivity to native-language (L1) distinctions, accompanied by a decline in sensitivity to many non-native distinctions (Werker & Tees, 1984; for reviews, see Werker, 1989; Kuhl, 2004). These results have led to the development of theories in which perceptual reorganization is understood as resulting from the acquisition of the specific inventory of native-language phonetic categories,[1] and the end-state is a reconfigured ("warped") perceptual space, where innate perceptual sensitivity along natural auditory boundaries is replaced by sensitivity along boundaries of phonetic categories in the learner's native language (Kuhl, 1991, 2000).

As a consequence, the long-held assumption underlying the research on non-native speech perception has been that non-native speech is necessarily "filtered" through listeners' L1 phonetic category inventory. The "L1-category filter" metaphor can be traced back to Trubetzkoy (1939/1969), and the essence of this idea is present in current theories of non-native speech perception and learning: the Native Language Magnet model (NLM, Kuhl, 1992, 1994, 2000; Kuhl & Iverson, 1995; Kuhl et al., 2008), the Speech Learning Model (SLM, Flege, 1988, 1992, 1995), and the Perceptual Assimilation Model (PAM and PAM-L2, Best, 1993, 1994, 1995; Best & Tyler, 2007). These theories, while different in several respects, preserve the basic insight captured in the "L1-category filter" metaphor: that the perceptual space warped in accordance with the L1 phonetic category inventory – the end-result of perceptual reorganization in infancy – acts as a perceptual filter when processing non-native languages. Specifically, according to these theories, naïve-listener and second-language (L2) learner discrimination of non-native sounds is determined by their mapping onto

* Corresponding author. Present address: Department of Brain & Cognitive Sciences, Meliora Hall, Box 270268, University of Rochester, Rochester, NY 14627-0268, USA. Tel.: +1 585 275 7171; fax: +1 585 442 9216.
E-mail address: bpajak@bcs.rochester.edu (B. Pajak).

[1] Throughout the paper we use the term "phonetic categories", not making an explicit distinction between phonemes and allophones. For discussion on the relationship between phonetic and phonological levels in perception and learning of sound inventories, see, for example, Best and Tyler (2007) and Dillon, Dunbar, and Idsardi (2013).

http://dx.doi.org/10.1016/j.wocn.2014.07.001

specific L1 phonetic categories that are acoustically or articulatorily most similar, if such categories are available. Broadly speaking, discrimination of non-native contrasts is thought to be impaired when the stimuli are mapped (i.e., perceptually assimilated) onto the same L1 category (with varying performance depending on the goodness of fit to that category), relative to when they are mapped onto differing categories.

These classic theories have been very successful in explaining a wide range of perceptual difficulties in non-native speech perception and learning (Best & Hallé, 2010; Best & Strange, 1992; Best, McRoberts, & Goodell, 2001; Flege & Eefting, 1987; Hallé, Best, & Levitt, 1999; McAllister, Flege, & Piske, 2002; Miyawaki et al., 1975; Polka, 1991, 1992, among others; for a review see Strange & Shafer, 2008), showing that the degree of similarity between native and non-native sounds – as assessed through acoustic and articulatory comparisons or direct measures of perceived similarity – can predict performance on discrimination of non-native sound pairs. That is, if two non-native sounds are both assessed as highly similar to a single L1 category, their discrimination is impaired. On the other hand, if each sound in the non-native pair is highly similar to a distinct L1 category, then their discrimination is facilitated. A widely cited example is the difficulty of L1-Japanese speakers in discriminating the English [ɹ]-[l] distinction, which is generally attributed to Japanese only having one phonetic category in the same acoustic-phonetic range (Goto, 1971; Miyawaki et al., 1975; Strange & Dittmann, 1984). This type of example has been used as evidence supporting the classic theories since perceptual difficulties can in this case be explained by L1-Japanese listeners' assimilating both of the non-native sounds onto a single L1 category.

However, recent evidence suggests that the theories of non-native speech perception might need to be extended to accommodate discrimination patterns that cannot be explained by specific L1 phonetic category inventories. In particular, it has been shown that native French, Danish, and German listeners outperform English native speakers on discriminating the English [w]-[j] contrast (Bohn & Best, 2012; Hallé et al., 1999). These results are extremely surprising for at least two reasons: (1) native speakers performed more poorly on discriminating sounds from their native language than did non-native listeners, and (2) this non-native perceptual advantage was observed for speakers of languages that do not even have [w] in their inventory (Danish and German). Bohn and Best suggested that these results could be explained by the influence of more general characteristics of the L1 inventory on the listener's perceptual system: French, Danish, and German have a relatively rich vowel inventory, and – unlike English – use lip rounding contrastively for vowels; since lip rounding is one of the cues to distinguish [w] from [j], the practice with lip rounding to distinguish between L1 vowels might boost French, Danish, and German listeners' performance on discriminating the [w]-[j] contrast. In other words, phonological principles – such as whether or not L1 uses a given feature contrastively – may affect perception of non-native contrasts.

Bohn and Best's proposal echoes prior suggestions that phonological distinctive features may affect non-native speech perception and learning (e.g., Brown, 1997, 2000; Hancin-Bhatt, 1994; McAllister et al., 2002). For example, McAllister et al. (2002) developed one of the components of the SLM into the “feature hypothesis” stating that “[phonological] features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature” (p. 230). That is, under the feature hypothesis, the difficulty in learning a given L2 contrast based on feature X is determined by the role of feature X in the learner's L1. In particular, forming an L2 phonetic category may be more difficult (or even blocked) if it requires attending to a feature that is not exploited in the learner's L1. As support for their hypothesis, McAllister and colleagues showed that L2 learners of Swedish are more successful at acquiring the short vs. long vowel distinctions if their L1 also employs the length feature to distinguish between vowels (as for L1-Estonian learners), relative to the case when vowel contrasts in the learners' L1 either only use duration as a secondary cue (as for L1-English learners) or do not make use of duration (as for L1-Spanish learners). In a similar vein, Brown (1997, 2000) argued for a model of non-native speech perception in which any phonological distinctive features used contrastively in L1 would be transferred to L2, which would in turn facilitate discrimination of any non-native sounds that are contrasted by those features.

Converging proposals can be found in the general literature on categorization and perceptual learning (e.g., Goldstone, 1993, 1994; Nosofsky, 1986), where native-language phonetic learning is viewed as shifting attention to relevant acoustic-phonetic cues (e.g., Best, 1994; Jusczyk, 1985, 1986, 1994, 1997; Nusbaum & Goodman, 1994; Nittrouer & Miller, 1997; Pisoni, Lively, & Logan, 1994). Furthermore, there is a growing body of work showing that the difficulty of non-native sound perception and learning is modulated by perceptual weighting of the relevant acoustic and articulatory cues in L1 (e.g., Escudero, 2009; Escudero & Boersma, 2004; Escudero, Benders, & Lipski, 2009; Francis & Nusbaum, 2002; Iverson, Hazan, & Bannister, 2005; Iverson et al., 2003; Kondaurova & Francis, 2008; Kondaurova & Francis, 2010; Lim & Holt, 2011). However, while these approaches emphasize the role of individual acoustic-phonetic cues in speech perception and learning, they largely focus on explaining the *difficulties* in discriminating and learning the contrasts that differ along unattended dimensions. It is in fact unclear whether they would predict any facilitatory effects on discriminating non-native sounds (e.g., [w]-[j]) that differ along dimensions attended in L1 but occurring in different contexts (such as the lip rounding cue distinguishing between the high front unrounded vowel [i] and the high front rounded vowel [y]). This is because selective attention is generally thought to operate on particular configurations of cues that are largely context-specific: that is, specific to particular segmental contexts, word positions, talkers, etc. (e.g., Francis, Baldwin, & Nusbaum, 2000; Francis & Nusbaum, 2002; Logan, Lively, & Pisoni, 1991; Nusbaum & Goodman, 1994; Pisoni et al., 1994; Reinisch, Wozny, Mitterer, & Holt, 2014). Therefore, these approaches do not immediately capture the potential facilitation in non-native speech perception coming from attending to cues that are relevant in L1, but occur in very different acoustic-phonetic contexts.

Some support for the idea that more general phonological principles might affect non-native speech perception comes from the literature on speech perception and learning by infant learners. In particular, it has been shown that exposing infants to particular sound categories leads not only to their enhanced sensitivity to those specific categories, but also to generally enhanced sensitivity to