"Similar or different?" – The phonetics and phonology of similarity in non-native vowel perception

Nadja Kerschhofer-Puhalo[1]
[1]*University of Vienna*

Due to the high dynamics and variability in the speech signal, perceptual vowel identification cannot be predicted directly from the acoustic properties of the signal but is determined by the listener's attentional tuning to specific acoustic cues and perceptual dimensions varying language-specifically and even individually. Acoustically different vowel sounds are perceived at some cognitive level as more or less "*similar*" or "distinct" from others, belonging either to the same or to different *categories*. In experimental identification tasks, listeners' perception of acoustically different sounds as belonging to the "same" or a "different" category are conditioned by signal-inherent as well as external factors from the listeners' language knowledge, context and expectations. The categorization of sounds is based on relationships of "*similarity*" between single items and mental categories of a given language. Language-specific or more general physical and cognitive *biases* determine the categorization of a given input as "*similar*" to items belonging to mental categories. *Similarity* has been used as one of the central concepts in many models of second language speech perception and acquisition (e.g. Best 1995; Flege 1987, 1995; Kuhl 1992, 1993). Yet, the operationalization of this construct has so far not deserved sufficient attention.

Based on response material from a prior study on non-native vowel identification (Kerschhofer-Puhalo 2014, in print), this presentation will discuss empirically grounded ways of *operationalizing perceptual similarity* in L2 in terms of (1) *acoustic similarity* scores and (2) *psychological similarity* and distances. In a vowel identification task, 15 German vowel types were presented in non-words with differing consonantal context to 173 L2 learners of German from 10 L1 subsamples (+ a native control group). Participants were asked to match the input stimuli with response categories, which consisted of all full vowel categories of German. The L2 listeners' responses were summarized in confusion matrices and were – together with data from the acoustic analysis of the input stimuli – subject to higher level statistical analysis. Multidimensional Scaling (MDS) was applied to visualize (intra-lingual) similarity of German vowel phonemes in a geometric 2- or 3-dimensional spatial representation of the L2 vowel space (Shepard 1972; Terbeek 1977; Johnson 2012). Several previous experiments using MDS solutions have postulated a high correspondence between spatial distances in MDS solutions and acoustic-phonetic properties such as vowel formants (Kewley-Port & Atal 1989; Iverson & Kuhl 1995; Fox, Flege & Munro 1995; Francis & Nusbaum 2002). Alternatively to more traditional mono-directional conceptions of similarity between L1 and L2 sounds and the claim that statistical correlations with acoustic properties are sufficient to understand *perceived similarity in L2*, we favour a *cross-linguistic influence*-approach focussing on *biases* associated with properties of *stimuli* (acoustic-phonetic) as well as *responses* (phonological) to account for ease and difficulty, preferences and avoidance in L2 perception experiments. A cross-language comparison of the L1 subsamples shows that – rather than predicting perceptual similarity directly from acoustic phonetic properties – perceptual *similarity* s_{ij} between vowel categories of the target language has to be modelled as the result of the complex interaction of (1) *phonetic proximity* p_{ij}, (2) *stimuli* biases b_i and (3) response *biases* b_j ($s_{ij} = p_{ij} * b_i * b_j$). We will show that *biases* vary according to characteristics of the acoustic signal, the set of stimuli and response categories presented in the experimental setting as well as to the L2 learners' language experience (in L1, L2, Ln), L2 proficiency and their individual conception of the target language vowel system.

[1] Best, C. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, pp. 171–204.
[2] Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, pp. 47–65.
[3] Flege, J. 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W. (ed.). Speech Perception and Linguistic Experience: Issues in Cross-Language Research. Timonium, MD: York Press, pp. 233–277.
[4] Fox, R., Flege, J., Munro, M. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA*, 97, pp. 2540–2551.
[5] Francis, A. & Nusbaum, H. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28 (2), pp. 349–366.
[6] Hillenbrand, J. 2013. Static and dynamic approaches to vowel perception. In: Morrison, G. & Assmann, P. (eds.). *Vowel inherent Spectral Change. Modern Acoustics and Signal Processing*. Berlin, Heidelberg: Springer, p. 9–30.
[7] Iverson, P. & Kuhl, P. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *JASA*, 97, pp. 553–562.
[8] Johnson, K. 2012. *Acoustic and Auditory Phonetics*. 3rd ed. Wiley-Blackwell.
[9] Kerschhofer-Puhalo, N. 2014. *Similarity, Cross-linguistic Influence and Preferences in Non-Native Vowel Perception. An experimental cross-language comparison of German vowel identification by non-native listeners*. Dissertation, University of Vienna, DOI: https://doi.org/10.25365/phaidra.51.
[10] Kerschhofer-Puhalo, N. (in print). Notions of similarity in non-native vowel perception. *Proceedings der P&P14. 14. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, University of Vienna, Sept 6-7, 2018.
[11] Kewley-Port, D. & Atal, B. 1989. Perceptual distances between vowels located in a limited phonetic space. *JASA*, 85, pp. 1726–1740.
[12] Kuhl, P. 1992. Speech prototypes: studies in the nature, function, ontogeny and phylogeny of the 'centers' of speech categories. In: Tohkura, Y., Vatikiotis-Bateson, W., Sagisaka, Y. (eds.). *Speech Perception, Production and Linguistic Structure*. Tokyo: Ohmsha, pp. 197–219.
[13] Kuhl, P. 1993. An examination of the 'perceptual magnet' effect. *JASA*, 93, p. 2423.
[14] Shepard, R. 1972. Psychological representation of speech sounds. In: David, E. & Denes, P. (ed.). *Human Communication: A Unified View*. New York: McGraw-Hill, pp. 67–113. [http://courses.cit.cornell.edu/ling4476/shepard-on-speech-sounds.pdf]
[15] Shepard, R. 1980. Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, 210, pp. 390–398. [http://serendip.brynmawr.edu/local/scisoc/emergence/shepard80.pdf]
[16] Terbeek, D. 1977. *A Cross-Language Multi-Dimensional Scaling Study of Vowel Perception*. PhD Dissertation. University of California, Los Angeles. [http://escholarship.org/uc/item/3nx4g138?display=all#page-1]