Quantifying cue trading in word decoding tasks

Louis ten Bosch[1]*, Odette Scharenborg*[2]
[1]*CLST, Radboud University Nijmegen*
[2]*Max Planck Institute for Psycholinguistics, Nijmegen*

Listeners can make use of multiple acoustic cues for each phonological contrast. It is well known that the absence of some cues may be compensated by the presence of other cues. In this paper, we investigate cue trading in the broader context of speech processing by using a computational model of human word processing (cf. Werker & Curtin, 2005). Cue trading has been considered an explanatory mechanism for phoneme perception, see e.g., the Fuzzy Logical Model of Perception (FLMP; Massaro & Oden, 1980) and normal *a posteriori* probability (NAPP) models (Nearey, 1997). Both NAPP and FLMP deal with probabilistic phone classification and treat cue weighting as a category-dependent process. This, however, leaves open the question to what extent cue trading plays a role in the context of word or speech processing – which is a broader context than speech sound categorization which has been the more conventional context in which cue trading has been studied. The here presented approach allows a precise quantification of the amount of cue trading as observed during speech decoding on a speech corpus.

Cue trading must be learned. It therefore makes sense to seek for mechanisms that explain cue integration and weighting as a result of an acquisition process. Toscano & McMurray (2010) show that cue-weighting provides a good fit to the perceptual data, but only when the weights emerged through the dynamics of learning. In line with Toscano & McMurray (2010), we address cue trading as a result of learning. We developed a method to quantify cue trading between articulatory features (AFs, e.g. Browman & Goldstein, 1992) as operational during a word decoding task. AFs describe the speech signal in terms of estimated values of, e.g., *manner* and *place of articulation* (see Table I). This representation allows more freedom in the description of the speech signal than the phoneme description.

The model used is HMM-based. In this model, the phone models were conventionally defined as Hidden Markov Models and lexical items were defined in terms of sequences of phones. In contrast with conventional ASR training, however, the phone models were initiated (without training) by using canonical articulatory feature definitions according to table I. The HMM paradigm enables us to adapt these parameters during an actual decoding task, such that the resulting parameters can be interpreted as cue weights (cf. McMurray, Aslin, Toscano, 2009). The cue weights are directly interpretable as measures of sensitivity to changes in any of the features. This method relates to the way Clayards, Tanenhaus, Aslin, and Jacobs (2008) demonstrated (for a different task) that artificially manipulating the variance of an acoustic cue changes how listeners weight it perceptually.

The model was applied on 2000 Dutch utterances from the database CAREGIVER (Altosaar et al., 2010). To that end, these utterances were represented as sequences of vectors with AFs. Figure 1 shows the found optimal phone-dependent cue weighting for each of the 33 features, in six situations: without any training and after each of in total 5 adaptations. Of all the AFs considered, *manner* and *place* are the most relevant ones (as shown by the higher weights in Figure 1) in terms of their contribution during word competition and word decoding.

In summary, this model is able to find the cue trading within the AF representation by using actual speech, and in a psycholinguistically interpretable way. It will be used in an update of Fine-Tracker (Scharenborg, 2010).

Altosaar, T., ten Bosch, L., Aimetti, G., Koniaris, C., Demuynck, K., & van den Heuvel, H. (2010). A speech corpus for modeling language acquisition: CAREGIVER. In *Proceedings of the international conference on language resources and evaluation (LREC)* (pp. 1062–1068), Malta.

Browman, C., Goldstein, L., (1992). Articulatory phonology: an overview. Phonetica 49, 155-180.

Clayards M, Tanenhaus M.K., Aslin R.N., Jacobs R.A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition.* 2008;108:804–809.

Massaro D.W., Oden G.C. (1980) Evaluation and integration of acoustic features in speech perception. *The Journal of the Acoustical Society of America.* 1980;67:996–1013

McMurray, B, Aslin R.N., Toscano J,C. (2009) Statistical learning of phonetic categories: Computational insights and limitations. *Developmental Science.* 2009;12:369–378.

Nearey T.M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America.* 1997; 101:3241–3256

Scharenborg, O. (2010). Modeling the use of durational information in human spoken-word recognition. *Journal of the Acoustical Society of America*, 127 (6), 3758-3770.

Toscano, J.C., and McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. Cogn.Sci. 34(3): 434–464.

Werker J.F., Curtin S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning & Development.* 2005;1:197–234.

Table I. Specification of the articulatory features. Nil denotes 'non-applicable'.

AF	**AF value**
manner	plosive, fricative, nasal, glide, liquid, vowel, retroflex, silence
place	bilabial, labiodental, alveolar, palatal, velar, glottal, nil, silence
voice	+voice, -voice
fr-back	front, central, back, nil
round	+round, -round, nil
height	high, mid, low, nil
dur-diph	long, short, diphthong, silence

.

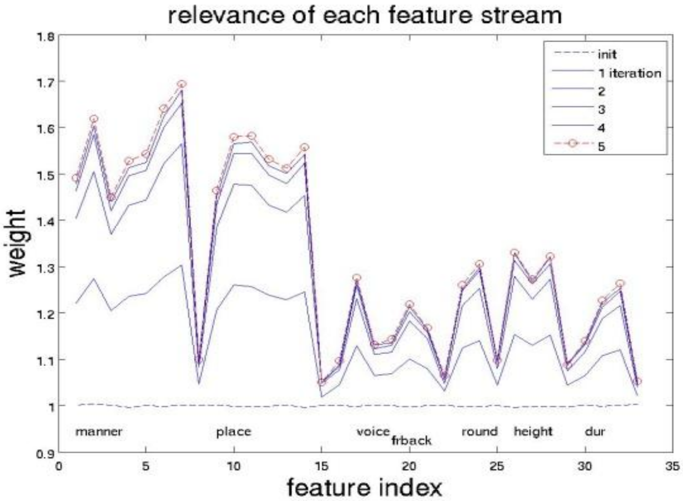

Figure 1. Cue trading (weights) as a result of learning. Relevance (weight) of the AF components, shown for the baseline (dashed curve) and after N iterations, where N=1 to 5.