Perceptual model evaluation using non-categorical data: A case study with Japanese sibilants

The goal of this study is to demonstrate how non-categorical speech perception data can be used in the evaluation of models of speech perception. Speech perception is inherently categorical. The endpoints of perception are distinct, contrastive categories such as words, syllables and segments. In some cases, however, our perceptual system do not fully converge towards a clear-cut interpretation of the signal. Since the ambiguity in the sound is normally resolved contextually, ambiguous sounds usually go unnoticed. However, some experiments explicitly aim to elicit eliciting non-categorical judgements, for example, by using ambiguous, synthetic stimuli. Non-categorical membership judgements, fitted into a sigmoid curve, are often used to determine perceptual boundaries (e.g., Mann & Repp 1980). This type of fuzzy data, however, is usually not employed in model verification. Namely, perceptual models are evaluated against categorical data. For example, a perceptual model of Japanese sibilants would be evaluated against unambiguous [s] and [sh] sounds. Ambiguous data would be removed from the evaluation set. The present study demonstrates how unambiguous data can contribute to the evaluation of perceptual models.

The non-categorical perceptual data for this study was created by eliciting /s/ and /sh/ judgments for a set of synthetic stimuli ranging between /s/ and /sh/ in 7 steps (S1-S7). The synthetic sibilants were incorporated into the carrier sentence of *kono ka__*. An XAB task was used to collect responses, where X was the synthesized stimuli, A and B were natural utterances of *kas* and *kash*—truncated from *kasa* and *kasha*. The 7 stimuli was presented 6 times with both XAB and XBA orderings. The responses resulted in an expected sigmoid curve (Fig. 1). This sigmoid was used for verification of perceptual models.

In order to create models of sibilant perception speech data was collected from 8 native speakers of Japanese through a citation task. In sum 119 [ʃ] and 114 [s] instances were collected, labelled and segmentally aligned. For each sibilant a 24 dimensional MFCC vector was calculated using the middle 50ms portion of the fricative. Based on these feature vectors Gaussian Mixture Models (GMMs) were trained for the two sibilants. No delta or delta-delta values were used. Models with various covariance types and mixture numbers were trained and tested (Fig. 2).

The trained models are to be evaluated against the non-categorical data from the perceptual experiment. However, the comparison of perceptual results and the output of GMM classifier is not a trivial task. The perceptual data, represented by the sigmoid in Fig. 1, is expressed by probability values interpreted in a binary decision task, while the output of GMMs are likelihood estimations. In order to make the two measurements comparable, the output of the perception task was translated into log probability ratios (Fig 3-4). Since the output of GMMs are log probabilities, the difference between model estimations for /s/ and /ʃ/ equals to the log of the ratios for individual probabilities (Fig. 5). This value is directly comparable with the log probability ratios calculated from the perceptual data (Fig. 4).

Although the precise numerical evaluation of the models against the non-categorical data is not presented here, from Fig. 5 it is apparent that GMMs using spherical covariance are preferred over ones with diagonal covariance--as they follow the almost monotonously increasing tendency present in the perceptual data (Fig 4.). While some GMMs with diagonal covariance has good predictions for categorical cases (namely S1 and S7), they fail to capture the general shape of Fig. 4. Although the small data size does not allow for far-reaching generalizations about models of sibilant perception, the case study demonstrates how non-categorical data can be used for model evaluation.

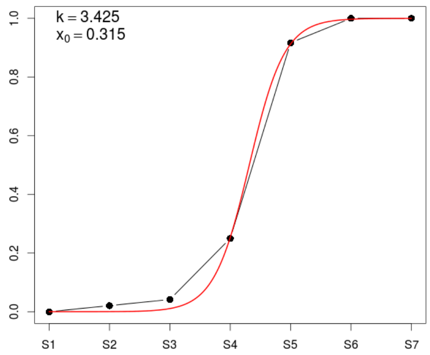

Fig. 1. Probability of /s/ response as a function of stimuli S1-S7

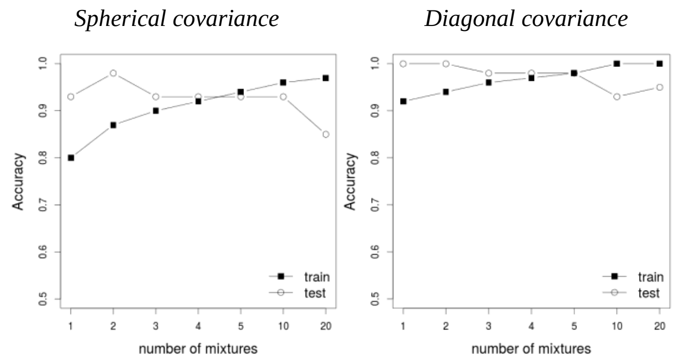

Fig 2. Test and training set accuracies for GMMs with spherical (left) and diagonal (right) covariance.

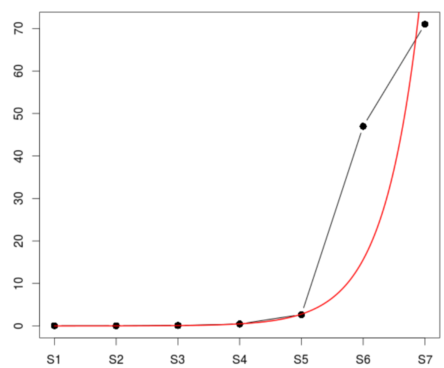

Fig. 3. Probability ratios for values in Fig 1.

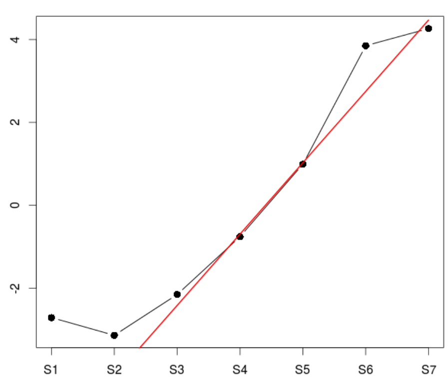

Fig. 4. Log probability ratios for values in Fig 1.

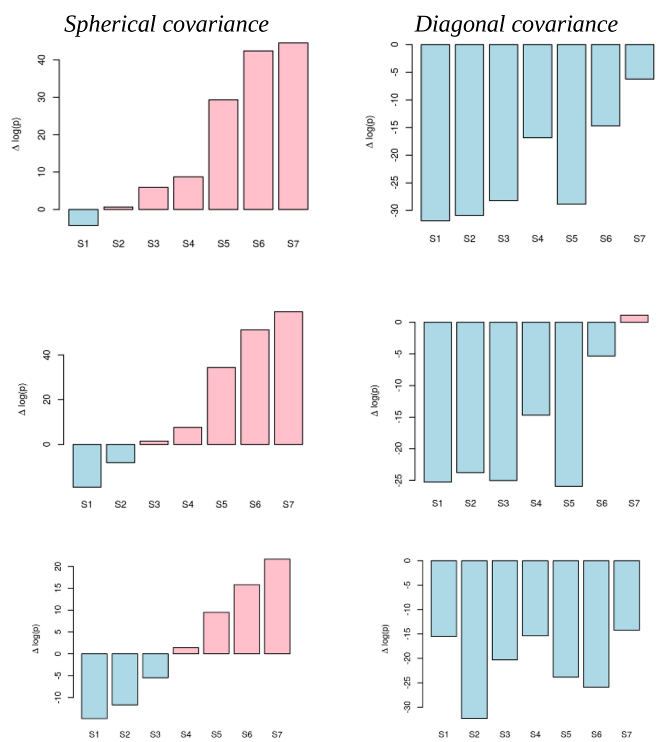

Fig. 5. Log probability ratios calculated as difference between GMMs: $log(p(S)) - log(p(SH)) = log\,\frac{p(S)}{p(SH)}$