LANGUAGE AND COGNITIVE PROCESSES, 2001, *16* (5/6), 683–690

# Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorisation

Joanne L. Miller
*Department of Psychology, Northeastern University, Boston, USA*

## INTRODUCTION

The early stages of speech perception are often characterised in terms of a perceptual mapping between acoustic signal and prelexical phonetic representations. Although early research focused on the abstract categorical nature of these representations, more recent findings have shown that phonetic categories are internally structured in a graded fashion, with some members of the category perceived as better exemplars than others (e.g., Kuhl, 1991; Miller, 1994; Oden & Massaro, 1978; Samuel, 1982). In this paper I present highlights from two recent investigations in our laboratory that examine this structure and its role in processing. The first focuses on how different types of contextual factors influence internal category structure and the second focuses on the role of internal category structure in speeded phonetic categorisation.

## CONTEXT EFFECTS

A widespread finding in the study of speech perception is that the mapping from acoustic signal to phonetic category is highly context-dependent. One form this context dependency takes involves the boundaries between phonetic categories. Numerous studies have shown that both acoustic-phonetic factors and higher-level linguistic factors systematically alter the

---

Requests for reprints should be addressed to Joanne L. Miller, Department of Psychology, Northeastern University, Boston, MA 02115. E-mail: jlmiller@neu.edu

This research was supported by NIH Grant DC 00130 to J. L. Miller and by a Minority Fellowship from the Acoustical Society of America to Sean Allen. The author thanks Sean Allen and Larry Brancazio for their helpful comments on an earlier version of the paper.



http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000065

location of such category boundaries along acoustic continua (see Repp & Liberman, 1987). Over the past few years we have been investigating whether one of these types of contextual factors, namely acoustic-phonetic factors, alters not only the location of phonetic category boundaries, but also which stimuli within a category are perceived as the best category exemplars. Our basic strategy for these studies has been to create extended speech series that deliberately include poor as well as good exemplars of a category, and ask listeners to rate the stimuli in terms of category goodness. The critical issue is how variation in a given contextual factor affects which stimuli along the acoustic continuum are perceived as the best category exemplars.

For example, in one of our early studies on the acoustic-phonetic contextual factor of speaking rate (Miller & Volaitis, 1989), we created two voice-onset-time (VOT) series. Each series ranged from short VOT values appropriate for /bi/ through longer VOT values appropriate for /pi/ to very long VOT values beyond those appropriate for /pi/. The /p/s with very long VOT values (which, for purposes of explication, we label */p/) sounded breathy and exaggerated. The two series differed from each other in speaking rate, specified by syllable duration. In a preliminary two-choice identification experiment, we found that as speaking rate slowed, the /b/-/p/ category boundary shifted toward longer VOT values, as expected from previous research. In the main experiment, which used a category-goodness rating task, we found that the location of the best-exemplar range within the category (defined as the range of stimuli receiving ratings that were at least 90% of that given to the highest-rated stimulus in the series; see Miller & Volaitis (1989) for details) shifted to longer VOT values with the slower speaking rate. In subsequent studies, we have found similar effects on best-exemplar location for other acoustic-phonetic contextual factors, including sentential speaking rate (Wayland, Miller, & Volaitis, 1994) and phonetic-feature context (Volaitis & Miller, 1992) (and see Hodgson & Miller (1996) for related findings). Taken together, these studies demonstrate that changes in acoustic-phonetic context can alter the location of the best exemplars of a category along an acoustic continuum.

In a recent investigation (Allen & Miller, in press) we asked whether a shift in best-exemplar location would occur if a higher-level linguistic contextual factor, rather than an acoustic-phonetic contextual factor, was manipulated. We focused on the higher-level contextual factor of lexical status, i.e., whether a sequence of phonetic segments forms a word of the language or a nonword. It is well known that lexical status affects category boundary location (Ganong, 1980). For example, if listeners are presented stimuli from the two VOT series *beef-peef* and *beace-peace*, they tend to identify stimuli with mid-range VOT values along each series so as to form a real word of the language, *beef* in the first series and *peace* in the second series. As a consequence, the /b/-/p/ category boundary is located at a longer VOT value for the *beef-peef* series than the *beace-peace* series. The question we addressed was whether the effect of lexical status is limited to the region of the category boundary, where there is ambiguity in category membership, or whether the effect pervades the category, altering which stimuli are perceived as the best exemplars.

To answer this question, we created two matched extended VOT series. One ranged from a word to a nonword (*beef-peef-*peef*) and the other ranged from a nonword to a word (*beace-peace-*peace*). In a preliminary two-choice identification experiment we found, as expected, that the /b/-/p/ boundary was located at a longer VOT value for the *beef-peef-*peef* series than the *beace-peace-*peace* series. In the main experiment, listeners were asked to judge the goodness of the initial consonant of each token from the two series as an instance of /p/ using a 1–7 rating scale. We expected to obtain non-monotonic goodness functions for both series. The critical question was whether the stimuli judged to be the best exemplars of the /p/ category would be shifted to longer VOT values for the *beef-peef-*peef* series compared to the *beace-peace-*peace* series.

The main finding, confirmed by statistical analysis, was that the change in lexical status did not shift the location of the entire best-exemplar range, as occurs with acoustic-phonetic contextual factors, but instead shifted only that part of the range bordering the category-boundary region. This is illustrated in the left panel of Figure 1. Initial inspection of the group functions (based on smoothed individual functions) revealed an overall difference in height between the functions, with the tokens receiving somewhat higher ratings when the /p/ was judged in the context of the word (*peace*) than the nonword (*peef*). To adjust for this height difference,

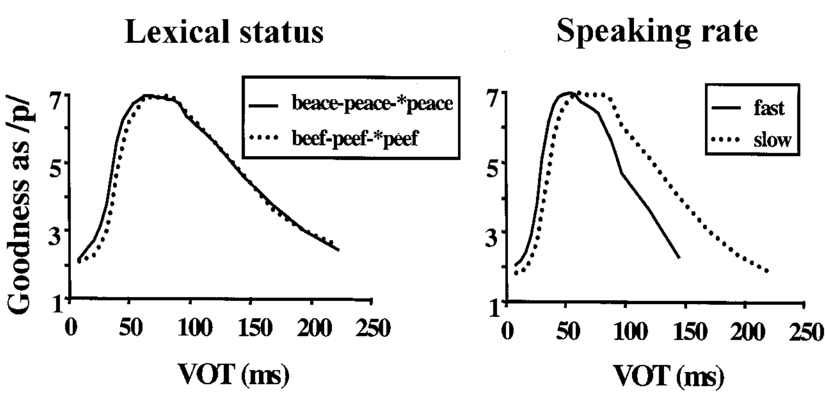


**Figure 1.** Normalised group goodness functions showing the effect of lexical status (left panel) and speaking rate (right panel) on the internal structure of the /p/ category. (Adapted from Figures 1 and 2 of Allen & Miller, in press.)

which is not relevant to our main question, we normalised the two group functions so that they both peaked at a value of 7, and it is these normalised functions that are shown in the figure. As can be seen, the location of the best-exemplar range remained relatively unchanged with the change in lexical status, a pattern which contrasts with the change in best-exemplar location that is found for acoustic-phonetic contextual factors, such as speaking rate. To ensure that this qualitatively different pattern was not due to some idiosyncrasy of our stimuli, we conducted a control experiment in which we altered the speaking rate of the stimuli in the *beace-peace-*peace* series, and tested a new group of listeners using the same procedures on the fast and slow versions of the stimuli. The normalised data, which are displayed in the right panel of Figure 1, show that the variation in speaking rate did alter the location of the entire best-exemplar range, as expected from our previous research, and statistical analysis confirmed this observation.

These findings suggest a dissociation between the effect of lexical status and the effect of acoustic-phonetic factors such as speaking rate on the location of a category's best exemplars. Acoustic-phonetic factors, but not the higher-order linguistic factor of lexical status, alter the location of the entire best-exemplar range of the category. We suggest that this dissociation derives from the listener's sensitivity to the differential effects of these contextual factors in speech production. Acoustic-phonetic contextual factors alter the production of the critical segmental information, and we propose that listeners take this variation into account by altering the location of the perceived best category exemplars. In contrast, lexical status does not alter VOT values in production: in a companion production study (Allen & Miller, in press) we found no systematic effect of lexical status on the VOT values of voiceless consonants in words (such as *peace*) versus matched nonwords (such as *peef*). Thus there is no systematic variation for the perceptual system to take into account and, as our perceptual data show, lexical status does not substantially alter the location of the best-exemplar range for the voiceless category in perception. Note that according to this proposal, although many contextual factors alter the location of the boundaries between categories, only those contextual factors that systematically change the way in which the critical segmental information is produced will substantially affect the location of the category's perceived best exemplars.

## SPEEDED CATEGORISATION

The dissociation between different types of contextual factors reported above underscores the importance of considering the internal structure of categories, as well as the boundaries between categories, when studying phonetic perception. Accordingly, in a related set of studies (Miller & Cox, in prep), we have been examining how internal category structure, as reflected in category goodness ratings, affects how quickly listeners can map the acoustic signal onto phonetic categories during on-line processing. For this investigation, we created nine different extended VOT series, each based on a different word pair. For three of the series, the word-initial consonant was labial (as in the series *bat-pat-*pat*), for three it was alveolar (as in the series *den-ten-*ten*), and for three it was velar (as in the series *ghost-coast-*coast*). In a preliminary category-goodness rating task, in which listeners were asked to rate each token for the goodness of its initial voiceless consonant (/p/, /t/, or /k/) using a 1–7 rating scale, we obtained the expected non-monotonic goodness function for each series.

For the first of two speeded categorisation experiments, we selected tokens from each series that covered a wide range of VOT values and varied considerably in category goodness. We presented these tokens to three groups of listeners. Each group heard the tokens from the three series with a given place of articulation randomised together. The listeners' task was to decide as quickly as possible whether each word began with /b/ or /p/ (labial group), /d/ or /t/ (alveolar group), or /g/ or /k/ (velar group) and to indicate their response by pressing one of two labelled keys (e.g., B or P). The main question was how quickly listeners could perform this two-choice categorisation. On the basis of the literature (e.g., Pisoni & Tash, 1974), we expected that response time would be relatively slow for the poor voiceless category exemplars near the voiced-voiceless category boundary, and relatively fast for the best voiceless category exemplars. At issue was the response time for poor voiceless category exemplars with VOT values longer than those of the best exemplars. If response time is a direct function of category goodness, then response time for these stimuli should be relatively slow. However, if response time depends more on the position of the stimulus vis-à-vis the category boundary, then response time should be relatively fast in that these stimuli are far from the voiced-voiceless boundary and are clearly identified as members of the voiceless category.

Overall, similar findings were obtained for all nine stimulus series. Representative data are shown for one of the series, *den-ten-*ten*, in the left panel of Figure 2, where the response time function for the relevant stimuli (those ranging from the voiced-voiceless boundary region to the longest VOT values) is presented along with the goodness function for the series. The main finding across the nine series, confirmed by statistical analysis, was that response time was relatively slow for poor voiceless exemplars near the voiced-voiceless category boundary, but relatively fast for poor voiceless exemplars with long VOT values—overall, poor voiceless exemplars with long VOT values were categorised as quickly

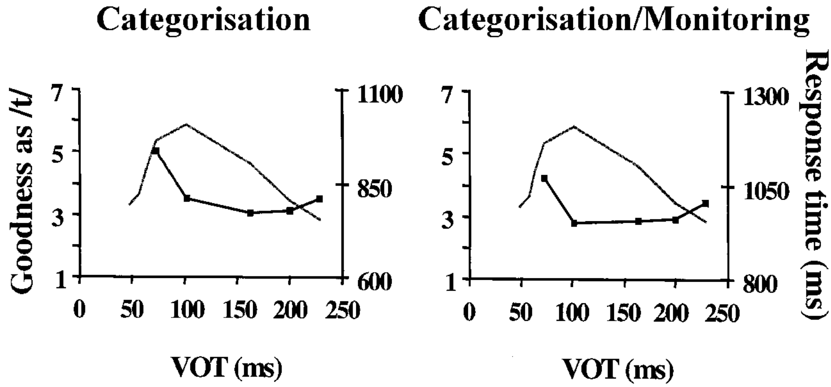


**Figure 2.** Group response time functions (thick lines with symbols) for the *den-ten-*ten* series from a categorisation task (left panel) and a categorisation/monitoring task (right panel). The group goodness function (thin line) from the preliminary category-goodness rating task is overlaid on the response time function in each panel.

as the best exemplars themselves. These findings provide support for the view that categorisation time is more dependent on position vis-à-vis the category boundary than on perceived category goodness.

Note that in the experiment just described listeners in a given group made two-choice speeded categorisation judgements on tokens that all had a single place of articulation. Given this design, it is possible that listeners were not fully analysing the poor exemplars with long VOT values, but simply quickly categorising them as voiceless by default, due to their breathy, exaggerated quality. To control for this possibility, we conducted a second experiment with new listeners using the same stimuli, but a task that involved monitoring as well as categorisation. Again, three groups of listeners were tested. However, this time each group heard the stimuli from all nine series randomised together. The listeners' task was to monitor for two initial consonants, /b/ or /p/ (labial group), /d/ or /t/ (alveolar group), or /g/ or /k/ (velar group). When they heard one of the target consonants they were to respond as quickly as possible by pressing one of two labelled keys (e.g., B or P); if the word did not begin with one of the target consonants, they were to refrain from responding. With this task, listeners could not use a default strategy of simply pressing the key corresponding to the voiceless consonant whenever they heard a poor exemplar that was breathy and exaggerated, as the task required listeners to analyse the stimuli for place of articulation as well as voicing. The question was whether with this new task, we would also find that poor exemplars of the voiceless category with long VOT values were categorised quickly. The answer was yes. Overall, response times were longer than in the first experiment (as to be expected, given the additional monitoring task), but the general pattern of response times was the same. The data for the *den-ten-*ten* series are shown in the right panel of Figure 2. Statistical analysis across the nine series confirmed that, as in the first experiment, response time was relatively slow for poor voiceless exemplars near the voiced-voiceless category boundary, but relatively fast for poor voiceless exemplars with long VOT values—overall, poor exemplars with long VOT values were categorised as quickly as the best exemplars themselves. Taken together, the data from the two tasks demonstrate that categorisation time is not a direct function of perceived category goodness. Instead, the time it takes to map the critical acoustic information onto a phonetic category appears to depend more on the exemplar's location in perceptual space vis-à-vis a boundary with a competing category.

## SUMMARY AND CONCLUSIONS

The findings described in this paper provide additional support for the view that phonetic categories have a rich internal structure, and they provide new information on the role of this structure in speech processing. The findings on *context effects* show that although some contextual factors that alter the location of a phonetic category boundary also alter the location of a category's best exemplars, other contextual factors have a more limited effect, primarily confined to the boundary region. Of considerable importance will be to determine just which contextual factors affect the entire structure of a phonetic category, and which are limited to the boundary region. As noted earlier, we propose that the way in which a contextual factor alters the mapping between acoustic signal and phonetic category depends on the role of that factor in speech production; for further discussion of this point, see Allen and Miller (in press). The findings on *speeded categorisation* show that the speed with which a listener maps the acoustic signal onto a given phonetic category depends more on the exemplar's position in perceptual space vis-à-vis a boundary with a competing category than on its perceived category goodness. Thus even though listeners are exquisitely sensitive to fine gradations in category goodness, revealing richly structured phonetic representations, these differences in perceived category goodness do not translate directly into categorisation speed. The nature of the categorisation process that yields this response pattern remains to be determined. More generally, although many questions about the mapping between acoustic signal and prelexical phonetic representation remain unanswered, the current findings underscore the importance of considering both the boundaries between categories, and the graded internal structure of categories, when building models of the early stages of speech processing.

## REFERENCES

Allen, J.S., & Miller, J.L. (in press). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception and Psychophysics*.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

Hodgson, P., & Miller, J.L. (1996). Internal structure of phonetic categories: Evidence for within-category trading relations. *Journal of the Acoustical Society of America*, *100*, 565–576.

Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*, 93–107.

Miller, J.L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*, 271–285.

Miller, J.L., & Cox, E. (in prep). Phonetic categories: The relation between perceived category goodness and categorization time.

Miller, J.L., & Volaitis, L.E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, *46*, 505–512.

Oden, G.C., & Massaro, D.W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*, 172–191.

Pisoni, D.B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, *15*, 285–290.

Repp, B.H., & Liberman, A.M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), Categorical perception: The groundwork of cognition. Cambridge: Cambridge University Press.

Samuel, A.G. (1982). Phonetic prototypes. *Perception and Psychophysics*, *31*, 307–314.

Volaitis, L.E., & Miller, J.L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*, 723–735.

Wayland, S.C., Miller, J.L., & Volaitis, L.E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, *95*, 2694–2701.