AN INVESTIGATION INTO EXEMPLAR EFFECTS IN THE PERCEPTUAL LEARNING PARADIGM

Johannah O'Mahony and Bernd Möbius

Department of Language Science and Technology, Saarland University
omahony@coli.uni-saarland.de and moebius@coli.uni-saarland.de

ABSTRACT

Variance in the speech signal is caused by many factors. Traditionally, it was assumed that variation is removed prior to lexical access. Evidence for abstraction has come from perceptual learning experiments. Exemplar Theory does not assume stored abstractions, but the storage of exemplars, as evidenced by same-speaker priming effects in which a listener is faster to react to a word when repeated in the same voice. This study combined the perceptual learning paradigm with a priming experiment to investigate whether exemplar effects are found on items which induce perceptual learning compared to canonically realised counterparts. After exposure to an ambiguous fricative in German between [f] and [s], categorisation changed as a function of condition, replicating previous studies. In the priming block, no difference in reaction-time was found between conditions, suggesting exemplars were not stored during perceptual learning. We discuss the results in relation to models of speech perception.

Keywords: Perceptual learning, exemplar effects, lexical decision, speaker normalisation

1. INTRODUCTION

Speech is inherently variable due to factors such as systematic dialectal differences, anatomical differences, and even incidental features such as having an object in one's mouth [14]. In abstractionist frameworks of speech perception, the variability found in speech was treated as irrelevant noise, removed by the perceptual system before lexical access, a process termed *speaker normalisation*. As analysis of speech outside the laboratory has increased, it has become apparent that idealised categories might be unrealistic due to a lack of one-to-one mappings between cues and categories [13, 4].

Evidence of same-speaker processing advantage within Exemplar Theory has shown that listeners may retain surface features of speech, which is not predicted by purely abstractionist accounts [8, 16, 20, 22]. Exemplar frameworks, however, cannot account for the productivity of phonological rules, as evidenced in perceptual learning studies [17]. As both approaches fall short of predicting crucial phenomena [7], hybrid solutions have been proposed [9, 12]. What is clear is that adaptation to new speakers is productive in nature, but is also based on experience with *exemplars* of the new speaker. The question of this study is, therefore, whether we find same-speaker effects on items that induce perceptual learning compared to canonical realisations.

1.1. Perceptual Learning and Exemplar Effects

Perceptual learning paradigms involve participants being exposed to an idiosyncratic variant of a phoneme, found on a continuum between two phonemes, e.g. [f] and [s]. After exposure to the ambiguous phoneme, categorisation of items on the continuum should change depending on which lexical context the subject was exposed to. This change in categorisation is productive, extending beyond lexical items from exposure, providing evidence for abstraction [17]. There have been conflicting results on automaticity in perceptual learning, with [1] reporting no difference in perceptual learning when a demanding distractor task was used versus without a second task, but [5, 6] found that processing words with an ambiguous phoneme can lead to longer reaction times. Finally, when the source of the ambiguous phoneme is unknown, listeners can delay a change in categorisation until the source of variation has been disambiguated [14], which provides evidence for the storage of contextual information.

Storage of contextual and phonetic information is a key point of Exemplar Theory and has been found in the form of same-speaker effects in voice priming studies. These effects have not been found in all cases. The Time-Course Hypothesis predicts that they will arise when processing is slow, as evidenced by longer reaction times [16], e.g when processing foreign accents [15] or speech in noise [19]. If exemplar effects are found when processing is slow, and items that induce perceptual learning lead to slower processing, then we can hypothesise that we may find exemplar effects on ambiguous items.

The link between the presence of exemplar effects on items which induce perceptual learning can be explained by both the Complementary Systems Model [9] and The Ideal Adapter Framework [12], which incorporate both exemplars and abstractions. In order to create generative models for new speakers we encounter, we require a certain amount of phonetic evidence from which we can extract the distributional properties in the signal in order to aid future processing. In both approaches, when a listener encounters a new speaker, we draw on prior experience, but also develop new generative models based on the new deviating input. We therefore store multiple extractions and need exemplars to develop these. The specificity of the abstractions means we can account simultaneously for specificity effects, as well as productive phonological knowledge.

2. THE CURRENT STUDY

2.1. Method

2.1.1. Participants

Forty-nine participants took part in the main experiment (30 females, mean age 25.8, SD=5.16) and ten in the pre-test (6 females, mean age = 23.9, SD=3.54). All participants were German native speakers and reported no hearing impairments.

2.1.2. Materials

To construct the [f]–[s] continuum, a male speaker of standard German recorded the syllables [ɛf] and [ɛs]. A 41 step continuum was constructed following [21] by adapting the ratio of each of the fricatives in a step-wise manner. The resulting fricatives were spliced onto a natural [ɛ] vowel which was recorded in an [ɛf] condition (duration 123 ms). Following [21], a pre-test was carried out to find the most ambiguous step and the categorisation boundary. Seventeen steps on the continuum were chosen. The experiment consisted of eight blocks, in which each of the steps was presented in a random order. Participants indicated by clicking [f] or [s] which phoneme they heard. The most ambiguous step was found between 19 and 21, and therefore step 20 was chosen.

Following the pre-test, 40 German words (20 [f]-final, e.g. Schaf *sheep*, 20 [s]-final, e.g. Glas *glass*) were chosen. Target words did not contain [f] or [s] in any other position. Log-frequency of the target words was measured using the deWaC corpus [2]. Mean frequency of the [f]-words was 8.37 (SD=1.85), and 8.27 (SD=1.40) for the [s]-words. Pairs were matched for stress, final vowel, and number of syllables. Overall three versions were created, the natural realisation of the male and female speaker, and the ambiguous male form in which the fricative was replaced by the ambiguous fricative [?]. Note that in this case the [s]-words were recorded as [f]-ending to control for coarticulatory effects. Finally 100 fillers and 150 phonotactically legal non-words (e.g. *kapitat*) were recorded by both speakers. The natural and ambiguous target items of the male speaker were used both as exposure items for perceptual learning, and as primes for the exemplar test. The female target items were used as the *different* voice in the exemplar test.

2.1.3. Design

The design consisted of three parts (1) an Exposure/Priming phase (2) a Perceptual learning Categorisation Test and (3) an Exemplar Test. Part (1) was a lexical decision task consisting of 40 targets (natural and ambiguous e.g. Gla[s] or Gla[?] in the male voice), 60 fillers (10 male voice, 50 female voice) and 100 non-words (50 per voice). This functioned as the exposure phase for the perceptual learning categorisation test, and as the priming block for the exemplar test. Part (2) was a categorisation task consisting of five steps (14,17,20,23,26) of the continuum from the pretest following [21, 18, 6], each repeated 6 times. This tested the difference in categorisation of the between-subject variable FS-Condition, i.e. the lexical context of the ambiguous phoneme. Part (3) was the Exemplar Test. In this section, half of the targets from Part (1) (ambiguous and natural [f/s]/[?]) were presented in a different female voice, and half the targets in the same voice as Part (1). This constituted a within-subjects design with two independent variables 1) whether the prime in Part (1) contained the natural or ambiguous phoneme and 2) whether the target was repeated in the same or different voice in the exemplar block. Furthermore, the final lexical decision consisted of 60 fillers and 100 non-words in both voices.

2.1.4. Procedure

Participants were tested individually in a quiet room. Each participant was assigned to one of four lists (two per FS-Condition). The instructions were presented on screen and clarified by the experimenter. Participants were told that they would hear different words, German words or fake words, spoken by a male and female speaker. They were instructed to respond quickly and accurately whether or not the word was an existing word on a Cedrus button box. The participants did not know that the first lexical

decision task would be followed by the categorisation test and exemplar test. After the exposure block the experimenter placed a label EF and ES on the button box for the categorisation test. Participants were told that they would hear different sounds spoken by the *male speaker*. Their task was to choose whether the sound was [ɛs] or [ɛf]. Finally, the exemplar test appeared with the same instructions as the exposure block. After the experiment, participants filled out a post-questionnaire following [6].

2.1.5. Statistical Analyses

For the analysis of reaction-time (RT) data in the exposure and target block, we employed linear mixed models. We used the log-transformed RT and only included correct responses. RT was measured from word offset because the phoneme manipulation occurred word-finally. For binary response results of accuracy and in the categorisation task, we used mixed-logit models following [11]. In order to assess the significance, we used pairwise comparisons starting from the most complex model. Experimentally manipulated variables were always included in the model regardless of significance, as we were performing confirmatory hypothesis testing. Random slopes for word and subject were added where appropriate. Covariates (Log Frequency, Preceding RT and RT of prime) were kept in the model if they significantly improved fit. Outliers were removed by model criticism, removing residuals more than 2.5 SD away from mean, and the results reported are based on refit models.

2.2. Results

2.2.1. Exposure Block

Based on the post-questionnaire, four participants were removed because they recognised one of the speakers. Participants were also removed based on two measures, overall accuracy of real words or non-words (<70%, 1 participant), as well as accuracy of target words (< 50% following [21], 2 participants). Overall accuracy of real words was 96% (SD 2%), while non-word accuracy was 93% (SD 5%).

For analysis of accuracy, we used two fixed effects in this model, viz. phoneme (ending in [f] or [s]) and the between-subjects condition FS-Condition, as well as the co-variate log frequency. Of importance was the interaction between the two. Both the between-subjects condition FS-Condition ($\beta_{FS-Condition} = 0.20$ $SE = 0.45$ $t = 0.45$ $p = 0.66$) and within-subjects condition Phoneme ($\beta_{phoneme} = -0.53$ $SE = 0.65$ $t = -0.81$ $p = 0.42$) did not reach significance. The interaction between the two was significant ($\beta_{interaction} = -1.20$ $SE = 0.50$ $t = -2.38$ $p = 0.02$), predicting a reduction of accuracy for an ambiguous target item ending in [s]. Finally, the addition of log frequency was significant ($\beta_{logfrequency} = 0.49$ $SE = 0.19$ $t = 2.61$ $p = 0.009$)

For the reaction time analysis, both the between-subjects fixed effect FS-Condition ($\beta_{FS-Condition} = -0.04$ $SE = 0.13$ $t = -0.30$ $p = 0.77$) and within-subjects condition Phoneme ($\beta_{phoneme} = 0.07$ $SE = 0.10$ $t = 0.67$ $p = 0.51$) showed no significance. Similarly, the interaction between the two was not significant ($\beta_{interaction} = 0.03$ $SE = 0.05$ $t = 0.62$ $p = 0.53$). This suggests that there was no difference between the natural and ambiguous items, unlike the results found in [6, 18]. Preceding RT, however, was significant ($\beta_{PrecedingRT} = 0.00008$ $SE = 0.00003$ $t = 2.84$ $p = 0.005$).

2.2.2. Categorisation Test

The between-subjects FS-condition was significant. The estimate of this effect shows that when the condition is [s], the estimated proportion of [f] responses drops significantly. We can see this by the negative estimated effect ($\beta_{FS-Condition} = -5.34$ $SE = 1.22$ $z = -4.36$ $p = 0.000013$).

2.2.3. Exemplar Block

Results from the FS-Condition were collapsed into one group with two IVs, Prime (natural/ambiguous target word from exposure e.g. containing [f/s]/[?]), and Voice (same/different voice in exemplar block). The fixed effects Prime and Voice were not significant predictors ($\beta_{Prime} = 0.02$ $SE = 0.05$ $t = 0.45$ $p = 0.65$, $\beta_{Voice} = 0.06$ $SE = 0.05$ $t = 1.19$ $p = 0.23$). The interaction was not significant ($\beta_{Interaction} = -0.07$ $SE = 0.07$ $t = -1.03$ $p = 0.30$). This suggests that there was no difference in voice priming between the ambiguous and natural items. Two control factors were significant predictors of reaction time, ($\beta_{LogPrecedingRT} = 0.14$ $SE = 0.03$ $t = 4.18$ $p = 0.00003$) and ($\beta_{ExposureBlockRT} = 0.21$ $SE = 0.03$ $t = 7.06$ $p < 0.0001$); note that this control factor was added to account for baseline differences in reaction time in the exposure block.

3. DISCUSSION

The results from the perceptual learning task replicated previous studies showing that listeners change their phonetic category boundaries after exposure to deviant phonetic input from a new speaker. Participants in the [f]-condition categorised more stim-

	Natural Primes		Ambiguous Primes	
Voice in Exemplar Block	Same	Different	Same	Different
Mean RT (ms)	378 (SD=334)	371 (SD=335)	388 (SD=392)	351 (SD=280)
Accuracy (%)	97.6%	95.6%	95.1%	97.1%

Table 1: Summary of mean reaction time and accuracy in the Target items in Exemplar Test.

uli as [f] in the test continuum than participants in the [s]-condition. This effect was found even when the exposure block consisted of two speakers. The main question of this study however, was whether exemplar effects would be found on items which contained the ambiguous phoneme compared to the naturally realised items. The results of the exemplar test indicate that there was no difference between conditions, suggesting no exemplar effects were found. This result may therefore indicate, following previous findings, that perceptual learning is a relatively automatic process [1] and does not require increased cognitive effort or processing. This indeed would explain the lack of exemplar effects, which have been found on items with higher processing costs. In fact, unlike previous studies on perceptual learning [18, 3, 5], reaction times were not significantly longer for ambiguous items, indicating that they were processed as quickly as natural items. The absence of exemplar effects is predicted by purely abstractionist theories, because they postulate speaker normalisation, and therefore any priming will take place after speaker normalisation, on the lexical level. Furthermore, despite the addition of another speaker, perceptual learning occurred, providing more evidence that perceptual learning is a rather automatic process, as the addition of a second speaker would possibly increase cognitive demand, having to track multiple distributional cues.

The result above may pose problems for purely exemplar frameworks, which assume the storage of exemplars, but the results may still be consolidated in a hybrid model. One of the main predictions of the Ideal Adapter Framework is that listeners will also draw on prior experience when they encounter a new speaker. From the answers in the post-questionnaire, and similar to the findings of [6], many participants reported that the male speaker had a lisp. This shows their awareness of the source of variability, so much so that there is a term for it. This suggests they encountered speakers with the same quality of frication before. This means, as predicted by the IAF, that listeners would not need to create a novel abstraction for this speaker, but rather draw on a previous abstraction. Due to this, a future improvement would be to use a more novel phonemic variant in the target items. One interesting result, however, which the IAF cannot completely account for, is the reduction in accuracy in items with lower frequency in exposure for [s]-ambiguous words. If listeners already had prior experience with this sound, and therefore a prior abstraction, then we should have found similarly high accuracy for low and high frequency words. It is still unclear how the IAF deals with frequency effects, which are a key point in exemplar frameworks.

There are of course methodological considerations which might have led to a lack of same-speaker effects, and indeed this result could be due to a Type II error. For example, there has been a lot of disparity in the literature with regards to number of exemplars used in experiments testing for same-speaker effects, with successful experiment often employing extremely low numbers of stimuli (e.g. [16]). It is possible that the current experiment, with 400 lexical items in total, was too long and did not contain a sufficient ratio of target items. This, however, highlights that exemplar effects may not be robust in more realistic situations [10]. This does not mean that exemplars are not of importance or stored, but that the method of measurement may not be sufficiently sensitive; cf. [20] who found results in EEG but not in the behavioural task. Due to the instability of the results in the literature, we conclude that this priming task alone may not be a sensitive enough measure, and that future work should employ more time-sensitive methods, such as eye-tracking or EEG, as well as testing more target items and participants to increase power.

In conclusion, this study sought to test for the presence of exemplar effects on items which induce perceptual learning, in order to investigate the role of exemplars in creating abstractions using a novel paradigm. While a perceptual learning effect was found between groups, no evidence for exemplar effects was found. This may provide evidence for the automaticity of perceptual learning, but we cannot rule our a type II error. The results can also be reconciled within the IAF because listeners may have drawn on previous abstractions. Further studies should aim to revisit this paradigm with the above mentioned methodological changes.

4. REFERENCES

[1] Baart, M., Vroomen, J. 2010. Phonetic recalibration does not depend on working memory. *Experimental Brain Research* 203(3), 575–582.

[2] Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation* 43(3), 209–226.

[3] Clarke-Davidson, C. M., Luce, P. A., Sawusch, J. R. 2008. Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics* 70(4), 604–618.

[4] Drouin, J., Monto, N. R., Theodore, R. 2017. Talker-specificity effects in spoken language processing: Now you see them, now you don't. In: Lahiri, A., Kotzor, S., (eds), *The Speech Processing Lexicon. Neurocognitive and Behavioural Approaches.* Berlin, Boston: De Gruyter Mouton chapter 6, 107–128.

[5] Drozdova, P., Hout, R., Scharenborg, O. 09 2016. Processing and Adaptation to Ambiguous Sounds during the Course of Perceptual Learning. *Proc. Interspeech 2016* 2811–2815.

[6] Eisner, F., McQueen, J. M. 2005. The specificity of perceptual learning in speech processing. *Perception & psychophysics* 67(2), 224–238.

[7] Ernestus, M. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142, 27–41.

[8] Goldinger, S. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5), 1166–1183.

[9] Goldinger, S. D. 2007. A complementary-systems approach to abstract and episodic speech perception. Trouvain, J., Barry, W., (eds), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007).*

[10] Hanique, I., Aalders, E., Ernestus, M. 2013. How robust are exemplar effects in word comprehension? *The Mental Lexicon* 8(3), 269–294.

[11] Jaeger, T. F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4).

[12] Kleinschmidt, D., Jaeger, F. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review* 122(2), 148–203.

[13] Kleinschmidt, D., Jaeger, T. F. 2011. A Bayesian belief updating model of phonetic recalibration and selective adaptation. *2nd ACL Workshop on Cognitive Modeling and Computational Linguistics.* 10–19.

[14] Liu, L., Jaeger, T. F. 2018. Inferring causes during speech perception. *Cognition* 174, 55–70.

[15] McLennan, C. T., González, J. 2012. Examining talker effects in the perception of native- and foreign-accented speech. *Attention, Perception, & Psychophysics* 74(5), 824–830.

[16] McLennan, C. T., Luce, P. A. 2005. Examining the Time Course of Indexical Specificity Effects in Spoken Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(2), 306–321.

[17] McQueen, J. M., Cutler, A., Norris, D. 2006. Phonological abstraction in the mental lexicon. *Cognitive Science* 30(6), 1113–1126.

[18] McQueen, J. M., Norris, D., Cutler, A. 2006. The dynamic nature of speech perception. *Language and Speech* 49(1), 101–112.

[19] Nijveld, A., ten Bosch, L., Ernestus, M. 2015. Exemplar Effects Arise in a Lexical Decision Task, but Only Under Adverse Listening Conditions. Wolters, M., Livingstone, J., Beattie, B., Smith, R., MacMahon, M., Stuart-Smith, J., Scobbie, J., (eds), *In Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)* Glasgow.

[20] Nijveld, A., Mulder, K., ten Bosch, L., Ernestus, M. 2017. ERPs reveal that exemplar effects are driven by episodic memory instead of the mental lexicon. *Poster presented at the Workshop Conversational speech and lexical representations, Nijmegen, The Netherlands.* 2–3.

[21] Norris, D., McQueen, J. M., Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology* 47(2), 204–238.

[22] Theodore, R. M., Blumstein, S. E., Luthra, S. 2015. Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Attention, perception & psychophysics* 77(5), 1674–84.