

Processing and Adaption to Ambiguous Sounds during the Course of Perceptual Learning

Polina Drozdova¹, Roeland van Hout¹, Odette Scharenborg¹

¹ Centre for Language Studies, Radboud University Nijmegen, The Netherlands

P.Drozdova@let.ru.nl, R.vanHout@let.ru.nl, O.Scharenborg@let.ru.nl

Abstract

Listeners use their lexical knowledge to interpret ambiguous sounds, and retune their phonetic categories to include this ambiguous sound. Although there is ample evidence for lexically-guided retuning, the adaptation process is not fully understood. Using a lexical decision task with an embedded auditory semantic priming task, the present study investigates whether words containing an ambiguous sound are processed in the same way as „natural“ words and whether adaptation to the ambiguous sound tends to equalize the processing of „ambiguous“ and natural words. Analyses of the yes/no responses and reaction times to natural and „ambiguous“ words showed that words containing an ambiguous sound were accepted as words less often and were processed slower than the same words without ambiguity. The difference in acceptance disappeared after exposure to approximately 15 ambiguous items. Interestingly, lower acceptance rates and slower processing did not have an effect on the processing of semantic information of the following word. However, lower acceptance rates of ambiguous primes predict slower reaction times of these primes, suggesting an important role of stimulus-specific characteristics in triggering lexically-guided perceptual learning.

Index Terms: adaptation, perceptual learning, word recognition

1. Introduction

Previous research has demonstrated the ability of the human perceptual system to quickly adapt to ambiguously sounding items [1]. Norris et al. were the first to show that listeners use their lexical knowledge to interpret ambiguous sounds, e.g., an ambiguous final sound between /f/ and /s/ in *giraf[s]* will be interpreted as an /f/ since *giraffe* is an existing English word and *giras* is not, while the same ambiguous sound in *bof[s]* will be interpreted as an /s/, since *boss* is an existing word and *bof* is not [2]. Listeners adjust their phonetic category boundaries to include this ambiguous sound in their sound system [3]. This mechanism is referred to as lexically-guided perceptual learning and is argued to aid listeners in adapting to unfamiliar speakers and accents [2,4].

Lexically-guided perceptual learning has been demonstrated using an exposure-test paradigm. In the exposure phase, participants listen to the ambiguous items, e.g., while performing a lexical decision task [2], and typically demonstrate learning in a subsequent phonetic-categorization task. For lexically-guided perceptual learning to occur, ambiguous sounds should be embedded in real words [2] or phonotactically legal sequences [5]. Moreover, it

has been shown that listeners, who accept more ambiguous items as real words, show a stronger learning effect [6]. This suggests that items with ambiguous sounds should be perceived as real words for learning to occur.

Although only a few studies have specifically looked at the time course of lexically-guided perceptual learning, it has been shown to be fast: exposure to as few as ten ambiguous items yields a stable learning effect [7,8]. Learning seems to occur in a step-wise manner: after exposure to ten items retuning did not get stronger with more exposure [8]. The process of lexically-guided perceptual learning was further investigated by [6] who showed that listeners increase their acceptance of words with an ambiguous sound as real words during the course of the exposure. The present study investigates in how far items containing ambiguous sounds are indeed perceived and processed as real, natural words. We do so by looking at the time-course of accepting words containing an ambiguous sound as a word, and by investigating the spreading of activation to semantically-related words by words containing an ambiguous sound.

We use an auditory semantic priming paradigm within a standard lexical decision task as the exposure phase of a lexically-guided perceptual learning study. Multiple studies (e.g., [9]) have demonstrated that processing of a word (target) is facilitated when it is preceded by a semantically-related prime. Primes in the present experiment contained an ambiguous sound [f/s], which either replaced all /s/ sounds while the /f/ sounds remained unchanged or replaced all /f/ sounds while all /s/ sounds remained unchanged. This set-up allowed us to compare reaction times and hit rates of words with ambiguous and natural sounds to study the recognition and (semantic) processing of „ambiguous“ words in comparison to that of natural words. As mentioned by [10], studying the effect of the mismatch at the acoustic level in primes on the processing of the semantically-related targets can demonstrate differential activation of these words within the lexicon itself.

Since substitution of only one sound in words, or a mismatch in phonetic detail, hampers word processing (see [11] for an overview), we predict that words containing an ambiguous sound are accepted less often as real words and responded to slower than the same words with natural sounds. Moreover, we expect to find the same pattern of difference for the semantically-related target words (directly following the primes) due to a reduced semantic spreading by the ambiguous prime words. Additionally following [6], we predict that listeners demonstrating more “natural-like” processing of ambiguous words exhibit more learning.

In order to investigate the time-course of accepting ambiguous items as real and natural words, we compare the difference in processing speed and recognition accuracy

between words containing ambiguous and natural sounds in different parts of the exposure phase. We hypothesize that processing and recognition of the manipulated items will become more like the processing and recognition of their non-manipulated counterparts by the end of the exposure.

2. Method

2.1. Participants

Forty seven native Dutch participants (10 males, $M_{age}=20.9$, $SD=2.0$), recruited from the Radboud University Nijmegen subject pool, took part in the main experiment. Additionally, 11 native Dutch listeners (2 males, $M_{age}=20.5$, $SD=0.5$) participated in the pilot test of the stimuli. None of the pilot test participants took part in the main experiment.

2.2. Materials

For the exposure phase, 40 semantically related word-pairs were chosen from the Dutch Word Association Database [12]. Crucially, the prime word of the pair contained either a word-final /f/ (e.g., *bankroof* – bank robbery; 20 words) or a word-final /s/ sound (e.g., *paleis* – palace; 20 words), while the target member of the word pair was semantically (highly) related to it (e.g., *geld* – money and *koning* – king for *bankroof* and *paleis*, respectively). Apart from the primes, no other words in the stimulus list contained /s/ or /f/. The lists of 20 /f/ and 20 /s/ prime words contained 8 one-syllable words, 8 two-syllable words, and 4 three-syllable words each. The same distribution was used for the target items. The chosen pairs were based on the “cue lookup” search mode in [12], which shows the ten most frequently generated associations for the cue word as well as the strength of the association. We used the highest associated word from the ten options which satisfied our constraints (word-final /s/ or /f/ in primes but none in targets, similar distribution of number of syllables per word, and the semantic association) as the target word. Due to the restrictions on the prime and target words, it was not possible to find all stimuli in the database. Another four word-pairs fitting the criteria were created and added to the stimulus set.

In addition, 60 Dutch words and 140 non-words were selected as fillers. The distribution of syllables was matched in both the critical items and fillers (i.e., 40% mono-, 40% bi- and 20% trisyllabic words). We divided the total number of stimuli into 14 blocks, each containing 20 items: three prime-target word-pairs, four filler words and ten non-words, except for the last block which contained one target word-pair. Each block contained more than twice as many filler items as critical items to hide the associative relations in the prime-target word-pairs. This set up is similar to the one used by [13]. All the items were produced by a female native Dutch speaker in a sound-attenuated booth at 44 kHz. The same speaker also recorded four minimal word-pairs for the test phase of the experiment: *brief-bries* (letter-breeze), *graf-gras* (grave-grass), *leef-lees* (live-read), and *lof-los* (praise-loose). Additionally, in order to create the ambiguous sound [f/s], 12 isolated syllable-pairs containing /s/ or /f/ with a vowel context identical to the vowel contexts in the primes were recorded (e.g., *eef-ees*).

2.3. Creating ambiguous stimuli

To create ambiguous versions of the prime words, the /s/ or /f/ sounds were excised from each recorded syllable and zero-padded with 25 ms of silence using a PRAAT [14] script and subsequently morphed using STRAIGHT [15] in Matlab. As a result of the morphing, an 11-step [f-s] continuum was created for each prime word separately, where step 0 was the most /f/-like and step 11 was the most /s/-like. To reduce an /s/-bias in some of the continua, sounds from these continua were remorphed using the original /f/ and step 7 of the created continuum. The most ambiguous sound between /f/ and /s/ was chosen on the basis of a pilot test with 11 native Dutch listeners. For the pilot test, the ambiguous sounds were spliced back to both members of the syllable-pairs (to avoid bias towards the /f/ or /s/ interpretation of the syllable). The pilot test included 240 items (five steps of each continuum presented four times). Items were presented to the participants binaurally through headphones in a sound-proof booth. Participants’ task was to indicate whether the presented item contained an /f/ or an /s/ sound and press the corresponding button on the button box. The most ambiguous step was the step that received approximately 50% of /s/ and /f/ responses. This step of the sound was then spliced back into the prime words and used in the exposure phase in the main experiment. For the words in the test phase of the experiment, five versions were created using the most ambiguous step and the two steps preceding and following it.

2.4. Procedure

Two experimental lists were created for the exposure phase: in one list all primes with an /s/ sound were natural and all primes with an /f/ sound contained an ambiguous sound [f/s], while in the second list, all primes with an /s/ sound were ambiguous. The order of items in both lists was constant, and the same words served as primes in both lists. Primes that were ambiguous in one list were in their natural form in the other list, therefore providing a baseline for the comparison.

In the first part of the experiment, participants performed the lexical decision task. Stimuli were presented to the participants through headphones at a fixed mean intensity level of 70 dB. Listeners were instructed to react as fast as possible, and press the right button on a button-box if they thought the item they just heard was an existing Dutch word, and the left if they thought this word did not exist in Dutch.

The subsequent phonetic categorization task consisted of 120 items, in which each ambiguous step of each minimal pair was presented 6 times. Listeners had to press the right button when hearing a word ending in an /s/-sound, and the left button if they heard a word ending in an /f/-sound. The /f/-interpretation of the minimal pair was shown on the left of the computer screen, and the /s/-interpretation of the minimal pair on the right side. The whole experiment took approximately 20 minutes.

3. Results

3.1. Phonetic categorization task

To investigate the processing of words with an ambiguous sound during lexically-guided perceptual learning, it is necessary to first establish whether lexically-guided perceptual learning occurred. Responses of the listeners in the

phonetic-categorization task were analyzed using generalized linear mixed effect models [16]. The dependent variable was the number of /s/-responses. The analysis started with the model including Exposure Condition (whether participants were exposed to /s/-ambiguous or /f/-ambiguous tokens), Step on the /f/ to /s/ continuum (as a nominal variable) and their interaction as fixed predictors. Subject and Word were included as random factors.

Figure 1 shows the proportion of /s/ responses in the phonetic-categorization task for the five test steps. The responses of the participants exposed to the items where all /f/ sounds were ambiguous are plotted with the dotted line with squares, the responses of the other group with the solid line with circles. The difference between the two lines represents the lexically-guided perceptual learning effect.

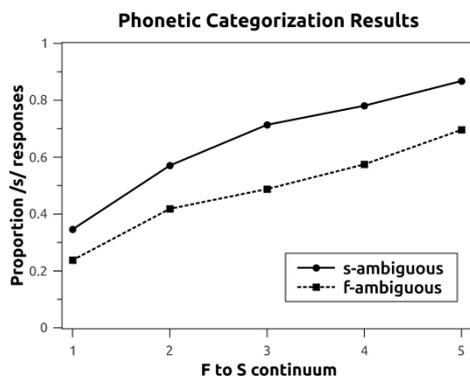


Figure 1: Proportion of /s/ responses of the two exposure conditions in the phonetic-categorization task.

As shown in Figure 1, participants exposed to the words with an ambiguous /s/-sound gave significantly more /s/ responses in the phonetic-categorization task than the participants exposed to the words with an ambiguous /f/- sound. This observation was confirmed by the statistical analysis which showed a significant interaction between Exposure Condition and Step 3 ($\beta=0.783$, $SE=0.226$, $p<.001$), Step 4 ($\beta=0.784$, $SE=0.232$, $p<.001$), and Step 5 ($\beta=0.850$, $SE=0.251$, $p<.001$). Note that the main effect of Exposure Condition was marginally significant ($\beta = 0.718$, $SE = 0.430$, $p = 0.095$). These systematic differences between the two exposure groups indicate that the listeners showed lexically- guided perceptual learning, and thus that the ambiguous sound was included in the sound system of the listeners.

3.2. Lexical decision task

To investigate the extent to which items with an ambiguous sound are processed and recognized as real words, responses of the listeners to the primes and targets in the lexical decision task were analyzed. Recognition of the ambiguous primes was investigated by comparing the hit rates (number of “yes” responses) and reaction times to the manipulated primes to those of the natural primes. Semantic spreading of activation of the ambiguous primes was investigated by comparing the hit rates and reaction times on the target items preceded by the manipulated and by the natural semantically-related prime. To investigate the time course of lexically-guided perceptual learning, the 40 prime-target pairs were subdivided into 4 equal-sized bins. Finally, following [6], the number of hits was used as a predictor in a separate analysis

to investigate whether more „natural“-like processing resulted in more learning.

3.2.1. Analyses of hit rates

One word-pair, *poos-tijd* (a while - time), was excluded from the analyses, since even the natural variant of the word *poos* was accepted as a real word less than 50% of the time. Table 1 provides the mean proportions of the “yes” responses and their standard deviations (in brackets) for the natural and manipulated primes and their semantically-related targets.

Table 1. Mean proportions of “yes” responses for different types of primes and targets.

Prime type	Hits (primes)	Hits (targets)
Natural	0.96 (0.19)	0.996 (0.07)
Manipulated	0.91 (0.29)	0.993 (0.08)

Generalized linear mixed effect models [16] were used to analyze the hit rates with either hit rates for primes or targets as a dependent variable. Manipulation of the prime (whether the prime contained an ambiguous sound), Bin Number and the interaction between them were used as fixed factors, and Subject and Item were added as random factors.

In the hit rate analysis for primes, only Manipulation was shown to be a significant predictor of the number of “yes” responses: participants accepted fewer manipulated than natural items as real words ($\beta=-1.020$, $SE=0.225$, $p< 0.001$). Bin Number and its interaction with Manipulation did not reach significance. When Bin Number was added to the model as a nominal variable with Bin Number 4 on the intercept, there was a significant interaction between Manipulation and Bin Number 2 ($\beta=-1.493$, $SE=0.617$, $p=0.016$) and 3 ($\beta=-1.543$, $SE=0.632$, $p=0.015$). Although the interaction between Bin Number 1 and Manipulation did not reach significance ($\beta=-1.073$, $SE=0.724$, $p=0.138$), it was in the same direction. Thus, where no significant difference was observed between the manipulated and natural primes in Bin Number 4, there were significant differences in acceptance rates for primes with the ambiguous sound and the same primes with the natural sound in the earlier bins. Recognition of the manipulated prime thus became more natural in the last Bin (i.e., with the last 10 items).

In the hit rates analysis for the targets, no fixed factors reached significance. Throughout the lexical decision task, recognition of the target words was high, irrespective of the type of the prime. A final analysis with the number of learning-consistent responses, i.e., responses given in accordance with the exposure condition (e.g. /s/ responses for the participants exposed to the /s/-ambiguous list), as a dependent variable and Acceptance Rate (i.e., the proportion of hits for the manipulated version of each prime) as fixed factor showed that a higher acceptance rate of ambiguous words as real words leads to a larger learning effect. ($\beta=-3.480$, $SE=1.6812$, $p=0.055$).

3.2.2. Analyses of reaction times

Table 2 provides the average reaction times for primes and targets with their standard deviations. All reaction times deviating more than two standard deviations from the mean were excluded, and only reaction times of the primes and

targets which were accepted as real words were analyzed. Log transformed reaction times for either prime or targets were used as a dependent variable in the analyses. In addition to the factors mentioned in Section 3.2.1., the Acceptance Rate and its interaction with Manipulation were included in the analysis to investigate whether ambiguous primes which were less easily accepted as real words also exhibited larger differences in reaction times between their manipulated and non-manipulated versions and less spreading of activation to the target words.

In the reaction times analysis of the primes, again a main effect of Manipulation ($\beta=0.288$, $SE=0.059$, $t=4.94$) was observed. Primes with manipulated sounds were reacted to slower than primes with natural sounds. Moreover, the interaction between Acceptance Rate and Manipulation reached significance ($\beta=-0.243$, $SE=0.063$, $t=-3.86$): lower acceptance rates of ambiguous primes related to larger differences in reaction times from their natural counterparts. The difference in reaction times between primes with an ambiguous and natural sound did not change during the task, the factor Bin Number and its interaction with Manipulation did not reach significance.

Similar to the hit rate analysis, none of the factors in the reaction time analysis of the targets reached significance. There were no differences in the speed of processing of the targets, irrespective of whether they were preceded by natural or manipulated primes. Moreover, processing of the target words preceded by ambiguous and natural primes remained similar throughout the task. Finally, reaction time differences between the manipulated and natural versions of the words did not predict perceptual learning.

Table 2. Mean reaction times for different types of primes and targets.

Prime type	Prime RTs	Target RTs
Natural	955.02 (178.98)	839.08(187.14)
Manipulated	1014.47 (184.95)	833.62 (185.81)

4. General Discussion and Conclusions

The present study investigated the perception and processing of words containing ambiguous sounds during the course of lexically-guided perceptual learning using an auditory semantic priming paradigm. Two questions were studied: whether words with an ambiguous sound are recognized and processed in the same way as their non-ambiguous, natural counterparts, and what the time course is of the adaptation to an ambiguous sound.

Our hypothesis that primes containing an ambiguous sound will be accepted less often as real words than their natural counterparts and will be reacted to slower was confirmed. The manipulation of a sound in a word led to an increase in reaction times and fewer “yes” responses. This finding is in line with existing literature showing that changes in phonetic details of phonemes interfere with word recognition (e.g., [11], [17]).

Despite the differences in the processing of ambiguous primes compared to natural primes, there were no differences in acceptance rates or reaction times between target words preceded by ambiguous primes and targets preceded by

natural primes. This suggests that although the manipulated primes were perceived as less natural and processed slower than the natural primes, this did not have an effect on the processing of semantic information. Possibly, the presence of an ambiguous sound slows down the build-up of the activation of the word so that the threshold for word recognition is reached later. The build-up is, however, fast enough so that activation can spread to semantically-related words. To specifically tap into the priming effect, the priming effect of ambiguous and natural primes could be further investigated by including a set of word pairs, where the target items are preceded by a non-related word (similar to [10]).

We hypothesized that processing and recognition of primes with an ambiguous sound would differ from that of natural primes at the start of the exposure phase and would become more like the processing of natural words towards the end of the exposure phase. The results showed that although processing of ambiguous words remained slower than that of natural words, recognition did become more natural-like. Similar to [6], listeners increase their acceptance of words with an ambiguous sound as real words during the course of the exposure phase. Participants’ recognition of the prime containing the ambiguous sound changed to natural-like after exposure to approximately 15 items. This is in line with [7] and [8] who observed learning after exposure to 10 items. Moreover, this learning seemed to occur in a step-wise manner, like was found by [8], as no convergence in the hit rates of the natural and ambiguous words was observed in the bins prior to the final bin. In line with [6], listeners who accepted more ambiguous words as real words showed a larger learning effect. The significance of the effect was only marginal, but it can be probably explained with the fact that our participants have been exposed to 20 ambiguous items, while in [6] there were 40 items with an ambiguous sound.

Ambiguous primes with lower acceptance rates were found to yield relatively longer processing times than their natural counterparts, as shown by the significant interaction between Manipulation and acceptance of the ambiguous version of the prime in the reaction time analysis. Some ambiguous items were thus more difficult to process and recognize than other ambiguous items, despite being manipulated in a similar way. Supposing that only ambiguous stimuli recognized as real words induce retuning, this is an important finding suggesting that stimulus-specific characteristics (e.g., the size of the lexical neighborhood) may influence lexically-guided perceptual learning. This factor of word characteristics adds to a growing list of factors known to influence lexically-guided perceptual learning, including listener-related factors (e.g., listeners’ acceptance of an ambiguous item as a word [6], attention-switching control [18]), or environment-induced factors (e.g., pen in the mouth of the speaker [19], noise [20,21]).

In conclusion, there are clear differences between the processing and recognition of words containing an ambiguous sound and the same words with a natural sound. The slower and less accurate processing of ambiguous words, however, does not interfere with semantic processing of the ambiguous words. Moreover, adaptation to the ambiguous sounds is (again shown to be) fast and quickly results in a recognition process that is similar to that of natural words.

5. Acknowledgements

This work was sponsored by a Vidi-grant from the Netherlands Organisation for Scientific Research (grant number: 276-89-003) to Odette Scharenborg.

6. References

- [1] Samuel, A. G., & Kraljic, T. (2009) Perceptual learning in speech perception. *Attention, Perception & Psychophysics*, 71, 1207-1218.
- [2] Norris, D., McQueen, J. M., Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204-238.
- [3] Clarke-Davidson, C., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, 70, 604–618.
- [4] Reinisch, E., & Holt, L. L. (2013). Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0034409
- [5] Cutler, A., McQueen, J.M., Butterfield, S., Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. *Proceedings of Interspeech 2008*, 2056-2056.
- [6] Scharenborg, O., & Janse, E. (2013). Comparing lexically-guided perceptual learning in younger and older listeners. *Attention, Perception, and Psychophysics*. 75(3), 525-36. doi: 10.3758/s13414-013-0422-4.
- [7] Kraljic T., Samuel, A.G., Perceptual adjustments to multiple speakers, *Journal of Memory and Language*, 56, 1-15, 2007.
- [8] Poellmann, K., McQueen, J.M., Mitterer, H. (2011). The time course of perceptual learning. *Proceedings of ICPHS*.
- [9] Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. doi:10.1037//0033-295X.82.6.407
- [10] Andruski, J.E., Blumstein, S. E., Burton, S. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163-187
- [11] McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (), *The Oxford handbook of psycholinguistics* (pp. 37-53). Oxford: Oxford University Press.
- [12] De Deyne, S., & Storms, G. (2008). Word associations: norms for 1,424 Dutch words in a continuous task. *Behavior research methods*, 40(1), 198–205. doi:10.3758/BRM.40.1.198.
- [13] Schmidt, J., Scharenborg, O., Jansse E. (2015) Semantic processing of spoken words under cognitive load in older listeners. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0253.
- [14] Boersma, P., Weenink, D (2005) Praat. Doing phonetics by computer (Version 5.1).
- [15] Kawahara, H., Masuda-Katsuse, I., Cheveigne, A (1999) Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds, *Speech Communication*, 27:187-207.
- [16] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- [17] Whalen, D.H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*. 50(4), 351-360
- [18] Scharenborg, O., Weber, A., Janse, E (2015). The role of attentional abilities in lexical-guided perceptual learning by older listeners. *Attention, Perception and Psychophysics*. 77(2), 493-507
- [19] Kraljic, T., Samuel, A.G., & Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19(4), 332-338
- [20] Drozdova, P., Hout, R.W.N.M. van & Scharenborg, O.E. (2015). The effect of non-nativeness and background noise on lexical retuning. In *The Scottish Consortium for ICPHS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0717
- [21] Zhang, X., & Samuel, A.G., (2014) Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception & Performance*, 40, (1), 200–217. DOI: 10.1037/a0033182.