Implicit learning leads to familiarity effects for intonation but not for voice

Ann-Kathrin Grohe, Bettina Braun

Department of Linguistics, University of Konstanz

ann-kathrin.grohe@uni-konstanz.de, bettina.braun@uni-konstanz.de

Abstract

Previous studies have shown that speech processing is accelerated for familiar voices in contrast to unfamiliar ones (e.g. [1]), and for familiar intonation in contrast to unfamiliar intonation [2]. The present experiments probed these effects in a single experiment and tested whether they also occur with short, implicit familiarization. Results of two auditory lexical decision tasks (Experiment 1 with a task-based familiarization phase and Experiment 2 with a passive listening familiarization phase), showed that familiarity with the intonation (rise vs. fall) affected reaction times but that familiarity with the voice (speaker A vs. B) did not. Our results suggest that intonation (which contributes to utterance interpretation) is stored in the mental lexicon, but voice information is not.

Index Terms: voice, intonation, lexical decision, familiarity, question intonation, declarative intonation

1. Introduction

Several accounts of speech comprehension deal with the processing of indexical information, such as a speaker's voice. The abstractionist approach [3] considers the mental lexicon as an accumulation of phonemic representations that reflect only abstract linguistic information. The episodic view [1] on the other hand states that the mental lexicon consists of episodes. Episodes are holistic and contain both lexical and indexical information, including voice information. Episodic and abstractionist models hence make different predictions on the effect of voice information on word recognition: pure abstractionist models cannot account for familiarity effects of indexical details, as any storage of these details is denied. Episodic approaches, on the other hand, predict such effects. However, as lexical and indexical information for one item are stored in the same episode, an effect of voice is predicted only for lexical items that were heard in the same voice before.

We follow the view of hybrid abstractionist-episodic models [4], which integrate both abstract representations and episodes. Lexical information is stored in abstract representations whereas indexical information is stored in episodes. This view does not specify properties of storage of prosodic details. The prosodically-structured view [5, 6] extends the principles of the hybrid approach. It states structures of voice episodes in terms of prosodic properties. This account illustrates voice familiarity effects for different items and it specifies the relationship between voice and prosodic properties, such as intonation.

Previous studies have shown that the familiarity with a voice facilitates speech processing [1, 7, 8]. Such voice familiarity effects were observed even when there were some minutes intervening between learning and testing, which leads to the conclusion that voice information is stored for longer periods of time. However, most studies on voice familiarity effects investigated the processing for identical lexical items in familiarization and test. Goldinger [1], for example, conducted a familiarization session in which voices were made familiar. In a subsequent word shadowing task items presented in the same voice as in the familiarization phase were shadowed faster and more accurately; however, items were identical in training and test.

In a few studies, voice familiarity effects were observed with test words that differed from the familiarized words [9]. In [9], the familiarization phase was explicit and very intense (nine familiarization sessions, each taking one hour, taking place within two weeks). Participants moreover learned to assign a common name to the learned voice, so that an explicit labeling of the learned voice took place. Therefore, it is still an open question whether shorter, and less explicit familiarization phases lead to similar voice familiarity effects as well.

In contrast to voice familiarity, findings of intonation familiarity effects on speech processing are sparse. Church and Schacter [2] found that the familiarity of intonation leads to higher recognition rates in an implicit auditory recognition task. Low-passed-filtered words presented in the same or a different phrasal intonation (rising vs. falling) as in a preceding familiarization phase and had to be identified as same or different. Their results showed longer reaction times for same-intonation trials, suggesting the storage of intonational information in memory.

In sum, long term familiarity effects for both voice and intonation were found, but with different methods, materials and participants. Since intonation and voice information are both generated at the larynx, it is likely that voice and intonation familiarity effects interact with each other. In this paper, we therefore directly compared familiarity effects of voice and intonation within the same experiments. In Experiment 1, familiarization with a speaker's voice and intonation was achieved by an orthographic task, in Experiment 2 by passive listening. Importantly, in both experiments the familiarization phase only lasted a couple of minutes and items were different in the familiarization and test phase. Furthermore, learning and testing took place in different experimental blocks so that longer term memory effects could be examined.

2. Experiment 1

Experiment 1 tested whether voice and intonation familiarity effects occur when using a familiarization phase with an explicit, orthographic task.

2.1. Methods

In a short familiarization session, participants were familiarized with one voice and one kind of sentential intonation. Half the participants heard the voice pronouncing items in a question intonation; the other half heard the voice pronouncing items in a declarative intonation. The testing phase consisted of an auditory lexical decision task. Items varied in voice familiarity (familiar vs. unfamiliar) and in familiarity with intonation (familiar vs. unfamiliar).

2.1.1. Participants

Forty-eight native German speakers (19-32 years, 19 male) took part for a small fee. None had any history of hearing problems and all were unaware of the purpose of the experiment.

2.1.2. Stimuli

The experimental stimuli consisted of 48 low-frequency trisyllabic monomorphemic German words, stressed on the second, open syllable. Lexical frequencies ranged between 0 and 0.07 occurrences per million, according to the CELEX word form dictionary [10]. For each of these words, we constructed 48 non-words with the same vowel sequences and the same onset; stress remained on the second syllable. The non-words could be identified as such after the second or third syllable. Examples for words and their corresponding non-words are *das Casino (the casino) – das Califo; die Anode (the anode) – die Ajone; der Mongole (the mongol) – der Mojone.*

All words and non-words were recorded by two female voices. Both speakers were linguistically trained, female native speakers of Standard German. Words were recorded digitally in a sound-attenuated chamber (44.1kHz, 16Bit). Each speaker read the words and non-words with two different intonation contours: a question intonation, i.e. a low stressed syllable ending in a rise, and a declarative intonation, i.e. a high stressed syllable ending in a fall. Each item was recorded several times in each intonation condition, and we selected the exemplar with the most extensive f0-movements. The voices differed significantly in f0-minima (181.6 Hz vs. 195.4 Hz, $t(95)=7.1$, $p < 0.001$) but not f0-maxima ($p > 0.1$).

The familiarization stimuli were also trisyllabic monomorphemic German words with stress on the second, open syllable. Lexical frequencies were generally higher than those of the experimental items (between 0 and 25 occurrences per million). Half of the items contained the letter "a" in their orthography, half did not. The familiarization stimuli were read by only one of the two speakers, both in the declarative and the interrogative intonation.

2.1.3. Design

We created a 2x2x2 design, manipulating voice familiarity (familiar vs. unfamiliar), intonation familiarity (familiar vs. unfamiliar), and intonation (question vs. declarative).

Experimental items were divided into four blocks, each containing 12 words and 12 non-words (the non-words were derived from the words of another block, see description above). The words in these blocks were balanced for word onset and vowel sequence: there was never the same word onset in one block, identical vowel sequences did not appear more than twice and were separated by two invervening trials with a different vowel sequence. Additionally, we balanced words for lexical frequency so that all blocks had the same distribution of high and low frequency words. Each block was then assigned to one of the four familiarity conditions (familiar/unfamiliar voice; familiar/unfamiliar intonation). Each block was used twice, once for each intonation condition (question vs. statement). Each participant was assigned to one intonation condition.

The order of trials in each list was pseudo-randomized with certain constraints: the same intonation or voice was not presented more often than three times in a row and not more than three words or non-words were presented in a row.

The familiarization items were randomized into two familiarization lists, one in which the items were presented in a question intonation (rising intonation), and one, in which the items were presented in a declarative intonation (falling intonation). Therefore, in the first condition, the question intonation of one of the two speakers was made familiar, and in the other, the declarative intonation of this speaker. Each experimental list was combined with each familiarization condition, which resulted in eight different conditions.

2.1.4. Procedure

Participants were tested individually in a quiet room. They were seated in front of a screen, wore headphones and had a button box with two buttons in front of them. Any potentially distracting circumstances were avoided so that participants' attention was fully turned to the experimental task. Each participant was assigned randomly to one intonation-/voice-familiarity condition and to one experimental list (six participants for each of the eight lists).

In the familiarization session, participants heard the familiarization items and were instructed to decide whether the item contained an “a” in its orthography or not. Right-handed participants pressed the right button for *yes*, left-handed participants the left button. Each trial started with a fixation cross that was shown for 500 ms. Then the stimuli was played. After the button press, there was a 500 ms inter-trial interval before the next trial started. There was no time-out and no feedback. The familiarization phase lasted for about three minutes. Neither reaction times nor hit rates were recorded in the familiarization session.

The test session consisted of a lexical decision task with 96 trials (48 words and 48 non-words), which were presented in different voices and intonation contours. The timing of the trials was identical to the familiarization phase. Right-handed participants used their right hand for a yes-response, left-handed participants their left hand. Reaction times were measured relative to the offset of the auditory stimulus.

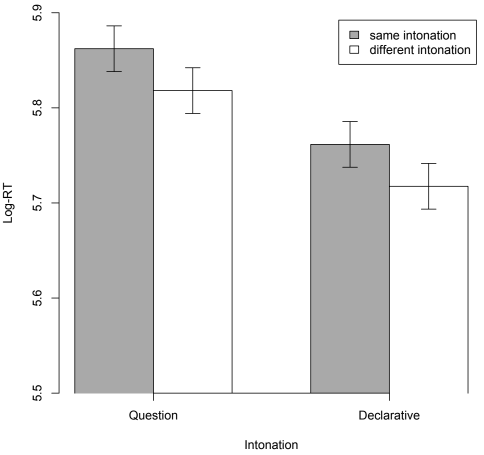
2.2. Results

Reaction times faster than 100 ms slower than 1100 ms and (6.0 % of the data) were excluded from analysis. Furthermore, the following items were excluded because of high error rates: *der Tamile* (85% errors), *die Soutane* (70% errors), *der Rhapsode* (89% errors), *die Pagode* (49% errors). One participant had to be excluded, as reaction times were remarkably slow and hit rates low (20% correct responses).

Participants' correctness was not affected by experimental conditions (all p-values > 0.4). Reaction times for correct responses were log-normalized and analyzed using linear-mixed effects regression models with *voice familiarity*, *intonation familiarity* and *intonation* as fixed factors as well as *subjects* and *items* as crossed random factors (allowing for random intercepts and slopes, cf. [11, 12]). Factors that were not significant were removed if they did not result in significant interactions and if this did not deteriorate the fit of the model, as tested by comparing the Akaike Information Criterion [13]. The most parsimonious model was validated by removing outliers with residuals beyond 2.5 SDs from the mean and the model was refitted. t-values $> |2|$ indicate a significant effect at $\alpha = 0.05$.

Results showed no effect of *voice familiarity* (t < 0.7), an effect of *intonation* (ß = 0.10, SE = 0.02, t = 4.1) and no interactions (all t-values < |0.6|). The effect of *intonation familiarity* approached significance (ß = 0.04, SE = 0.02, t = 1.8), see Figure 1.

Figure 1. *Average reaction times as derived from the statistical model. Whiskers represent standard errors.*



2.3. Discussion

Unlike previous experiments, we found no effect of voice familiarity in this auditory lexical decision task. In those studies that found familiarity effects for voice, stimuli were either identical in familiarization and test [1], or used a very intense and explicit familiarization session for voice information [9]. We hypothesize that by varying the lexical context of familiarization and test items and by using a short and implicit familiarization phase, representations for a speakers' voice could not be established well enough to have an effect on the test items.

Intonation familiarity, on the other hand, approached significance. Stimuli presented in the same intonation as in the familiarization led to slower reaction times than items in a different intonation. If this effect was genuine, it would mean that intonation is stored as independent representations in the long term memory. However, contrary to prior findings [2], a familiar intonation did not accelerate reaction times, but it slowed participants down. We hypothesize that this deceleration was due to the change in tasks between the familiarization ("a" spotting) and test (lexical decision). When same intonation items were presented in the test phase, the task of spotting an "a" may have been activated, which might have slowed down the reaction times. We observed a second effect related to intonation, which was clearly significant: questions were responded to more slowly than declaratives, suggesting a difference in the processing of question contours compared to declarative contours.

3. Experiment 2

As the difference in tasks of the familiarization and the test phase may have led to a slower processing of familiar intonation items, we conducted a second experiment in which the familiarization did not include any task.

3.1. Methods

3.1.1. Participants

Fourty-eight native speakers of German (17-32 years, 19 male), different from those of Experiment 1, participated in the experiment. None of them was aware of the purpose of the experiment and none reported a history of hearing problems

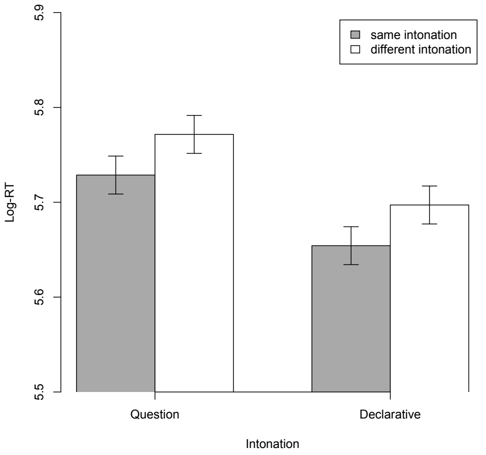
3.1.2. Materials and Procedure

The experimental lists, stimuli, and testing procedure were identical to Experiment 1. The only difference lies in the familiarization session: in Experiment 2, participants did not have to accomplish any task. They were told to listen attentively to the presented words, but not memorize them.

3.2. Results

As in Experiment 1, reaction times slower than 1100 ms and faster than 100 ms (5.8 % of the data) were excluded from the analysis. Participants' correctness was not affected by experimental conditions (all p-values > 0.3). Reaction times for correct trials were analyzed in the same way as for Experiment 1. Results showed a main effect of *intonation familiarity* (ß= 0.04, SE = 0.02, t = 2.2) and of *intonation* (ß= 0.07, SE = 0.02, t=3.1), see Figure 2. There was no effect of *voice familiarity* (t < |0.7|) and no interactions between any of the factors (all t-values < |0.3|).

Figure 2. *Average reaction times in Experiment 2.*



3.3. Discussion

In Experiment 2, the familiarization phase was a passive listening condition and did not involve any task. Under these circumstances, the familiarity with an intonation contour in the test phase led to significantly faster reaction times, in line with Church and Schacter [2]. This result supports our assumption that intonation is stored independently from lexical information. Furthermore, as in Experiment 1, items presented in a question intonation resulted in significantly slower reaction times than items produced in a declarative intonation. This constitutes evidence for different processing mechanisms for different intonation contours even in a lexical task.

As in Experiment 1, no effect for voice familiarity was observed, nor an interaction with intonation familiarity. As discussed in section 2.3, the familiarization phases apparently need to be longer and more explicit for voice familiarity effects to surface.

A possible reason for why familiarity with intonation, and not familiarity with a voice, had an effect on reaction times, is that intonation contours contribute to utterance meaning whereas voice information does not. Our results are in line with [14] who demonstrated the importance of meaning in storing pitch information. Dutch listeners were more attentive to pitch movements that signaled meaningful information (question vs. declarative) than to similar pitch movements that did not. Our results can be interpreted along these lines as well: intonation, which is communicatively meaningful, is stored, but voice information, which is less communicatively meaningful in German, is not. These results have important implications for the possible abstractions in hybrid abstractionist-episodic models [4].

4. General Discussion

The present auditory lexical decision experiments tested the combined effects of voice familiarity, intonation familiarity and intonation contour on response latencies. The results from both experiments demonstrate that the familiarity with a certain voice does not affect response times when the learning phase is short and implicit. This finding stands in contrast with earlier findings on facilitatory effects of voice familiarity (e.g., [1]), which occurred with longer or more explicit familiarization phases. Apart from the difference in exposure duration, the lack of a voice familiarity effect might have also been caused by the manipulation of intonation in the same experiment, which heavily relies on the larynx as well. Therefore, the combined manipulation of voice and intonation might have rendered the effect of voice less important, at least compared to the familiarity effect of intonation. This weighting of familiarity effects for voice and intonation may be explained by the functional relevance of the two sorts of information. While voice information signals mostly indexical information in German, intonation signals a wide variety of communicative functions (e.g. [15]). Therefore, when a familiarization phase is short and implicit, voice information does not seem to be stored as long term representations in the mental lexicon, whereas intonation undergoes longer term storage. This familiarization with an intonation contour was stronger when the familiarization phase did not involve a explicit, potentially distracting, task (Experiment 2) compared to when it involved an extralinguistic task ("a" monitoring in Experiment 1).

Both experiments further resulted in longer lexical decision latencies when the stimuli were presented in a rising, interrogative contour compared to a falling, declarative contour. Possibly, the hearer-orientation and uncertainty of these contours [16] made it more difficult for listeners to make the lexical decision.

Our results reflect a clear dichotomy between the storage of voice and intonation. We found evidence against the storage of voice information, which is structured for prosodic features, as suggested by the prosodically-structured view [5, 6]. We observed abstract representations for intonation, which can generalize over different lexical items. These representations may speed up (Exp. 2) or slow down (Exp. 1) lexical processing, depending on the familiarization conditions. On the other hand, if learning sessions are short, voice information apparently cannot be stored in independent representations. In line with the hybrid abstractionist-episodic view [4], we assume abstract lexical representations that include concrete voice episodes for those voices. For new lexical contexts, voice information has no effect, which indicates that there is no abstraction of voice information to other lexical entries. Future research may address the nature of storage of other prosodic units.

5. Conclusions

Our auditory lexical decision tasks have shown familiarity effects for intonation after only minimal exposure, but no familiarity effects for a speaker's voice. We argued that it is the communicative relevance of intonation compared to voice that explains this difference in outcomes.

6. References

[1] Goldinger, S. D., "Echoes of echoes? An episodic theory of lexical access", Psychological Review, 105(2):251, 1998.

[2] Church, B. A. and Schacter, D. L., "Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency", Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(3):521–533, 1994.

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

[4] McQueen, J. M., Cutler, A. and Norris, D., "Phonological Abstraction in the Mental Lexicon", Cognitive Science, 30(6):1113–1126, 2006.

[5] Hawkins, S. and Smith R.H., "Polysp: a polysystemic, phonetically-rich approach to speech understanding", Italian Journal of Linguistics-Rivista di Linguistica, 13:99–188, 2001.

[6] Hawkins, S., "Roles and representations of systematic fine phonetic detail in speech understanding", Journal of Phonetics, 31(3):373–405, 2003.

[7] Palmeri, T. J., Goldinger, S. D. and Pisoni, D. B., "Episodic encoding of voice attributes and recognition memory for spoken words", Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2):309–328, 1993.

[8] Goldinger, S. D., "Words and voices: Implicit and explicit memory for spoken words: Research on Speech Perception Technical Report No. 7", Bloomington, IN: Indiana University, 1992.

[9] Nygaard, L. C. and Pisoni, D., "Talker-specific learning in speech perception", Attention, Perception, & Psychophysics, 60(3):355–376, 1998.

[10] Baayen, H. R., Piepenbrock, R. and Gulikers, L., "The CELEX lexical data base [CD-ROM]", Philadelphia, PA: University of Pennsylvania, 1995.

[11] Cunnings, I., "An overview of mixed-effects statistical models for second language researchers", Second Language Research, 28(3):369–382, 2012.

[12] Barr, D. J., Levy, R., Scheepers, C. and Tily, H. J., "Random effects structure for confirmatory hypothesis testing: Keep it maximal", Journal of Memory and Language, 68(3):255–278, 2013.

[13] Akaike, H., "A new look at the statistical model identification", IEEE Transactions on Automatic Control, 19(6):716–723, 1974.

[14] Braun, B. and Johnson, E. K., "Question or tone 2? How language experience and linguistic function guide pitch processing", Journal of Phonetics, 39(4):585–594, 2011.

[15] Ladd, D. R., "Intonational phonology", 2nd ed., Cambridge: Cambridge University Press, 2008.

[16] Pierrehumbert, J. and Hirschberg, J., "The Meaning of Intonational Contours in the Interpretation of Discourse", in P. R. Cohen, J. Morgan, and M.E. Pollack [Eds], Intentions in Communication, 271–311, Cambridge: MIT Press, 1990.