

# LET THE MUSIC PLAY!

## *A Short-Term but No Long-Term Detrimental Effect of Vocal Background Music with Familiar Language Lyrics on Foreign Language Vocabulary Learning*

Annette M. B. de Groot and Hilde E. Smedinga  
*University of Amsterdam*

---

Participants learned foreign vocabulary by means of the paired-associates learning procedure in three conditions: (a) in silence, (b) with vocal music with lyrics in a familiar language playing in the background, or (c) with vocal music with lyrics in an unfamiliar language playing in the background. The vocabulary to learn varied in concreteness (concrete vs. abstract) and phonological typicality of the foreign words' forms (typical vs. atypical). When tested during and immediately after training, learning outcomes were poorer in the familiar language music condition than in the unfamiliar language music and silence conditions, but this effect was short-lived, as shown in a delayed test 1 week after training, on which the effect was no longer found. Learning outcomes were better for concrete words than for abstract words and better for typical foreign forms than for atypical ones. Contrary to the adverse effects of familiar language music, the effects of concreteness and typicality were lasting. We explain the interference effect in the familiar language

We thank Jasper Wijnen for programming the experiment and all students who contributed to collecting the data: Suzanne Allart, Laura Cox, Iris Kleinhout, Lizet van Knippenberg, Ronald Koekenbier, Anne Maren Krüse, Noémie te Rietmolen, Irene Sormani, and Lottie Stipdonk. We are grateful to Norbert Schmitt for suggesting this article's main title and for providing comments on an earlier version of this article.

Correspondence concerning this article should be addressed to the first author: Annette M. B. de Groot, University of Amsterdam, Department of Psychology, Weesperplein 4, 1018 XA Amsterdam, the Netherlands. E-mail: a.m.b.degroot@uva.nl

music condition in terms of the workings of the phonological loop, a component of working memory involved in vocabulary acquisition (e.g., Gathercole & Baddeley, 1990).

---

On witnessing their children doing their homework while vocal pop music is playing in the room, many parents may wonder about the effect this music may have. They may, for instance, worry that it detracts attention from the homework, thus having an adverse effect. Alternatively, they may suspect that the pleasant environment created by the music motivates their offspring to spend more time on their assignments than they would when working in silence, thus ultimately improving the outcomes. The present study aims to answer the question of what role vocal background music plays in one common homework assignment: foreign vocabulary learning by means of paired-associates learning.

## **BACKGROUND**

### **Background Music and Human Cognitive Performance**

The substantial body of literature on the effect of background music on human performance does not provide the answer to this question because the effect of background music is task specific and specific to the type of music being played. In other words, results obtained with, for example, a memory task do not necessarily generalize to a reading comprehension task. Similarly, performance in one of these categories of tasks may be adversely affected by one type of music (e.g., vocal pop music), whereas it may benefit from another type of music (e.g., classical instrumental music). Furthermore, individual differences between learners have been shown to play a role as well. Depending on the exact combination of task, type of music, and learner characteristics, one specific cause (out of a larger number of possible ones) of any effect of the music to be obtained may apply.

To date, in studies that focus on the effect of background music on cognitive activities (as opposed to, for example, its effect on sports or sorting goods on a conveyer belt), the following variables have been shown to play a role: the type of cognitive activity (e.g., Angel, Polzella, & Elvers, 2010), the tempo of the music (Kämpfe, Sedlmeier, & Renkewitz, 2011; Thompson, Schellenberg, & Letnic, 2012), the music's arousal potential (Cassidy & MacDonald, 2007), the music's complexity (Furnham & Allass, 1999), the loudness of the music (Thompson et al., 2012; Wolfe, 1983), whether the music is vocal or

instrumental (Belsham & Harman, 1977; Salamé & Baddeley, 1989), the listener's musical preference (Daoussis & McKelvie, 1986; Johansson, Holmqvist, Mossberg, & Lindgren, 2012; Parente, 1976), and whether the listener is introverted or extroverted (Cassidy & MacDonald, 2007; Daoussis & McKelvie, 1986; Furnham & Allass, 1999; Furnham & Bradley, 1997). The proposed causes of observed detrimental or beneficial effects are that background music detracts attention from the task (Cassidy & MacDonald, 2007), interferes with phonological short-term memory (Salamé & Baddeley, 1989), boosts the level of cortical arousal to an optimal level in some task performers but beyond this level in others (e.g., Cassidy & MacDonald, 2007; Daoussis & McKelvie, 1986; Furnham & Allass, 1999; Furnham & Bradley, 1997), or creates a state of "relaxed alertness" or a pleasant mood that is somehow conducive to performance (see Felix, 1993, p. 279, who reviews studies done within suggestopedia, a teaching method in which background music, especially classical music, is used to enhance learning).

Given this multitude of possible relevant factors, it comes as no surprise that a recent meta-analysis (Kämpfe et al., 2011) of 97 studies that examined the effect of background music on widely divergent types of behavior (classified as mundane behavior, cognitive judgment, cognitive achievement, and emotion) showed an overall null effect of background music. This null effect may have resulted from adverse and beneficial effects of background music under specific circumstances canceling one another out. Indeed, a detailed analysis of a subset of the studies suggested that background music improves achievement in sports, has a positive effect on emotional reactions, has a small adverse effect on memory, and disturbs reading. The authors therefore appealed for the use of specific theories that predict specific effects of specific types of background music under specific circumstances.

The present study exemplifies the approach recommended by Kämpfe et al. (2011). It examines the effect of vocal pop background music (that is, a specific type of background music) on the learning of foreign vocabulary through paired-associates learning (PAL; a specific task) and exploits (a specific component of) the working memory model developed by Baddeley and his colleagues (e.g., Baddeley 1986, 2000; Gathercole & Baddeley, 1993) as a theoretical framework in accounting for such an effect. The language of the lyrics of the vocal music was either familiar or unfamiliar to the learners. The reason for this manipulation is detailed in the section describing the present study. Another critical feature of our study is that we examine not only the effect of music in the short term but also the longevity of the effect, because the goal of foreign vocabulary learning is to establish durable, not transient, representations in memory.

Paired-associates learning is an approved and effective way to bootstrap foreign language (FL) vocabulary learning. In its most common form,

PAL links the forms of novel FL words to extant meanings in the learner's long-term memory by presenting a FL word to learn together with its translation in the learner's native language (L1). Once a connection between the FL word and the extant meaning has been established in memory, the precise meaning of the FL word (which is seldom, if ever, the exact equivalent of the meaning of the corresponding L1 word) can subsequently develop gradually through extensive exposure to the FL—that is, through contextual learning. The effectiveness of PAL has been demonstrated in several studies that compared retention scores following PAL with those following other, typically more time-consuming learning methods, such as learning FL words through the keyword method or learning them in context. Long-term retention is at least as good, but sometimes better, following PAL than following keyword learning (Thomas & Wang, 1996; Wang, Thomas, & Ouellette, 1992), particularly in experienced FL learners (Van Hell & Candia Mahn, 1997), and retention scores are substantially higher following learning through PAL than when the meanings of the FL words have to be inferred from context (Hulstijn, Hollander, & Greidanus, 1996; Mondria, 2003).

The effect of background music on FL vocabulary learning by means of PAL has been examined before by De Groot (2006), who showed a beneficial effect of music as compared with silence. That study also indicated that background music might not affect the learning of all types of PAL stimuli to an equal extent (the positive effect of background music was larger for PAL stimuli containing infrequent L1 words than for those containing frequent L1 words) and that it might affect PAL differently in different stages of learning (the effect was relatively small early on in the learning session). De Groot used a piece of classical baroque instrumental music as background music (a rhythmical part of the fourth *Brandenburg Concerto* by J. S. Bach involving “two flutes and a violin set against an ensemble of string instruments,” De Groot, 2006, p. 475). This is the type of music that Felix (1993) concluded to be particularly conducive to learning. Given the fact that the effect of music also depends on the exact type of music used (e.g., instrumental or vocal, classical or modern), one cannot just generalize the results of De Groot's study, assuming that any kind of background music, including the vocal pop music we used in the present study, improves FL vocabulary learning through PAL. As will be seen, the present study shows that such an assumption would be wrong indeed. That instrumental and vocal music may affect task performance differently is also suggested by a couple of previous studies, theoretically related to the present one, that used digit recall (Salamé & Baddeley, 1989) and reading comprehension (Martin, Wogalter, & Forlano, 1988) instead of foreign vocabulary learning through PAL as the task to be performed by the participants (see the Phonological Short-Term Memory, Background Speech, and Background Music section for details).

## Foreign Vocabulary Learning and Phonological Short-Term Memory

There is evidence that a component of working memory, the so-called *phonological loop* system (e.g., Baddeley, 1986, 2000, 2003; Baddeley & Hitch, 1974; Gathercole & Baddeley, 1993; Salamé & Baddeley, 1982) is operative in acquiring vocabulary, both native and foreign. The phonological loop transforms spoken and written verbal input into phonological codes by means of a process of subvocalization, and the codes thus formed are held in phonological short-term memory. A rehearsal process refreshes the stored codes, preventing them from fading from this memory store. Meanwhile, more permanent memory representations are being constructed.

One source of evidence that such a system is operative in acquiring vocabulary comes from studies in which the relation between children's ability to repeat verbal materials (an activity that is assumed to involve the phonological loop) and their ability to learn native or foreign vocabulary was examined (Gathercole & Baddeley, 1989, 1990; Hu, 2003; Service, 1992). These studies have shown a positive correlation between repetition ability and vocabulary learning skill. A second source of evidence is the finding that the learning of nonsense words by means of PAL is impossible if phonological short-term memory is impaired (Baddeley, Papagno, & Vallar, 1988). A third source is the finding that foreign vocabulary learning through PAL is impaired when the learners are required to continuously utter some nonsensical sound during learning (Papagno, Valentine, & Baddeley, 1991). In terms of the theory, uttering the nonsensical sound engages the phonological loop system, and this detracts from its involvement in the primary learning task. On the basis of this and related evidence, Baddeley and his collaborators proposed that the phonological loop evolved to facilitate language acquisition (Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998).

### Phonological Short-Term Memory, Background Speech, and Background Music

Using a different verbal learning task than PAL, Salamé and Baddeley have shown that phonological short-term memory is disrupted by background speech (Salamé & Baddeley, 1982, 1989) and background music (Salamé & Baddeley, 1989). The task used in both studies concerned the visual presentation of series of digits, nine digits per series, that subsequently had to be recalled in their order of presentation. The earliest of this pair of studies showed that participants produced more errors (an error being defined as a failure to recall the correct item in the

appropriate serial position) when digit presentation was accompanied by to-be-ignored spoken words or wordlike nonsense words than when the digits were presented in silence. This finding indicates that concurrent speech hinders recall. A condition with concurrent bursts of white noise did not adversely affect performance. The words and nonsense words produced equally large interference effects, suggesting that the effect was not due to semantic processing of the background speech. Furthermore, the interference effect was larger when the concurrently spoken materials were phonologically similar to the (spoken forms of the) visually presented digits that participants were asked to recall than when the spoken materials were phonologically dissimilar from them. The researchers interpreted this adverse effect of concurrent speech in terms of the workings of the phonological loop system: They suggested that the digits to recall were phonologically coded by the participants to store them in phonological short-term memory but that, meanwhile, the concurrently spoken words and nonsense words, but not noise, automatically gained access to this memory store as well, thus interfering with the digit-recall task.

In their follow-up study, Salamé and Baddeley (1989) conceptualized the results from the 1982 study in terms of a mental filter that allows speech, but not noise, to pass through the filter and gain access to phonological short-term memory. They then wondered what characteristic of speech allows it to pass through the filter: the fact that a human voice is involved or that speech is structured and patterned. If the structured nature of speech is the critical feature, then background music, both vocal and instrumental, may also hinder immediate recall of series of digits because, just like speech, music is highly structured. However, if the involvement of the human voice is crucial, vocal but not instrumental music should have a deleterious effect on recall. Salamé and Baddeley (1989) examined these questions in three experiments that together compared recall in a silent condition with recall in a vocal music condition, an instrumental music condition, a condition with background speech, and one with background noise. Recall was impaired by vocal and instrumental background music and by background speech, but the disruptive effect of instrumental music was relatively small. The presence of background noise did not impair performance. On the basis of these results, Salamé and Baddeley suggested that both instrumental and vocal music (as well as speech), but not noise, pass the assumed mental filter but that vocal music subsequently has a more disruptive effect on the information in phonological short-term memory than instrumental music because it has more acoustic features in common with the information in this memory store.

As mentioned, in Salamé and Baddeley (1982), background speech consisting of spoken words (that is, meaningful stimuli) and spoken nonsense words (meaningless stimuli) disrupted digit recall equally, which

led to the conclusion that the disrupting feature of concurrent speech on phonological short-term memory is its phonology, not its semantics. However, a study by Martin et al. (1988) suggested that semantics does play a role in the disruptive effect of background speech. These authors examined reading comprehension, a task that is thought to also involve phonological short-term memory (e.g., Gathercole & Baddeley, 1993). The participants read different text passages on varied topics in silence and with background speech and background music. Degree of reading comprehension in each condition was assessed by means of a set of short-answer comprehension questions that followed the reading of each text passage. Martin and colleagues found that background speech but not instrumental music impaired reading comprehension. The effect of speech also occurred when the speech was sung. But of special interest here was the observation that, if the background speech was in a language unknown to the participants (i.e., Russian; in the other background speech conditions it was in the participants' native English), the interference effect was reduced substantially, suggesting a role for semantics in the effect of background speech on reading comprehension.<sup>1</sup> Furthermore—and contrary to what Salamé and Baddeley (1982) found in their serial digit-recall task—it was shown that background speech consisting of words impaired performance substantially more than background speech consisting of nonsense words, again suggesting a role for the semantic content of the background speech. The authors attributed the different patterns of results in theirs and Salamé and Baddeley's study to the different requirements of the specific tasks used: In serial recall, the participants must remember the presented items verbatim, whereas, in reading comprehension, the participants must understand the text. Hence, in serial recall, the phonological properties of the background materials interfere with the phonological representations of the items to recall, whereas, in reading comprehension, the semantic properties of the background materials interfere with the text's semantics.

## **THE PRESENT STUDY**

The present study examines the effect of background music on foreign vocabulary learning by means of PAL. The type of music used is vocal pop, and the language of the lyrics is known or unknown to the learners. As discussed, Martin et al. (1988) showed that especially meaningful background speech hinders reading comprehension, a task that involves the processing of meaning. There is reason to believe that PAL also involves the processing of meaning: Recall is better for foreign words that are paired with concrete L1 words during training than for those paired with abstract L1 words, even when the forms of the foreign words paired

with concrete L1 words are matched in difficulty (e.g., with regard to length and phonological structure) with those paired with abstract L1 words (De Groot, 2006; De Groot & Keijzer, 2000; Ellis & Beaton, 1993a, 1993b; Van Hell & Candia Mahn, 1997). In other words, a variable that involves native language words' meaning (concreteness) affects the speed with which new forms can be linked onto the native words. This finding implies that, while linking the new form (the foreign word to learn) to the old form (the corresponding native word), the extant meaning of the native word is processed as well and influences learning. Plausibly, then, just like in Martin and colleagues' study, meaningful background stimuli in particular may have a deleterious effect on performance, in this case in PAL.

We examined this hypothesis by manipulating the language of the lyrics of the background songs. In the familiar language condition, the songs' lyrics were either in Dutch (i.e., the native language of our participants) or in English, a language they all had also mastered, although their level of mastery was slightly less than that of native Dutch. In the unfamiliar language condition, the songs' lyrics were in Greek, a language unknown to all participants. In a third condition, no background music was played.

In addition to manipulating the language familiarity of the background songs, we varied the learning materials: We included the concreteness manipulation mentioned previously, pairing the foreign words to learn with either concrete or abstract Dutch words. This enabled us to replicate the concreteness effect described previously and to thus once again demonstrate that foreign vocabulary learning by means of PAL involves the processing of meaning. We furthermore varied the phonological typicality (typicality, in short) of the foreign words to learn by including both foreign words with a typical phonological structure—that is, a structure that agrees with the phonological system of the learner's native Dutch—and foreign words with an atypical phonological structure. Earlier studies have shown that typical foreign words are easier to learn than atypical ones (De Groot, 2006; Ellis & Beaton, 1993b; Service & Craik, 1993). A likely cause of this effect is that typicality influences the workings of the phonological loop: Subvocalization of atypical word forms is plausibly harder than that of typical word forms; the subvocalization of atypical word forms frustrates a smooth operation of the loop. The typicality effect thus provides converging evidence that foreign vocabulary learning through PAL involves the phonological loop system.

In addition to several immediate-recall tests—that is, recall tests taken during and immediately after the learning session—the present study also includes a delayed-recall test 1 week later. This is a crucial component in the study's design because it allows us to test

whether any effect of background music to be obtained is durable. The relevance of doing so is suggested by evidence that a specific learning condition that leads toward poorer recall scores than a second learning condition during immediate testing may show equal or better recall scores on a delayed test, and vice versa (e.g., Schneider, Healy, & Bourne, 1998, 2002; Thomas & Wang, 1996; Wang & Thomas, 1995). Therefore, if an adverse effect of vocal background music occurs on the immediate tests, it may not last or may even reverse into a beneficial effect on the delayed test. It is interesting to note that, whereas specific learning conditions have only shown short-lived effects or opposite effects in the short and long term, the aforementioned effects of word type (concreteness and typicality) occurred both initially and at some later point in time (De Groot, 2006; De Groot & Keijzer, 2000).

To summarize, this study examines whether vocal pop background music influences the learning of foreign vocabulary by means of PAL and the longevity of such influence, if it occurs at all. To become better informed on the cause of any effect of this type of background music to be obtained (e.g., an adverse effect due to interference with phonological short-term memory), the language familiarity of the vocal music's lyrics is manipulated, and different types of PAL stimuli are presented for learning: The L1 terms in the PAL stimuli are concrete or abstract, and the FL terms have a typical or an atypical phonological structure. We predict that, as in previous studies, lasting effects of concreteness and typicality will occur. New information is provided regarding the following main questions:

1. Does vocal pop background music, contrary to the baroque instrumental background music used in De Groot (2006), adversely affect FL PAL?
2. If so, does this hold irrespective of the meaningfulness of the music's lyrics? That is, is the effect as large when the lyrics are in a familiar language as when the lyrics are in a language unfamiliar to the learner?
3. Does any effect of music to be observed survive a time delay of 1 week between initial learning and retesting?

An additional but secondary question that this study potentially answers is whether any effect of vocal pop music affects different types of words equally or differently (if an adverse effect of music is obtained, it plausibly affects FL words that are difficult to learn in the first place more than words that are easier to learn). A further question is whether or not any effect of music remains constant in size over the course of the experiment. These two additional questions were motivated by the findings of De Groot (2006) that PAL stimuli containing infrequent L1 words were affected by the baroque instrumental music used in that study to a greater degree than those containing frequent L1 words

and that the effect of the background music varied across the different test sessions. Whereas in De Groot (2006) music's beneficial effect was larger during later recall tests than in recall tests held early on during learning (as if the participants at first attempted to focus fully on the learning task and only gradually allowed the music to penetrate task performance), an adverse effect of vocal pop music might gradually diminish over the course of learning (due to the participants learning to ignore the music).

## METHOD

### Participants

The participants were 41 psychology students at the University of Amsterdam, all native speakers of Dutch with a high level of proficiency in English, a language they had all been trained in through secondary school and the language of most of the textbooks and articles they read during their university schooling. In contrast, none of the participants knew Greek. All participants performed the vocabulary learning task in silence and with songs with Greek lyrics played in the background. Additionally, they all performed the learning task with songs in a familiar language played in the background. For 20 participants this language was English, for the remaining 21 it was Dutch. The data of these two groups of participants were gathered in two different periods with 1 year in between.<sup>2</sup> In a questionnaire administered after data collection, the participants in the English familiar language condition were asked to indicate their speaking and reading ability in English relative to native Dutch on a 7-point scale, on which 7 indicated equal proficiency in English and Dutch. The mean scores for speaking and reading were 5.20 ( $SD = 1.06$ ) and 6.20 ( $SD = 0.77$ ), respectively. The participants received course credit or a small monetary payment for participating.

### Materials

**Vocabulary to Learn.** Seventy-two paired-associates (PA) stimuli were used, each consisting of a Dutch word and a nonword (the FL word to learn).<sup>3</sup> The Dutch word and the paired FL word were always noncognates of each other; that is, their forms were dissimilar. Excepting two, all Dutch and FL words consisted of four, five, or six letters. The 72 PA stimuli varied with respect to the concreteness of their Dutch words and the phonological typicality of the FL words. Thirty-six of the Dutch

words had concrete referents; the remaining 36 had abstract referents. Of both the 36 concrete Dutch words and the 36 abstract ones, 18 were paired with a typical FL word, whereas the remaining 18 were paired with an atypical FL word. In total, then, four groups of 18 PA stimuli each were created this way. Because the frequency of occurrence of the familiar element in a PA stimulus has previously been shown to affect learning (De Groot, 2006; Lotto & De Groot, 1998), the frequency of the Dutch words in the stimuli was matched across these four groups, making use of the CELEX corpus (Baayen, Piepenbrock, & Van Rijn, 1993). Further details about the PA stimuli and examples of the four groups of stimuli are given in the Appendix.

The 72 PA stimuli were divided into three matched lists of 24 stimuli each, with each list consisting of six stimuli of each of the four different types (e.g., a group of six in which the Dutch words were all concrete and the FL words were all typical). A series of nonparametric tests (Kruskal-Wallis) was performed to verify that the concreteness and typicality scores for the four different types of stimuli did not differ among the three lists. No differences between lists were obtained ( $p > .16$  in all cases). Table 1 shows the average concreteness and typicality scores for the four different types of stimuli in all three lists. The corresponding frequency scores (matched between the lists and stimulus types) are also provided.

**Background Songs.** All songs used in the music conditions were quiet, easy-listening pop songs, five with Greek lyrics (presented to both groups of participants in the unfamiliar language condition), five with Dutch lyrics (presented in the familiar language condition to one group of participants), and five with English lyrics (presented in the familiar language condition to the other group of participants). All selected songs were judged to be fairly or totally unfamiliar to the participants, as judged by the research assistants who ran the experiments and belonged to the same population as the participants. Further details about the songs used are given in the Appendix.

## Apparatus

The experiment was run on a standard Windows PC to which two monitors (Dell E2209W widescreen LCD) were connected. The experimenter faced one monitor, and the participant the other. The experiments were programmed in Presentation (Version 14.6). The music was played from a standard stereo set at an average volume of 65–70 dB (measured with a Voltcraft Sound Level Model number 33-2050 from the participant's seat).

**Table 1.** Mean frequency, concreteness, and typicality scores for the four types of PAL stimuli in the three matched lists of materials (standard deviations in parentheses)

List	Frequency	Concreteness	Typicality
1			
Concrete/typical	3.40 (0.77)	6.25 (0.15)	39,283 (12,267)
Abstract/typical	3.31 (0.96)	2.24 (0.80)	40,730 (13,803)
Concrete/atypical	3.31 (0.96)	6.44 (0.16)	3,449 (2,203)
Abstract/atypical	3.44 (0.97)	2.43 (0.47)	3,493 (2,322)
2			
Concrete/typical	3.50 (0.74)	6.43 (0.11)	38,785 (11,911)
Abstract/typical	3.31 (0.96)	2.69 (0.41)	39,019 (10,735)
Concrete/atypical	3.47 (0.78)	6.35 (0.15)	4,746 (3,091)
Abstract/atypical	3.36 (0.74)	2.53 (0.76)	4,377 (2,317)
3			
Concrete/typical	3.30 (0.77)	6.36 (0.13)	35,211 (12,193)
Abstract/typical	3.22 (0.61)	2.50 (0.41)	39,630 (11,748)
Concrete/atypical	3.39 (0.75)	6.36 (0.21)	4,422 (1,606)
Abstract/atypical	3.26 (0.55)	2.71 (0.50)	4,139 (1,233)

*Note.* Frequency scores concern the logarithm of the Dutch words' frequency in the CELEX corpus (based on a count of 42.5 million words). Concreteness scores concern the average score of 26 participants who rated Dutch words on imageability on a 7-point scale (1 = the word does not evoke a mental image or only with great difficulty; 7 = the word evokes a mental image quickly and easily). The typicality scores concern the cumulative positional letter frequencies of the FL words (see the Appendix for further information).

## Procedure

The experiment consisted of a practice session, three learning + recall sessions, the completion of two questionnaires, and a delayed-recall test taken 1 week later.

**Practice Session.** Prior to the start of the training session, a brief practice session was held to familiarize the participants with the procedure used in the recall tests. In this session, participants were presented with 10 English words, one by one, and were asked to say out loud the Dutch translation of each of them. The English word was preceded by a fixation stimulus (an asterisk), which appeared in the middle of the screen for 1 s. Next, the English word replaced the fixation stimulus. The moment the participant spoke its Dutch translation out loud, the experimenter pressed a key to discontinue stimulus presentation and typed the participant's response. What was being typed appeared simultaneously on the experimenter's monitor but not on the participant's monitor. Subsequently, the experimenter started the next trial by pressing the *Enter* key. If the participant failed to come up with a

response, the stimulus word disappeared 5 s after its onset, and the experimenter typed a code for a nonresponse and then started the next trial by pressing the *Enter* key. The practice session was held without background music.

**Learning + Recall Sessions.** Following the practice session, the actual experiment started. It consisted of three experimental blocks, each containing three learning + recall rounds. In each block the participant studied one of the three matched lists of 24 PA stimuli in one of the three music conditions (silence, familiar language music, and unfamiliar language music), and their recall of the newly learned words was tested. The presentation of each specific list rotated across participants and music conditions so that any effect of the music manipulation to emerge could indeed be attributed to the music manipulation rather than to any accidental difference between the three lists in terms of the learning difficulty of the materials that may have remained despite the careful matching procedure. For instance, in Block 1, one of the three lists of 24 PA stimuli was presented for learning three times with Greek background music; each of these learning rounds was immediately followed by a round of recall. In Blocks 2 and 3 this procedure was repeated, once with learning in silence and a second time with background music in one of the two familiar languages (English for one group and Dutch for the other group). In all, there were nine learning + recall rounds per participant: three per music condition/block. The order in which the data for the three music conditions were obtained rotated between the participants.

During learning, the two terms of each PA stimulus were presented simultaneously in the middle of the screen next to each other: the Dutch word to the left and the corresponding FL word to the right. This presentation order—from the known to the unknown element—has been shown to lead to better learning than the reverse order (Griffin & Harley, 1996). The PA stimuli were presented for 10 s each in a different random order for each participant and each presentation round. The participants were allowed to use a learning strategy of their own choice. They were not explicitly asked to ignore the background music, or to attend to it, for reasons that the provision of such instruction would detract from the naturalness of the task.

The recall test that was administered after each learning round involved receptive testing, using the same procedure as in the practice session previously described: The newly acquired FL words appeared as stimuli, one by one and in random order, and the participants were asked to produce the Dutch translations of the FL words. Recall always took place in silence. This mimics the way FL vocabulary tests are usually administered in the classroom. The running of all nine learning + recall sessions took about 1.5 hr per participant.

**Questionnaires.** Immediately following the nine learning + recall rounds, each participant filled out a questionnaire that was intended to determine his or her position on the introvert–extrovert dimension. Additionally, the participants answered a small set of other questions; these sets differed slightly for the two groups of participants. For instance, all participants were asked to indicate on a 7-point scale how often they learned with music playing in the background ( $1 = \text{never}$ ;  $7 = \text{always}$ ), and the participants in the Dutch familiar language condition were asked whether they had recognized any of the songs.<sup>4</sup>

**Delayed Recall Session.** One week ( $\pm 1$  day) later, a delayed-recall test was administered in which all 72 FL words were presented once again, and the participants were asked to name the corresponding Dutch words. The presentation of the 72 FL words was split up in the same three sets as used during the recall tests the week before (i.e., the immediate-recall tests). The order in which these sets were presented rotated between the participants. Just as with the immediate-recall tests, this delayed test took place in silence. It took about 15 min.

## Data Analysis

Because each participant learned four different types of FL words (concrete/typical, concrete/atypical, abstract/typical, and abstract/atypical) and was tested four times (on three immediate tests and on one delayed test) in each of three music conditions (silence; lyrics in a familiar language, Dutch or English; and lyrics in unfamiliar Greek), in total there were 48 conditions in the experiment ( $4 \times 4 \times 3$ ), of which 36 concerned the training session ( $4 \times 3 \times 3$ ) and the remaining 12 the delayed-recall test ( $4 \times 1 \times 3$ ). For each participant, we calculated 48 recall scores: one for each of these 48 conditions. Each score was the proportion of correct responses of the participant to the six test stimuli of a particular type (e.g., concrete/typical) within one of the four recall tests, within one of the three music conditions. These scores could thus vary between 0 and 1, where 0 would mean that none of the six test stimuli was responded to correctly, and 1 would mean that all of them were responded to correctly.

On these recall scores two factorial repeated-measures ANOVAs were run. The first analysis (the learning analysis) examined the effects of the music variable and the word-type variables during learning and was based on the 36 recall scores calculated for each participant on the data obtained in the training session. This analysis had the

following structure: 3 (Music: silence vs. familiar language lyrics vs. unfamiliar language lyrics)  $\times$  2 (Concreteness: concrete vs. abstract)  $\times$  2 (typicality: typical vs. atypical)  $\times$  3 (Test: Test 1 vs. Test 2 vs. Test 3)  $\times$  2 (Group: English vs. Dutch as the language of the lyrics in the familiar language music condition). The variable group was treated as a between-subjects variable. All remaining variables were within-subjects variables.

The second analysis was based on the 12 recall scores calculated for each participant on the data of the delayed-recall test. The purpose of this analysis was to discover whether any effects of the music and word-type manipulations obtained during learning were long lasting. This analysis had the structure 3 (Music)  $\times$  2 (Concreteness)  $\times$  2 (Typicality)  $\times$  2 (Group). Note that the three levels of music concern the manipulation during learning the week before. After all, testing always took place in silence.<sup>5</sup> For all analyses, the threshold for significance was set at  $p < .05$ , and effect sizes (partial eta-squared,  $\eta_p^2$ ) are provided for all significant effects.

## RESULTS

### Learning

**Main Effects.** The main effect of music was statistically significant,  $F(2, 78) = 6.60$ ,  $MSE = .114$ ,  $p = .002$ ,  $\eta_p^2 = .14$  (see Table 2 for the mean proportion of correct scores). The contrasts between the silence and familiar language conditions and between the familiar and unfamiliar language conditions were significant ( $p = .007$  and  $p = .028$ , respectively), but the contrast between the silence and unfamiliar language conditions was not ( $p = 1.00$ ). These results show that, as compared with silence, music with familiar language lyrics impairs foreign vocabulary learning through PAL, whereas such learning is not affected by music with unfamiliar language lyrics.

The main effects of concreteness, typicality, and test were also significant—concreteness:  $F(1, 39) = 99.94$ ,  $MSE = .049$ ,  $p < .001$ ,  $\eta_p^2 = .72$ ; typicality:  $F(1, 39) = 62.66$ ,  $MSE = .026$ ,  $p < .001$ ,  $\eta_p^2 = .62$ ; test:  $F(2, 78) = 433.26$ ,  $MSE = .076$ ,  $p < .001$ ,  $\eta_p^2 = .92$  (all three contrasts relating to the variable test were statistically significant;  $p < .001$ ). The main effect of group was not significant ( $p = .375$ ). Table 2 presents the mean proportion correct scores for the different levels of the variables that show significant effects. As can be seen, the present results replicate previous findings that the learning of foreign vocabulary is easier for concrete words than for abstract words and that foreign words with typical phonology are easier to learn than those with atypical phonology. Additionally, the results show that recall performance improves over the course of learning.

**Table 2.** Mean proportion correct recall and corresponding standard errors (in parentheses) for the significant main effects on the learning analysis

Main effect	Proportion correct ( <i>SE</i> )
Music	
Silence	.70 (.023)
Unfamiliar	.69 (.026)
Familiar	.63 (.029)
Concreteness	
Concrete	.73 (.021)
Abstract	.61 (.026)
Typicality	
Typical	.70 (.022)
Atypical	.64 (.024)
Test	
1	.39 (.030)
2	.74 (.027)
3	.89 (.016)

**Interactions.** The introductory section concluded by mentioning two secondary questions this study may answer—namely, whether any effect of music to be obtained would affect all types of words equally and whether this effect’s magnitude would remain constant over the course of the experiment. Answers to these two questions can be found by examining the interactions between the variable music on the one hand and the remaining variables on the other. With the exception of one, all 16 first-, second-, and third-order interactions between music and the other variables failed to reach statistical significance, suggesting that, in general, vocal pop music with lyrics in a familiar language hinders the learning of different types of FL words similarly and that its effect remains constant over successive learning sessions. The one exception concerned the first-order interaction between music and concreteness,  $F(2, 78) = 4.61$ ,  $MSE = .031$ ,  $p = .013$ ,  $\eta_p^2 = .11$ . The associated

**Table 3.** Mean proportion correct recall and corresponding standard errors (in parentheses) for the significant interaction of concreteness and music on the learning analysis

Concreteness	Music		
	Silence	Unfamiliar	Familiar
Concrete	.74 (.022)	.75 (.024)	.69 (.027)
Abstract	.66 (.027)	.62 (.031)	.56 (.032)

**Table 4.** Mean proportion correct recall and corresponding standard errors (in parentheses) for the significant interaction of concreteness and test on the learning analysis

Concreteness	Test		
	1	2	3
Concrete	.46 (.032)	.81 (.024)	.92 (.014)
Abstract	.31 (.030)	.66 (.033)	.86 (.020)

proportion correct scores, shown in Table 3, indicated that not only lyrics in a familiar language but also those in unfamiliar Greek had a detrimental effect (but a relatively small one) on the learning of abstract (but not concrete) words.

Of the remaining interactions (those not involving the variable music), two more were significant—namely, the one between test and concreteness, shown in Table 4,  $F(2, 78) = 14.47$ ,  $MSE = .022$ ,  $p < .001$ ,  $\eta_p^2 = .27$ , and between test and typicality, shown in Table 5,  $F(2, 78) = 5.30$ ,  $MSE = .022$ ,  $p = .007$ ,  $\eta_p^2 = .12$ . The cell means associated with these interactions (see Tables 4 and 5) indicate that the difficult words (abstract, atypical) catch up with the easy ones (concrete, typical) over the course of learning. None of the interactions between the variable group and the remaining variables approached significance (all  $p$  values  $> .19$ ),<sup>6</sup> which suggests that the adverse effect of vocal music with familiar language lyrics was equally large with Dutch and English lyrics.

## Delayed Recall

**Main Effects.** The main effects of concreteness and typicality were again significant—concreteness:  $F(1, 39) = 32.60$ ,  $MSE = .054$ ,  $p < .001$ ,

**Table 5.** Mean proportion correct recall and corresponding standard errors (in parentheses) for the significant interaction of typicality and test on the learning analysis

Typicality	Test		
	1	2	3
Typical	.43 (.033)	.78 (.026)	.91 (.015)
Atypical	.35 (.029)	.69 (.030)	.88 (.019)

$\eta_p^2 = .46$ ; typicality:  $F(1, 39) = 32.23$ ,  $MSE = .039$ ,  $p < .001$ ,  $\eta_p^2 = .45$ . The mean proportion correct score was .39 for concrete stimuli ( $SE = .029$ ) and .27 for abstract stimuli ( $SE = .026$ ); it was .38 for typical foreign words ( $SE = .029$ ) and .28 for atypical ones ( $SE = .025$ ). The main effect of music was now nonsignificant ( $F < 1$ ), and the corresponding effect size was small ( $\eta_p^2 = .01$ ). The proportion correct score was .35 for the silence condition ( $SE = .031$ ), .33 for the condition with unfamiliar language lyrics ( $SE = .033$ ), and .32 for the condition with familiar language lyrics ( $SE = .029$ ). These results indicate that the word-type manipulations have long-lasting effects, whereas the adverse effect of familiar language background music as obtained during learning does not last. The main effect of group was not significant ( $p = .20$ ), indicating that, statistically, the degree of long-term retention was the same for the two different familiar language groups.

**Interactions.** The interaction between concreteness and typicality was significant:  $F(1, 39) = 3.95$ ,  $MSE = .028$ ,  $p = .05$ ,  $\eta_p^2 = .09$ . The proportion correct score was .24 ( $SE = .026$ ), .31 ( $SE = .031$ ), .33 ( $SE = .030$ ), and .46 ( $SE = .032$ ) for the abstract/atypical, abstract/typical, concrete/atypical, and concrete/typical stimuli, respectively, thus indicating that the degree of long-term retention is especially good for stimuli that are both concrete and typical. None of the remaining interactions was significant.<sup>7</sup>

## SUMMARY AND DISCUSSION

The primary goal of this study was to see whether vocal pop background music influences the learning of foreign vocabulary through PAL and, if so, whether the meaningfulness of the background music played a role, what the nature of the effect would be (detrimental or facilitative), and whether the effect would be durable. A secondary goal was to see whether earlier findings that the concreteness and typicality of the learning materials influence the recall scores, both during and immediately after learning and (about) 1 week later, would be obtained again. Furthermore, we were interested in seeing whether any effect of vocal pop music on learning obtained would affect the different types of PAL stimuli equally and whether it would stay the same size over the course of learning.

To these ends, and to become informed on the specific cause of music's effect, we manipulated the language familiarity (meaningfulness) of the music's lyrics and the type of foreign words presented for learning. Additionally, we included multiple recall tests during initial learning and had our participants come back to the laboratory for an additional recall test 1 week after training.

The results showed impaired recall on the immediate tests when the language of the songs' lyrics was familiar to the learner but not when it was unfamiliar. The adverse effect of familiar language lyrics was statistically equally large for the four different stimulus types, and its magnitude did not change over the course of learning. But a particularly noteworthy finding is that this detrimental effect was short-lived: On a delayed-recall test held 1 week after learning, the recall scores were statistically equally large for words learned in silence and for those learned with either type of background music. This result agrees with earlier demonstrations that a specific learning condition that leads to poorer recall than a second learning condition during immediate testing may no longer show this adverse effect when testing takes place at some later point in time (Schneider et al., 1998, 2002; Thomas & Wang, 1996; Wang & Thomas, 1995).

It is furthermore noteworthy that the specific effect of background music that we obtained here was negative, whereas the classical, instrumental background music used in De Groot (2006) exerted a positive effect on learning. This strongly suggests that the effects of background music in the two studies have different causes. As detailed later in this section, a plausible cause of the detrimental effect of vocal pop music with familiar language lyrics is interference with the operation of phonological short-term memory. In contrast, the positive effect of classical instrumental background music may be due to a physiological state of relaxed alertness that it causes, to the induction of a pleasant mood (Felix, 1993), or to the music boosting cortical arousal to a level that is beneficial for learning (e.g., Cassidy & MacDonald, 2007; Furnham & Allass, 1999).

Whereas the effect of familiar language background music was short-lived, the effects of the word-type manipulations were long lasting. In agreement with previous studies (e.g., De Groot, 2006; De Groot & Keijzer, 2000; Ellis & Beaton, 1993b; Van Hell & Candia Mahn, 1997), the recall scores were larger for the concrete stimuli than for the abstract ones, and this held for both the immediate and delayed-recall tests. Similarly, and again in agreement with previous findings (De Groot, 2006; Ellis & Beaton, 1993b; Service & Craik, 1993), the recall scores were larger for foreign words with typical phonology than for those with atypical phonology, a finding that again held for both immediate and delayed recall.

In sum, the present study indicates that vocal pop background music with lyrics in a familiar language hinders foreign vocabulary learning through PAL. But, importantly, it also shows that this adverse effect is short-lived. Additionally, the present study replicated the word-type effects obtained in earlier studies, including the finding that, unlike the specific effect of background music, these effects are long lasting.

As mentioned in the introductory section, the occurrence of a typicality effect in foreign vocabulary learning through PAL constitutes evidence that the phonological loop system, including phonological short-term memory, is involved in the task. Given that a typicality effect clearly showed up in this study and, therefore, short-term memory was likely involved, a logical starting point in accounting for the present specific effect of vocal background music on PAL vocabulary learning is to assume that it was caused by the different types of background music affecting short-term memory differently. On this assumption, which of the various views on the permeability of short-term memory by background stimuli presented in the introduction is supported by the present data? One of the hypotheses considered was that background stimuli pass through a mental filter, accessing short-term memory if they are structured but not when they are unstructured. This account would predict a detrimental effect of both vocal and instrumental music (Salamé & Baddeley, 1989). Our study did not include an instrumental music condition and, therefore, provides no direct means to evaluate this hypothesis. Nevertheless, it is ruled out by the fact that the background music in our two music conditions was equally structured and yet only one of them hampered performance. It is also inconsistent with the earlier finding (De Groot, 2006) that instrumental background music helps rather than hinders foreign vocabulary learning through PAL.

A second hypothesis was that background information involving a human voice has privileged access to short-term memory (Salamé & Baddeley, 1989). This hypothesis can be rejected on the grounds that, if correct, both of our two music conditions should have shown equally large, adverse effects on learning performance. Yet, only vocal pop music with lyrics in a familiar language did so.

A third hypothesis was that the degree of phonological similarity between the background information and the materials presented on the actual task played a role, the interfering effect being especially large when the background materials and the materials of the task to be performed were phonologically similar (Salamé & Baddeley, 1982). Because Greek is likely to be phonologically dissimilar to Dutch, and because Dutch was the language of the majority of the learning materials (it was the language of one of the two terms in all PAL stimuli; additionally, half of the foreign words in the pairs were Dutchlike), this account could explain the fact that vocal music with Greek lyrics produced a null effect. However, it runs into trouble when considering the fact that the statistical analysis revealed that the two groups of participants, those with either Dutch or English as the language of the music's lyrics in the familiar language condition, showed an equally large adverse effect of familiar language background music, despite the fact that the Dutch lyrics were likely more similar to the task

materials than the English lyrics. On the basis of a phonological similarity account, one could have expected a larger effect with Dutch than with English lyrics.

A final hypothesis, based on Martin et al.'s (1988) study, was that meaningful, but not meaningless, background stimuli interfere with performance on tasks that involve semantic processing. The present concreteness effect once again demonstrated that foreign vocabulary learning by means of PAL indeed involves semantic processing. According to this account, learning should be hampered when the lyrics of the vocal background music are in a familiar language (i.e., are meaningful) but not when they are in an unfamiliar language (i.e., are meaningless). The present data are consistent with this account.

To summarize, it appears that it was not the structured nature of the background music that caused the interference in the familiar language music condition nor was it the human voice in the music or the phonological similarity between the background stimuli and the materials to learn. Instead, just as in Martin et al.'s (1988) study on reading comprehension, it seems that the meaningfulness of the background stimuli in this specific condition hindered task performance, plausibly because meaningful stimuli gain access to short-term memory. This then obstructs learning, possibly because the meaningful background stimuli that have accessed short-term memory detract attention from the learning task. Attentional capacity being finite, this leaves less attention to dedicate to learning than with background stimuli that do not access short-term memory.

But in view of the obvious goal of foreign language vocabulary learning to store permanent, not transient, representations of the learned words in memory, perhaps the most noteworthy finding of this study is that the present inhibitory effect of background music with familiar language lyrics on foreign vocabulary learning did not last. This finding motivated the present article's title: "Let the music play!" Plausibly, similar results would have been obtained with study tasks other than the present vocabulary learning task. Nevertheless, prudence in generalizing our conclusions is called for because, in agreement with Kämpfe et al.'s (2011) conclusion, the present finding that vocal background music with familiar language lyrics but not with unfamiliar language lyrics negatively affects learning (and the earlier finding that classical instrumental background music has a positive effect) also shows that the effect of background music on performance can be very specific.

*Received 18 June 2012*

*Accepted 16 July 2013*

*Final Version Received 13 August 2013*

## NOTES

1. Even though Salamé and Baddeley (1989) did not explicitly examine the role of the meaningfulness of the background speech, their study included a manipulation that could have revealed an effect of meaningfulness if this variable had been implemented differently. Specifically, in one of the two experiments that included a vocal music condition, the music concerned modern French chansons, the songs' lyrics being in the participants' native language (French). In the second experiment, it concerned classical opera arias in languages foreign to the participants (German and Italian). Had the native language/foreign language manipulation not been confounded by type of music (modern vs. classical) and had the difference in performance between these two vocal music conditions actually been tested statistically, a possible role of semantics in the disruptive effect of vocal background music on phonological short-term memory could have been revealed.

2. We considered reporting the data of the participants with English as the language of the lyrics in the familiar language condition and of those for whom Dutch served this role as two separate experiments rather than two separate conditions of one and the same experiment, because that would do justice to our original goal: to replicate our first set of findings with a slightly modified implementation of the familiar language condition. We decided against this because presenting the combined data in a single experiment allows a more concise presentation of the results.

3. The use of mock FL words in examining foreign vocabulary learning by means of PAL is quite common and enables the careful manipulation and control of certain word-type variables such as phonological typicality and cognate status and the ruling out of effects of prior knowledge about the language to learn.

4. The participants were asked to fill out the extroversion questionnaire because earlier studies (Cassidy & MacDonald, 2007; Daoussis & McKelvie, 1986; Furnham & Allass, 1999; Furnham & Bradley, 1997) have suggested that especially introverts may be hindered by background music. The question of how often the participants listened to background music while studying was asked because their answer to this question might relate to the learning scores to be obtained: Participants who often learn with background music may do so because they feel the music does not hamper their learning or even improves it, whereas those who usually learn in silence may experience an adverse effect of background music while studying. The question of whether any of the songs had been recognized was posed because familiarity with the background music per se, instead of familiarity with the language of its lyrics, may also affect task performance. One participant recognized (just) one of the Dutch songs, and none recognized any of the Greek songs, thus confirming the experimenters' judgment that the songs were generally unfamiliar to the participants.

5. Repeated-measures ANOVA requires homogeneous variances. Whether this requirement is met can be tested with Levene's test of homogeneity, which we did. According to Levene's test, in the analysis of the data obtained during learning, only 2 out of the 36 combinations of levels of the within-subjects variables had heterogeneous variances across the two groups, and, in the analysis of the data obtained in the delayed test, heterogeneity held for only 1 out of the 12 combinations. Despite the fact that heterogeneity held in only very few cases, we performed an arcsine transformation of our data to stabilize the variances between the groups. We then repeated the analyses, which showed results similar to the analyses just described and reported in the main text: All effects that were significant in the analysis of the nontransformed data were significant in the analysis of the transformed data, and all effects that were nonsignificant in the former analysis were nonsignificant in the latter analysis.

6. Music  $\times$  Group,  $p = .422$ ; Test  $\times$  Group,  $p = .834$ ; Concreteness  $\times$  Group,  $p = .926$ ; Typicality  $\times$  Group,  $p = .590$ ; Music  $\times$  Test  $\times$  Group,  $p = .716$ ; Music  $\times$  Concreteness  $\times$  Group,  $p = .463$ ; Test  $\times$  Concreteness  $\times$  Group,  $p = .460$ ; Music  $\times$  Test  $\times$  Concreteness  $\times$  Group,  $p = .954$ ; Music  $\times$  Typicality  $\times$  Group,  $p = .982$ ; Test  $\times$  Typicality  $\times$  Group,  $p = .610$ ; Music  $\times$  Test  $\times$  Typicality  $\times$  Group,  $p = .609$ ; Concreteness  $\times$  Typicality  $\times$  Group,  $p = .403$ ; Music  $\times$  Concreteness  $\times$  Typicality  $\times$  Group,  $p = .751$ ; Test  $\times$  Concreteness  $\times$  Typicality  $\times$  Group,  $p = .197$ ; Music  $\times$  Test  $\times$  Concreteness  $\times$  Typicality  $\times$  Group,  $p = .572$ .

7. We performed post hoc analyses to see whether especially introverts were affected by background music (see Note 4). The only variables included in these analyses, one on

the learning data and a second on the data of the delayed-recall test, were music and extroversion. The means for the interaction between these variables as obtained in the analysis of the learning data indeed suggested that music with familiar language lyrics hindered only the introverts: The proportion correct score was .60 for the introverts and .68 for the extroverts (introverts and extroverts performed similarly in the unfamiliar language and silence conditions; all four proportion correct scores were between .69 and .72). However, the interaction was not significant,  $F(2, 72) = 2.54$ ,  $MSE = .028$ ,  $p = .09$ ,  $\eta_p^2 = .07$ . This also held for the analysis performed on the data of the delayed-recall test (Music  $\times$  Extroversion:  $F < 1$ ).

We also performed a post hoc analysis to see whether people who usually learn in a silent environment are especially negatively affected by background music (see Note 4). To examine this possibility, for each participant we calculated the difference between the recall scores for the silence and familiar-language-lyrics conditions in Test 3. Next, we fit a linear regression model on the resulting difference scores, with the participants' scores on the question of how often they learn with background music as the predictor variable. The results indicated that the score on this question could not significantly predict the difference score,  $F(1, 38) = 3.49$ ,  $p = .07$ .

## REFERENCES

- Angel, L. A., Polzella, D. J., & Elvers, G. C. (2010). Background music and cognitive performance. *Perceptual and Motor Skills, 110*, 1059–1064.
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Neuroscience, 4*, 829–839.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586–595.
- Belsham, R. L., & Harman, D. W. (1977). Effect of vocal vs non-vocal music on visual recall. *Perceptual and Motor Skills, 44*, 857–858.
- Cassidy, G., & MacDonald, R. A. R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. *Psychology of Music, 35*, 517–537.
- Daoussis, L., & McKelvie, S. J. (1986). Musical preferences and effects of music on a reading comprehension test for extraverts and introverts. *Perceptual and Motor Skills, 62*, 283–289.
- De Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning, 56*, 463–506.
- De Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language, 33*, 600–629.
- De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*, 1–56.
- Ellis, N., & Beaton, A. (1993a). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 46*, 533–558.
- Ellis, N. C., & Beaton, A. (1993b). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*, 559–617.
- Felix, U. (1993). The contribution of background music to the enhancement of learning in Suggestopedia: A critical review of the literature. *Journal of the Society for Accelerative Learning and Teaching, 18*, 277–303.

- Furnham, A., & Allass, K. (1999). The influence of musical distraction of varying complexity on the cognitive performance of extroverts and introverts. *European Journal of Personality, 13*, 27–38.
- Furnham, A., & Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology, 11*, 445–455.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213.
- Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439–454.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Erlbaum.
- Grace, G. (2008). *Glennis* [CD]. Netherlands: CNR.
- Grace, G. (2008). Zeg me nee [Say no to me]. On *Glennis* [CD]. Netherlands: CNR.
- Griffin, G., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics, 17*, 443–460.
- Hu, C.-F. (2003). Phonological memory, phonological awareness, and foreign language word learning. *Language Learning, 53*, 429–462.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80*, 327–339.
- Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2012). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of Music, 40*, 339–356.
- Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music, 39*, 424–448.
- Lotto, L., & De Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48*, 31–69.
- Marlin, L. (1999). *Playing my game* [CD]. Europe: Virgin.
- Marlin, L. (1999). Unforgivable sinner. On *Playing my game* [CD]. Europe: Virgin.
- Marlin, L. (1999). Where I'm headed. On *Playing my game* [CD]. Europe: Virgin.
- Marlin, L. (2003). *Another day* [CD]. Norway: EMI Music.
- Marlin, L. (2003). Another day. On *Another day* [CD]. Norway: EMI Music.
- Marlin, L. (2003). You weren't there. On *Another day* [CD]. Norway: EMI Music.
- Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language, 27*, 382–398.
- Mondria, J. A. (2003). The effects of inferring, verifying, and memorizing on the retention of L2 word meanings. *Studies in Second Language Acquisition, 25*, 473–499.
- Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*, 331–337.
- Parente, J. A. (1976). Music preference as a factor of musical distraction. *Perceptual and Motor Skills, 43*, 337–338.
- Pausini, L. (2002). *From the inside* [CD]. Europe: Atlantic.
- Pausini, L. (2002). I need love. On *From the inside* [CD]. Europe: Atlantic.
- Presentation (Version 14.6) [Computer software]. Albany, CA: Neurobehavioral Systems.
- Rolf, P. C., & Rijnsoever, Van, R. J. (1984). *Positionele letter-frekventies van het Nederlands* [Positional letter frequencies of Dutch]. Lisse, the Netherlands: Swets & Zeitlinger.
- Rombley, E. (1997). Ik doe wat ik wil [I do what I want]. On *Thuis* [CD]. Netherlands: TWF Music.
- Rombley, E. (1997). *Thuis* [Home] [CD]. Netherlands: TWF Music.
- Rombley, E. (2007). Een keer meer dan jij [One time more than you]. On *Meer dan ooit* [CD]. Europe: Universal Music.
- Rombley, E. (2007). Ik hou je vast [I am holding you]. On *Meer dan ooit* [CD]. Europe: Universal Music.
- Rombley, E. (2007). Laat me gaan [Let me go]. On *Meer dan ooit* [CD]. Europe: Universal Music.
- Rombley, E. (2007). *Meer dan ooit* [One more time than you] [CD]. Europe: Universal Music.
- Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150–164.

- Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *41*, 107–122.
- Schneider, V. I., Healy, A. F., & Bourne, L. E. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 77–90). Mahwah, NJ: Erlbaum.
- Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419–440.
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *45*, 21–50.
- Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, *32*, 608–623.
- Tamta. (2006). Den S' Afino [I am not leaving you]. On *Tamta* [CD]. Greece: Minos EMI.
- Tamta. (2006). Den Teleionei Etsi I Agapi [This is not how our love should end]. On *Tamta* [CD]. Greece: Minos EMI.
- Tamta. (2006). Oti Teleionei Pona [Whatever ends, hurts]. On *Tamta* [CD]. Greece: Minos EMI.
- Tamta. (2006). *Tamta* [CD]. Greece: Minos EMI.
- Tamta. (2006). Ti Sou Ftaio Ego Pou Toso S' Agapo [Why are you blaming me for loving you so much]. On *Tamta* [CD]. Greece: Minos EMI.
- Tamta. (2006). Tora Pos Na Sviso [How can I erase now]. On *Tamta* [CD]. Greece: Minos EMI.
- Thomas, M. H., & Wang, A. Y. (1996). Learning by the keyword mnemonic: Looking for long-term benefits. *Journal of Experimental Psychology: Applied*, *2*, 330–342.
- Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*, 700–708.
- Van Dale Nieuw Handwoordenboek der Nederlandse Taal* [Van Dale New Desk Dictionary of the Dutch Language] (7th ed.). (1972). 'S-Gravenhage, the Netherlands: Martinus Nijhoff.
- Van Hell, J. G., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, *47*, 507–546.
- Wang, A. Y., & Thomas, M. H. (1995). Effect of keywords on long-term retention: Help or hindrance? *Journal of Educational Psychology*, *87*, 468–475.
- Wang, A. Y., Thomas, M. H., & Ouellette, J. A. (1992). Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology*, *84*, 520–528.
- Wolfe, D. E. (1983). Effects of music loudness on task performance and self-report of college-aged students. *Journal of Research in Music Education*, *31*, 191–201.

## APPENDIX

### ADDITIONAL INFORMATION ON THE PA STIMULI AND BACKGROUND SONGS

#### Selection of the PA Stimuli

As a proxy of the word concreteness variable, we used word imageability, a measure of how easily a word evokes a mental image of its referent. This variable correlates extremely highly ( $r \approx .95$ ) with word

concreteness, a variable that reflects the degree to which a word's referent can be experienced by the senses: Words that are easy to imagine are typically concrete; words hard to imagine are usually abstract. The selected Dutch words' imageability scores were derived from a study by De Groot, Dannenburg, and Van Hell (1994).

The typicality scores of the constructed FL words were derived from a corpus that contains counts of how often a specific letter occurs in a specific position in Dutch words of a specific length (Rolf & Van Rijnsoever, 1984). The atypical FL words were constructed from letters that occur infrequently in their respective positions in Dutch words of this specific length. For instance, *obfa* is an atypical FL word because the *o*, *b*, *f*, and *a* occur relatively infrequently in the first, second, third, and fourth position, respectively, of Dutch four-letter words. In contrast, *voje* is a typical FL word because the *v*, *o*, *j*, and *e* occur relatively frequently in the first, second, third, and fourth position, respectively, of Dutch four-letter words. Consultation of a Dutch dictionary (*Van Dale Nieuw Handwoordenboek der Nederlandse Taal*, 1972) confirmed that the thus-constructed FL words did not occur as words in Dutch.

### Examples of the PA Stimuli

Concrete/typical: *kamer-niek* "room"; *tabak-giedar* "tobacco"; *mond-daak* "mouth"; *wiel-wreeg* "wheel"; *vader-moraan* "father"; *haai-kodet* "shark"

Abstract/typical: *manier-taar* "manner"; *blaam-gadet* "blame"; *geval-helet* "case"; *lening-soren* "loan"; *eeuw-mije* "century"; *wrok-aalnee* "grudge"

Concrete/atypical: *vrouw-isfo* "woman"; *erwt-umvo* "pea"; *hoofd-pjubo* "head"; *bruid-izubi* "bride"; *boek-ajwuc* "book"; *peper-psiz* "pepper"

Abstract/atypical: *kans-ibwa* "chance"; *rente-uspo* "rent"; *gevoel-iffux* "feeling"; *deugd-ubbuh* "virtue"; *indruk-fsiwi* "impression"; *spraak-isboj* "speech"

It is important to note that the left and right elements in each PA stimulus are the Dutch word and the FL word to learn, respectively. The English translation of the Dutch word is in quotation marks. All words are nouns.

### Background Songs

All Greeks songs were from the album *Tamta* (Tamta, 2006, tracks 2, 6, 8, 10, and 11) of the Greek singer Tamta. The English songs were from the album *From the Inside* (Pausini, 2002, track 1) and from *Playing My Game* and *Another Day* by Lene Marlin (Marlin, 1999, tracks 3 and 8; Marlin, 2003, tracks 1 and 3). A selection from different albums was made to match the type of songs used in the English condition with those presented in the Greek condition (all songs from Tamta's album are quiet ones, whereas the albums by Marlin and Pausini also contain

some that are noisier; only quiet, easy-listening songs from the Pausini and Marlin albums were selected). The Dutch songs were from the album *Glennis* (Grace, 2008, track 2) and from two albums by Edsilia Rombley: three from *Meer dan ooit* (Rombley, 2007, tracks 2, 3, and 12) and one from *Thuis* (Rombley, 1997, track 2); again all songs were selected such that they matched the type of songs used in the Greek condition.