



HAL
open science

Is loudness part of a sound recognition process?

Patrick Susini, Olivier Houix, Lou Seropian, Guillaume Lemaitre

► **To cite this version:**

Patrick Susini, Olivier Houix, Lou Seropian, Guillaume Lemaitre. Is loudness part of a sound recognition process?. *Journal of the Acoustical Society of America*, 2019, 146 (2), pp.EL172-EL176. 10.1121/1.5121562 . hal-02270897

HAL Id: hal-02270897

<https://hal.science/hal-02270897v1>

Submitted on 2 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Is loudness part of a sound recognition process?

Patrick Susini,^{a)} Olivier Houix, Lou Seropian, and Guillaume Lemaitre

STMS Lab (Sciences et Technologies de la Musique et du Son, Ircam-CNRS-SU),

1 place Igor Stravinsky, 75004 Paris, France

patrick.susini@ircam.fr, olivier.houix@ircam.fr, louseropian@gmail.com,

guillaumelemaitre@gmail.com

Abstract: Influence of loudness on sound recognition was investigated in an explicit memory experiment based on a conscious recollection—test phase—of previously encoded information—study phase. Three encoding conditions were compared: semantic (sounds were sorted in three different categories), sensory (sounds were rated in loudness), and control (participants were solely asked to listen to the sounds). Results revealed a significant study-to-test change effect: loudness change between the study and the test phases affects recognition. The effect was not specific to the encoding condition (semantic vs sensory) suggesting that loudness is an important hint for everyday sounds recognition.

[Q-JF]

Date Received: May 21, 2019 **Date Accepted:** July 23, 2019

1. Introduction

Everyday experience shows that loudness can change the perception and recognition of everyday sounds. For example, a recorded sound played with a different loudness can mislead listeners to a wrong identification; an explosion played at a very low sound level could be confused with the fizzling sound of oil in a hot stove; an explosion is a powerful event, and thus, cannot be associated with a soft sound pressure level in people's minds. On the contrary, some sound sources usually at a very low level, for example a crumpled paper, played at too high a level may be perceived louder than sources normally louder—such as water flow—or simply may become unidentifiable because of an acoustical zooming effect. This leads to the present question: Is loudness part of a sound recognition process? Several studies have shown that loudness might overpower less salient auditory attributes and thus strongly influence timbre perception (Melara and Marks, 1990; Susini *et al.*, 2015). Results from different studies using a sorting task all revealed an effect of loudness on the obtained groupings of sounds (Maffiolo, 1999; Giordano *et al.*, 2010; Aldrich *et al.*, 2009). In recent studies on everyday sounds (Lemaitre *et al.*, 2010; Houix *et al.*, 2012), an “ecological adjustment of sound levels” was performed in order to avoid any influence of the loudness on participants' identification during a sorting task experiment; loudness was adjusted for each sound to reproduce the “usual” or “ecological” level of the sounds as it will be in an everyday life situation. The objective of the present work was to evaluate the influence of loudness on sound recognition. The procedure used was based on an explicit memory experiment which entails two successive phases, as it was proposed in Schacter and Church (1992), to test auditory identification: first, the study phase, during which information is encoded, and then the test phase, consisting of a conscious recollection of the information. In the study phase, participants were first required to listen to the sounds and then perform a specific task to encode the information. In the test phase, participants were instructed to make an explicit recognition judgment about the same sounds, mixed with new sounds and presented at a different sound pressure level (SPL). Accordingly, in the present study we examined whether explicit memory, involved during the test phase, is sensitive to a change in SPL between the test phase and the study phase; in other words, the effect of the “study-to-test change” in SPL was evaluated. To accomplish this objective, several sounds were presented to three groups of participants. During the study phase, one group of participants performed a semantic task consisting of making category judgments, one group performed a sensory task consisting of making loudness judgments, and one group just listened to the sounds (control condition). After a brief delay, participants were asked during the test phase to make a yes or no recognition judgment about the target sounds mixed with new sounds; the target sounds were presented with the same SPL as during the study

^{a)}Author to whom correspondence should be addressed.

phase and with a different SPL. Our main hypothesis was that the study-to-test change in SPL might be observed after an encoding task that focuses participants on sensory characteristics, such as loudness, with larger amplitude than after a semantic encoding task (Graf and Ryan, 1990). Our second hypothesis was that participants in the semantic encoding condition will be more accurate than those in the sensory encoding condition, because explicit memory is supposed to be higher when priming is based on semantic information (Schacter and Church, 1992). However, it was expected that participants in the sensory encoding condition will be better than those in the control condition.

2. Experiments

2.1 Participants

Three groups, 18, 11, and 11 participants ranging in age from 20 to 43 years, respectively, with a mean of 26.8, participated in the experiment. Each group performed one of the three variants of the first part (study phase) of the experiment, then they all performed the second part (test phase). Participants were randomly assigned to each group. All reported normal hearing. They were paid for their participation, and were all naive with respect to the hypotheses being tested.

2.2 Stimuli

Two sets of 42 and 12 everyday sounds were selected from two previous studies (Lemaitre *et al.*, 2010; Houix *et al.*, 2012) and used in the current experiments as targets and distractors, respectively. The sounds had approximately the same duration; most of the sounds lasted about 1.5 s, the shortest one lasted 0.5 s. In order to cover a broad range of everyday sounds, the 42 target sounds were balanced across three identifiable sounds: 14 sounds made by liquids (“liquid sounds”), 14 sounds of machines (“machine sounds”), and 14 sounds made by solid objects (“solid sounds”). The liquid sounds were sounds such as water drops, flows in a water pipe, or a container filled with water. Machine sounds were mainly engine or electric sounds. The solid sounds were composed of sounds made with different materials and produced by different actions such as opening/closing, compressing, cutting, and rubbing. The stimuli were selected on the basis of their SPL in an everyday context; half of the target sounds (21) were chosen because they are sounds with an expected low SPL in an everyday life context (for example, the sound of a fridge) compared to the other half of the target sounds (21), which were selected because they are usually heard with a higher SPL (for example, the sound of a vacuum cleaner). In order to respect this criteria, a typical SPL of 55 dB was associated with the quiet sounds (7 liquid, 7 machine, and 7 solid sounds), and similarly a typical SPL of 70 dB was associated with the loud sounds (7 liquid, 7 machine, and 7 solid sounds). The 12 distractors were also balanced over the same three categories and the two SPLs; a SPL of 55 dB was associated with 2 liquid, 2 machine, and 2 solid sounds, and a similar association was done for the SPL of 70 dB. The imbalance between targets and distractors is a compromise to maintain a reasonable duration of the experiment and to discourage attempts to memorize targets by presenting a large diversity of sounds, while presenting distractors.

2.3 Apparatus

Sounds were played by a Macintosh Mac Pro workstation with an RME Fireface 800 soundcard and presented diotically through headphones (Sennheiser HD 650, Sennheiser electronic GmbH & Co. KG, Wedemark, Germany). The experimental setup was calibrated using a Brüel&Kjær 2238 Mediator sound-level meter, coupled with the mounting plate provided for circumaural headphones. A 1-kHz pure-tone at a level of 70 dB SPL was used for the calibration. Each participant was tested in a double-walled IAC sound-insulated booth.

2.4 Procedure

The whole experiment consisted of six blocks. Each block contained two successive phases; a study phase followed by a test phase. For each block, seven sounds were randomly selected without replacement among the set of targets; during the study phase, they were played in a random order at their typical level, some at 55 and some at 70 dB SPL; during the test phase, the same seven target sounds were played also in a random order at both levels, 55 and 70 dB SPL. In addition, two sounds randomly selected among the set of distractors were also presented at 55 and 70 dB SPL during the test phase. The two distractors were new to all participants, as they were not presented before during the study phase. In summary, during the test phase, seven targets

and two distractors were played in a random order at both levels. This procedure was the same for the six blocks, and therefore, the 42 targets and the 12 distractors were presented throughout the six blocks. During the study phase within one block, the three groups of participants had to listen to the target sounds. Each group of participants received different instructions: the first group was instructed to listen to the sounds, and then to sort them in three different categories labeled liquid, machine, and solid (semantic condition); the second group was also asked to listen to the sounds and then to rate their loudness on a 9-point scale (sensory condition); finally, the third group was solely asked to listen to the sounds (control condition). The semantic and sensory conditions were designed to encourage participants to concentrate on two types of sound characteristics related to category judgments and loudness judgments, respectively. For each sound presented during the test phase, the three groups of participants were asked to decide as quickly and as accurately as possible whether it was presented during the study phase, i.e., if it was recognized; they had to press the spacebar of the keyboard for a positive answer, otherwise they were asked to do nothing. If the participant pressed the key after hearing a target sound—whatever the level was, same or different between the study and the test phases, the response was registered as a hit (Hit). If the participant pressed the key after hearing a distractor, the response was registered as a false alarm (FA). The next trial was 1 s after the end of the sound.

The test phase used a $3 \times [3 \times 2]$ factorial design; there was one between-subjects variable, encoding task (semantic, sensory, control), and two within-subjects variables: the two levels of the target sounds (same vs different), and the type of sounds (liquid, machine, solid). The experiment was completely counterbalanced such that each sound was randomly presented across blocks and participants. All six blocks lasted less than 40 min.

2.5 Results

For the study phase, one group of participants (11 participants) was instructed to sort the sounds into the three categories (semantic condition). The judgments are in agreement with the three categories: scores for the liquid, machine, and solid sounds are 93.5%, 92.9%, and 86.0%. Another group of participants (18 participants) was instructed to estimate the loudness of each sound on a 9-point scale (sensory condition). Ratings are consistent with the two values of the SPL (55 and 70 dB SPL, respectively) without any difference between the three categories of sounds: average ratings are 3.26 and 6.60 for the liquid sounds, 3.51 and 6.48 for the machine sounds, and 2.87 and 6.59 for the solid sounds. In both conditions, the results thus indicate that participants could readily discriminate sounds according to, on the one hand, the three categories of sounds (liquid, machine, and solid, respectively), and on the other hand, the two values of the SPL (55 and 70 dB SPL, respectively).

For the test phase, participants' d' were calculated over all trials. The distribution was almost entirely above zero, indicating a feasible task for the participants except for one ($d' < 0$) in the sensory condition. For this participant, hits and FAs are very low (8.3%) and high (96%), respectively, indicating that instructions were misunderstood. For this reason, this participant was excluded from the sensory group. Table 1 displays the proportion of the target sounds, presented in the study phase, given yes responses on the recognition test (i.e., Hits), as a function of encoding task (semantic, sensory, control) and SPL (same vs different). The effect of the study-to-test change in SPL (i.e., proportion of same minus proportion of different) on recognition accuracy is higher for the sensory condition (13%) and lower for the control condition (8%). Table 1 displays also the proportion of yes responses for the distractors (i.e., FAs), and thus, the sensitivity index d' for the three encoding tasks. This value provides a measure of the magnitude

Table 1. Proportion (%) of Hits and FAs on the recognition test as a function of SPL and the Encoding Task. S = Same; D = Different; M = Mean.

Encoding Task	SPL		M	FA	d'
	S	D			
Semantic (Listen + Category)	75	65	70	9	1.8
Sensory (Listen + Loudness)	64	51	57	6	1.7
Control (Listen)	70	62	66	15	1.4
Mean	70	59	66	10	

of the encoding task effect, which is slightly similar for the semantic (1.8) and the sensory (1.7) conditions, and is lower for the control condition (1.4).

A mixed-analysis of variance (ANOVA), 1-between and 2-within factors, was conducted on Hit scores. The main result is the significant effect of the study-to-test change in SPL [$F(1, 36) = 34.6, p < 0.001, \varepsilon = 1, \eta^2 = 0.49$], which means that the mean proportion of Hits for the same SPL is significantly higher (70% vs 59%). In other words, recognition is significantly better when target sounds are presented with the same level in the study phase and in the test phase. The analysis did not reveal any interaction with encoding tasks, which means that the significant effect of the study-to-test change in SPL is valid whatever are the encoding task and the type of sounds. Finally, the ANOVA reveals a small effect for the type of sounds [$F(2, 72) = 4.16, p < 0.05, \varepsilon = 0.99, \eta^2 = 0.1$]; a *post hoc t*-tests with Bonferroni corrections for repeating comparisons shows that only the difference between liquid and solid categories is significant (62% vs 68%, $p < 0.01$); Hit rates for the machine category is in between (63%).

3. Discussion

The present study has yielded a number of new experimental facts about explicit memory in the auditory domain. Especially, the influence of loudness on sound recognition based on an explicit memory experiment was evaluated. Results revealed that recognition scores are significantly different when sounds are presented with the same or different SPL between the study and the test phases; recognition is more accurate when there is no study-to-test change in SPL. In other words, recognition is significantly better when target sounds are presented with the same typical level in the study phase and in the test phase. “Typical” means here that the sounds were assigned to a specific SPL corresponding to their usual level in an everyday life context. What would happen if they were presented with an atypical level (i.e., unusual level in a real context) during the study phase? For example, what would happen if a low SPL were associated with the loud sounds. Would that change the result? Would recognition still be better for target sounds with the same level in the study phase and in the test phase even if it is an atypical level? If it were the case, it would mean that participants are able, within the study-to-test experimental procedure, to associate and memorize the sound level even for an atypical sound-level association. Similarly, what would happen if sounds were presented with their typical and atypical SPLs during the test phase without any study phase? In that case, participants’ judgment would be based on their daily experience with the sounds presented during the test phase. If the result would be the same as in the current study—recognition would be better when target sounds were presented with their typical level—it would reveal that the typical level related to an everyday life context is an important ecological parameter to take into account in a recognition experiment which has to be controlled as it was done in [Lemaitre *et al.* \(2010\)](#) and [Houix *et al.* \(2012\)](#). Those questions need to be investigated in a future study.

The main result in the present study—better recognition scores when there is no study-to-test change in SPL—was obtained for the two encoding tasks used in the present experiment (semantic or sensory), and also when participants were just required to listen to the sounds. Thus, our hypothesis that study-to-test change in SPL might be observed is confirmed, but it is not specific to the encoding task that focuses participants on loudness; although there was a numerical trend in the sensory condition, it did not approach statistical significance. Furthermore, the recognition task involved in the present study was based on explicit memory, thus a larger effect of semantic versus sensory encoding manipulations on recognition scores was expected; however, the obtained scores are similar to whatever the encoding task is, although there was a trend in favour of the semantic encoding task. It is worth noting that when participants were just required to listen to the sounds, without any specific task to perform, the recognition scores were lower. [Halpern and Müllensiefen \(2008\)](#) evaluated effects of timbre and tempo change on memory for music; results revealed that timbre and tempo change both impaired explicit memory. In addition, results suggested that the encoding task made no difference. These results go in the same direction as ours. In another study in the auditory domain, [Schacter and Church \(1992\)](#) manipulated speaker’s voice in a study-to-test procedure; the voice was the same or different between the encoding task and the auditory-word identification task. However, their results did not reveal any significant voice change effect following a semantic encoding task and a sensory encoding task. To explain their results, the authors suggested participants tend to rely on semantic strategies overriding the potential importance of voice characteristics in a recognition process, even after a task that focuses on voice characteristics. In the present study, a significant loudness change effect was obtained, but as in [Schacter and](#)

Church (1992), there was no significant difference between the semantic encoding task and the sensory one; the effect was not specific to the task that focused participants on loudness. Thus, it can be speculated that some sort of loudness encoding is part of the sound recognition process. It can be speculated also that recognition is impaired because timbre is affected by loudness modifications. In all cases, it appears that loudness is an important hint for everyday sounds recognition.

References and links

- Aldrich, K. M., Hellier, E. J., and Edworthy, J. (2009). "What determines auditory similarity? The effect of stimulus group and methodology," *Quarterly J. Exper. Psychol.* **62**, 63–83.
- Giordano, B. L., McDonnell, J., and McAdams, S. (2010). "Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds," *Brain Cognition.* **73**, 7–19.
- Graf, P., and Ryan, L. (1990). "Transfer-appropriate processing for implicit and explicit memory," *J. Exp. Psychol.* **16**(6), 978–992.
- Halpern, A. R., and Müllensiefen, D. (2008). "Effects of timbre and tempo change on memory for music," *Quarterly J. Exp. Psychol.* **61**(9), 1371–1384.
- Houix, O., Lemaitre, G., Misdariis, N., Susini, P., and Urdapilleta, I. (2012). "A lexical analysis of environmental sound categories," *J. Exp. Psychol.* **18**, 52–80.
- Lemaitre, G., Houix, O., Misdariis, N., and Susini, P. (2010). "Listener expertise and sound identification influence the categorization of environmental sounds," *J. Exp. Psychol.* **16**, 16–32.
- Maffiolo, V. (1999). "De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain: Structuration des représentations mentales et influence sur l'appréciation qualitative. Application aux ambiances sonores de Paris" ("The semantic and acoustic characterization of the sound quality of the urban environment: mental representations and influence on qualitative judgments. Application to Paris soundscapes"), Université du Maine, Le Mans, 1999.
- Melara, R. D., and Marks, L. E. (1990). "Interaction among auditory dimensions: Timbre, pitch, and loudness," *Percept. Psychophys.* **48**, 169–178.
- Schacter, D. L., and Church, B. A. (1992). "Auditory priming: Implicit and explicit memory for words and voices," *J. Exp. Psychol.* **18**(5), 915–930.
- Susini, P., Houix, O., and St Pierre, G. (2015). "The effect of loudness on the perceptual representation of sounds with similar timbre," *Acta Acust. Acust.* **101**(6), 1174–1184.