scientific reports

Check for updates

OPEN

Visual consequent stimulus complexity affects performance in audiovisual associative learning

Kálmán Tót[1], Gabriella Eördegh[2], Ádám Kiss[1], András Kelemen[3], Gábor Braunitzer[4], Szabolcs Kéri[1,4], Balázs Bodosi[1] & Attila Nagy[1]✉

In associative learning (AL), cues and/or outcome events are coupled together. AL is typically tested in visual learning paradigms. Recently, our group developed various AL tests based on the Rutgers Acquired Equivalence Test (RAET), both visual and audiovisual, keeping the structure and logic of RAET but with different stimuli. In this study, 55 volunteers were tested in two of our audiovisual tests, SoundFace (SF) and SoundPolygon (SP). The antecedent stimuli in both tests are sounds, and the consequent stimuli are images. The consequents in SF are cartoon faces, while in SP, they are simple geometric shapes. The aim was to test how the complexity of the applied consequent stimuli influences performance regarding the various aspects of learning the tests assess (stimulus pair learning, retrieval, and generalization of the previously learned associations to new but predictable stimulus pairs). In SP, behavioral performance was significantly poorer than in SF, and the reaction times were significantly longer, for all phases of the test. The results suggest that audiovisual associative learning is significantly influenced by the complexity of the consequent stimuli.

Associative learning is a basic cognitive function, in which different stimuli, cues, and/or outcome events are coupled. This learning type includes cognitive tasks like probabilistic learning[1,2], latent inhibition[3] and sensory preconditioning[4], and equivalence learning[5,6]. Equivalence learning is a specific kind of associative learning in which two discrete and often different percepts (antecedents) are linked together based on a shared outcome (consequent). A visually guided paradigm, the Rutgers Acquired Equivalence Test (RAET), was developed by Myers et al.[7] to investigate equivalence learning in humans. The test is computer-based and divided into two main phases: the acquisition and the test phases. In the acquisition phase, the subject's task is to associate two different visual stimuli based on feedback information about the correctness of the choices. This way, the rule of pairing is acquired. In the subsequent test phase, the subject must recall the already learned associations (retrieval) and build new, hitherto not seen but predictable associations (generalization or transfer). Regarding the neural correlates, both the original study of the Myers group[7] and subsequent investigations[8–13] demonstrated that the acquisition phase is linked to the fronto-striatal loops, while the test phase is linked primarily to the hippocampi and the medial temporal lobe[7,14–19]. The basal ganglia and the hippocampi are structures of key importance in equivalence learning, and they are also involved in multisensory processing[20–23].

Several studies reported that stimulus complexity influences auditory guided associative learning and more complex stimuli cause better responses and greater cortical activation[24–26]. It has also been demonstrated in studies from the cellular to the behavioral level that responses are quicker and more precise in the case of multimodal stimuli[22,27–30]. A recent study by our research group[31], applying a modified but structurally identical version of RAET, showed that the complexity of the applied visual stimuli could also strongly influence the efficiency of associative learning: simple visual stimuli (antecedents: white, light gray, dark gray and black circles; consequents: colorless triangle, square, rhombus, and concave deltoid) with restricted semantic and color information allowed significantly poorer equivalence acquisition than more complex stimuli (antecedents: cartoon faces of a woman, a man, a boy and a girl; consequents: green, yellow, red and blue fish) without such feature restrictions. However, stimulus complexity did not affect retrieval and generalization (transfer).

Given that the key neural structures associated with RAET also play a role in multisensory processing, the question arises whether the complexity of the applied visual stimuli can also influence the effectiveness of multisensory (audiovisually guided) associative learning. In this study, we sought to answer this question. For this,

[1]Department of Physiology, Faculty of Medicine, University of Szeged, Dóm Tér 10, Szeged 6720, Hungary. [2]Faculty of Health Sciences and Social Studies, University of Szeged, Szeged, Hungary. [3]Department of Applied Informatics, University of Szeged, Szeged, Hungary. [4]Nyírő Gyula National Institute of Psychiatry and Addictions, Budapest, Hungary. ✉email: nagy.attila.1@med.u-szeged.hu

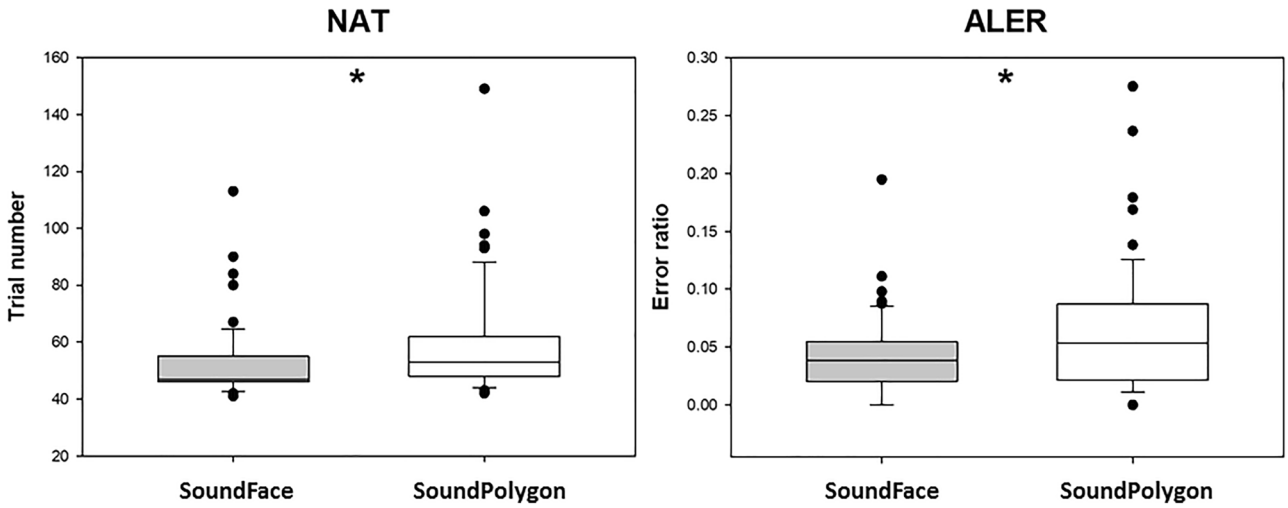

Figure 1. Performance in the acquisition phase in the two tests. NAT: the number of trials needed to complete the acquisition phase. ALER: error ratios in the acquisition phase. The lower margin of the boxes indicates the 25th percentile, the upper margin the 75th percentile, while the line within the boxes marks the median. The error bars (whiskers) above and below the boxes are the 90th and 10th percentiles, respectively. The dots over and under the whiskers represent the extreme outliers. Asterisk (*) indicates a significant difference at the level $p < 0.05$.

we used two audiovisual tests, both of which follow the RAET paradigm: SoundFace (SF) and SoundPolygon (SP). These tests have been developed in our laboratory. Both tests use sounds as antecedent stimuli, but SF uses cartoon faces and SP uses simple geometric shapes as consequents. That is, the consequent stimuli in SF (colored cartoon faces of a boy, a girl, a man and a woman) are relatively complex in the sense that they have well-defined, readily detectable and readily verbalizable distinctive features (e.g., colors, gender, age), while the consequents in SP lack such features. By comparing our volunteers' performance on these tests, we sought to test if consequent stimulus complexity influences audiovisual associative learning at all. It is important to note that the goal of this study was not to analyze how the gradual extraction of the different features of the consequent stimuli influences audiovisual associative learning (and to learn this way which features are more important, and which are less important for audiovisual associative learning). Instead, we chose to use the cartoon faces of RAET and a set of completely different, simple, non-face stimuli that lack all the specific features associated with the cartoon faces merely to establish the lack or existence of an effect.

Results

All 55 volunteers completed both tests. The analysis of their performance data is presented according to the two main phases of the test paradigm (acquisition and test).

Acquisition phase. The median number of trials in the acquisition phase (NAT) in SF was 47 (range: 41–113), and in SP, it was 53 (range: 42–149). The participants needed significantly more trials to learn the associations in SP ($Z = 2.417$, $p = 0.016$) (Fig. 1).

In SF, the median of error ratios in the acquisition (ALER) was 0.038 (range: 0.00–0.19), and in SP, it was 0.058 (range: 0.00–0.28). The difference between the two tests was significant ($Z = 2.213$, $p = 0.027$) (Fig. 1). The median reaction time (RT) for the acquisition trials in SF was 1611.619 ms (range: 1095.022–4016.42), and in SP, it was 1834.810 ms (range: 1204.867–4762.33). The difference was significant ($Z = 3.703$, $p = 0.0002$).

Test phase. The median of retrieval error ratio (RER) in SF was 0.00 (range: 0.00–0.25), and in SP, it was 0.028 (range: 0.00–0.39). The difference was significant ($Z = 2.727$, $p = 0.0064$) (Fig. 2). As for the reaction times, the median RT for the retrieval trials in SF was 1586.39 ms (range: 980.056–3802.11), and in SP, it was 2000.42 ms (range: 1222.33–4790.44). The difference was significant ($Z = 4.994$, $p = 0.000001$).

The median of generalization error ratio (GER) in SF was 0.00 (range: 0.00–0.58), and in SP, it was 0.083 (range: 0.00–1.00). The difference was significant ($Z = 3.085$, $p = 0.002$) (Fig. 2). The median RT for the generalization trials in SF was 1769.17 ms (range: 1145.75–5722.83), and in SP, it was 2544.25 ms (range: 1282.58–18,381.00). The difference was significant ($Z = 3.938$, $p = 0.00008$).

Discussion

To our knowledge, this is the first study to investigate the effect of visual stimulus complexity on multisensory (audiovisual) guided associative learning. The same set of auditory stimuli were applied with two different series of visual stimuli in two tests based on the same paradigm. The two sets of visual stimuli differed in their feature richness and complexity. Multisensory guided equivalence learning and subsequent retrieval and generalization

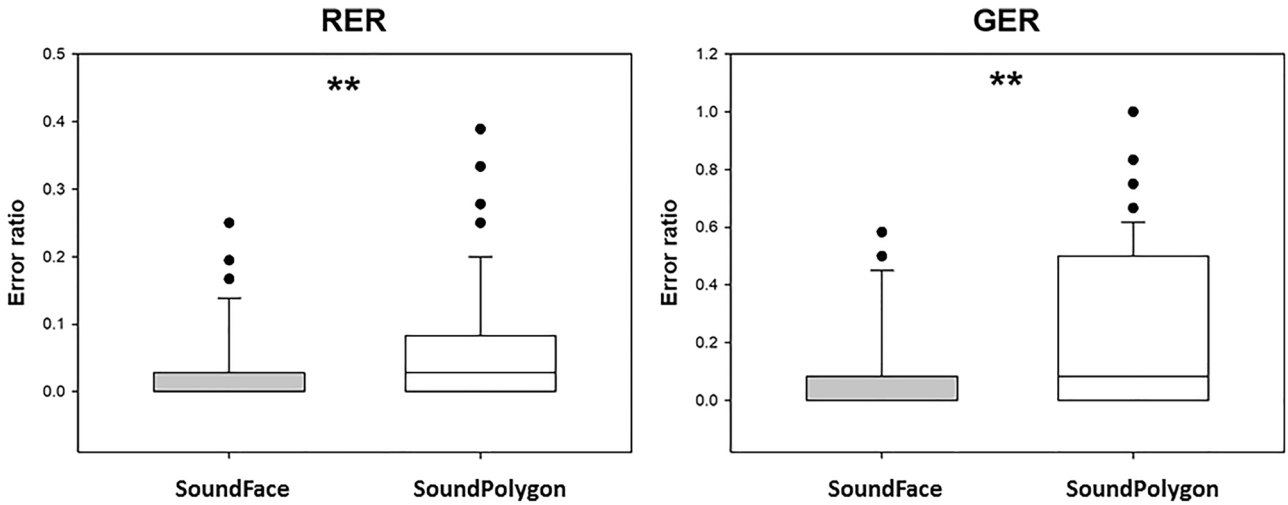

Figure 2. Performance in the test phase in the two tests. RER: retrieval error ratio, GER: generalization error ratio. Asterisk (*) indicates significant difference at the level $p < 0.01$. Otherwise, the conventions are the same as in Fig. 1.

were all influenced markedly by the complexity of the visual stimuli. The difference also showed in significantly shorter reaction times when the more complex, feature-rich visual stimuli were used.

In this study, we used two audiovisual tests that were developed in our laboratory, based on RAET, a visually guided equivalence learning test designed by Catherine E. Myers and colleagues at Rutgers University[7]. The original paradigm tests visually guided pair learning, the retrieval of the already learned stimulus pairs and the ability to apply the previously learned associations to build new stimulus pairs. The key brain structures associated with this task (the hippocampi and the basal ganglia) are also known for their role in multisensory processing[22,23,32,33]. Therefore, we developed SF that uses cartoon faces as consequents with auditory antecedents[34]. SF was administered to healthy adult subjects and in psychiatric patient populations and the results were compared to those of the original, visually guided RAET test[13,34]. The comparison indicated that the fact alone that the task had become multisensory did not influence the volunteers' performance to a significant degree, which led us to the conclusion that multimodality itself does not interfere with the efficiency of associative learning, retrieval, or generalization.

The visual stimuli both in the visual RAET and the audiovisual SoundFace are complex, colored stimuli with the potential to evoke associations and emotional responses, which, in turn, can serve as extra clues that recruit various cortical areas to enhance performance[35–37]. Such clues can thus mask the contribution of subcortical structures. We developed a new visually guided test, Polygon[31] to reduce this effect. Polygon uses simple geometric shapes both as antecedents and consequents. Such simple shapes are relatively meaningless in themselves, and they can hardly evoke emotions. Therefore, we hypothesized that the use of geometric shapes would allow us to minimize cortical contributions to task performance and thus allow a better assessment of subcortical contributions. Indeed, the first study with Polygon[31] revealed a specific pattern: it took significantly more trials for the volunteers to learn the stimulus pairs (and they made significantly more mistakes), but the reduced complexity of the stimuli had no significant effect on either the retrieval or the generalization part of the test phase.

The next logical step was to investigate what effect visual stimulus complexity might have on multisensory guided (audiovisual) equivalence learning. For this purpose, we combined the antecedent sounds of SF[31] and the geometric shapes of Polygon[31] into a new test (SP) and compared volunteers' performance on this test to their performance on SF. This comparison is presented in this study. In contrast to what was found when tests using visual stimuli only were compared (RAET vs. Polygon)[31], in the case of these multisensory tests, decreased stimulus complexity affected not only the acquisition phase, but the entire test phase, including retrieval and generalization. That is, it seems that when only visual stimuli are used, decreased stimulus complexity makes learning difficult, but if learning has been successful, retrieval and generalization are spared. Such a sparing does not seem to occur when visual stimulus complexity is decreased in an audiovisual (multimodal or multisensory) learning environment. While it comes as no surprise (in fact, it is somewhat intuitive) that stimulus complexity influences the efficiency of associative learning[24–26], it is difficult to tell why decreased stimulus complexity affects performance in all phases of the audiovisual version of the test paradigm, while in the visual version, only acquisition is affected.

In an earlier study[38], based on developmental data, we argued that learning and memory in this specific paradigm might be best described by the integrative encoding account of associative learning[39,40]. This account concentrates on two specific neural loops, the substantia nigra (SN)- striatum loop and the ventral tegmentum (VT)-hippocampus loop, which can be activated in parallel. While the SN- striatum loop supports primarily the voluntary learning of stimulus pairs with the help of feedback, the VT- hippocampus loop transfers information to the hippocampi, where a network of all encountered stimuli is constructed, with their connections and overlaps[41–46]. Then, this hippocampal network is activated in the test phase, which makes both retrieval and generalization possible. Based on this account, it is possible that the hippocampi, even if they are typically

discussed as structures of key importance in explicit memory[39,47], can support implicit functions as well. That is why, as we earlier argued[38,46], children can generalize at a high level with poor acquisition and retrieval. In other words, they have the information, and they can use it as long as no conscious effort is involved. It must be noted that the integrative encoding account was developed based on visual (non-multisensory) learning paradigms, and we have no information whatsoever if it can be applied to multisensory learning as well. Based on the information available at this point, it might be hypothesized that decreased stimulus complexity affects all phases of the multisensory test (but not the visual test, as demonstrated earlier) because hippocampal compensation is either specific to the visual stimulus modality or it works only if stimuli of the same modality are used. This, however, is only a crude hypothesis, which is made by inference from the literature.

At the same time, it must be also noted that a direct comparison between SF/SP and the purely visual version of RAET is not possible as the latter uses colored cartoon fish as consequent stimuli. While it is not entirely obvious how this could contribute to the observed difference, the confounding effect of this methodological factor cannot be ruled out. Another possible explanation is that in SF, equivalence might be established between the face consequents too, based on their various features, which makes learning easier in SF than in SP, where consequents do not share such readily detectable features. However, assuming that the integrative encoding account[39,40] (see above) can be applied in an audiovisual context too, this should affect only the acquisition phase. It may be that the lack of equivalence between the consequents in SP can explain poorer acquisition, but it does not seem to be a good explanation for poorer transfer. In the sense of the integrative encoding account, the fact that the stimulus pairs are more difficult to learn (which shows as more errors and a longer acquisition phase in RAET and its various versions) does not necessarily imply poor transfer. This is exactly what we saw when we administered the original version of RAET to small children.[46].

Beside the poorer performance in all phases, reaction times were also significantly longer in SP. This is consistent with earlier findings, where, in a face-name association task, subjects' reaction times were significantly shorter in response to more complex, high-salience colored faces expressing emotions than to less distinctive, grayscale face stimuli.[48] This shows that complexity facilitates decision making, while in the (relative) lack of distinctive features, the facilitating effect is absent.

As for the limitations, we would like to point out the following.

First, this study is best understood as an exploratory study that sought to establish if the complexity of visual consequent stimuli has any effect on subjects' performance in an audiovisual associative learning paradigm. By complexity, we simply meant how rich the applied stimuli were in well-described, readily identifiable (and possibly verbalizable) features that can be used as cues for learning. The cartoon faces are relatively rich in these: age, gender, hair color and facial expression are all such cues. In contrast, the polygons are colorless and they definitely do not have age, gender or any feature that is even close to a facial expression. This is a crude comparison, and it does not allow a finer analysis of the difference; all it allows is the conclusion that the presence or lack of such easily identifiable and obvious features does make a difference. Whether this is because these cues are easy to verbalize or because they are characteristically human features (which activate additional neural circuits) or for some other reason should be addressed in studies designed for that purpose. A logical next step would be to generate several sets of the cartoon face consequents with gradually decreasing complexity (cue content) and repeat the measurements with all the sets.

Second, it is a limitation of SP sepcifically that it is possible that the female voice and the female face are matched, which could provide an extra cue for learning. While it is not always the case (the stimulus pairings are randomly generated for each session), this could interfere with the results in some cases, even if not to a major extent.

In summary, in this study we have demonstrated that in a multisensory associative learning paradigm, where the antecedent stimuli are sounds, the complexity of the consequent visual stimuli has a significant effect on both learning and generalization.

Methods

Subjects. Fifty-five healthy adult volunteers participated in the presented study (27 females and 28 males, age mean: 31.36 ± 14.56 years, range: 18–69; five participants were over the age of 60). The estimated minimum sample size was 47, assuming $p < 0.05$, $1 - \beta = 0.95$ and an effect size of 0.5. The sample size estimation was performed in G*Power 3.1.9.2 (Düsseldorf, Germany). The volunteers received no compensation and were free to quit without any negative consequence. The volunteers were informed about the study's background, goals, and procedures and gave their written informed consent. Any psychiatric, neurological, otological or ophthalmological condition that could interfere with the participant's performance was an exclusion criterion. Before each testing session, the participants were shown the stimuli of the tests one by one (each stimulus once) to make sure that they could see and hear them correctly. The study protocol conformed to the Declaration of Helsinki in all respects and was approved by the Regional Research Ethics Committee for Medical Research at the University of Szeged, Hungary (27/2020-SZTE).

The applied multisensory tests. Two audiovisual tests of our own development were administered (SoundFace and SoundPolygon, see below). Both tests were run on laptops (Lenovo ThinkBook 15-IIL, Lenovo, China), and the auditory stimuli were administered through over-ear headphones (Sennheiser HD439, Sennheiser, Germany). The volunteers were tested one-by-one in a quiet room, sitting at a comfortable distance (57 cm) from the screen. No forced quick responses were expected to avoid performance anxiety, but the participants were instructed to respond as quickly as possible. This way, explicit time pressure could be avoided, yet, the participants were aware that it was desirable that they spent a limited amount of time with each trial. The

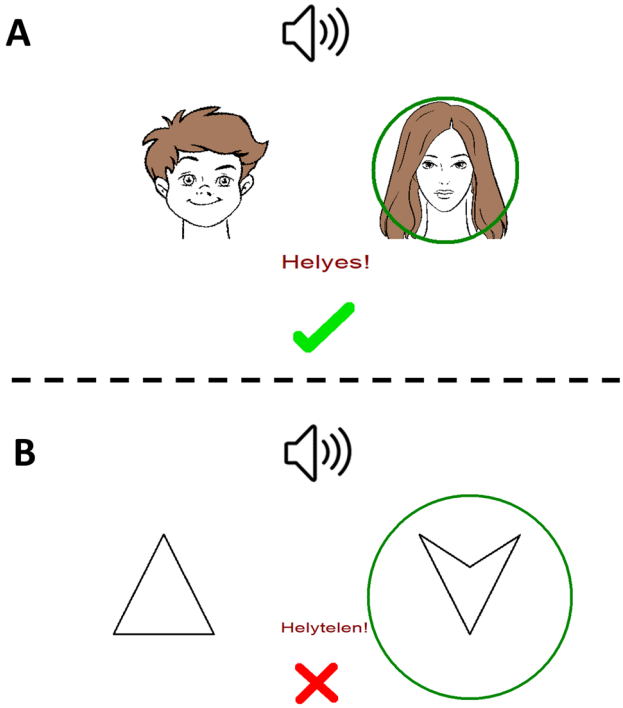

Figure 3. A trial in the acquisition phase of SoundFace (**A**) and SoundPolygon (**B**). In each trial, the subject simultaneously hears a sound (antecedent) and sees two faces (consequents) on the left and right side of the screen. Then the subject guesses which consequent belongs to the given antecedent sound by pressing the "left" or "right" button. The choice is indicated with the green circle. Immediate visual feedback is given. A green checkmark with the word Helyes! (Correct!) indicates a correct guess and a red X with the word Helytelen! (Incorrect!) indicates an incorrect guess.

volunteers completed both tests immediately one after another, in a pseudorandom order to avoid the carry-over effect.

SoundFace is based on RAET as described in Myers et al.[7,34]. The structure of RAET was kept, but it was translated into the Hungarian language and transformed into an audiovisual paradigm in Assembly for Windows. These modifications were performed with the written permission of Catherine E. Myers. In SoundFace, the subject is first asked to learn associations through trial and error. There are four sounds as antecedents and four possible faces as consequents. The antecedent stimuli are different and clearly distinguishable sounds: a cat (A1), a guitar note (A2), the sound of a vehicle (B1) and a woman's voice (B2). The consequents are different cartoon faces: an adult male (X1), a girl (X2), an adult woman (Y1), and a boy (Y2). The auditory and visual stimuli were semantically incongruent (except for the case when a woman's voice is matched with a woman's face, but this is not always the case). In each trial, the subject simultaneously hears a sound and sees two faces on the left and right sides of the screen. The subject is instructed to guess which face belongs to the given sound and indicate his or her guess by pressing either the „left" or the „right" button. The duration of the auditory stimulus was consistently 1.5 s. The visual stimuli lasted until the participant made the decision with the pressing of the "left" or "right" button. In this respect, SoundFace and SoundPolygon are identical (Fig. 3). The pairs are randomly generated by the software for each subject.

The paradigm is divided into two main phases: the acquisition and test phases. In the acquisition phase, visual feedback was given about the correctness of the choice. In the initial part of the acquisition phase, the subject learns through trial-and-error that if sounds A1 or A2 are presented, the correct response is to choose face X1 over Y1. Similarly, if sounds B1 or B2 are presented, the correct response is to choose face Y1 over X1. This way, it is learned that in terms of their consequents, A1 = A2 and B1 = B2. Once this has been established, new stimulus pairs are added. This time, the subject learns that if sound A1 is presented, the correct response is to choose X2 over Y2, and if sound B1 is presented, the correct response is to choose Y2 over X2. This way, antecedents A1 and B1 gain additional consequents. At this point, the subject knows that A1: X1, X2 and B1:Y1, Y2. Six items are presented in the acquisition phase from the eight possible stimulus pairs. A2:X2 and B2:Y2 are not presented, but it is implied by the connection A1 = A2 and B1 = B2. After each newly introduced stimulus pair, the participant must give a certain number of subsequent correct answers (4, 6, 8, 10, 12 after each new association, respectively) to accomplish the acquisition phase. Because of this, the number of trials in this phase is not constant, and it depends on how efficiently the given individual learns.

Once having completed the acquisition phase, the participant continues with the test phase, where feedback is no longer given about the correctness of the responses. In this phase, the retrieval and generalization are tested. Retrieval refers to the recall of the already known (learned) stimulus pairs, while generalization refers to

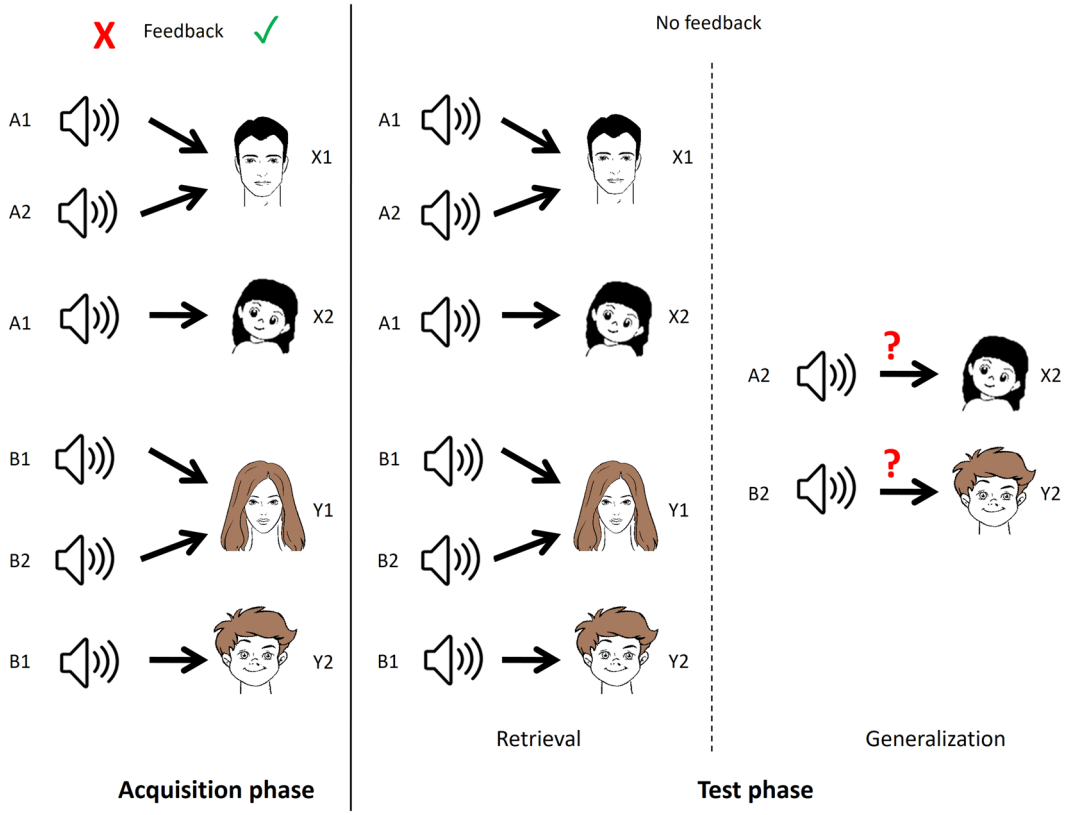

Figure 4. Overview of the structure of the SoundFace test. The antecedent stimuli are sounds of a cat (A1), a guitar note (A2), the sound of a vehicle (B1), and a woman's voice (B2). The consequents are cartoon faces of a man (X1), a girl (X2), a woman (Y1), and a boy (Y2).

Acquisition			Test	
Shaping	**Equivalence training**	**New consequents**	**Retrieval**	**Generalization**
A1—>X1	A1—>X1	A1—>X1	A1—>X1	
	A2—>X1	A2—>X1	A2—>X1	
		A1—>X2	A1—>X2	
				A2—>X2
B1—>Y1	B1—>Y1	B1—>Y1	B1—>Y1	
	B2—>Y1	B2—>Y1	B2—>Y1	
		B1—>Y2	B1—>Y2	
				B2—>Y2

Table 1. A summary of the audiovisual associative learning paradigms. A,B: antecedents (the same sounds in both tests), X,Y: consequents (faces in SoundFace and simple geometric forms in SoundPolygon tests).

making the A2:X2 and B2:Y2 stimulus pairs not presented in the acquisition phase but implied by the connection A1 = A2 and B1 = B2. If the subject has successfully acquired the said associations, he or she will choose X2 when A2 is presented and Y2 when B2 is presented, even if he or she has not seen these pairs before. The subject is not informed that new stimulus pairs are to be expected in the test phase. The number of trials in the test phase is constant. There are altogether 48 trials, of which 36 are retrieval trials, and 12 are generalization trials. These are mixed in random order. The overview of the paradigm is given in Fig. 4. See also Table 1 for clarification.

SoundPolygon has the same structure as SoundFace, but with simplified visual stimuli. Instead of cartoon faces, simple geometric shapes are used as consequents: a triangle (X1), a square (X2), a rhombus (Y1), and a concave deltoid (Y2). The auditory stimuli are the same as in SoundFace. Figure 5 summarizes SoundPolygon.

Data analysis. The performance of the participants was characterized by four main parameters: the number of trials necessary for the completion of the acquisition phase (NAT), association learning error ratio (the ratio of incorrect choices during the acquisition trials, ALER), retrieval error ratio (RER), and generalization error

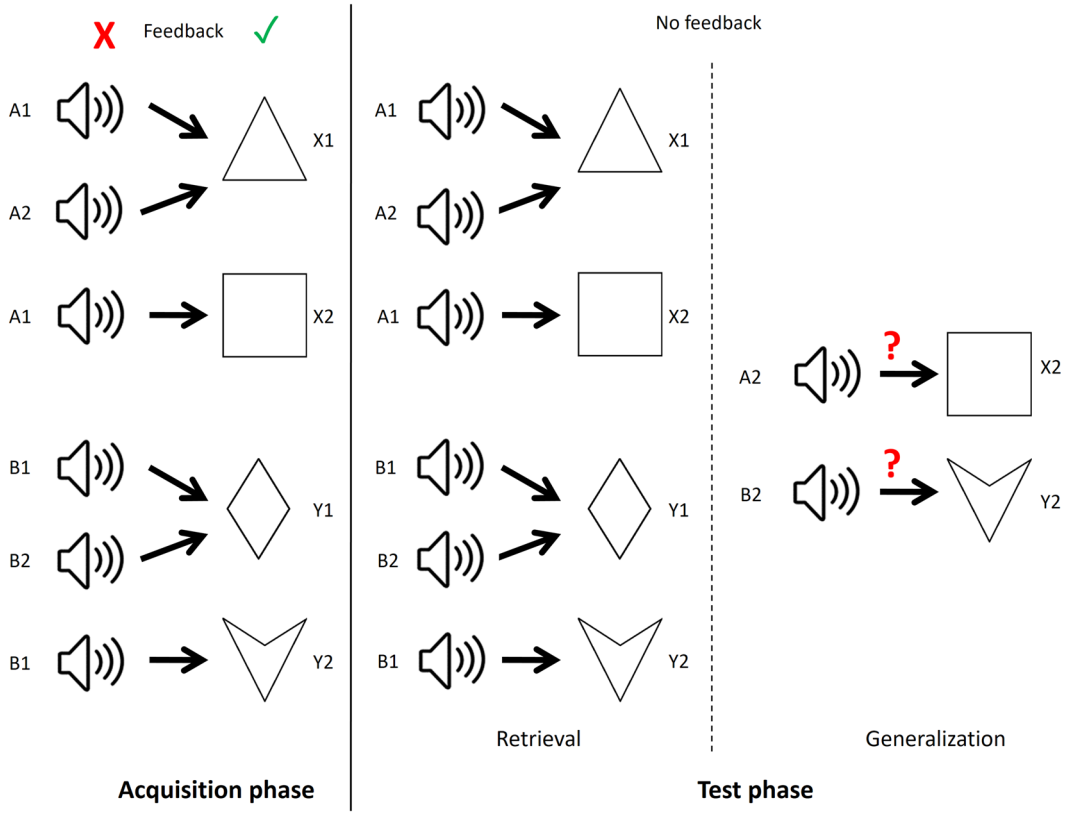

Figure 5. Overview of the structure of the SoundPolygon test. The consequents are simple geometric shapes: a triangle (X1), a square (X2), a rhombus (Y1), and a concave deltoid (Y2).

ratio (GER). NAT and ALER are performance parameters of the acquisition phase. RER and GER are performance parameters of the test phase. Error ratios were calculated by dividing the number of incorrect responses by the total number of trials. Reaction times were recorded for each trial, and they were analyzed for the acquisition, retrieval, and generalization trials separately. Reaction time was defined as the time elapsed between the appearance of the stimuli and the subject's response. Only RTs of the correct choices were included, and values over 3SD were excluded.

Statistical analysis was performed in Statistica 13.4.0.14 (TIBCO Software Inc., USA). NAT, ALER, RER, and GER were compared between the two paradigms. As the data were non-normally distributed (Shapiro–Wilk $p < 0.05$), the Wilcoxon matched-pairs test was used for the hypothesis tests.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 3 June 2022; Accepted: 20 October 2022
Published online: 22 October 2022

References

1. Shohamy, D., Myers, C. E., Hopkins, R. O., Sage, J. & Gluck, M. A. Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *J. Cogn. Neurosci.* **21**, 1821–1833. https://doi.org/10.1162/jocn.2009.21138 (2009).
2. Sáringer, S., Fehér, Á., Sáry, G. & Kaposvári, P. Online measurement of learning temporal statistical structure in categorization tasks. *Mem. Cognit.* https://doi.org/10.3758/s13421-022-01302-5 (2022).
3. Weiss, K. R. & Brown, B. L. Latent inhibition: A review and a new hypothesis. *Acta Neurobiol. Exp. (Wars)* **34**, 301–316 (1974).
4. Rescorla, R. A. Simultaneous and successive associations in sensory preconditioning. *J. Exp. Psychol. Anim. Behav. Process* **6**, 207–216. https://doi.org/10.1037//0097-7403.6.3.207 (1980).
5. Ward-Robinson, J. & Hall, G. The role of mediated conditioning in acquired equivalence. *Q. J. Exp. Psychol. B* **52**, 335–350. https://doi.org/10.1080/027249999393031 (1999).
6. Molet, M., Miller, H. & Zentall, T. R. Acquired equivalence of cues by presentation in a common context in rats. *Anim. Cogn.* **15**, 143–147. https://doi.org/10.1007/s10071-011-0431-4 (2012).
7. Myers, C. E. *et al.* Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J. Cogn. Neurosci.* **15**, 185–193. https://doi.org/10.1162/089892903321208123 (2003).
8. Coutureau, E. *et al.* Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. *J. Exp. Psychol. Anim. Behav. Process* **28**, 388–396 (2002).
9. Bódi, N., Csibri, E., Myers, C. E., Gluck, M. A. & Kéri, S. Associative learning, acquired equivalence, and flexible generalization of knowledge in mild Alzheimer disease. *Cogn. Behav. Neurol.* **22**, 89–94. https://doi.org/10.1097/WNN.0b013e318192ccf0 (2009).

10. Myers, C. E. *et al.* Learning and generalization deficits in patients with memory impairments due to anterior communicating artery aneurysm rupture or hypoxic brain injury. *Neuropsychology* **22**, 681–686. https://doi.org/10.1037/0894-4105.22.5.681 (2008).
11. Kéri, S., Nagy, O., Kelemen, O., Myers, C. E. & Gluck, M. A. Dissociation between medial temporal lobe and basal ganglia memory systems in schizophrenia. *Schizophr. Res.* **77**, 321–328. https://doi.org/10.1016/j.schres.2005.03.024 (2005).
12. Eördegh, G. *et al.* Impairment of visually guided associative learning in children with Tourette syndrome. *PLoS ONE* **15**, e0234724. https://doi.org/10.1371/journal.pone.0234724 (2020).
13. Pertich, Á. *et al.* Maintained visual-, auditory-, and multisensory-guided associative learning functions in children with obsessive-compulsive disorder. *Front. Psychiatry* **11**, 571053. https://doi.org/10.3389/fpsyt.2020.571053 (2020).
14. Cohen, N. J. *et al.* Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus* **9**, 83–98. https://doi.org/10.1002/(SICI)1098-1063(1999)9:1%3c83::AID-HIPO9%3e3.0.CO;2-7 (1999).
15. Gogtay, N. *et al.* Dynamic mapping of normal human hippocampal development. *Hippocampus* **16**, 664–672. https://doi.org/10.1002/hipo.20193 (2006).
16. Larsen, B. & Luna, B. In vivo evidence of neurophysiological maturation of the human adolescent striatum. *Dev. Cogn. Neurosci.* **12**, 74–85. https://doi.org/10.1016/j.dcn.2014.12.003 (2015).
17. Moustafa, A. A., Myers, C. E. & Gluck, M. A. A neurocomputational model of classical conditioning phenomena: A putative role for the hippocampal region in associative learning. *Brain Res.* **1276**, 180–195. https://doi.org/10.1016/j.brainres.2009.04.020 (2009).
18. Persson, J. *et al.* Sex differences in volume and structural covariance of the anterior and posterior hippocampus. *Neuroimage* **99**, 215–225. https://doi.org/10.1016/j.neuroimage.2014.05.038 (2014).
19 Porter, J. N. *et al.* Age-related changes in the intrinsic functional connectivity of the human ventral vs. dorsal striatum from childhood to middle age. *Dev. Cogn. Neurosci.* **11**, 83–95. https://doi.org/10.1016/j.dcn.2014.08.011 (2015).
20. Chudler, E. H., Sugiyama, K. & Dong, W. K. Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Res.* **674**, 33–45. https://doi.org/10.1016/0006-8993(94)01427-j (1995).
21. Schwarz, M., Sontag, K. H. & Wand, P. Sensory-motor processing in substantia nigra pars reticulata in conscious cats. *J. Physiol.* **347**, 129–147. https://doi.org/10.1113/jphysiol.1984.sp015057 (1984).
22. Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z. & Benedek, G. Multisensory integration in the basal ganglia. *Eur. J. Neurosci.* **24**, 917–924. https://doi.org/10.1111/j.1460-9568.2006.04942.x (2006).
23. Bates, S. L. & Wolbers, T. How cognitive aging affects multisensory integration of navigational cues. *Neurobiol. Aging* **35**, 2761–2769. https://doi.org/10.1016/j.neurobiolaging.2014.04.003 (2014).
24. Güçlütürk, Y., Güçlü, U., van Gerven, M. & van Lier, R. Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. *Sci. Rep.* **8**, 3439. https://doi.org/10.1038/s41598-018-21636-y (2018).
25. Staib, M. & Bach, D. R. Stimulus-invariant auditory cortex threat encoding during fear conditioning with simple and complex sounds. *Neuroimage* **166**, 276–284. https://doi.org/10.1016/j.neuroimage.2017.11.009 (2018).
26. Maor, I. *et al.* Neural correlates of learning pure tones or natural sounds in the auditory cortex. *Front. Neural Circuits* **13**, 82. https://doi.org/10.3389/fncir.2019.00082 (2019).
27. Frens, M. A. & Van Opstal, A. J. Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res. Bull.* **46**, 211–224. https://doi.org/10.1016/s0361-9230(98)00007-0 (1998).
28. Harrington, L. K. & Peck, C. K. Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp. Brain Res.* **122**, 247–252. https://doi.org/10.1007/s002210050512 (1998).
29. Giard, M. H. & Peronnet, F. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cogn. Neurosci.* **11**, 473–490. https://doi.org/10.1162/089892999563544 (1999).
30. Patching, G. R. & Quinlan, P. T. Cross-modal integration of simple auditory and visual events. *Percept. Psychophys.* **66**, 131–140. https://doi.org/10.3758/bf03194867 (2004).
31. Eördegh, G. *et al.* The influence of stimulus complexity on the effectiveness of visual associative learning. *Neuroscience* **487**, 26–34. https://doi.org/10.1016/j.neuroscience.2022.01.022 (2022).
32. Nagy, A., Paróczy, Z., Norita, M. & Benedek, G. Multisensory responses and receptive field properties of neurons in the substantia nigra and in the caudate nucleus. *Eur. J. Neurosci.* **22**, 419–424. https://doi.org/10.1111/j.1460-9568.2005.04211.x (2005).
33. Ravassard, P. *et al.* Multisensory control of hippocampal spatiotemporal selectivity. *Science* **340**, 1342–1346. https://doi.org/10.1126/science.1232655 (2013).
34. Eördegh, G. *et al.* Multisensory guided associative learning in healthy humans. *PLoS ONE* **14**, e0213094. https://doi.org/10.1371/journal.pone.0213094 (2019).
35. Puszta, A. *et al.* Cortical power-density changes of different frequency bands in visually guided associative learning: A human EEG-study. *Front. Hum. Neurosci.* **12**, 188. https://doi.org/10.3389/fnhum.2018.00188 (2018).
36. Puszta, A. *et al.* Power-spectra and cross-frequency coupling changes in visual and audio-visual acquired equivalence learning. *Sci. Rep.* **9**, 9444. https://doi.org/10.1038/s41598-019-45978-3 (2019).
37. Puszta, A. *et al.* Predicting stimulus modality and working memory load during visual- and audiovisual-acquired equivalence learning. *Front. Hum. Neurosci.* **14**, 569142. https://doi.org/10.3389/fnhum.2020.569142 (2020).
38. Öze, A. *et al.* Acquired equivalence and related memory processes in migraine without aura. *Cephalalgia* **37**, 532–540. https://doi.org/10.1177/0333102416651286 (2017).
39. Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41–50. https://doi.org/10.1038/35036213 (2000).
40. Shohamy, D. & Wagner, A. D. Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378–389. https://doi.org/10.1016/j.neuron.2008.09.023 (2008).
41. Daniel, R. & Pollmann, S. A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiol. Learn. Mem.* **114**, 90–100. https://doi.org/10.1016/j.nlm.2014.05.002 (2014).
42. Hiebert, N. M. *et al.* Striatum in stimulus-response learning via feedback and in decision making. *Neuroimage* **101**, 448–457. https://doi.org/10.1016/j.neuroimage.2014.07.013 (2014).
43. Balleine, B. W., Delgado, M. R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* **27**, 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007 (2007).
44. Scimeca, J. M. & Badre, D. Striatal contributions to declarative memory retrieval. *Neuron* **75**, 380–392. https://doi.org/10.1016/j.neuron.2012.07.014 (2012).
45. Delgado, M. R., Miller, M. M., Inati, S. & Phelps, E. A. An fMRI study of reward-related probability learning. *Neuroimage* **24**, 862–873. https://doi.org/10.1016/j.neuroimage.2004.10.002 (2005).
46. Braunitzer, G. *et al.* The development of acquired equivalence from childhood to adulthood—A cross-sectional study of 265 subjects. *PLoS ONE* **12**, e0179525. https://doi.org/10.1371/journal.pone.0179525 (2017).
47. Tulving, E. & Markowitsch, H. J. Episodic and declarative memory: Role of the hippocampus. *Hippocampus* **8**, 198–204. https://doi.org/10.1002/(SICI)1098-1063(1998)8:3%3c198::AID-HIPO2%3e3.0.CO;2-G (1998).
48. Bender, A. R., Naveh-Benjamin, M., Amann, K. & Raz, N. The role of stimulus complexity and salience in memory for face-name associations in healthy adults: Friend or foe?. *Psychol. Aging* **32**, 489–505. https://doi.org/10.1037/pag0000185 (2017).

Acknowledgements
The authors thank Anna Lazsádi and Ábel Hertelendy for their help in conducting the investigation and data collection. This work was supported by a grant from SZTE ÁOK-KKA Grant No. 2019/270-62-2. KT and ÁK were supported by the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the Source of the National Research, Development and Innovation Fund. GE was supported by EFOP-3.6.1-16-2016-00008 grant.

Author contributions
A.N., Sz.K., G.E. G.B., conceived the study conception and design. Data collection was made by K.T., G.E., Á.K. and B.B. Data analysis were performed by K.T., A.N., Á.K. and A.K. The manuscript was written by A.N., G.B. and K.T. All authors discussed data analysis and interpretation. All authors reviewed/edited the manuscript and approved the final version. Funding acquisition was made by A.N.

Funding
Open access funding provided by University of Szeged.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22880-z.

Correspondence and requests for materials should be addressed to A.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022