Consciousness and Cognition 77 (2020) 102862

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Pupillometric decoding of high-level musical imagery

Olivia Kang*, Mahzarin R. Banaji

Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138, United States

ARTICLE INFO

Keywords:
Pupillometry
Pupil dilation
Mental imagery
Auditory imagination
Music
Norepinephrine
LC-NE system
Attention

ABSTRACT

Humans report imagining sound where no physical sound is present: we replay conversations, practice speeches, and "hear" music all within the confines of our minds. Research has identified neural substrates underlying auditory imagery; yet deciphering its explicit contents has been elusive. Here we present a novel pupillometric method for decoding what individuals hear "inside their heads". Independent of light, pupils dilate and constrict in response to noradrenergic activity. Hence, stimuli evoking unique and reliable patterns of attention and arousal even when imagined should concurrently produce identifiable patterns of pupil-size dynamics (PSDs). Participants listened to and then silently imagined music while eye-tracked. Using machine learning algorithms, we decoded the imagined songs within- and across-participants following classifier-training on PSDs collected during both imagination and perception. Echoing findings in vision, cross-domain decoding accuracy increased with imagery strength. These data suggest that light-independent PSDs are a neural signature sensitive enough to decode imagination.

1. Introduction

At any given moment, we exist in two worlds: a shared external world that exists outside our bodies, and the private internal world we create within our minds. Inside these private worlds, we relive past memories, process present experience, and plan future actions – rendering "mind-reading" an essential skill for a social species. Indeed, humans have evolved to be experts at inferring covert cognitions from subtle signals. We glean attention from gaze (Frischen, Bayliss, & Tipper, 2007), and read emotions from facial expressions (Ekman, 1987) and vocal prosody (Scherer, Johnstone, & Klasmeyer, 2003). Anecdotally, we also "catch" when those around us traverse between worlds, redirecting attention from their external to internal environments ("mind wandering"). Yet only recently have we begun decoding the explicit contents of covert mental experience.

Research conducted over the past 50 years shows that physiological, neural, and reaction time data can be used to draw inferences about the objects and operations we imagine. For instance, effortful calculations result in pupil dilation (Hess & Polt, 1964), neural activity reveals whether we are imagining sight or sound (Cichy, Heinzle, & Haynes, 2011; Kraemer, Macrae, Green, & Kelley, 2005), and reaction-time increases proportionally to the degree we mentally rotate an object (Cooper & Shepard, 1973). Advances in neuroimaging have further expanded this ability to understand covert cognitions and operations. Capitalizing on unique information encoded in higher-level patterns of brain activity, researchers have demonstrated rudimentary reconstruction of imagined visual categories (Reddy, Naotsugu, & Serre, 2010), images (Koenig-Robert & Pearson, 2019), and dreams (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013). While much of mental imagery research has centered on the visual domain (leaving open questions of generalizability to other kinds of "sensory thought", Pearson, 2019), intracranial recordings have provided some evidence for the decoding of imagined auditory features, e.g. spectrotemporal features of speech (Martin et al., 2014). Here we report a novel and non-invasive

---

* Corresponding author at: 1540 William Hames Hall, 33 Kirkland St., Cambridge, MA 02138, United States.
E-mail address: oliviakang@fas.harvard.edu (O. Kang).

https://doi.org/10.1016/j.concog.2019.102862
Received 22 November 2019; Received in revised form 11 December 2019; Accepted 12 December 2019

method for decoding high-level auditory mental imagery.

Across two experiments, we investigate whether patterns of pupil-size dynamics (PSDs) can sensitively encode the music we hear "inside our heads". We first test the robustness of this signal (i.e., how well it is conserved) within and across individuals. We additionally assess how well this signal is conserved across perception and imagination domains, following findings in the vision literature that mental imagery can function like a weak form of perception (Pearson, Naselaris, Holmes, & Kosslyn, 2015). We hypothesize that pupillometric decoding of imagination should be possible in all these conditions due to known relationships between pupil size and the locus coeruleus-norepinephrine system.

1.1. *Pupil size and the locus coeruleus-norepinephrine system*

In response to stress and salience, the locus coeruleus (LC) modulates attention and arousal via the release of norepinephrine (NE) to the forebrain (Berridge & Waterhouse, 2003). While the exact path of innervation remains unknown, studies conducted in humans and animals show evidence of a tight and reliable relationship between NE release and pupil dilation (e.g., Joshi, Li, Kalwani, & Gold, 2016; Sterpenich et al., 2006), such that under constant-light conditions, changes in pupil size are likely mediated near-exclusively by LC-NE activity (Koss, 1986). This association is true across modalities (e.g. touch – Chapman, Oka, Bradshaw, Jacobson, & Donaldson, 2003; smell – Schneider et al., 2009; sound – Partala & Surakka, 2003; sight – Hess & Polt, 1960), and this association is precise; Joshi and colleagues found that light-independent pupil dilations index NE activation on a "spatiotemporal scale" (i.e., amplitude and timing) as fine as single spikes from single-unit recordings (2016). These data suggest that light-independent PSDs reflect moment-to-moment changes in attention and arousal. While mental imagery and attention are dissociable processes (Thompson, Hsaio, & Kosslyn, 2011), attention is necessary to consciously generate internal representations; and if the generated percept is dynamic – unfolding in meaning over time – mental effort and physiological arousal are also likely to vary. Hence, we hypothesize that given the tight link between pupil dilation and noradrenergic firing by the LC, stimuli that evoke *unique and reliable* fluctuations in attention and arousal even when imagined will produce unique pupillary signatures capable of identifying that imagined stimulus.

Eye-tracking research has seen initial strides towards the idea of 'mind-reading': Stoll et al. leveraged known relationships between pupil size and mental effort to decipher binary "Yes/No" (nonverbal) answers from patients with locked-in syndrome (2013); the timing of maximum pupil size over a trial has been shown to identify the time at which a decision is reached (Einhäuser, Koch, & Carter, 2010); and Laeng and Sulutvedt found that the pupil adjusts to imagined brightness (Laeng & Sulutvedt, 2014), and with Mannix: imagined size and distance (Sulutvedt, Mannix, & Laeng, 2018), demonstrating physiological response to imagined percepts. We build on this literature to investigate what global patterns of pupillary responding can reveal about the high-level contents of auditory imagery. Across two experiments, we continuously measured PSDs of participants as they physically listened to and then silently imagined musical excerpts. We used these data to train machine learning algorithms and test their ability to correctly identify these songs when later imagined.

2. General methods and materials

2.1. *Participants*

The reported studies were approved by the Committee on the Use of Human Subjects at Harvard University. Participants were recruited through Harvard University's Study Pool website, and compensated for their participation with course credit. Written informed consent was obtained following description of the task. All participants had normal or corrected-to-normal vision and hearing. Participants under 18 years of age provided parental consent for their participation.

2.2. *Eye-tracking and quality control*

Eye-tracking was conducted using the SMI Red-n eye-tracker, a system with a large "head box" (i.e., tracking range) that allows participants to behave naturally during music listening (e.g., nodding their heads in time to a beat). Participants were seated approximately 30 in. from the eye-tracker. Participants were calibrated prior to the task and asked to keep their gaze somewhere on the computer screen for the duration of the experiment. Pupil diameter was recorded from the left eye at 30 Hz. Missing values due to excessive movement or eye-blinks were linearly interpolated. Trials requiring over 25% of data to be interpolated were discarded (see Kang & Wheatley, 2015). The resulting data was order 5 median-filtered and low-pass filtered (cutoff frequency 10 Hz) to remove spikes from the data, averaged into 100 ms bins, and detrended to correct for slow drift (see Smallwood et al., 2011; Wierda, van Rijn, Taatgen, & Martens, 2012).

2.3. *Rationale for using music stimuli*

We test the hypothesis that PSD patterns identifiably encode online mental experience. We chose musical stimuli for three reasons: first, we believe that inherent properties of rhythm and meter render music an easy stimulus to imagine vividly and on a reasonably stable time-scale within and across individuals. Second, music remains mentally and physiologically engaging over repeated representations (Laeng, Eidet, Sulutvedt, & Panksepp, 2016; Salimpoor, Benovoy, Larcher, Dagher, & Zatorre, 2011), presenting lower risk of mind-wandering. Finally, the pervasive role of music in everyday life allows reasonable likelihood that stimuli will be familiar and possible to mentally recreate by a broad population.

It is important to note that the LC-NE system responds to stress and salience regardless of source. Music contains many such sources: low-level features such as volume, pitch, or timbre, higher-level features such as attention-grabbing lyrics, and combinations of the two that may contribute to e.g. musical/emotional tension and resolution. In these experiments, we do not claim to know which features drive noradrenergic release (indeed, this is likely to vary across individuals); we test only whether the resultant pattern of NE release is uniquely associated with specific pieces of music and subsequently identifiable using machine learning methods.

2.4. Data analysis

Binary support vector machine (SVM) and dynamic time warping (DTW) classification algorithms were used to test whether PSDs uniquely and reliably encoded the contents of silent imagination. Classification data were then analyzed using intercept-only logistic regressions with random intercepts for participants, imagined song, and classifier (as determined by the model). Thus, all classification accuracies reported take participant-level and stimulus-level dependencies into account.

2.4.1. Between-participants classification

To investigate whether PSD patterns uniquely encode imagined songs across participants, we used binary SVM classifiers to conduct leave-one-out cross-validation in R using the *caret* package (Kuhn, 2017). For example, given the song-pair *Build Me Up Buttercup* (The Foundations) and *Chandelier* (Sia), a classifier was trained on the PSD patterns collected from all but one of the participants as they imagined both songs. After this learning, the classifier was given the remaining participant's imagination trial to decode. This was done for every possible song-pair for each participant.

2.4.2. Within-participants classification

We used DTW, a standard algorithm for detecting pattern similarities in signals that may be offset in time (Berndt & Clifford, 1992; Mueen & Keogh, 2016) for the within-participants analyses in Experiments 1 and 2. The DTW function uses a dynamic programming approach to time series via the stretching and compressing of the time axis within specified time windows. This temporal flexibility compensates for small deviations in timing that may occur e.g., when imagining music in silence. We specified 3-second windows to ensure that the window contained a local but meaningful portion of the signal (see Kang & Wheatley, 2015, 2017). DTW calculates the effort needed to align two signals and outputs this calculation as a "cost value", such that the higher this value, the more *dissimilar* the signals being compared.

In the current studies, DTW was used to compute pattern similarity of PSD patterns during the listening and imagination of songs. For each participant, DTW computed the cost value for signals collected as participants listened to and imagined the *same* song, and compared this value to the cost values of aligning that same imagination trial to every other listening trial. For instance, if participants imagined *Song 1,* DTW would align this pupillary signal to the signal collected as participants listened to *Songs 1, 2, 3,* and 4. The cost of these alignments would then be compared in a binary fashion (e.g., *1,1 vs. 1,2; 1,1 vs. 1,3, etc.)* to classify the imagined song.

3. Experiment 1: Decoding 10 s of musical imagery

3.1. Methods

3.1.1. Participants.

50 participants signed up for Experiment 1 during the recruiting window. 1 participant did not complete the experiment, 1 participant's data was lost due to technical error, and an additional 2 participants' data did not meet quality-control thresholds due to excessive movement and/or blinking *(2.2)*. Data from 46 participants were analyzed.

3.1.2. Musical stimuli

Experiment 1 presented 40-second excerpts from 4 popular songs: *Build Me Up Buttercup* (The Foundations), *Chandelier* (Sia), *I Want To Hold Your Hand* (The Beatles), and *Wannabe* (The Spice Girls). All songs contained vocals and instrumentals, and were rated as being "highly familiar" by a group of 36 independent raters (M = 4.39 on a 5-point familiarity scale, where "1" corresponded to "highly unfamiliar" and "5" corresponded to "highly familiar").

3.1.3. Procedure

Following calibration, participants were presented with four musical excerpts, presented on over-ear headphones. Participants listened to each excerpt in its entirety ("Listening" trials), and then to that same excerpt with 10 s removed at some point within the song ("Imagination" trials). During listening trials, participants were asked to listen carefully to the excerpt "as though you're trying to memorize it", and to redirect their attention back to the music should they notice their minds wandering. After the first listening trial for each song, participants indicated how well they knew the song on a 5-point familiarity scale (*3.1.2*). During imagination trials, participants' goal during the 10-second periods of silence was to "fill in the music as vividly and accurately possible" silently in their minds, with the goal of perfectly matching the music when it recommenced (see Fig. 1). This was done twice for each song. Songs were presented in random order.

Throughout the experiment, participants were asked to keep their gaze somewhere on the computer screen. To keep luminance

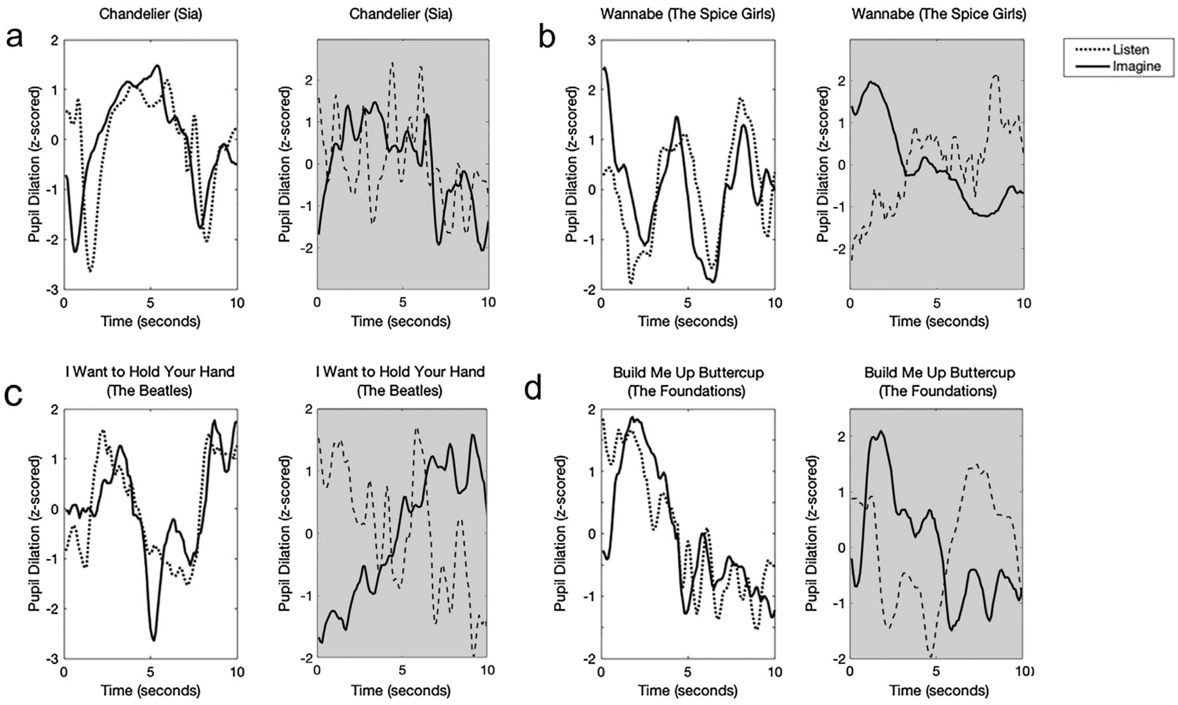

Fig. 1. Pupil dilation patterns during physical listening and silent imagining of musical excerpts. Examples of pupillometric time-series data from eight participants (1 per panel) as they physically listened to and silently imagined the musical excerpts presented in Experiment 1 (a-d) in constant-light conditions. For each song, we show sample trials from two participants: one whose imagination trial was correctly classified across domain (left) and one whose imagination trial was incorrectly classified. All panels display z-scored pupillometric time-series collected during 10 seconds of either music or silence occurring within a 40-second trial. As pupil-size amplitude at each moment reports noradrenergic release in response to stress and salience, pupillary time-courses yield a signature of participants' subjective experience of attention and physiological arousal on a trial-by-trial basis.

constant, screens displayed a solid black background. Prior to each trial, the word "LISTEN" or "IMAGINE" appeared onscreen to alert participants to the upcoming task.

3.2. Results

3.2.1. Classifier decoding of novel imagination trials following training on imagination PSDs (within-domain classification)

Pupil dilation time-series data collected over these 10 s of imagination were used to train leave-one-out binary SVM classifiers *(2.4.1)*. For each participant, classifiers were trained on all *other* participants' imagination data, and then given the participant's own imagination trial to decode. Classifiers correctly decoded the imagined song 74.1% of the time, taking into account participant-level dependencies (chance = 50%; $OR = 3.1$, $p < 0.0001$, *95% CI* [71.4%, 79.7%]). These results suggest that dynamic stimuli – even when silently imagined – can elicit unique patterns of NE release that are conserved across individuals. This performance cannot be explained solely by low-level features of the pupillary signal; a secondary set of classifiers trained only on the minimum, maximum, and average amplitude of pupils and their standard deviations were unable to identify the imagined songs better than chance ($p = 0.5$).

We additionally tested whether participants' *own* PSDs collected during music imagination could predict the songs they later imagined. As participants did not provide enough individual data to reasonably train SVM classifiers, we used a pattern recognition algorithm (DTW) to classify participants' imagination trials *(2.4.2)*. When computing pattern similarity between pairs of signals (each collected over a single imagination trial), DTW identified the correct song 64% of the time; taking into account participant-, song-, and classifier-level dependencies, the likelihood of correct decoding significantly exceeds chance accuracy ($OR = 1.91$, $p = 0.003$, *95% CI* [54.0% 75.7%], chance = 50%*)*. These accuracies fall into similar ranges seen in studies decoding visual imagery (Harrison & Tong, 2009; Koenig-Robert & Pearson, 2019; Reddy, Tsuchiya, & Serre, 2010).

3.2.2. Classifier decoding of novel imagination trials following training on listening PSDs (cross-domain classification)

We also conducted decoding across domains to interrogate similarities in noradrenergic response to externally-perceived versus internally-generated music. Findings in the vision literature suggest that mental processes associated with the perception of a stimulus overlap with but differ from mental processes associated with imagining that stimulus (Amedi, Malach, & Pascual-Leone, 2005; Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015; see Pearson et al., 2015 for a review), and that this overlap may be influenced

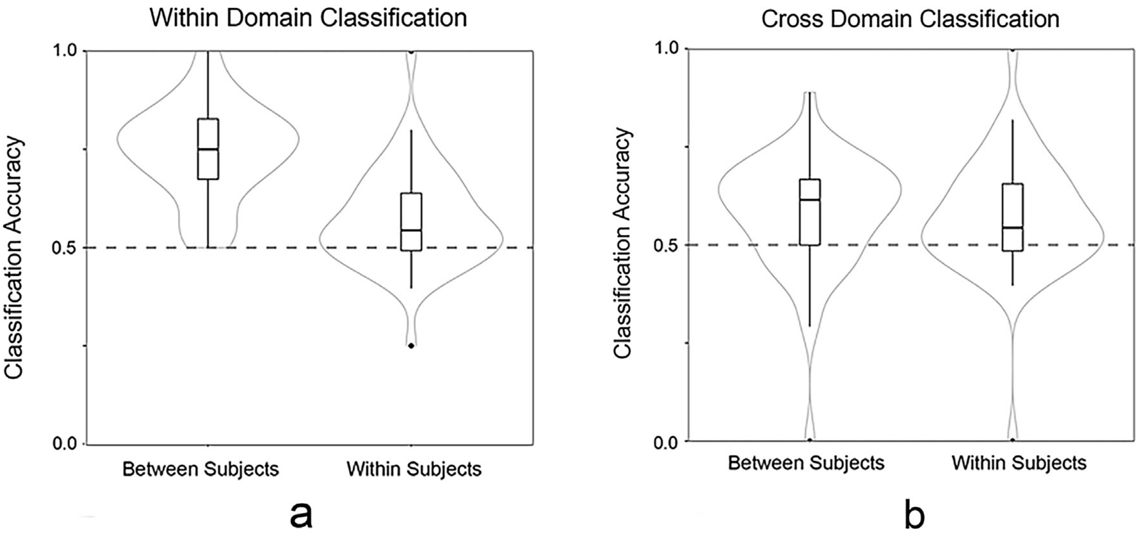

Fig. 2. Classification of imagined music given pupillometric training data collected during imagination (within-domain classification) and music-listening (cross-domain classification). Machine learning algorithms tasked with decoding the song participants imagined from eye-pupil data alone significantly exceeded chance accuracy following training on others' Imagination trials (Panel A: between subjects: 74.1%; within subject: 64%, all *p*'s < 0.005). Classifier performance was reduced when subjected to cross-domain classification (i.e., decoding Imagination trials following training on Listening trials, Panel B): performance exceeded chance accuracy between subjects (58.6%, $p < 0.02$) using SVM classifiers, but not within-subjects (56.2%, $p = 0.071$) using DTW algorithms. Strategies for boosting cross-domain classification are investigated in Study 2.

by idiosyncratic differences in the vividness of the imagined percept (Koenig-Robert & Pearson, 2019). To test this in the auditory domain, we additionally tasked classifiers with decoding imagination trials following training on *listening* trials. Taking into account participant- and song-level dependencies, the optimal model performed significantly better than chance ($OR = 1.4$, $p < 0.02$, *95% CI* [50.2%, 67.3%], chance = 50%) when decoding across perception and imagination domains, with expected decreases in overall accuracy (58.6% versus 71.4%). This was exacerbated further at the individual subject level when DTW algorithms were tasked with decoding each participant's imagination trials given their single-trial listening data. The imagined song was correctly identified 56.2% of the time; however, performance failed to reach significance ($p = 0.071$; *see* Fig. 2).

3.3. Discussion

Taken together, data from Study 1 suggest the following conclusions: (1) *patterns* of pupil dilation and constriction are conserved within and across individuals as they imagine the same piece of music; when a secondary set of classifiers was trained only low-level features of the pupillary signal (minimum, maximum, average amplitude and their standard deviations), they were unable to decode the imagined songs better than chance. This suggests that classifiers are decoding higher-level attention and arousal dynamics evoked by musical imagery (either in part e.g., by rhythm, volume, or by overall gestalt). (2) PSD patterns are more weakly conserved across perception and imagination domains: classifiers trained on "Listening" trials were able to decode "Imagination" trials across participants, but were not able to do so given single-trial data within individual subjects.

One possible reason for this difference in cross-domain decoding accuracy within- and across-participants is the greater amount of training/reference data available for the between-subjects analysis: leave-one-out classifiers were trained on the PSD patterns of all other participants (approximately 100 trials), whereas DTW calculated the pattern similarity of signal-pairs collected over single trials of listening and imagination.

Additionally, PSDs may reflect nuanced and idiosyncratic differences in the processes associated with listening to versus imagining music. Mentally recreating a sensory percept requires more effort and familiarity than the passive experiencing of that same percept; consequently, mental imagery in vision is often conceptualized as a weak or "noisy" form of perception. As PSD patterns reflect dynamic changes in attention, arousal, and mental effort, it is reasonable that classifier performance would suffer when decoding across domains, and that performance would be most affected in the within-subject analysis. Kosslyn theorized that the fidelity of an imagined percept is dependent on an individual's idiosyncratic ability to **activate stored memories of the percept,** and then to **recreate that percept** (1988), and research has supported the role of vividness in facilitating neural decoding of visual imagery (Koenig-Robert & Pearson, 2019). If auditory imagery functions similarly, PSD patterns should be more robust across domains (and decoding accuracy should improve) when familiarity with the imagined stimulus is maximized and the effort needed to recreate it is reduced. We test these possibilities in Experiment 2.

4. Experiment 2: Cross-domain decoding of 19 s of auditory imagination

In Experiment 2, we investigated whether PSDs collected as an individual *listened* to music would reliably identify the song they then *imagined* when (a) recollection and internal reconstruction of music is facilitated and (b) classifiers have access to richer referent data. As in the previous study, participants were eye-tracked as they physically listened to and then silently imagined music.

4.1. Methods

4.1.1. Participants

55 participants signed up for Experiment 2 during the recruiting window. 3 participants' data failed to pass quality-control thresholds, and 1 participant's data was lost due to technical error. Data from 51 participants were analyzed.

4.1.2. Facilitating memory activation and recreation of musical excerpts

Participants' familiarity ratings for the musical excerpts used in Experiment 1 indicated only middling experience with the musical stimuli (M = 2.99 on a 5-point scale where "1″ corresponded to "highly unfamiliar" and "5" corresponded to "highly familiar"). These ratings diverged from pre-ratings given by independent raters (M = 4.39 on the same scale). This lack of familiarity may have affected participants' ability to remember and mentally recreate music with high fidelity, distorting the associated PSD pattern. Additionally, idiosyncratic differences in participants' ability to keep mental time (i.e., to maintain a steady beat in sustained periods of silence) may have affected the temporal dimension of their pupillary signatures, stretching and compressing different points of the signal.

To increase participants' memory for the music, we presented participants with short excerpts from three well-known songs (the first verses of John Newton's *Amazing Grace* (19 s) and *Somewhere Over the Rainbow* from "The Wizard of Oz" (22 s), as well as the first 22 s of *The Imperial March (Darth Vader's Theme)* from "Star Wars") and asked them to choose the one song they knew best to subsequently imagine. We hypothesized that this would reduce mental effort associated with retrieving stored memories of the music later on.

In addition to their ubiquity, these three songs were selected for their comparable speeds. All three clips were tempo-matched at 85 bpm and overlaid with a steady metronome beat. We hypothesized that providing this beat during both music listening and imagination would reduce idiosyncrasies in participants' ability to keep steady time (minimizing mental effort associated with mental recreation), and time-lock PSD signatures across perception and imagination domains. The addition of the metronome is an extra-musical feature with the potential to influence the PSD response. However, because the same unchanging beat is present in all "Listening" and "Imagination" trials, any effect of beat alone would be applied across all songs, making accurate classification less likely.

4.1.3. Granting DTW algorithms access to more information for classification

In our between-subjects analyses, classifiers were given pupillometric time-series data from all other participants during the training period (approximately 100 training trials per song). Given the tight association between pupil size and attention (*1.1*) and the difficulty of asking participants to listen to the same song without mind-wandering enough times to train a classifier, we instead asked participants in Study 2 to listen to and imagine longer musical excerpts. Rather than the 10-second imagination trials classifiers were given in Experiment 1, data collected over 19 s of imagination (i.e. the maximum duration shared across all three songs) were analyzed in Experiment 2. This provided DTW algorithms with more differentiated and information-rich time-series data for classification.

4.1.4. Procedure

As in Experiment 1 (*3.1.3)*, participants were eye-tracked as they physically listened to and then silently imagined music. During the listening portion of the task, participants were asked to listen to each song carefully, and to redirect their attention to the music as necessary. During the imagination portion of the task, participants heard a metronome track beating at 85 bpm. They were told to wait 4 beats to establish tempo, and then begin imagining their chosen song as vividly and accurately as they could starting on the fifth beat (Fig. 3). Music was presented on over-ear headphones, and participants were instructed to keep their gaze on the computer screen for the duration of the task. To maintain constant luminance, a solid black screen was presented during listening and imagination trials.

4.2. Results

4.2.1. Within-subject classification across domains

Following procedures outlined by Kang and Wheatley (2017), DTW algorithms calculated the degree to which PSD patterns were conserved during the listening and imagination of songs *(2.4.2)*. Given participants' imagination trials, DTW algorithms calculated pattern similarity for every possible pair of time-series. The song participants chose to imagine was correctly identified 75.5% of the time ($OR = 5.12$, $p = 0.004$, *95% CI* [69.6%, 99.1%], chance = 50%) when given access to the full common dataset (19 s) and 62.8% of the time when given only the first 10s of dilation data during listening and imagination ($OR = 1.77$, $p = 0.025$, *95% CI* [52.6% 76.5%], chance = 50%), taking participant-level dependencies into account (Fig. 4).

Each of the three musical excerpts were chosen with roughly equivalent frequency, and all participants reported being very

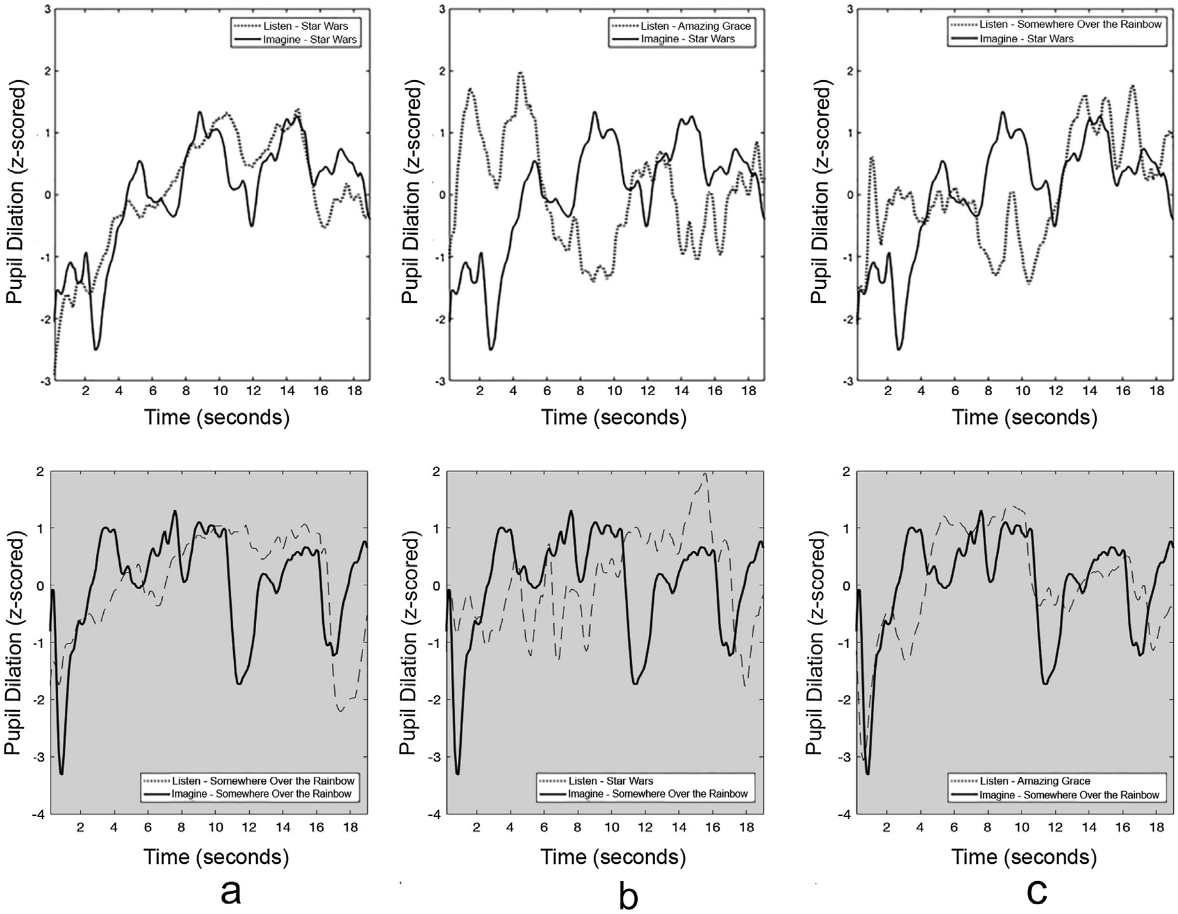

Fig. 3. Pupil dilation dynamics of two participants during listening and imagining of the same (a) and different (b, c) songs. Pupil dilation patterns of two participants as they imagined 19 s of their chosen song (Top Row: *The Imperial March (Darth Vader's Theme)* from "Star Wars"; Bottom Row: *Somewhere Over the Rainbow* from "The Wizard of Oz") were compared to dilation patterns collected as they listened to the same song (a) or different songs (b, c). The top panel illustrates a trial where the participant's imagined song was correctly decoded; the bottom panel illustrates a trial where the participant's imagined song was incorrectly decoded.

familiar with the song they imagined (M = 4.25 on the 5-point familiarity scale previously described).

4.3. Discussion

These data demonstrate pupillometric decoding of musical imagery within an individual across perception and imagination domains. The data additionally suggest that auditory mental imagery functions like a "noisy" form of perception, with idiosyncratic differences in imagery strength influencing the degree of overlap in mental processes associated with sensory perception and internal generation. Decoding performance further increased when algorithms were given access to more reference data. Decoding accuracies of PSD patterns are comparable to those seen in neural decoding research in the visual domain (e.g. Harrison & Tong, 2009; Koenig-Robert & Pearson, 2019; Reddy, Tsuchiya, et al., 2010).

5. Conclusions

Over the past 60 years, scientists have primarily used discrete changes in pupil size to broadly infer individuals' state of attention, arousal, and mental effort. Recently, psychological investigations have extended their consideration to the pupillometric time-series, demonstrating the ability of higher-level patterns of pupil constriction and dilation to reveal when and why these mental states shift. In the current work, we build on this knowledge to show that PSD patterns not only encode online changes in physiological experience, but that this signal can be sensitive and robust enough to identify high-level contents of mental experience.

Across two experiments, machine learning methods decoded imagined songs from the dynamic pupillary response with accuracy significantly better than chance in a variety of contexts: across individuals using classifiers trained on group-level data, within an individual given single-trial data, and – under conditions where the degree of cognitive demand is sufficiently similar– across domains of perception and imagination. This decoding was not possible when classifiers were trained only on low-level and discrete

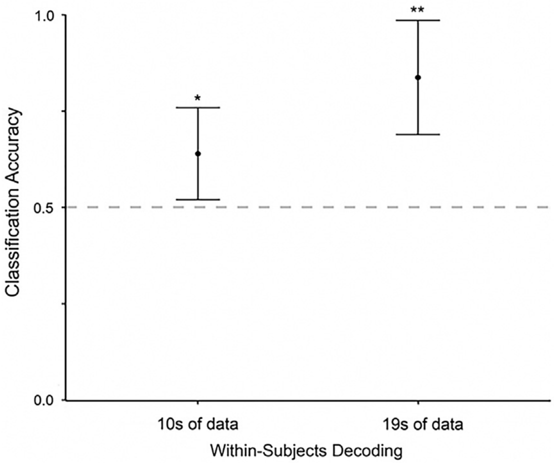

Fig. 4. Classification of imagined music given pupillometric training data collected during physical listening (cross-domain classification) of personally-familiar music. Participants listened to three tempo-matched musical excerpts accompanied by a metronome beat and chose the song they knew best to vividly imagine given only that metronome beat. Machine learning algorithms were tasked with decoding which of the three songs participants imagined from pupil-size time-series alone. The unique morphology of the pupillary response was conserved across perception and imagination such that DTW algorithms, given every possible song pair, identified the imagined song correctly 62.8% of the time when given 10-second time-series to decode (chance = 50%, *OR* = 1.77, p = 0.025, *95% CI* [52.6% 76.5%]), and 75.5% of the time given 19-second time-series to decode (*OR* = 5.12, p = 0.004, *95% CI* [69.6%, 99.1%]).

pupillary features. Taken together, these data show that the information encoded within pupillometric time-series data is rich, sensitive, and robust enough to allow classification. The current work focused on imagined audition, however LC-NE-mediated pupil size changes are modality-free. Paired with the myriad practical advantages of eye-tracking as a method for capturing neural activity (e.g. accessibility, mobility, ecological validity), these findings suggest broad potential for pupillometric methods to expand investigation and understanding of dynamic mental processes. To decode the contents of mental life is becoming increasingly possible.

CRediT authorship contribution statement

O. Kang: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **M.R. Banaji:** Conceptualization, Resources, Writing - review & editing.

Acknowledgements

Special thanks to B. Kurdi for statistical advice, S. Sanjay for research assistance, and T. Wheatley for use of eye-tracking equipment.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2019.102862.

References

Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron, 48*(5), 859–972.
Berndt, D. J., & Clifford, J. (1992). Using dynamic time warping to find patterns in time series. *KDD Workshop, 10*, 359–370.
Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews, 42*(1), 33–84.
Chapman, C. R., Oka, S., Bradshaw, D. H., Jacobson, R. C., & Donaldson, G. W. (2003). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain-evoked potentials and pain report. *Psychophysiology, 36*(1), 44–52.
Cichy, R. M., Heinzle, J., & Haynes, J. D. (2011). Imagery and perception share cortical representations of content and location. *Cerebral Cortex, 22*(2), 372–380.
Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.). *Visual information processing*. Oxford, England: Academic.
Einhäuser, W., Koch, C., & Carter, O. L. (2010). Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience, 4*, 1–9.
Ekman, P., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712.
Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694.
Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual area. *Nature, 458*, 632–635.
Hess, E. G., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*(3423), 349–350.
Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem- solving. *Science, 143*(3611), 1190–1192.
Horikawa, T., Tamaki, N., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science, 340*(6132), 639–642.
Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex.

Neuron, 89, 221–234.

Kang, O., & Wheatley, T. (2015). Pupil dilation patterns reflect the contents of consciousness. *Consciousness and Cognition, 35*, 128–135.

Kang, O., & Wheatley, T. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology: General, 146*(4), 569–576.

Koenig-Robert, R., & Pearson, J. (2019). Decoding the contents and strength of imagery before volitional engagement. *Scientific Reports, 9*(1), 3504.

Koss, M. C. (1986). Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *Journal of Pharmacological Methods, 15*(1), 1–19.

Kosslyn, S. M. (1988). Aspects of a cognitive neuroscience of mental imagery. *Science, 240*(4859), 1621–1626.

Kuhn, M. (2017). Caret: Classification and regression training. R package version 6.0-77. https://CRAN.R-project.org/package=caret.

Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: Sound of silence activates auditory cortex. *Nature, 434*(7030), 158.

Laeng, B., & Sulutvedt, U. (2014). The eye pupil adjusts to imaginary light. *Psychological Science, 25*(1), 188–197.

Laeng, B., Eidet, L. M., Sulutvedt, U., & Panksepp, J. (2016). Music chills: The eye pupil as a mirror to music's soul. *Consciousness and Cognition, 44*, 161–178.

Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N. E., … Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering, 7*. https://doi.org/10.3389/fneng.2014.00014.

Mueen, A., & Keogh, E. J. (2016). Extracting optimal performance from dynamic time warping. *Tutorial at KDD, 2016*, 2129–2130.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for eary visual areas decodes mental images of remembered scenes. *NeuroImage, 105*, 215–228.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *Human-Computer Studies, 59*, 185–198.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences, 19*(10), 590–602.

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual imagery. *Nature Reviews Neuroscience, 20*, 624–634.

Reddy, L., Naotsugu, T., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage, 50*(2), 818–825.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience, 14*(2), 257.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.). *Handbook of affective sciences* (pp. 433–456). Oxford University Press.

Schneider, C. M., Ziemssen, T., Schuster, B., Seo, H.-S., Haehner, A., & Hummel, T. (2009). Pupillary responses to intranasal trigeminal and olfactory stimulation. *Journal of Neural Transmission, 116*(7), 885–889.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., … Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE, 6* e18298.

Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., … Maquet, P. (2006). The locus ceruleus is involved in the successful retrieval of emotional memories in humans. *Journal of Neuroscience, 26*, 7416–7423.

Stoll, J., Chatelle, C., Carter, O., Koch, C., Laureys, S., & Einhäuser, W. (2013). Pupil responses allow communication in locked-in syndrome patients. *Current Biology, 23*(15), R647–R648.

Sulutvedt, U., Mannix, T. K., & Laeng, B. (2018). Gaze and the eye pupil adjust to imagined size and distance. *Cognitive Science, 42*(8), 3159–3176.

Thompson, W. L., Hsaio, Y., & Kosslyn, S. M. (2011). Dissociation between visual attention and visual mental imagery. *Journal of Cognitive Psychology, 23*(2), 256–263.

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America, 109*(22), 8456–8460.